

Welkom

Informatieportaal, PAD

Project Architectuur Document

Eigenaar: Thijs Berden
17 augustus 2016
Status: Concept

Inhoudsopgave

- 1. Programma Spoordata**
- 2. Project**
 - 2.1 Project doel
 - 2.2 Epic Value Statement
 - 2.3 Non-Functional Requirements
- 3. Business architectuur**
 - 3.1 As-Is en To-Be
 - 3.2 Functionele eisen
 - 3.3 Richtlijnen vanuit Business perspectief
 - 3.4 Aspecten van de te ontsluiten data
- 1. Informatie architectuur**
 - 4.1 Referentie architectuur
 - 4.2 Componenten
 - 4.3 Ontsluiten en integreren
 - 4.4 Publiceren (IA naar Integratie)
 - 4.5 Publicatiedatabase, Data
 - 4.6 Publicatie Documenten
 - 4.7 Publicatiedatabase, Relaties
 - 4.8 Geo functies
 - 4.9 Zoeken
 - 4.10 Front-End en API-services
 - 4.11 Databronnen
- 5. Informatiesysteem architectuur**
 - 5.1 Architectuur schets
 - 5.2 Architectuur richtlijnen
 - 5.3 Organisatie van het beheer
- 6. ICT Architectuur**
 - 6.1 ICT Architectuur
 - 6.2 ICT Architectuur richtlijnen
 - 6.3 Project overstijgende architectuurkeuzen
- 7. Implementatie**
 - 7.1 Beheer organisatie
 - 7.2 Informatiebeveiliging
 - 7.3 Migratie
 - 7.4 Uitmaken bestaande applicaties
- 8. Afwijkingen van de referentiearchitectuur**

2.1 Project doel

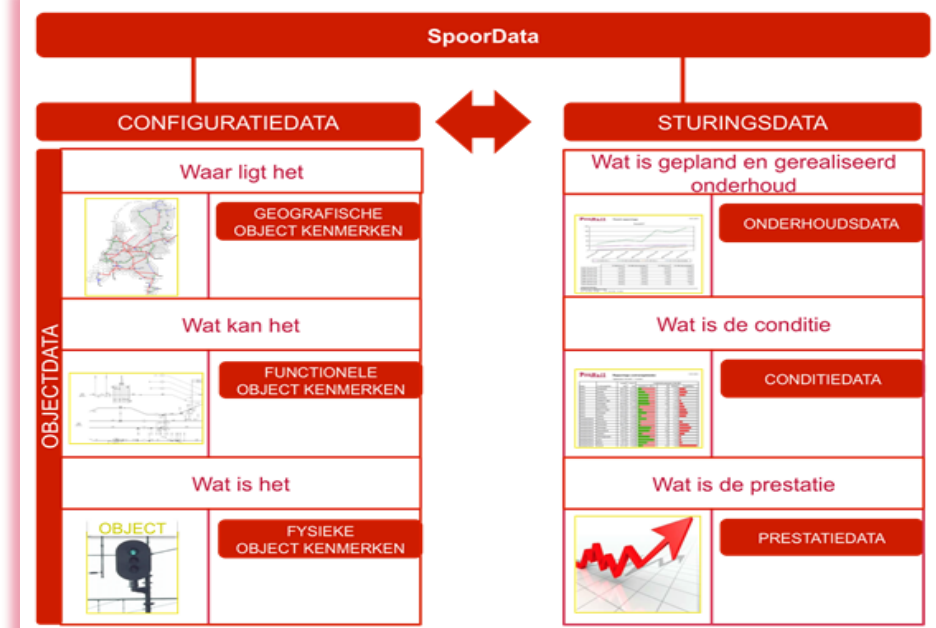
Het informatieportaal is onderdeel van het Integraal publiceren binnen het programma Spoordata, en geeft invulling aan de derde ontwikkelpijler:

Informatie toegankelijk.

De bredere doelstelling van dit programma is het voeden van kernprocessen van ProRail en partners met juiste, volledige, actuele en consistente data op de informatie domeinen Configuratie data (Geografische, Functionele en Fysieke kenmerken) en Sturings data (Conditie, Onderhoud en Prestatie).

De gegevens zijn eenduidig en onafhankelijk van plaats en tijd te ontsluiten voor alle betrokkenen in de spoorbranche. Daarmee draagt Spoordata bij aan een veilig, betrouwbaar, punctueel en duurzaam spoornet.

Het integraal ontsluiten van de Assetdata is de Taak van het Informatieportaal Spoordata.



2.2 Epic Value statement

Voor externe partijen uit de spoorbranche en interne ProRail medewerkers

die toegang nodig hebben tot asset informatie vanuit ProRail

is het Informatie Portaal

een manier om onafhankelijk van plaats en tijd snel toegang te hebben tot relevante, geïntegreerde asset informatie

zodat

- Vanuit 1 portaal sneller en op een intuïtieve wijze gezocht kan worden naar (gecombineerde) asset informatie en het IDC minder datavragen hoeft te verwerken
- Een level playing field ontstaat in de spoorbranche m.b.t. de beschikbaarheid van asset informatie vanuit ProRail
- Data inconsistenties tussen de individuele bronnen beter kunnen worden gesignaleerd
- Er een infrastructuur/platform ontstaat waar toekomstige initiatieven voor het ontsluiten van asset informatie op aan kunnen sluiten tegen lagere kosten

in tegenstelling tot de huidige situatie, waarin:

- gebruikers moeten inloggen op verschillende bronsystemen om gefragmenteerde asset informatie te raadplegen
- gevestigde aannemers een informatievoordeel hebben t.o.v. andere partijen
- gebruikers regelmatig werken met verouderde informatie (risico voor veiligheid / efficiëntie)
- het bewaken van consistentie in de asset informatie over de bronnen heen erg moeilijk is
- geen centraal platform bestaat voor het ontsluiten van diverse bronnen voor asset informatie

biedt onze oplossing een portaal om asset informatie via één laagdrempelige toegang te ontsluiten voor geautoriseerde interne en externe partijen, afgestemd op deze gebruikersgroepen

2.3 Non Functional Requirements

Non-Functional Requirement	NEN-ISO benaming	Omschrijving	Argumentatie waarom wel/niet belangrijk:	Top 5
Leerbaarheid	Learnability	De mate waarin een product of een systeem gebruikt kan worden door gespecificeerde gebruikers om gespecificeerde leerdoelen te bereiken met betrekking tot het gebruik van het product of systeem met effectiviteit, efficiëntie, vrijheid van risico en voldoening, in een gespecificeerde gebruikcontext.	Het portaal werkt intuïtief en is zonder training of gebruikershandleiding direct te gebruiken voor gebruikers met kennis van de spoorse asset informatie (aan te tonen door usability expert)	1
Beschikbaarheid	Availability	De mate waarin een systeem, product of component operationeel en toegankelijk is wanneer men het wil gebruiken.	Het informatieportaal moet 7*24 beschikbaar zijn, met een uptime van > 97,5 %. Het servicelevel is Brons-plus.	2
Snelheid	Time behavior	De mate waarin antwoord- en verwerkingstijden en doorvoersnelheid van een product of systeem, tijdens de uitvoer van zijn functies, voldoet aan de wensen (NB. Ook wel response tijden of performance genoemd).	Een adequate performance van de zoekfunctie is randvoorwaardelijk voor het succes van het portaal	3
Aanpasbaarheid	Adaptability	De mate waarin een product of systeem effectief en efficiënt aangepast kan worden voor andere of zich ontwikkelende hardware, software of andere operationele of gebruiksomgevingen. Hieronder valt ook de SCHAALBAARHEID van de oplossing !	Het ontwerp moet maximale uitbreidbaarheid faciliteren, zodat aantal bronnen, eindgebruikers en hoeveelheid/type data makkelijk uitbreidbaar is. Het informatieportaal moet wijzigingen in de (structuur) van de data kunnen publiceren.	4
Toepasbaarheid	Appropriateness	De mate waarin de functies het bereiken van gespecificeerde taken en doelen mogelijk maakt.	Het informatieportaal heeft als kerndoel integraal publiceren van data uit de relevante Assetmanagement systemen, met zoekfuncties.	5

Zie verder PAD Bijlage B Non Functional Requirements

2.4 Uitgangspunten

	Uitgangspunt
URA01	De ontsluiting van de data vanuit de Bronsystemen tot en met de Integratiedatabase in het datawarehouse wordt uitgevoerd onder regie van ccBI onder de architectuur richtlijnen van ccBI, deze architectuur is in deze PAD niet verder uitgewerkt.
URA02	De ontsluiting van geo data wordt uitgevoerd door applicatiebeheer GIS bij het competence center InfraData.
URA03	De gegevens en informatie die in de Spoordata.nl website en informatieportaal worden ontsloten en opgeslagen in daarvoor benodigde cloud diensten. Het betreft hier alleen publiceerbare gegevens die conform het ProRail cloud beleid kunnen worden opgeslagen in de Cloud;
URA04	Externe partijen (uit de spoorbranche) welke toegang moeten krijgen tot het informatieportaal zijn reeds gekoppeld via ADFS of hebben de beschikking over een EX account en hebben daarvoor de benodigde overeenkomst gesloten met ProRail
URA05	Het informatieportaal is te benaderen met reguliere kantoorautomatisering van ProRail op desk/laptop als ook op mobiele apparaten zoals tablet en smart Phone. Om hieraan te kunnen voldoen zal het ontwerp van de website en het informatieportaal moeten kunnen schalen naar de verschillende schermresoluties, het dient daarom een zgn. responsive design te zijn.
URA06	Gewijzigde data, documenten en tekeningen mogen maximaal één dag ouder zijn dan vastgelegd in het relevante bronsysteem. M.a.w. een gewijzigd object in het bron systeem is maximaal één dag na de wijziging beschikbaar in het informatieportaal.
URA07	Gebruik van Microsoft tenzij . . .
URA08	Gebruik van Azure Cloud voor het informatieportaal en de Publicatie database, tenzij . . . <i>dit uitgangspunt is meegegeven vanuit de Referentiearchitectuur.</i>
URA09	Bronnen zijn verantwoordelijk voor de kwaliteit (actualiteit en beschikbaarheid en dergelijke) van data. Het informatieportaal voert geen datakwaliteit verbetering uit en publiceert de data zoals deze is.

3.1 As-Is en To-Be

In de huidige situatie is er veel informatie beschikbaar, echter:

- Gefragmenteerd in verschillende systemen
- Niet eenduidig
- Niet geïntegreerd (vaak slecht aan elkaar te relateren zonder 'diepte' kennis)
- Niet voor iedereen gemakkelijk toegankelijk
- Veel accounts en inloggen

De gewenste situatie is

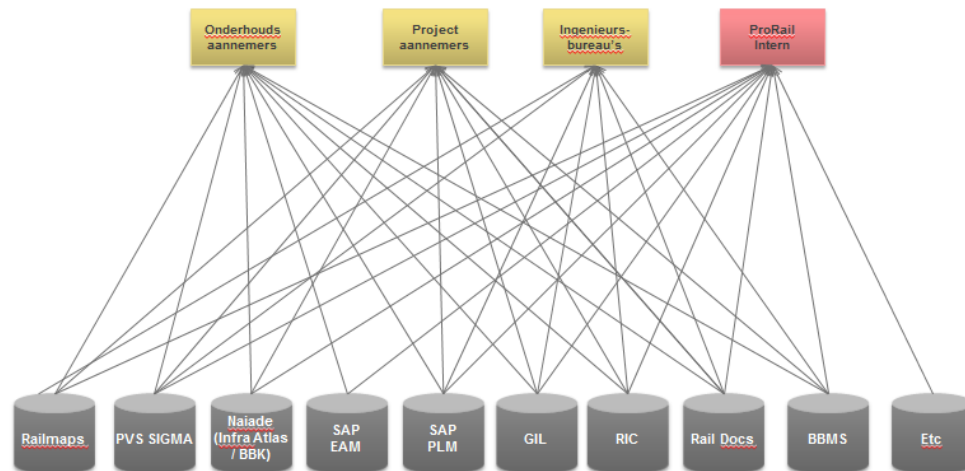
Eén loket:

- van en voor de spoorbranche
- afgestemd op de gebruikersgroepen
- voor het ontsluiten van asset informatie

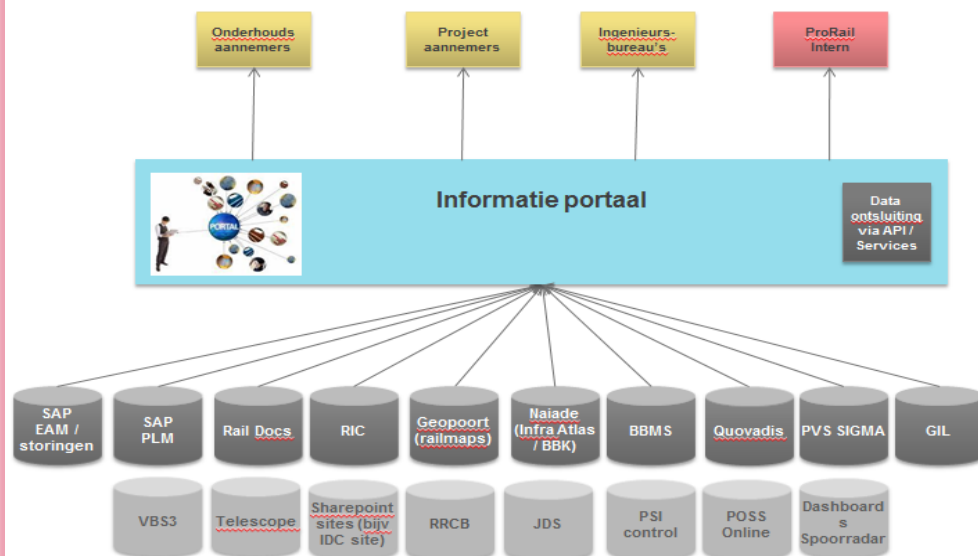
Het loket biedt:

- Eén 'portaal' die toegankelijk is via gangbare 'browsers' op gangbare devices
- Services (API) die partners of andere applicaties faciliteert in het ophalen van de data of gebruik maken van deze data in eigen applicaties.

Huidige situatie – gefragmenteerde ontsluiting

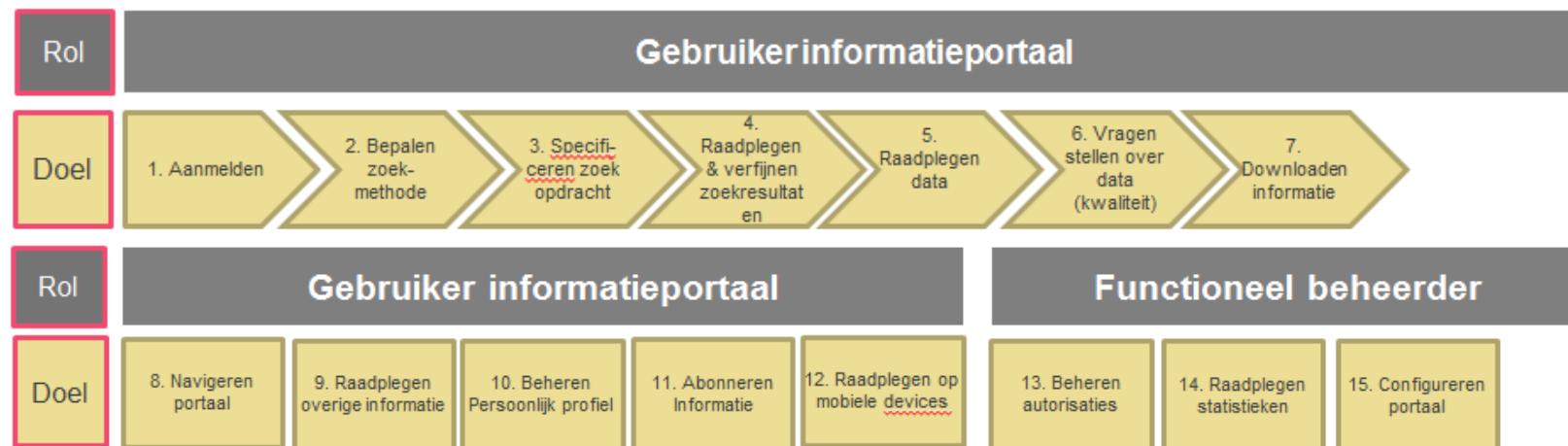


Gewenste situatie – Informatieportaal



3.2 Functionele Eisen

De Functionele Eisen zijn onderstaand samengevat.



Aan de zoekfunctie worden onderstaande eisen gesteld:

- Ongestructureerd zoeken binnen teksten van gegevensrecords, documenten en gedigitaliseerde tekeningen;
- Facet-gebaseerd zoeken;
- Gestructureerd zoeken in de publicatiedatabase;
- Geografisch zoeken met behulp van kaarten.

De ontsluiten data:

- Data uit de Asset informatiesystemen van ProRail;
- Documenten, tekeningen en beeldmateriaal die is gerelateerd aan de Assets;
- Geografische ligging en data.

Voor nadere details inzake de Functionele eisen zie:

- Informatieportaal Samenvatting Functionele Requirements, versie 1.0
- Informatieportaal, Initiële productbacklog (Excel)
- Informatieportaal PID 1.9, bijlage 3 van de voor de exacte Scope
- De paragraaf 'Te ontsluiten bronnen' in de PID Informatieportaal

3.3 Richtlijnen vanuit Business perspectief

Kerndoel van het informatieportaal is informatie uitwisseling met de spoorsector. Het informatieportaal zal daarom ten aanzien van de te ontsluiten data aan moeten sluiten op de standaarden, die nader worden uitgewerkt in de 2 andere pijlers van het programma Spoordata: Informatie standaarden en BIM Spoor Branche.

**INFORMATIE
STANDAARDEN**Nederlandse
CAD standaardInformatiemodel
SpoorObjecttype
Bibliotheek spoorProRail Unieke
Object
identificatie**BIM SPOOR
BRANCHE**Regie op
informatieDigitale
projectruimteBIM in de
spoorbranche

Voor het Informatieportaal zijn hierbij relevant Informatiemodel Spoor (IMSpoor) en Objecttype Bibliotheek voor spoor (OTL Spoor).

Deze standaarden zullen de komende jaren nog ontwikkeling zijn, dus zal het informatieportaal vloeiend moeten kunnen meebewegen met wijzigingen in deze standaarden. **Zie verder bijlage xxx voor een opsomming van de uitwisseling standaarden.**

Beperkende factor is dat de informatiesystemen van ProRail zijn nog niet aangepast naar de nieuwe standaarden. IMSpoor is nog niet doorgevoerd in de informatiesystemen van ProRail.

De ProRail Unieke Object identificatie (PUIC) is relevant voor het koppelen van de gegevens.

BR-1 Standaarden Bouwsector.

Het informatieportaal wordt ingericht om de informatie te ontsluiten conform de standaarden in de bouwsector.

BR-2 Absorberen wijzigende standaarden

Het informatieportaal moet wijzigingen in de standaarden kunnen absorberen met geen of zeer beperkte aanpassingen.

BR-3. IMSpoor datadefinities

De informatie in het informatieportaal wordt gebaseerd op de IMSpoor richtlijnen

BR-4 Huidige inhoud (Constraint)

Zolang de systemen niet zijn aangepast wordt de huidige inhoud van de kenmerken (attributen) gevoerd. Wel worden de kenmerkennamen vertaald naar de IMSpoor richtlijnen..

3.4 Aspecten van de te ontsluiten data

Objecten en kenmerken

De (asset) Objecten en de Kenmerken die van deze Objecten vastgelegd moeten worden zijn beschreven in Business Information Documenten (BID0001 en BID0008), zowel de Objecttypes als de Kenmerken die worden vastgelegd van deze objecttypes zijn continue in beweging. Deze variëteit en bewegelijkheid zet zich voort in de onderhoudsdata, conditiedata en prestatiedata.

Bron applicaties

Het informatieportaal onttrekt de Asset data uit 10 tot 30-tal bronsystemen.

- Deze applicaties zijn in de komende jaren onderhevig zijn aan wijzigingen of zelfs vervanging.
- Het informatieportaal wordt geconfronteerd met een groot aantal datasets met veel relaties tussen deze datasets, waarbij er kennis van de (transactie)systemen nodig is om door deze datasets te navigeren;
- Data uit systemen met hun eigen naamgeving, eigen datadefinities, enz
- Verschillende types informatie: data, geo, kaarten, tekeningen, documenten, fotomateriaal, filmpjes, enz.

Datakwaliteit

In de huidige situatie doen zich datakwaliteit issues voor zowel binnen één systeem als over de systemen heen, het informatieportaal moet hiermee kunnen omgaan.

BR-5 Wijzigingen volgen

Het informatieportaal moet wijzigingen in de aangeboden data van de bronapplicaties kunnen volgen zonder aanpassingen in de programmatuur.

BR-6 Inconsistente data

Het informatieportaal moet kunnen omgaan met inconsistente data en 'broken links' tussen de dataset uit de verschillende applicaties.

BR-7 Grote variëteit en dynamiek in kenmerken van de Objecten

Het informatieportaal moet dynamisch de variëteit en dynamiek in Objecten en Kenmerken kunnen volgen, zonder wijzigingen in de programmatuur.

BR-8 Omvangrijke aantal datasets uit veel informatiesystemen

Het informatieportaal moet veel data uit verschillende informatiesystemen kunnen ontsluiten en aan elkaar kunnen relateren.

4.1 Referentie architectuur

Onder verwijzing naar de Referentie architectuur een korte toelichting van de Informatie architectuur.

Presentatie

De presentatie laag biedt aan de gebruikers de zoekfuncties aan met de presentatie van de data en documenten, met waar relevant het tonen en zoeken op de kaart (Geo). Dit wordt uitgewerkt naar design en implementatie van de User-interface die wordt geïmplementeerd in de Front-End applicatielaag.

Publicatie

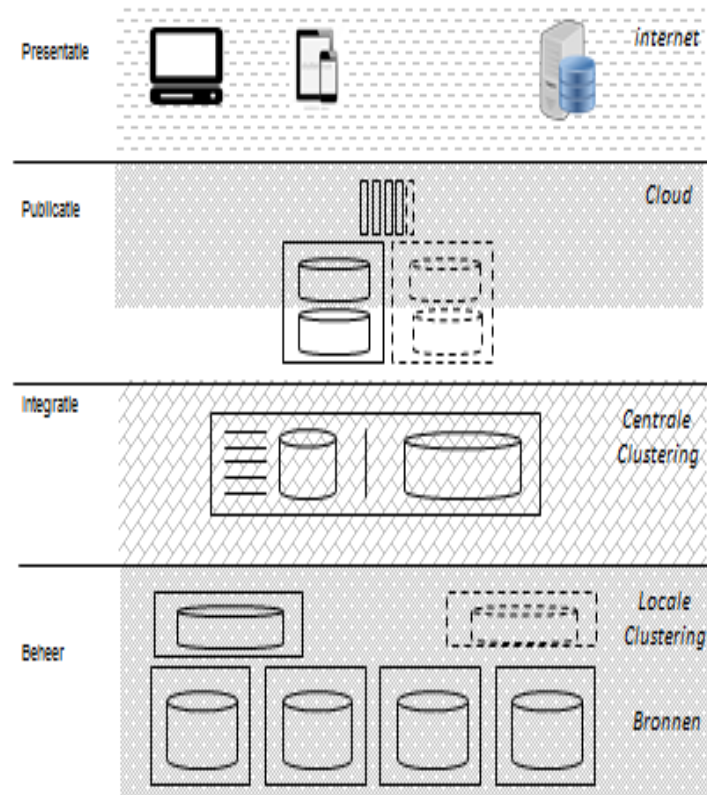
De publicatie laag bevat functies waarin logica is ondergebracht om de data te ontsluiten met een zoekfunctie, navigatiefunctie en data-retrieval functies. Ook bevat deze laag ondersteunende functies voor de Front-End applicatie. Daarnaast bevat de Publicatie de dataopslag voor zowel data als documenten, in een formaat die optimaal past bij de gebruikersvraag van Informatieportaal gebruikers.

Integratie

In de integratie laag wordt de data uit de afzonderlijke bronsystemen samengebracht en omgezet naar het IMSpoor formaat, waarbij de data ook wordt gekoppeld (geïntegreerd).

Beheer

Dit zijn de bronsystemen waarin de data wordt beheerd, deze bronnen zijn voor het Informatieportaal uitgangspunt en worden niet verder behandeld.



4.2 Componenten

Het hiernaast schema visualiseert – vereenvoudigd - de oplossing die in de navolgende paragrafen nader is beschreven.

De data wordt uit de bronnen ontsloten via het datawarehouse en van daaruit gepubliceerd en geladen in de Publicatiedatabase in de Cloud. De dataopslag van het informatieportaal heeft als eigenschap dat deze iedere aangeboden dataset kan opslaan zonder de datastructuur te kennen.

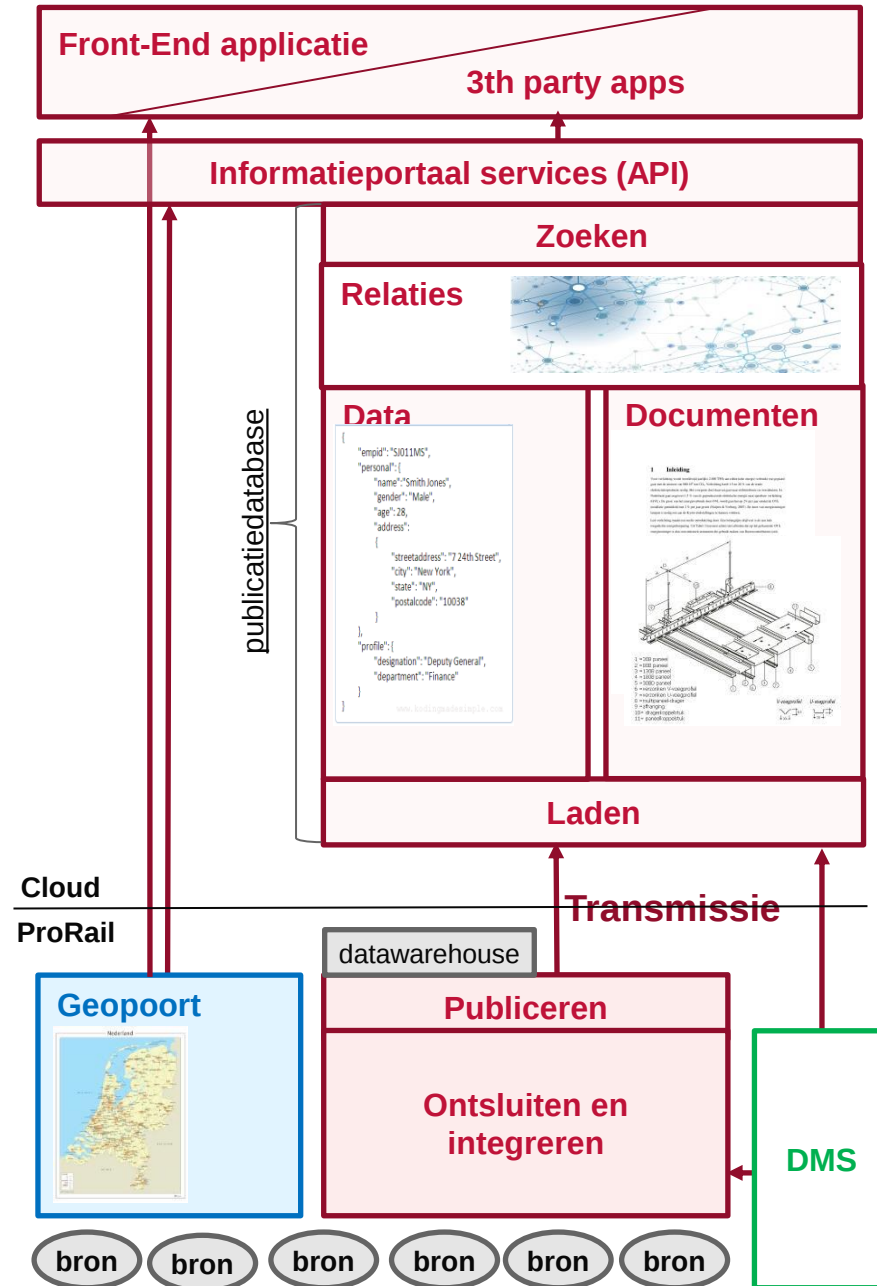
Documenten worden ontsloten vanuit de Document Management Systemen, zo nodig via opslag in de Publicatie opslag (blob). De kenmerken deze van documenten worden behandeld als een 'dataset' waaraan een document is gekoppeld en via het Datawarehouse geladen in de Publicatiedatabase.

Relaties worden aangelegd met semantische relaties: Ieder te ontsluiten subject kan een relatie van een bepaald type hebben met een ander te ontsluiten subject.

Geo (kaarten) wordt ontsloten via Geo-poort, waarbij in het Informatieportaal aanvullende services zijn geïmplementeerd voor het relateren van de data en documenten aan Geo.

Bovenstaand concept maakt het mogelijk dat het informatieportaal de ontsluiting van allerlei data kan faciliteren en deze data ook aan elkaar kan relateren zonder de inhoud te kennen.

In de navolgende sheets wordt e.e.a. nader uitgewerkt.

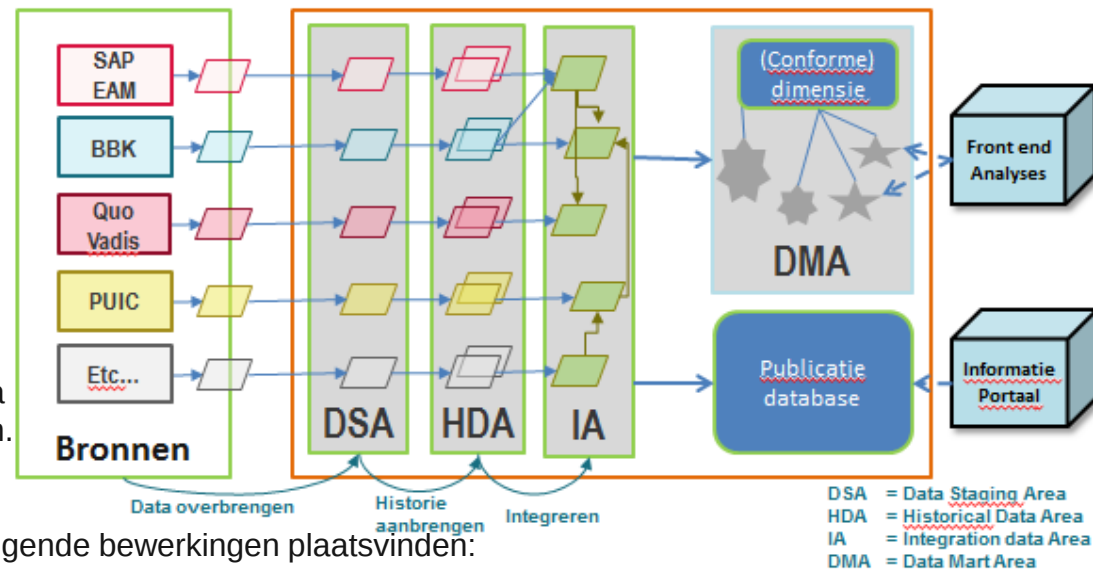


4.3 Ontsluiten en Integreren

De integratie database moet voldoen aan een aantal eisen:

- BR-3 IMSpoor datadefinities
- BR-4 Huidige inhoud (Constraint)
- BR-5 Wijzigingen volgen

Vanaf de bronsystemen wordt de data ontsloten via de gangbare datawarehouse aanpak: van bron naar staging (DSA) naar historie (HDA). In dit traject behoudt de data de structuur en inhoud van het bronsysteem.



Vanuit de historie wordt de data vervolgens geladen in de Integratie Area, waarbij de volgende bewerkingen plaatsvinden:

- De te publiceren data wordt omgezet naar de business definities en structuur (IMSpoor, OTL, BID), met in acht neming van het feit dat nog niet (alle) inhoud van de kenmerken kan worden getransformeerd (BR-4) ;
- De data wordt samengevoegd en/of gekoppeld (integreren).

Rationale:

Door de data te integreren wordt voldaan aan BR-3, bovendien wordt het eenvoudiger om wijzigingen in bronsystemen (BR-5) te absorberen omdat de data in de Integratie database al is omgezet naar ImSpoor dataformaat.

Implicatie:

- IN-01. De datadefinities van (IMSpoor, OTL en BIDxxx) worden geüniformeerd en waar nodig aangevuld.
- IN-02. Translatie naar IMSpoor datadefinities en structuur is een functie van de Integratielaag
- IN-03. Data die uit verschillende systemen komt wordt samengevoegd als deze betrekking heeft op dezelfde entiteit
- IN-04. De data wordt in de Integratiedatabase aan elkaar gekoppeld
- IN-05. Elk kenmerk heeft een master, de domeinwaarden van de master zij leidend.
- IN-06. De inhoud van kenmerken wordt voorsnog 1:1 overgenomen uit de bronsystemen en niet vertaald

Van IA naar de Publicatie database wordt beschreven onder Publiceren

4.4 Publiceren (IA naar Publicatie)

Extractie en Translatie naar publicatie formaat

Bij de Publicatie van de gegevens (van de Integratie database naar de Publicatie database) worden de gegevens gereed gezet in een formaat die optimaal aansluit op de eisen van de gebruikers van het Informatieportaal. Hierbij worden - waar relevant - de volgende bewerkingen uitgevoerd:

- De te publiceren wordt opgehaald uit de Integratiedatabase;
- Translatie van gegevens naar (externe) data-definities, zowel de kenmerknaam als inhoud;
- Translatie naar standaarden in de bouwsector;
- Samenvoegen van gegevens naar één dataset, dit kan zowel betrekking hebben op het combineren van gegevens uit meerdere bronnen als het combineren van gegevens uit meerdere tabellen naar één dataset.
- Publicatie in het een beschrijvend formaat die input is voor de Publicatie database (JSON of XML).

Rationale:

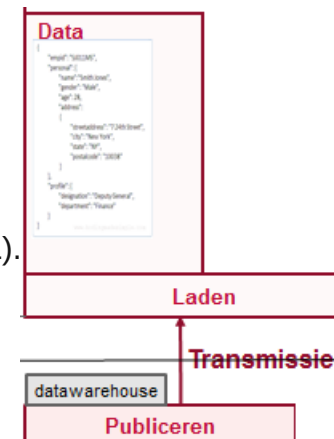
- De data wordt hier getransformeerd van de datastructuur en definities van de (interne) integratiedatabase naar een structuur en definities die optimaal aansluiten op de eisen van het Informatieportaal. Hiermee wordt voldaan aan BR-1 en BR-3.
- Het samenvoegen (de-normaliseren) van data draagt bij aan het kunnen aan voldoen BR-8.
- Het toepassen van het zelf beschrijvende overdracht formaat maakt het mogelijk om een grote variëteit, wijzigingen en dynamiek te absorberen, waardoor kan worden voldaan aan BR-5, BR-7 en ook aan BR-1 en BR-2.
- Wijzigingen in de te ontsluiten data of het publicatieformaat worden in deze publicatie opgelost, waar gebruik kan worden gemaakt van de ETL-tooling van het datawarehouse die hierin is gespecialiseerd. Dit verlaagt de beheerkosten en kosten van wijzigingen significant.

Implicatie:

- IN-07. Transformatie naar het formaat wat aansluit bij gebruikers eisen is een functie van Publiceren:
 - waar relevant translatie van kenmerknaam en/of inhoud;
 - omzetten naar de structuur die aansluit bij de gebruikerseisen van het informatieportaal.
- IN-08. De overdracht IA naar Publicatiedatabase vindt plaats via een zelf beschrijvend formaat (JSON of XML).
- IN-09. De data wordt met een laadfunctie geladen in de Publicatiedatabase.

Het informatieportaal moet voldoen aan een aantal business eisen:

- BR-1 Standaarden Bouwsector
- BR-2 Absorberen wijzigende standaarden
- BR-3 IMSpoor datadefinities
- BR-4 Huidige inhoud (constraint)
- BR-5 Wijzigingen volgen
- BR-6 Inconsistente data
- BR-7 Grote variëteit en dynamiek in kenmerken van de Objecten
- BR-8 Omvangrijke aantal datasets uit veel informatiesystemen



4.5 Publicatie database, data

noSQL database techniek

Om aan de eisen te voldoen wordt een database techniek ingezet die data op kan slaan waarbij:

- de datastructuur en de definities niet vooraf geconfigureerd hoeven te worden
- de datastructuur en definitie wordt overgenomen uit de aangeboden data die wordt aangeboden in een zelf beschrijvend formaat, zoals XML of JSON.

De noSQL databasetechniek van het type “Documents” ondersteunt dit.

Rationale:

- Wijzigingen in de data kunnen dynamisch worden gevolgd zonder aanpassingen in de software of database-inrichting waardoor wordt voldaan aan BR-2, BR-5 en BR-7. Iedere aangeboden dataset wordt verwerkt.
- In een noSQL database kan data gedenormaliseerd worden opgeslagen en gereproduceerd waarbij de structuur wordt ingebed in de gegevens-set die worden opgeslagen. Hierdoor kunnen gegevens die bij elkaar horen (bijv. Storing) als één dataset worden opgeslagen waardoor het aantal tabellen en het aantal daartussen te onderhouden relaties sterk terugloopt. Dit levert een grote bijdrage aan het voldoen aan BR-2, BR-5 en BR-8.
- Alle aangeboden data kan worden opgeslagen ongeacht de inhoud, wat bijdraagt aan BR-6.

Implicatie:

- IN-10. De te publiceren data wordt opgeslagen in een structuur die aansluiten op de eisen van de gebruikers in datasets waarin deze data als één set (gedenormaliseerd) wordt opgeslagen.
- IN-11 Om de gegevens aan elkaar te relateren moeten de kenmerknamen (attribuutnamen) worden gestandaardiseerd over alle datasets heen die worden gepubliceerd.
- IN-12 De te publiceren data moet worden aangeboden in een zelf-beschrijvend formaat (JSON, voorkeur of XML)
- IN-13 De publicatiedatabase kent geen vooraf gedefinieerd datamodel, doch zal wel voor de uitvoering van de informatieportaal taken wel een aantal ‘huishoudelijke’ aspecten hebben.

Het informatieportaal moet voldoen aan een aantal business eisen:

- BR-1 Standaarden Bouwsector
- BR-2 Absorberen wijzigende standaarden
- BR-3 IMSpoor datadefinities
- BR-4 Huidige inhoud (constraint)
- BR-5 Wijzigingen volgen
- BR-6 Inconsistente data
- BR-7 Grote variëteit en dynamiek in kenmerken van de Objecten
- BR-8 Omvangrijke aantal datasets uit veel informatiesystemen

Zie bijlage: noSQL voorbeelden voor nadere inzage een de noSQL dataopslag

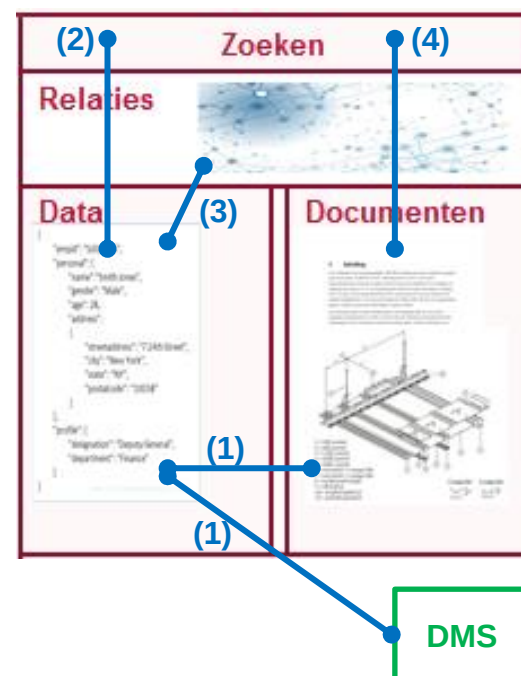
Zie bijlage: SQL versus noSQL voor de afwegingen noSQL

4.6 Publicatie documenten

Het informatieportaal moet documenten van verschillende aard kunnen opslaan en reproduceren: Tekstdocumenten, PDF, 2-D en 3-D tekeningen, Geodata, Foto's, Film, enz. Documenten worden ontsloten vanuit de Document Management Systemen, zo nodig via opslag in een Publicatie BLOB Storage.

De kenmerken (metadata) die bij de Documenten horen worden ontladen uit het Document Management Systeem en via het Datawarehouse als een Dataset geladen in de Publicatie database, met een verwijzing (1) naar het bijbehorende Document (of set documenten). Eventuele Geo-koppeling verloopt ook via deze datasets.

De documenten kunnen worden gevonden via het Zoeken op kenmerken (2), via de relaties (3) die zijn aangelegd met de data waar deze documenten bij horen of via een zoekindex van de inhoud (4) van de documenten.



Rationale:

- Documenten repliceren naar de Cloud vindt alleen plaats indien noodzakelijk (beperking in toegang of eisen om in de tekst van de documenten te zoeken)
- Door de kenmerken van de Documenten als Datasets te behandelen kan gebruik worden gemaakt van één mechanisme voor koppeling (relaties) van de data, documenten en geo.
- Voor zoeken in de inhoud wordt alleen gerepliceerd als het te ontsluiten bronsysteem op geen enkele wijze services in dit kader kan aanbieden.

Implicatie:

- IN-14 Documenten worden direct worden ontsloten uit het DMS onder de conditie dat dat DMS Single SignOn ondersteunt en benaderbaar is voor externe gebruikers.
- IN-15 Voor ieder document (of set documenten) wordt een dataset aangemaakt met kenmerken en een verwijzing naar het fysieke document
- IN-16 De document-dataset worden op dezelfde wijze als de datasets behandeld ten aanzien van zoeken, relaties leggen en geo-koppelingen.
- IN-17 Voor zoeken op inhoud van Documenten worden een (tekst)index opgebouwd.

4.7 Publicatie database, relaties

Semantische relaties

In de NoSQL database worden kaarten (datasets) opgeslagen. Deze gegevensgroepen hebben een relatie met elkaar, het informatieportaal zal de relaties tussen de gegevensgroepen inzichtelijk moeten maken en bovendien moeten navigeren over de relaties heen.

Om de te ontsluiten data en documenten aan elkaar te relateren wordt gebruik gemaakt van de semantische relaties zoals die ook worden toegepast in IMSpoor en de OTL.

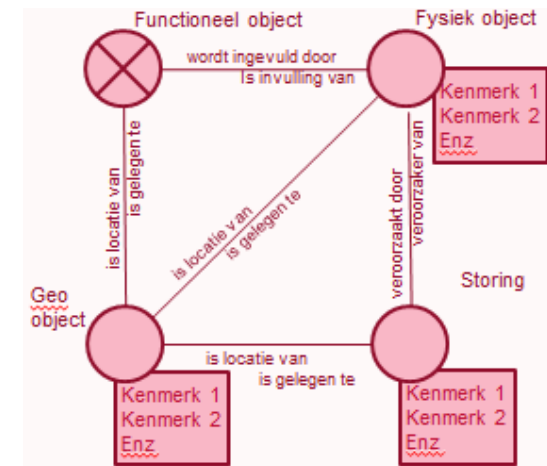
Een semantische relatie koppelt twee gegevensgroepen. ook wel nodes genoemd, en geeft deze koppeling een betekenis. Semantische relaties worden beschreven in de vorm subject-relatietype-object. Hierbij zijn subject en object nodes.

Rationale:

- Semantische relaties sluiten aan op IMSpoor (BR-3) en standaarden in de Bouwsector (BR-1).
- Door toepassing van semantische relaties worden alle relaties tussen datasets of documenten binnen één mechanisme ondergebracht waardoor het Informatieportaal relatief eenvoudig relaties van allerlei aard (data, geo, documenten, tekeningen enz) aan elkaar kan relateren en kan navigeren over veel datasets (BR-8).
- Wijzigingen in de data kunnen ook wijzigingen in relaties omvatten, omdat er één mechanisme is die alle relaties ondersteunt zullen wijzigingen weinig tot geen impact hebben op de datanavigatie, dit draagt bij aan BR-2. BR-5, BR-7.
- De semantische relatietechniek kan inconsistente data ondersteunen: verwijzingen naar andere gegevenssets die niet (meer) worden aangeleverd leiden tot een 'empty node'. De relatie wordt wel aangelegd met de aantekening dat de bijbehorende data nog niet is aangeleverd. Dit maakt het Informatieportaal resistent tegen Inconsistente data (BR-6) en a-synchrone aanlevering van data.

Het informatieportaal moet voldoen aan een aantal business eisen:

- BR-1 Standaarden Bouwsector
- BR-2 Absorberen wijzigende standaarden
- BR-3 IMSpoor datadefinities
- BR-4 Huidige inhoud (constraint)
- BR-5 Wijzigingen volgen
- BR-6 Inconsistente data
- BR-7 Grote variëteit en dynamiek in kenmerken van de Objecten
- BR-8 Omvangrijke aantal datasets uit veel informatiesystemen

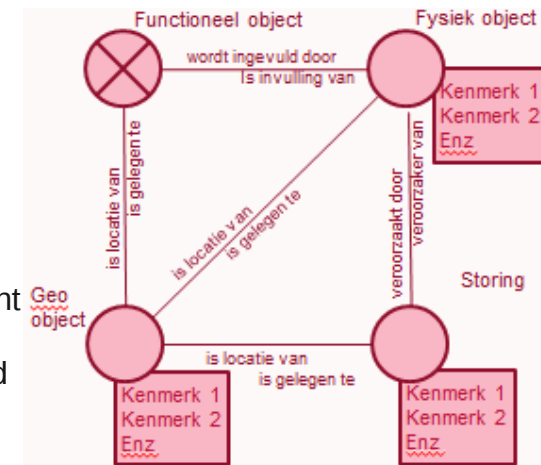


4.7 Publicatie database, relaties

Implicatie:

- IN-18 Alle relaties tussen datasets worden omgezet naar semantische relaties.
- IN-19 De relaties tussen de te beheren datasets en documenten zijn geconcentreerd in één mechanisme met nodes en relaties waardoor voor de navigatie geen applicatielogic meer nodig is.
- IN-20 De semantische relaties zoals gedefinieerd in IMSpoor en OTL zijn uitgangspunt waar nodig wordt de naamgeving naar de gebruiker aangepast waar nodig worden in de geest van IMSpoor en OTL nieuwe types gedefinieerd en ter toetsing overlegd bij IMSpoor.
- IN-21 In de Integratiedatabase zijn nog geen semantische relaties beschikbaar, het Informatieportaal bouwt zelf de semantische relaties op en onderhoudt deze.
- IN-22 Als wordt verwezen naar een Node waarvan (nog) geen data beschikbaar is dan wordt een lege node aangelegd met alleen de sleutel informatie.
- IN-23 Vanwege datakwaliteit problemen zijn lege nodes 'common business' waarmee het informatieportaal moet omgaan.

Voor de momenteel onderkende Features van het Informatieportaal volstaat een eenvoudige opslag en manipulatie van de Nodes en Semantische relaties. ProRail initieert momenteel binnen het BIM programma een onderzoek naar het gebruik van triple-stores (semantische relaties), mochten er in de toekomst zwaardere eisen aan het Informatieportaal worden gesteld (zoeken over veel relaties heen) dan is aansluiting op de uitkomst van dat onderzoek te overwegen.



4.8 Geo functies

Het informatieportaal ontstluit de kaartfuncties via Geopoort (3).

De Front-End (1) presenteert de kaarten en kaartlagen en navigeert binnen de kaart met behulp van de Esri SDK die direct met Geopoort is verbonden.

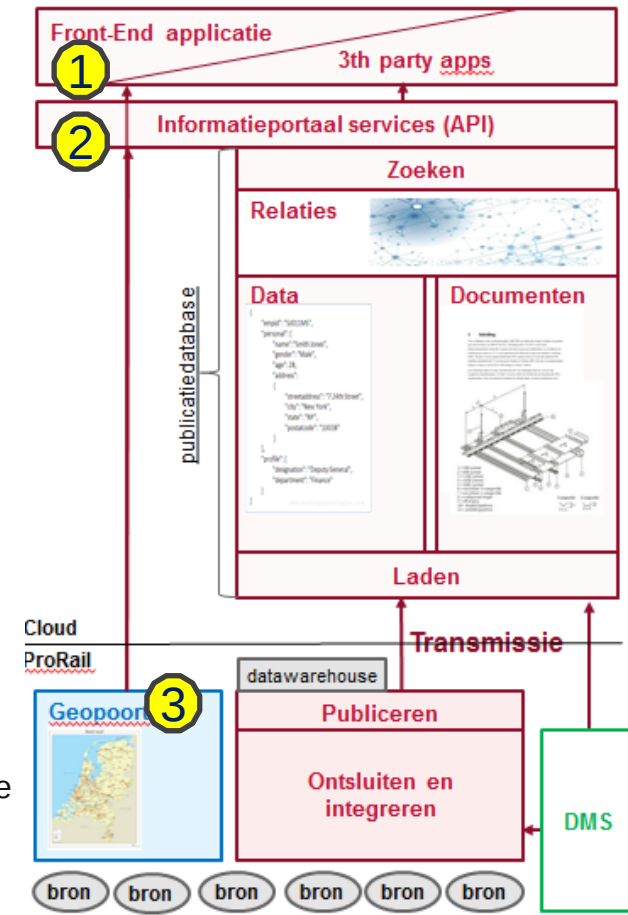
De logica die Geo-objecten koppelt met Data-objecten en Document objecten wordt geïmplementeerd in de Informatieportaal API (2) die daartoe ook met de Geopoort (3) communiceert.

Rationale:

- De Esri tooling in combinatie met Geopoort biedt de functionaliteit die nodig is voor het informatieportaal.
- De (complexere) logica om Geo met Data en Documenten te koppelen wordt op één plaats gerealiseerd en staat daardoor zowel ten dienste van het eigen portaal als derden die eigen applicaties willen realiseren.

Implicatie:

- IN-24 De logica die Geo koppelt met Data en Documenten wordt gerealiseerd in de Informatieportaal API.
- IN-25 Geopoort kent veel kaartlagen waarop één of meer objecttypes zijn getekend. De mapping van moet vanuit het perspectief van het informatie portaal flexibel zijn.



1. Zoeken op **vrije tekst in data**, met **facetgericht** zoeken / verfijnen
Wordt geïmplementeerd met een Search engine waarin de kenmerken
waarop vrije tekst zoeken van toepassing is worden geladen **(1)**.

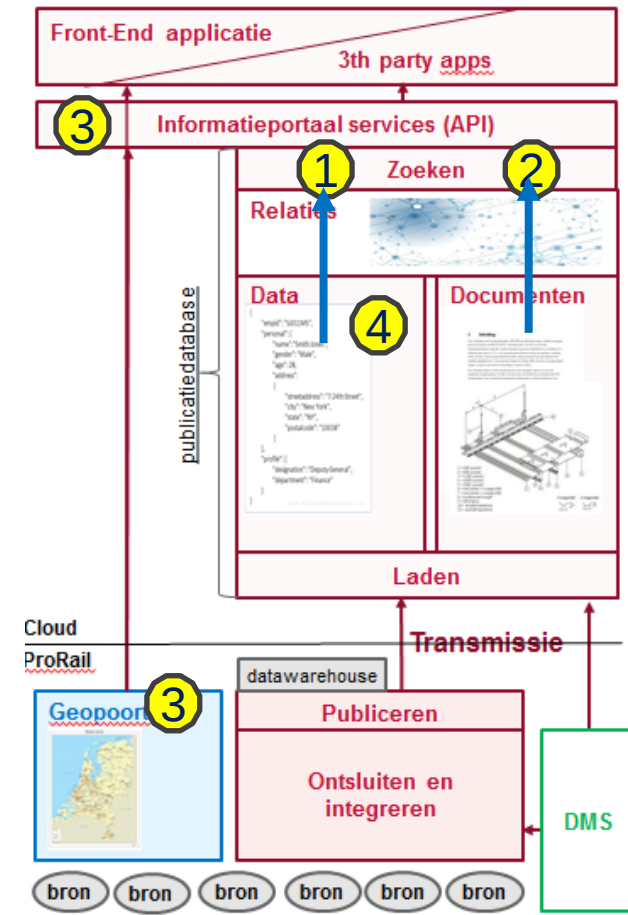
Wordt geïmplementeerd met een Search engine met een index waarin de inhoud van documenten wordt geïndexeerd **(2)**.

Wordt geïmplementeerd met de Mapservices van Geopoort **(3)** die met kaartlagen zoeken van objecten op de kaart ondersteunt. Additionele services in de Informatieportaal API koppelen vervolgens de gevonden objecten aan de data.

Wordt gerealiseerd door met zoekopdrachten op de noSQL database **(4)**, die als eigenschap heeft dat alle kolommen geïndexeerd zijn.

- De verschillende zoekmethodes bieden aan de (verschillende) gebruikersgroepen altijd een ingang die aansluit bij de behoefte.

- IN-26 Voor vrije tekst zoeken op kenmerken wordt een selectie van relevante kenmerken vastgesteld en geconfigureerd, waarbij de effectiviteit van dat kenmerk mee wordt gewogen.
- IN-27 In de user-interface worden de verschillende zoekmethodes aan de gebruiker vanuit een natuurlijke interactie aan de gebruiker aangeboden.



Opmerking: Indexeren en zoeken op de geïndexeerde inhoud van documenten is momenteel nog niet beschikbaar binnen Azure.

4.10 Front-end en API-services

Het informatieportaal wordt ontsloten via API-services, zowel naar de eigen Front-End applicatie als naar 3th party apps.

De API-services ontsluiten en de Front-End worden dusdanig opgezet dat zij de flexibele datastructuur ontsluit en de navigatie via de semantische relaties mogelijk maken.

De eigen Front-End applicatie wordt dusdanig opgezet dat data kan worden gepresenteerd ongeacht de inhoud. Er worden generieke 'frames' ontwikkeld om de data te tonen van de verschillende datasets, waarbij zowel de applicatiebeheerder als gebruikers kunnen aangeven welke data zij willen zien van welke datasets. De navigatie wordt gebaseerd op de semantische relaties.

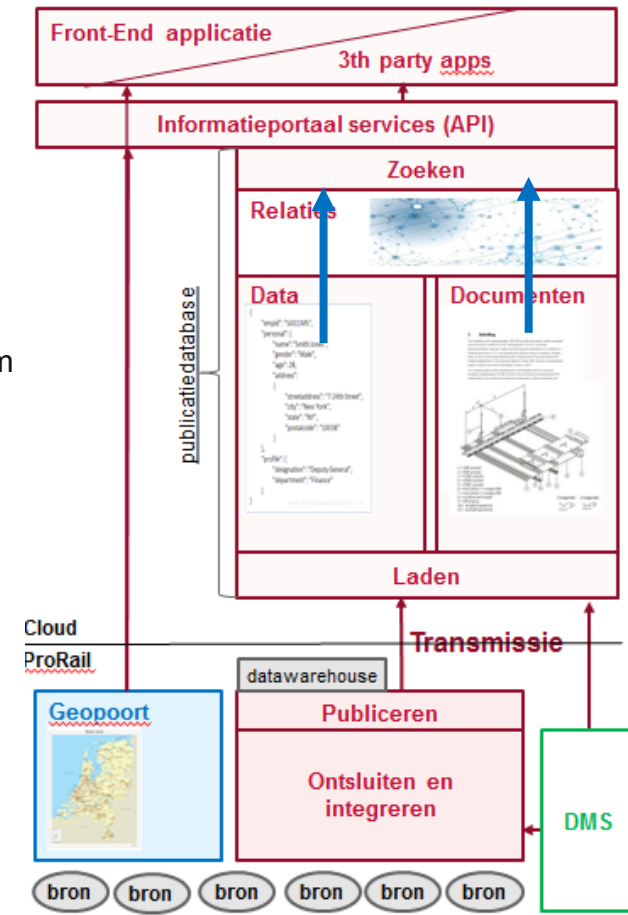
Rationale:

- De API-services moet de dynamiek in de data ondersteunen;
- Door de Front-End generieke en configureerbaar op te zetten kan de Front-End de dynamiek in de data en relaties (vergaand) volgen zonder aanpassingen in de software.

Implicatie:

IN-28 De API-services ontsluiten de data in een flexibel formaat (JSON)

IN-29 De Front-End wordt opgebouwd uit componenten die toepasbaar zijn voor de alle datasets en reageren dynamisch op de inhoud. De applicatiebeheerder kan de standaard instellingen configureren, de gebruikers kunnen eigen instellingen configureren.



4.11 Data bronnen

Onderstaand een opsomming van de te ontsluiten bronnen. Het betreffen zowel data als documenten, daarnaast zal het Informatieportaal ook Geografische functies moeten ontsluiten (tonen en zoeken op kaart).

Omschrijving	Bron (afkorting)	Nu al ontsloten via een DWH
Basis beheer kaarten	Geopoort (Railmaps)	Geopoort
Geodata (geografische locaties)	BBK, straks Naiade	
Configuratiedata Spoor (functioneel)	Infra Atlas, straks Naiade	
Objectdata (Fysieke object)	SAP-EAM Equipment	Geopoort
Storingen	SAP-EAM Storingen	Spoorkompas
Kabelgeul inhoudlijsten	GIL	Spoorkompas
Onderhoudsgegevens	SAP EAM (okt 2016)	
Inspectiedata	SAP, beschik datum onbekend	
Object-gerelateerde documenten	SAP EAM-PLM	
Bedrijfsvoorschriften over informatie uitwisseling, Contractering en Veiligheid, Railinfra generieke technische documenten.	RIC	
Spoor-beveiling: SWOD (OBE-, OS- en OR-bladen)	RailDocs. Let op: Sommige documenten hebben als bron SAP EAM-PLM	
Bedienvoorschriften Verkeersleidingposten (BVS en WVK bladen)		
Alignementen en Geometrische kenmerken spoornetwerk.	Sigma (momenteel nog PVS)	
Belasting Spoor	Quo Vadis	
Monitoringdata Spoor	BBMS	
Milieuvergunningen, plannormen geluid en EV	VBS3	Geopoort
Fysieke en logische informatie telecom	Telescope / Cocon	
V&G dossier	Geopoort	Geopoort
Historische documenten	Artiwin, nader onderzoeken	
Juridische documenten	JDS2	Geopoort
Preventief Onderhoud- en Storingsdiagnose	Possonline	
Wisselverwarming monitoring		
ATB Vv monitoring	ATB Vv monitoring	
Overige documenten	Scharepoint sites	

5.1 Architectuur richtlijnen

- **IS-1. Microsoft Azure**

Voor cloud diensten wordt eerst naar Microsoft Azure gekeken.

Rationale: Overeenkomstig uitgangspunt URA08.

- **IS-2. SaaS boven PaaS**

Bij het maken van nieuwe oplossingen geldt een voorkeur voor SaaS boven PaaS. IaaS wordt afgeraden.

Rationale: Dit is een enterprise-architectuur principe.

- **IS-3. Presentatie: NoSQL**

De presentatie laag gebruikt NoSQL als database.

Rationale: Zie 4.5

- **IS-4. Presentatie: DocumentDb**

De presentatie laag maakt gebruik van de NoSQL database DocumentDb, zolang er nog geen ProRail architectuur keuze is gemaakt voor een NoSQL database.

Rationale: Dit is een NoSQL SaaS oplossing in Microsoft Azure, wat voldoet aan IS-1, IS-2 en IS-3.

Deze database is snel toe te passen, en later relatief eenvoudig te vervangen.

- **IS-5. Umbraco**

Als Content Management Systeem (CMS) wordt Umbraco gebruikt.

Rationale: Dit is een SaaS CMS in Microsoft Azure, wat voldoet aan IS-1 en IS-2.

- **IS-6. Loosly-coupled SOA**

- Rationale: Voor Integraal Publiceren is gekozen voor het *loosely coupled* (SOA) architectuurprincipe.

Dit wordt gerealiseerd met een ETL-mechanisme met maximale ontkoppeling naar de Cloud.

- **IS-7. Horizontaal en Verticaal schaalbaar**

Alle componenten worden zo ontwikkeld dat zo zowel horizontaal als verticaal schaalbaar zijn.

5.2 Informatiesysteem architectuur

Webbrowser: Een gebruiker benadert de informatieportaal website via een browser. In de JavaScript code van de website wordt verwezen naar de publicatie API, de Geo JS SDK en de presentatie API.

Presentatie API: Ontsluit de niet-publicatie gegevens van de website, zoals het CMS en de presentatie database.

Presentatie database: Bevat overige gegevens, zoals gebruikersvoorkeuren / instellingen.

CMS: Bevat de statische content van de website.

Autorisatie API: Hierin wordt geconfigureerd welke rollen van welke organisaties toegang hebben tot welke gegevens.

IPM API: Bevat services voor het opvragen van data uit de publicatiedatabase, en het zoeken hiernaar. Voert ook autorisatietaken uit: afschermen toegang data. Zowel B2B services als de webbrowser maken hier gebruik van.

Search: Geïndexeerd zoeken in de publicatiedatabase en in de bestanden (blob) met ondersteuning van facetten.

Files (blob): Documenten, tekeningen, beeldmateriaal die moet worden ontsloten.

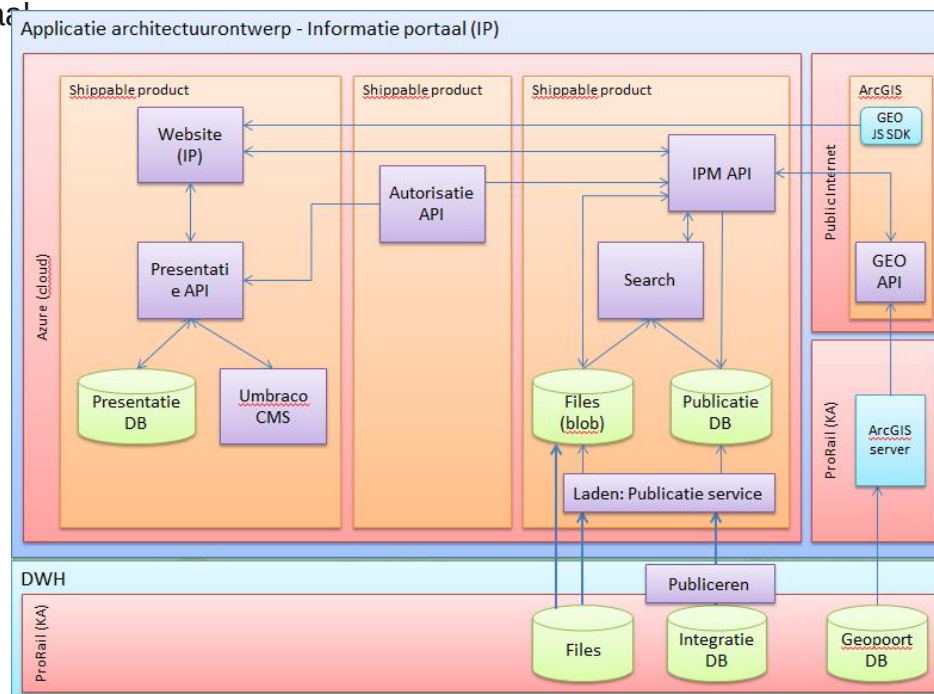
Publicatie DB: Bevat de gegevens die moeten worden ontsloten.

Laden: Publicatie service: Het publicatie-component ontvangt gegevens van de integratielaag, en schrijft deze gegevens weg in de publicatiedatabase, maakt tevens de semantische relaties aan.

Publiceren: Conversie van de datastructuur van de Integratielaag naar de datastructuur van de Publicatiedatabase.

Integratie DB: Datawarehouse waarin de gegevens uit de bronnen worden samengevoegd en geïntegreerd.

GEO API (Geopoort): Ontsluit de kaarten, geïmplementeerd op de **ArcGIS Server** die gebruik maakt van de **Geopoort DB**.



5.3 ICT applicatie componenten

Ref_ID	Applicatie component	Product	Beschrijving / functionaliteit	Type	Aantekening
ACM01	WebBrowser		Standaard webbrowser om HTML5 websites weer te geven	Niet van toepassing	
ACM02	Front End	Angular JS	JavaScript	Bestaand	
ACM03	CMS		Umbraco, Content Management System	Nieuw	SaaS
ACM04	Presentatie API IPM API Autorisatie API Laden: Publicatie service	Microsoft Visual Studio	Informatie portaal logica, geschreven in C#	Bestaand	
ACM05	Publicatie database Presentatie database	Documentdb	NoSQL database van het Document-type	Nieuw	SaaS
ACM06	Integreren Publiceren	Informatica ETL-tool	Extractie, Transitie, Laden voor zowel ontsluiten bronnen tm Integratie database als publiceren naar de Publicatiedatabase	Bestaand	ccBI
ACM07	Integratie database	SQL Server	Datawarehouse, relationele database.	Bestaand	ccBI
ACM08	Zoeken	Azure Search	Zoekmachine, vrije tekst zoeken met Facetten Geïndexeerd zoeken in blob en in de publicatiedatabase	Nieuw	SaaS
ACM09	Documenten	Azure BLOB	Document storage	Nieuw	SaaS
ACM10	Geo	Geo JS SDK Geo API ArcGIS Server	Presentatie en navigatie Esri kaartservices Geopoort Kaart (Geo) Services	Bestaand	
ACM11	VPN connectie	ExpressRoute	Veilige verbinding met dedicated bandbreedte tussen Azure Cloud en ProRail		
ACM12	Cloud	Microsoft Azure	Cloud services voor OTAP-omgevingen	Gewijzigd	PaaS
ACM13	Autorisatie	2FA	Twee-factor authenticatie	Bestaand	
ACM14	Autorisatie	AD	Microsoft Active Directory	Bestaand	
ACM15	Autorisatie	SSO	Single Sign-On	Bestaand	
ACM16	Autorisatie	AAD	Microsoft Azure AD	Bestaand	

6.1 ICT architectuur

Cloud

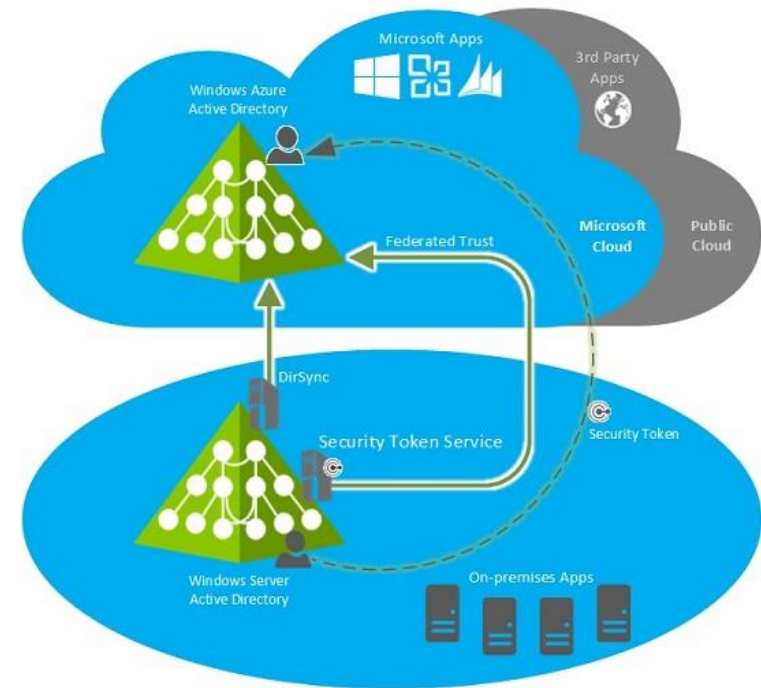
Een aantal generieke hoofdcomponenten bestaat al in de public Azure Cloud infrastructuur die door ProRail wordt afgenomen bij Microsoft.

Netwerk

De Azure diensten worden in dezelfde context worden ontsloten als SharePoint Online opgeleverd door het programma Fun2Share, zal de netwerkinrichting gebruikmaken van door Fun2Share opgeleverde netwerkconstructies en –beveiligingen : ExpressRoute.

Databases

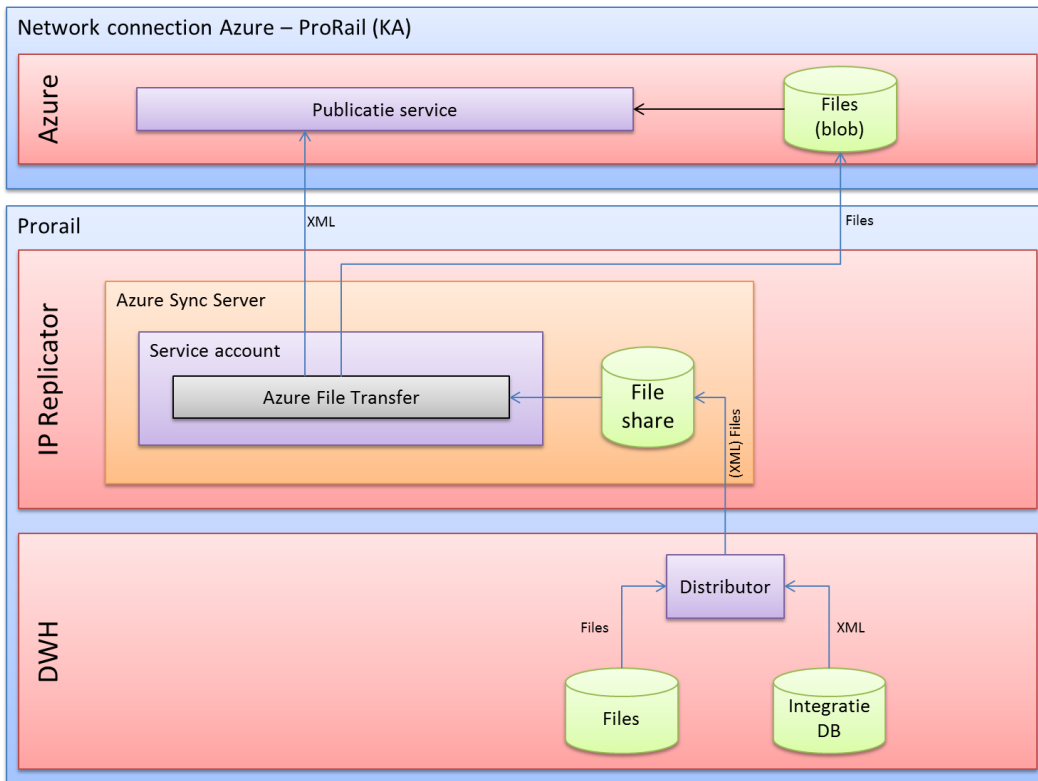
Er zijn verschillende types databases nodig.
Navolgende tabel somt op, op welk niveau de verschillende databases een rol spelen in de oplossing.



Database	Type	Noot
Integratiedatabase(s)	Relationeel	ccBI
Publicatiedatabase(s)	NoSQL, Documentdb	SaaS
Files: Integratie-files	Bestandssysteem	Standaard infrastructuur
Files: Publicatie-BLOB-opslag(en)	Azure BLOB storage	SaaS
CMS	SaaS, type is niet relevant	SaaS

Netwerkkoppeling Azure / ProRail

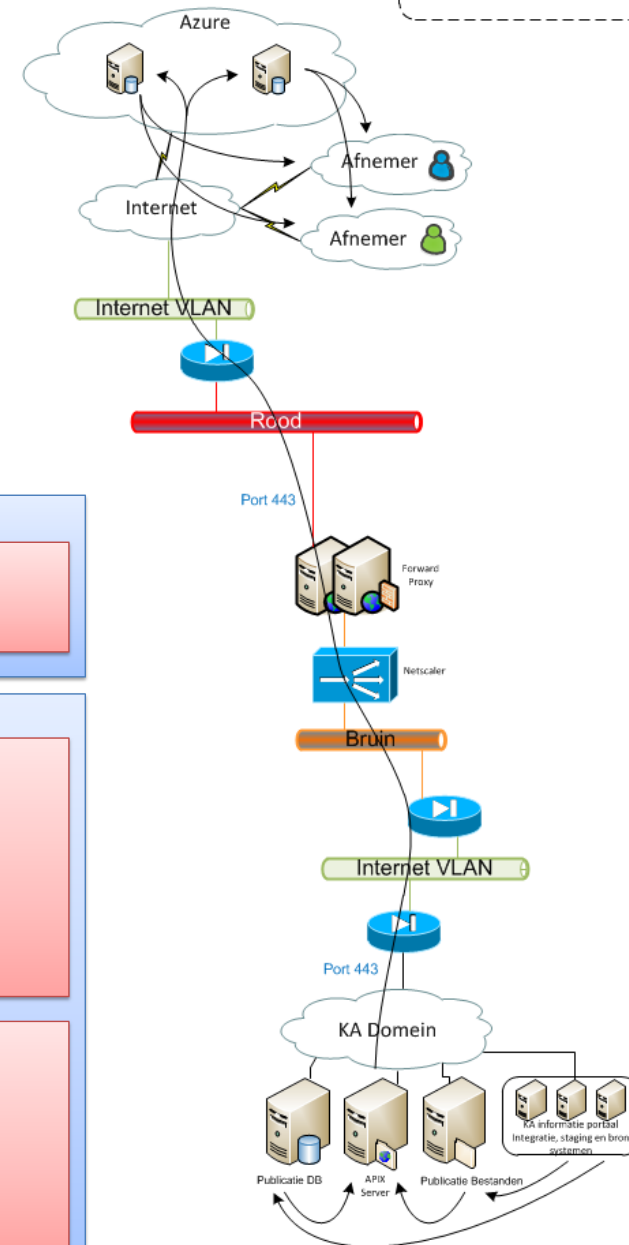
De afbeeldingen hiernaast en hieronder geven de koppeling tussen ProRail en Azure weer.



Verkeersmatrix

TMG -> Azure: tcp/443 (https uitgaand)

Informatielaag -> TMG: tcp/443 (https uitgaand)



6.2 ICT architectuur richtlijnen

In principe gelden alle ProRail richtlijnen voor een dergelijk project, zoals zowel SAP als Microsoft tenzij.

Verder gelden de volgende richtlijnen conform het ProRail Cloud beleid:

- De Cloud leverancier voldoet aan de ProRail eisen qua beveiliging en intellectueel eigendom.
- Een Cloud dienst biedt alleen extra functionaliteit.
- De koppeling van een Cloud dienst met de ProRail omgeving verloopt via een ESB koppeling, tenzij....
- De Cloud dienst moet passen in de ProRail architectuur (bijv. applicatie en ICT Infra architectuur)
- De authenticatie van gebruikers van de Cloud dienst maakt gebruik van ADFS
- De Cloud dienst is gebaseerd op internationale standaards en is voorbereid op een roll-back
- De Cloud dienst voldoet aan de gestelde beschikbaarheidseisen
- De Cloud dienst hanteert een helder back-up en restore strategie

6.3 Project overstijgende architectuurkeuzen

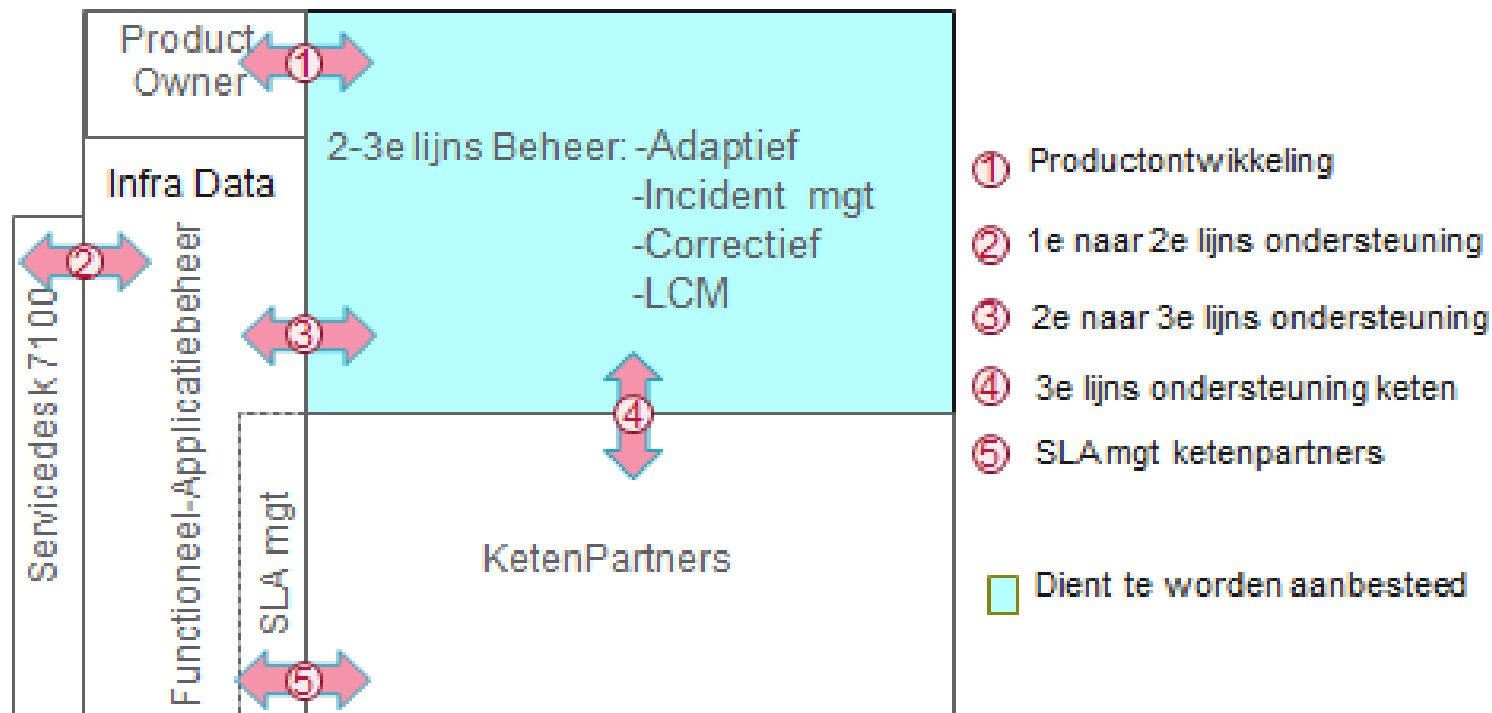
ExpressRoute

Er is gekozen voor ExpressRoute voor de communicatie van de ProRail omgeving naar Azure. Dit is een oplossing die door meerdere projecten wordt gedeeld.

7.1 Inrichting beheer organisatie

De Azure Cloud omgeving is nieuw voor ProRail. Ondergelegen is een denkkader opgenomen voor de inrichting van het beheer, met daarbij potentiële 'eigenaren', dit is momenteel (19-8) onderwerp van afstemming in het project.

Maatregel: in het project zelf is werkpakket opgenomen om dit aspect verder uit te werken.



7.2 Informatiebeveiliging

De Website dient uitsluitend de laagste mogelijke BIV-codering (LLL) te hebben.

Voor iedere te ontsluiten bron wordt met de eigenaar dit aspect afgestemd welke informatie mag worden ontsloten, dit wordt vastgelegd in een gedocumenteerd besluit en ter accordering nader afgestemd met de security officer.

In het Informatieportaal worden wel voorzieningen opgenomen om de informatie af te schermen voor bepaalde rollen en/of organisaties.

Toegang tot het informatieportaal wordt alleen gegeven aan geautoriseerde gebruikers die over een ProRail-account beschikken, danwel over een account beschikken van een van de partners van ProRail waar federatie mee toegestaan is. Authenticatie verloopt altijd via een ADFS server.

Conform ProRail-beveiligingsbeleid en -kaders is Twee-Factor Authenticatie (2FA) vereist. Indien gebruik van ADFS niet mogelijk is, dan moet ontheffing bij het Coördinatie Overleg Informatie Voorziening (COIV) worden aangevraagd.

Beveiligingsmechanisme	Integraal Publiceren Spoordata
Organisatie-accounts Azure Cloud gebruikers	Ja
Gefedereerde gebruikers van organisaties op basis van ADFS	Ja
EX-accounts	Ja
Third party identity providers	Nee

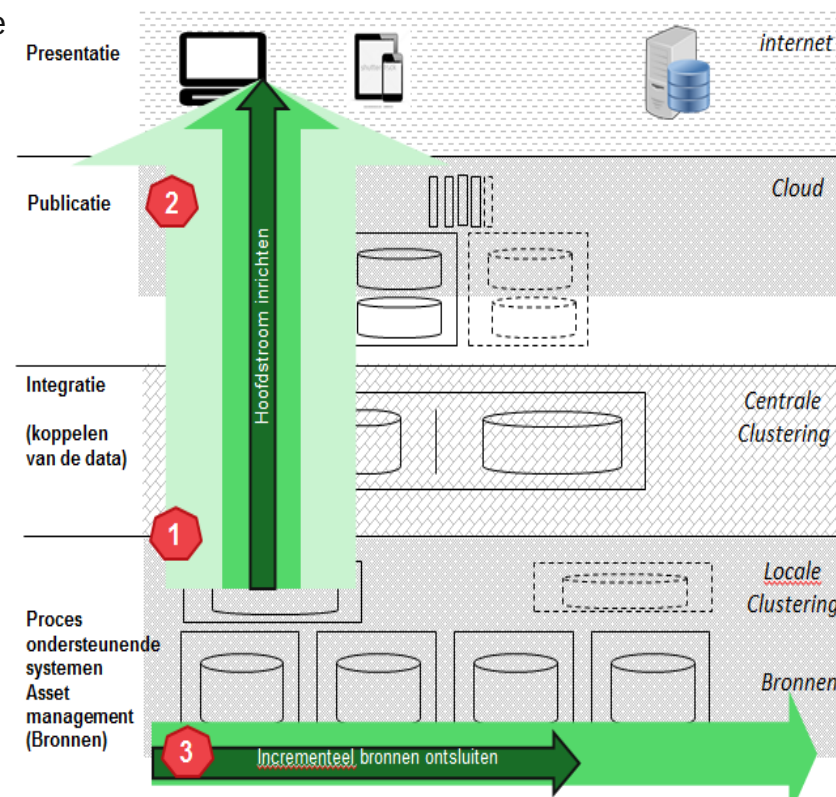
Een nieuw project moet zich gaan richten hoe externen toegang te geven tot de ProRail tenant, hierbij zal Twee-Factor Authenticatie (2FA) een vereiste blijven op basis van geldend ProRail beveiligingsbeleid en –kaders.

7.3 Migratie

De ontsluiting van de data vindt plaats via het datawarehouse waarbinnen de datastructuur en de inhoud wordt geüniformeerd en de data uit de verschillende applicaties wordt geïntegreerd. Vanuit de Integratedatabase wordt de Informatie vervolgens ontsloten.

De Migratie kent 3 stromen:

1. De inrichting van de gegevensontsluiting vanuit de Proces ondersteunende systemen naar het datawarehouse.
Binnen het datawarehouse wordt de Integratie uitgevoerd:
 - data uniformeren naar de business definities.
 - koppelen van de databronnen.
2. Het informatieportaal wordt ingericht:
 - inrichting van een publicatiedatabase waar de gegevens worden klaargezet zoals de doelgroepen deze nodig hebben.
 - zoeken, presenteren en 'downloaden' van de data.
3. Incrementeel meer bronnen integreren en ontsluiten.



8. Afwijkingen Referentiearchitectuur

De Solution architectuur wijkt voor een aantal aspecten af van de in voorgaande fasen gedefinieerde referentiearchitectuur

1. Geografische functies

Voor de geografische functionaliteit wordt gebruik gemaakt van Geopoort die vanuit de Front-End applicatie ten aanzien van het aspect presenteren en navigeren wordt aangesproken met de Esri SDK. Deze SDK werkt direct op de Esri Geopoort server en niet via de Publicatie API zoals was voorzien in de Referentie architectuurpublicatie database.

In de Informatieportaal API worden, conform de referentiearchitectuur, de services geïmplementeerd waarin de intelligentie is opgesloten om de koppelingen te leggen tussen de kaarten (Geo), Data en Documenten. Deze API spreekt voor deze taken ook de Esri Geopoort services aan.

2. Dataopslag, Publicatie database

Voor de dataopslag wordt een NoSQL database gebruikt: Documentdb, onderdeel van de Azure productsuite. De datasets in deze noSQL database worden aan elkaar gerelateerd met Semantische relaties.

Een aantal kanttekeningen:

- Documentdb heeft spatial functionaliteit die onvoldoende precies is, voor de uitvoering van de kerntaken van ProRail is een hogere precisie nodig. Voor de taken van het informatieportaal is de precisie van Documentdb werkbaar, bovendien geldt dat het informatieportaal voor Geo-functies primair gebruikt maakt van GeoPoort.
- ProRail overweegt een keuze te maken voor een noSQL die voldoet aan alle eisen. Dit kan leiden tot de noodzaak om Documentdb om te wisselen naar de dan gemaakte keuze.
Voor het informatieportaal geldt dat bij een eventuele overstap de 'technical debt' vele malen kleiner zal zijn vanuit Documentdb dan vanuit SQL.