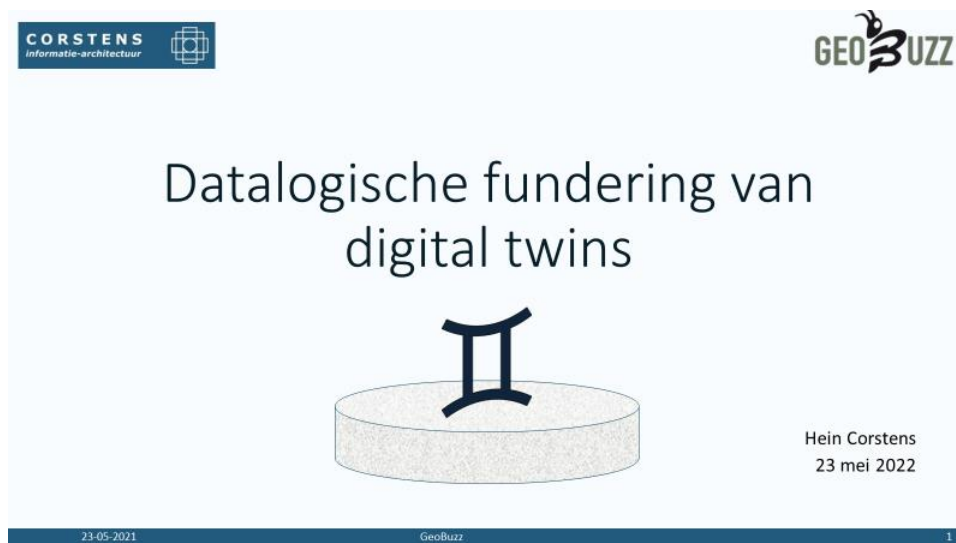


GeoBuzz 2022 23-05-2022

Hein Corstens:

Datalogische fundering van digital twins

1. Opening



(♊ staat voor Tweelingen (Gemini) (sterrenbeeld), genoemd naar de onsterfelijk tweeling Castor en Pollux, kinderen van Zeus en Leda, en nu twee sterren in Gemini.)



Hein Corstens



- onafhankelijk adviseur
- bouw- en geo-informatie
- i.h.b. gegevensmanagement en gegevensmodellering
- v/h oprichter/directeur URBIDATA en UDS
- nu Corstens informatiearchitectuur: www.corstens.nl
- contact: hein@corstens.nl (06)55382288

23-05-2021

GeoBbuzz

2



Datalogische fundering van digital twinning



- I. Digital twin: wat en waarom?
- II. Datalogische eisen aan digital twins
- III. Implementatievereisten voor digital twins
- IV. Conclusies

23-05-2021

GeoBbuzz

3



1. Digital twin: wat en waarom?

23-05-2021

GeoBbuzz

4

I. Digital twin: wat en waarom?

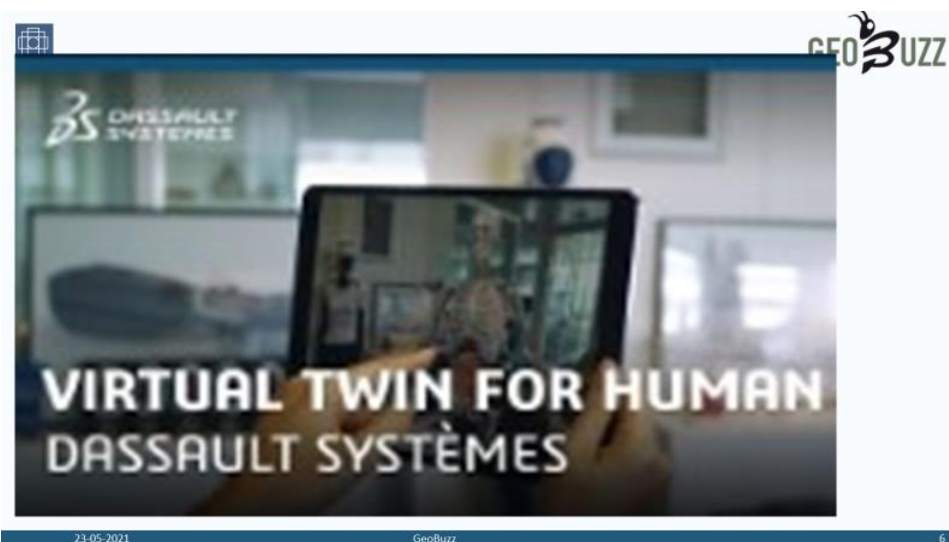


Ter introductie laat ik eerst een aantal voorbeelden van digital twins zien:

- industrie

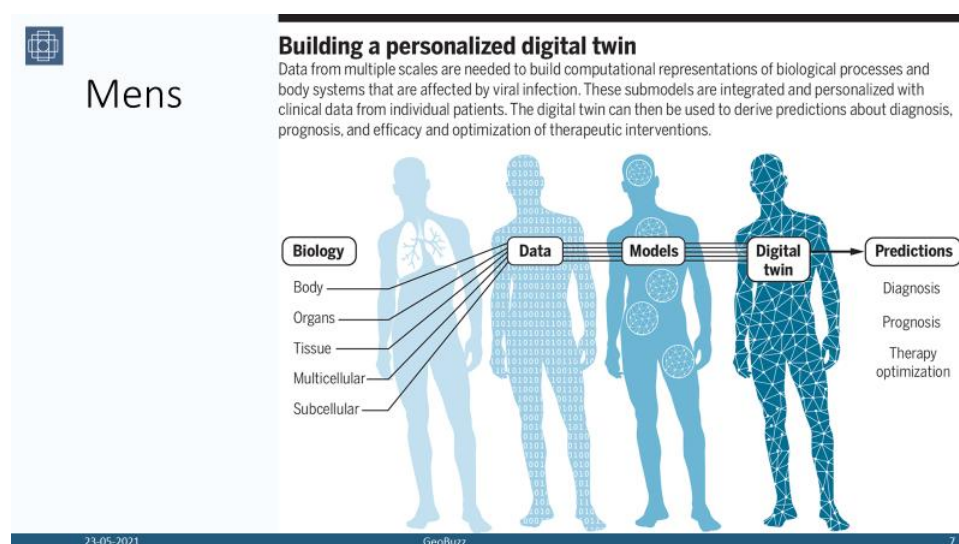


1:36



2:25

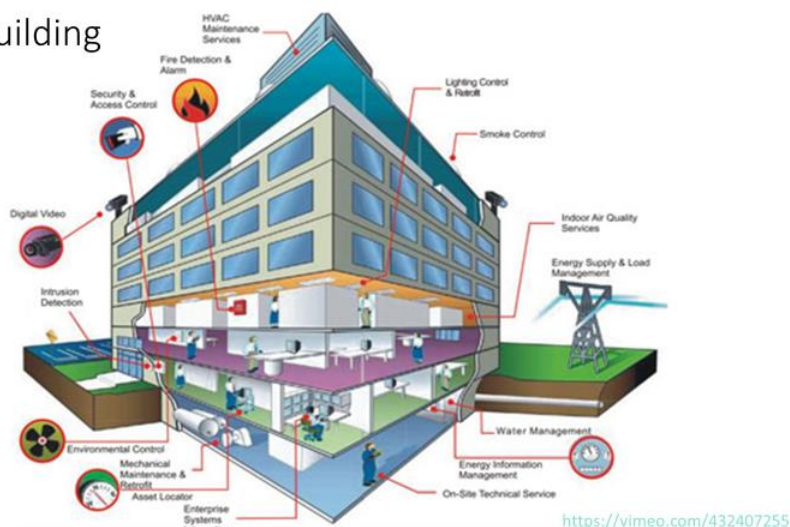
- medische sector



Als deze ontwikkeling doorzet, zie ik een ontwikkeling, waarbij een digital twin de basis is van het medische dossier: een soort 'binnenstebuitenkering', niet een dossier met verhalende tekst en eventueel een plaatje erbij, nee vanuit de samenstelling van je lichaam uit organen en systemen wordt een model opgebouwd, waarin de informatie geïntegreerd is en waaraan allerlei informatie gekoppeld kan worden.

- bouw

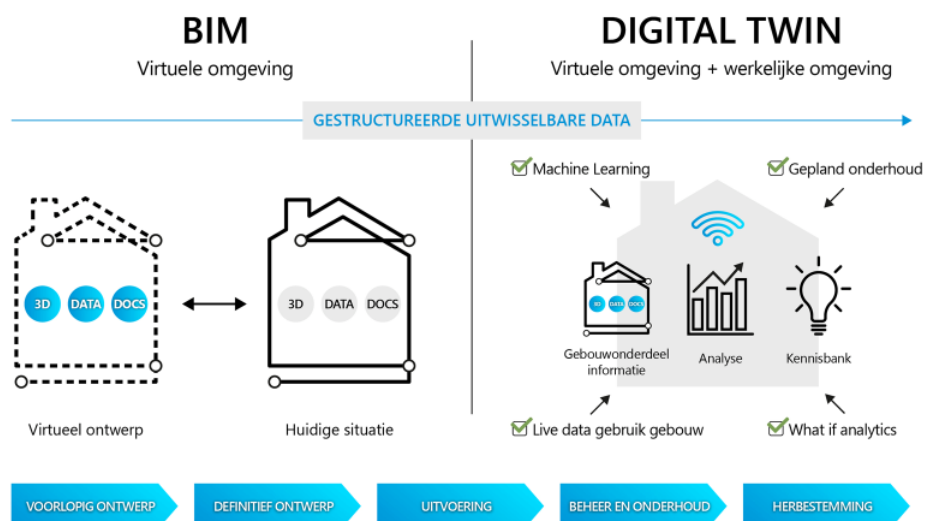
Smart building



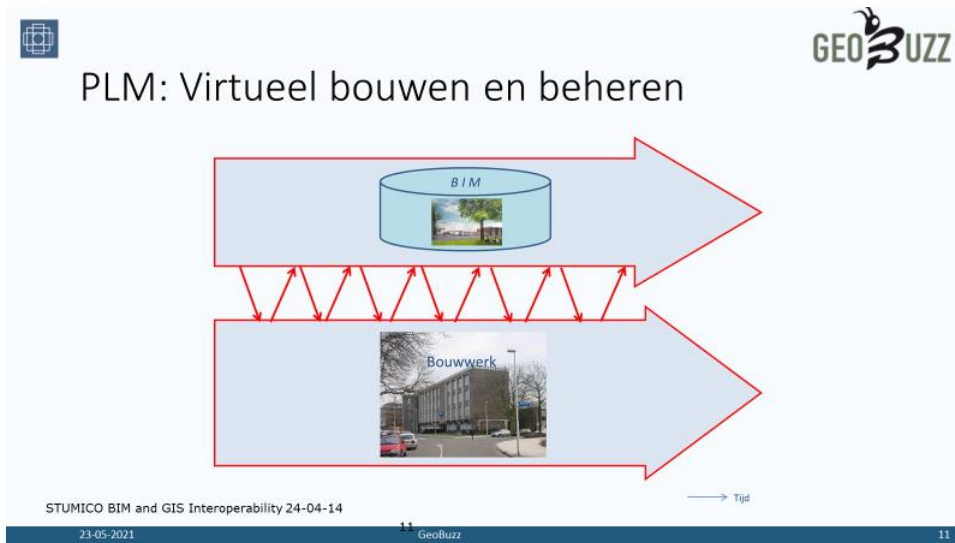
1:42

Gebouwen

Er worden steeds meer 'smart buildings' gerealiseerd, gericht op onder meer een beheersing van de milieubelasting en de operationele kosten. Een digital twin van een gebouw is een virtuele replica van de fysieke asset en die verzamelt en organiseert allerlei data over het gebouw, zijn gedrag, het gebruik, enzovoorts. Gericht op beter functioneren. Digital twinning is gestart bij de NASA, is in gebruik genomen in de industrie, de medische wereld, enzovoorts en is nu aangeland in de bouwwereld. Continu worden gegevens verzameld en gekoppeld aan een digitale kopie van een gebouw.



Uiteraard heeft de bouwwereld niet gewacht totdat de ruimtevaart en de industrie zover waren, immers al waren ze dan geen voorlopers, al geruime tijd wordt er in de bouw met BIM gewerkt, met een bouwwerkinformatiemodel¹.



In wezen zie ik geen essentieel verschil tussen een BIM en een digital twin: een goed BIM was al geschikt voor het volgen van een gebouw gedurende de gehele levenscyclus en niet alleen op statische, maar ook op dynamische aspecten gericht (zoals bijvoorbeeld verzakking, veroudering, enz. Voor een deel is digital twinning dus oude wijn in nieuwe zakken! Dit plaatje stamt uit 2014 en is gericht op Product Lifecycle Management: tijdens de hele levensloop van een gebouw houd je alle relevante informatie in een BIM bij.

(En ik zie de begripsinflatie al weer aankomen: het begrip BIM is in de loop van de tijd van bouwwerkinformatieMODEL via bouwwerkinformatie MODELLERING verwaterd tot bouwwerkinformatieMANAGEMENT. (Het zou me niet verbazen als, mede als gevolg van mislukkingen door een verkeerde, niet echt op het model gerichte, aanpak, hetzelfde met digital twin gaat gebeuren.))

- stad.

Gebouwen en gebieden

¹ een verkeerde term overigens: het is een model van een bouwwerk, dus een bouwwerkmodel en geen model van de informatie over het bouwwerk, dus geen informatiemodel. Wat dat betreft deed de industrie het beter met de term productmodel.



Smart city



<https://youtu.be/soAjc8caTRw>

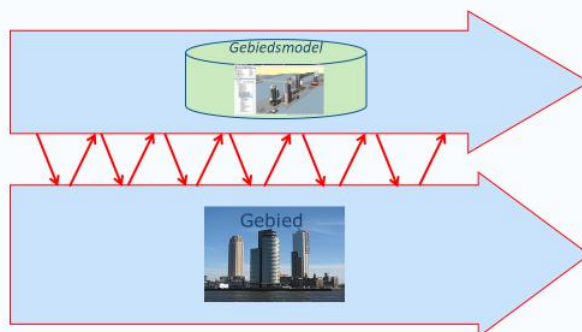
23-05-2021

GeoBuzz

8



Virtueel ontwikkelen en beheren



STUMICO BIM and GIS Interoperability 24-04-14

→ Tijd

23-05-2021

27 GeoBuzz

27



Nederlands voorbeelden

- Omgevingsserver; zie video: <https://www.spotinfo.nl/wp/wp-content/uploads/2021/02/SPOTinfo-managed-dataplatform.mp4>
- 3D-stadsmodellen Rotterdam, Den Haag, Groningen, Eindhoven, enz.
- Schiphol, Havenbedrijf Rotterdam
- en meer

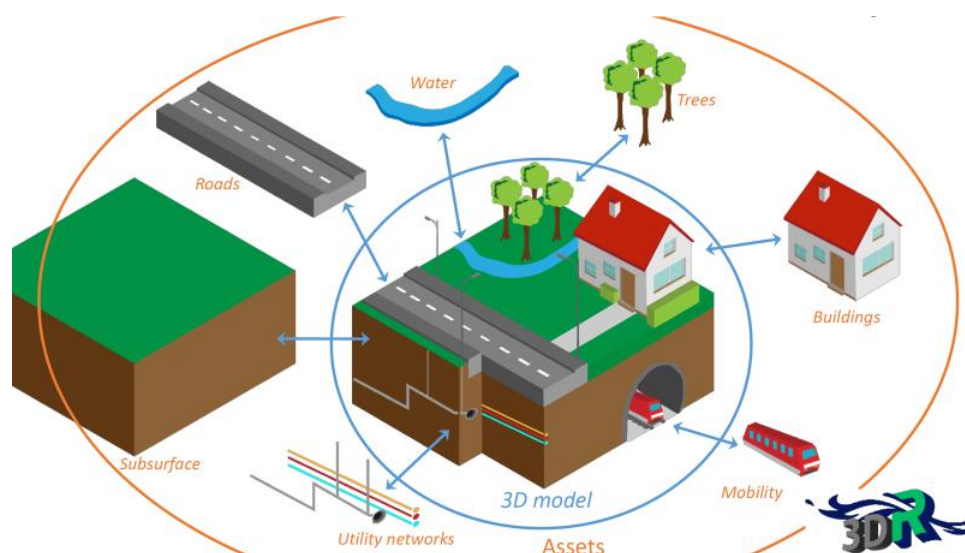
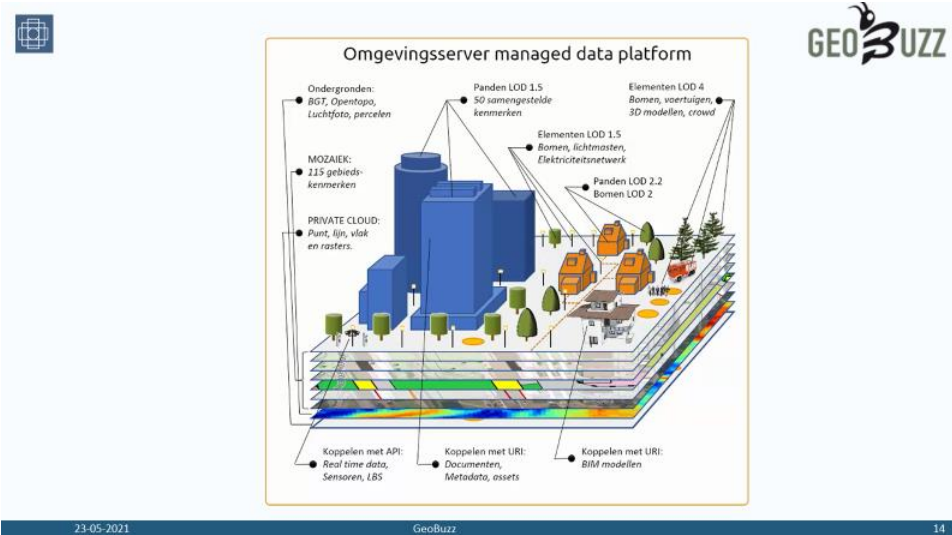
23-05-2021

GeoBbuzz

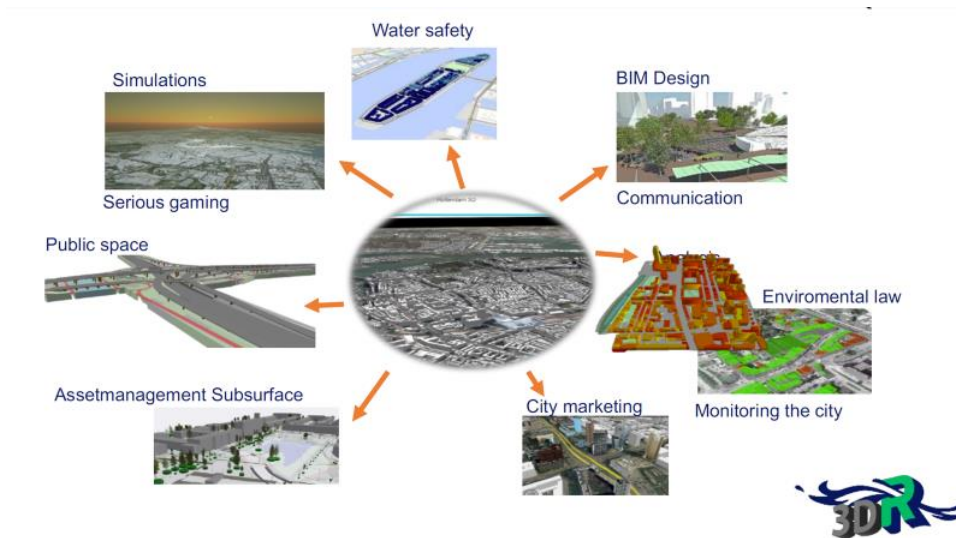
13

Voorbeelden NL

- Rotterdam
 - Den Haag (Stadsmodel, Living Lab)
 - Amsterdam
 - Eindhoven
 - Utrecht
 - Groningen
- en meer

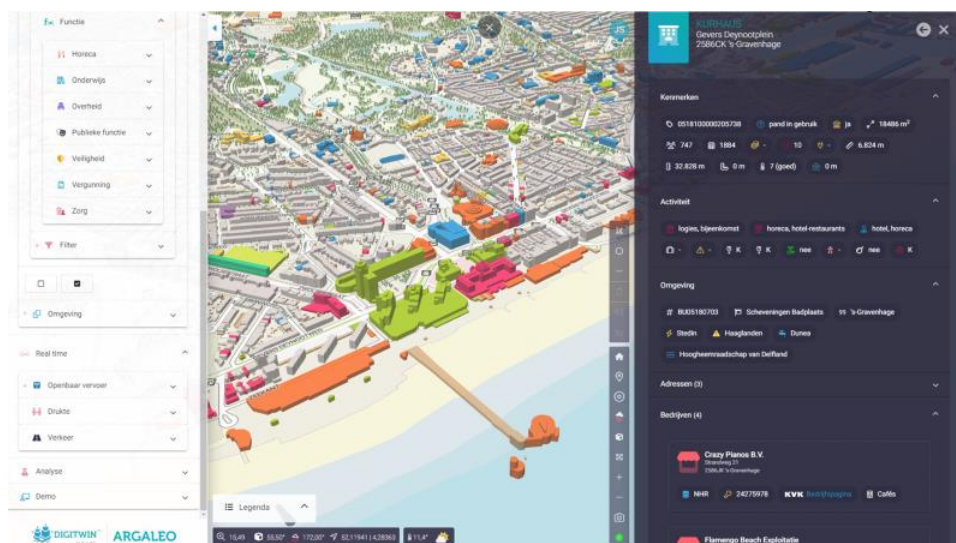


Ik ben zelf betrokken bij T3D – Totaaldriedimensionaal, een programma van de gemeenten Amsterdam, Rotterdam, Den Haag en de VNG, gericht op daadwerkelijke implementaties van de inwinning, registratie en beschikbaarstelling van driedimensionale objecten vanuit een 3D-stadsmodel. Plaatjes van Jane Hermans-van Ree.



(zie ook <https://w3s.nl/ons-werk/digitale-stad-rotterdam/>)

(Rotterdam/ Roland van der Heijden)



Het model van Den Haag wordt daadwerkelijk gebruikt bij allerlei experimenten in een 'Living Lab' in Scheveningen (het heeft de World Smart City Award 2021 gewonnen). Daarin wordt het model gekoppeld aan sensorgegevens, gegevens uit social media en meer, over drukte, parkeren, sentimenten, het weerbericht en ga maar door.

Welnu, wat zijn volgens jullie de essenties?

.....



Essenties van een digital twin

23-05-2021

GeoBbuzz

10



Digital twin: wat?

1. Een dt is een **model**
2. Een dt is een **dynamisch** model
3. Een dt is een model met **terugkoppeling**
4. Een dt is **integraal**

23-05-2021

GeoBbuzz

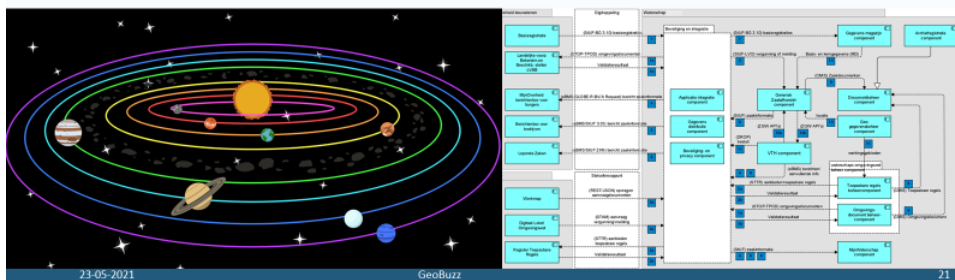
19

Wel, naar mijn idee gaat het over een viertal zaken:



Systeem

- Verzameling objecten die door hun onderlinge relaties een geheel vormen



23-05-2021

GeoBbuzz

21

1. Een Digital twin is een **model** van in het algemeen reëel systeem. Een model is een systeem dat gebruikt wordt om kennis te verkrijgen over een (ander) systeem. Dit door het hebben van een gelijkvormige structuur.

Voorbeelden van modellen

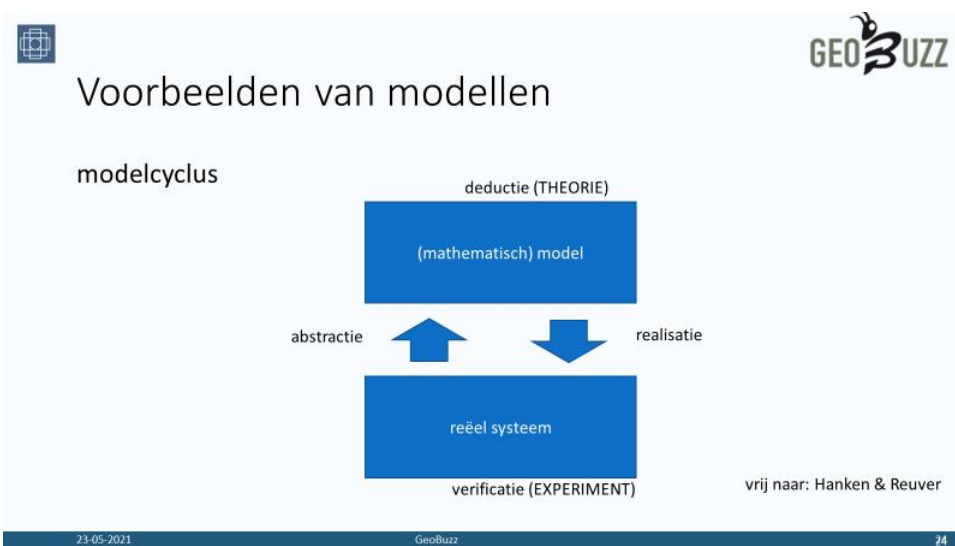
Maquette
transformatie
door
verscaling,
materiaal-
afbeelding, enz.



23-05-2021 GeoBuzz 25

Bijvoorbeeld een maquette.

Modellen zijn van belang in wetenschappelijk onderzoek. Je ontwikkelt een theoretisch model, waaruit hypothesen afgeleid worden, welke vervolgens getoetst worden in experimenten in een reëel systeem. Vervolgens wordt de theorie zo nodig bijgesteld of vervangen (wetenschappelijke revolutie).





Voorbeeld

Model: Derde wet van Newton: Actie = - Reactie

Systeem:



120

Opmerkelijk: zie je dat in dit geval het reële systeem een model is van de theorie? En dat dus niet altijd het model abstracter is dan het afgebeelde systeem?



Dit komt meer voor!

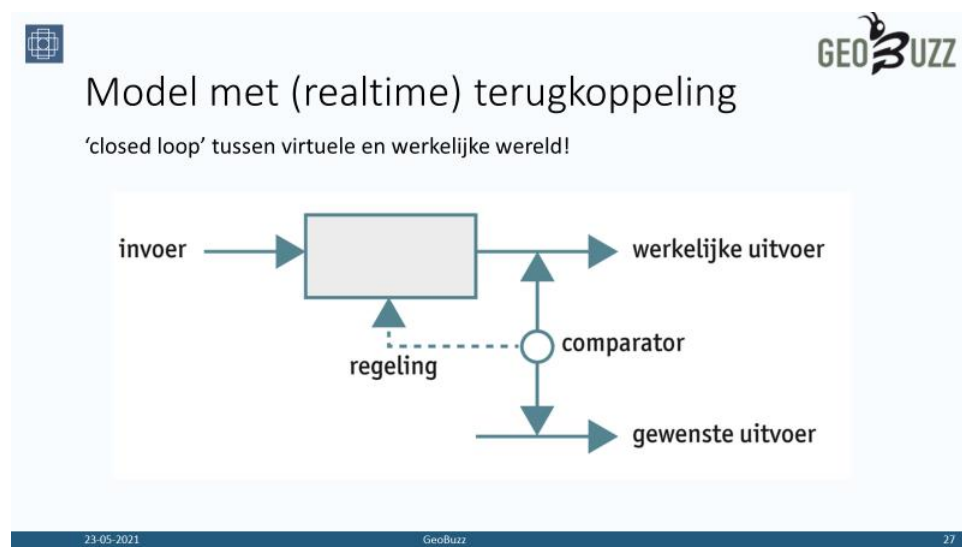
Een model is ook een VOORBEELD! In dit geval van een modetrend.

Samenvattend: een digital twin is een digitaal model van een systeem, dat een fysiek systeem afbeeldt. Voorbeelden zijn: een vliegtuig, een machine, een plant, een mens, de aarde, een gebouw, een stad.

2. Een dt is een **dynamisch** model.

In een dt is mijns inziens het tijdsaspect onmisbaar. Het kan zelfs in ruimtelijke zin tweedimensionaal zijn, maar zonder tijdsaspect is het niet echt bruikbaar als instrument voor analyse, voorspelling,

simulatie en dat soort dingen. Zie het voorbeeld (ook al uit 2014), waarin een gebied in de tijd gevolgd en bijgestuurd wordt met behulp van een digitaal model.



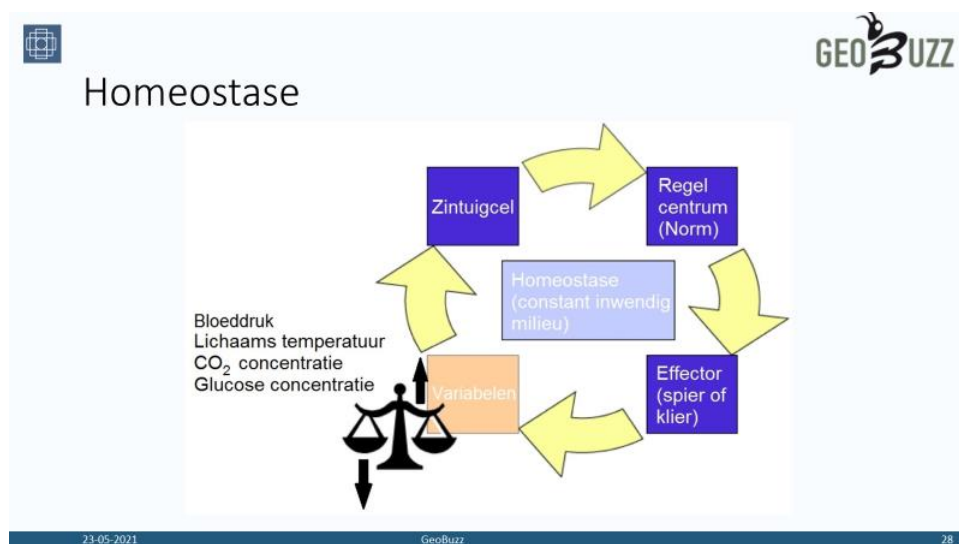
3. Een dt is een model met **terugkoppeling**, het is gericht op **besturing**

Je stelt een gewenste uitvoer in, bijvoorbeeld een gewenste snelheid bij cruise control. Het systeem vergelijkt die norm met de werkelijke snelheid en als die afwijkt wordt er gas bijgegeven óf gereduceerd, al dan niet met terugschakelen of ect remmen. (In een geavanceerd systeem wordt rekening gehouden met hellingen e.d.: 'adaptive cruise control')).



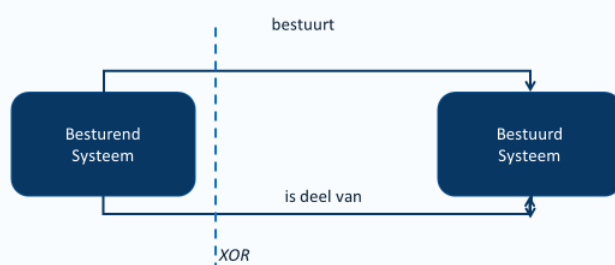
In een digital twin vindt iets vergelijkbaars plaats, alleen praten we hier over meer complexe systemen met dito besturing. Dus we willen bijvoorbeeld van een stedelijk gebied allerlei variabelen in samenhang monitoren, zowel op de korte als de lange termijn gericht. Bijvoorbeeld in Scheveningen: de drukte, het verkeer, het parkeren, de incidenten, de stemming, het weer, enz. en dat geheel van variabelen bepaalt je ingreep.

Er komt nog een factor bij: de systemen waarover we het hebben zijn in zekere mate zelfregulerend of 'homeostatisch'. Zo worden in ons lichaam allerlei waarden op peil gehouden (de temperatuur, de zuurgraad, de bloeddruk, de suikerspiegel, enz.) op peil gehouden via signaalstoffen, zoals hormonen. Ook in sociale systemen, zoals de stedelijke samenleving, is dat aan de hand: mensen en organisaties reageren autonoom op veranderingen, waarna correcties volgen. Daarmee is besturing complex: je bestuurt niet alleen een 'dom' fysiek systeem, maar een intelligent ecologisch stelsel, dat veelal van buitenaf alleen gecorrigeerd kan worden door indirecte prikkels. Bijvoorbeeld het verlenen van subsidie aan het vergroenen van tuinen teneinde de gevolgen van klimaatopwarming te mitigeren.





Herhaalde besturing



23-05-2021

GeoBuzz

29

NB In feite is homeostase natuurlijk een van de doelen van digital twins! Hiermee wordt dan het gehele systeem bestuurd. Binnen dat systeem heb je allerlei deelsystemen met een eigen besturingsorgaan! (Ik heb van toepassing op een stedelijk systeem hiervan nog geen goed voorbeeld kunnen vinden).

4. Een digital twin is *integraal*

Een digital twin is een model van een systeem – een apparaat, een product, een mens, een gebouw, een stad of wat dan ook. Een adequate digital twin moet alle aspecten die van belang zijn voor het functioneren van dat systeem in zich opnemen. Partiële digital twins bestaan niet, ik vergelijk die met 2D-doorsneden van een 3D-model. Het is een projectie, een platte kopie. Integraal moet niet verward worden met gedetailleerd: integraal wil zeggen dat alle relevante aspecten op een vergelijkbaar detailniveau worden opgenomen.

Intermezzo: datalogie.



Datalogie (gegevenskunde) (zoals het zou kunnen/ moeten bestaan)

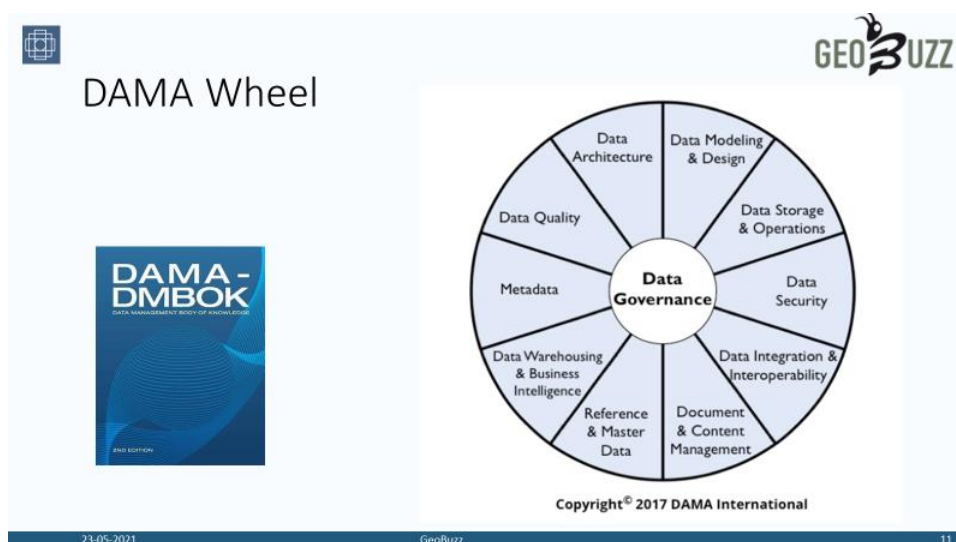
- Wetenschappelijke basis voor ontwikkeling, beheer en gebruik van gegevens en gegevensverzamelingen.
- Onderwerpen:
 - grondslagen (logica, verzamelingenleer, linguïstiek en meer)
 - technologie: databases, webtechnologie, algoritmen
 - gegevensmodellering
 - bouw en beheer van gegevenssystemen
 - gegevensmanagement, gegevenskwaliteit, organisatie
 - data-analyse
 - AI, neurale netwerken, enz.
 - gegevensrecht, beveiliging, ethiek
 - domeingerichte gegevenskunde, bijv. ruimtelijke gegevenskunde

23-05-2021

GeoBuzz

1

Net zoals er disciplines bestaan als bouwkunde, werktuigbouwkunde en bedrijfskunde zou er een discipline gegevenskunde ofwel datalogie moeten bestaan. Misschien bestaat die wel, maar dan ben ik wellicht onvoldoende bekend met de nieuwste ontwikkelingen. en ik kom te weinig mensen tegen die over een coherent geheel aan kennis beschikken betreffende gegevens. Misschien vult de Jheronimus Academy of Data Science (JADS) hier in Den Bosch wel de leemte! Datalogie is een wetenschappelijke basis voor ontwikkeling, beheer en gebruik van gegevens en gegevensverzamelingen. De inhoud zou moeten bestaan uit: grondslagen (logica, verzamelingenleer, linguïstiek en meer), technologie (databases, webtechnologie, algoritmen), gegevensmodellering, bouw en beheer van gegevenssystemen, gegevensmanagement, gegevenskwaliteit, organisatie, data-analyse, AI, neurale netwerken, enz., gegevensrecht, beveiliging, ethiek, domeingerichte gegevenskunde, bijv. ruimtelijke gegevenskunde.



Er is overigens al wel een praktisch handboek: DAMA-DMBOK: de Data Management Body of Knowledge, een uitgave van DAMA International, een vereniging van technische en bedrijfskundige professionals, gericht op de promotie van concepten en praktische toepassing op het gebied van datamanagement.

In het 'DAMA Wheel' zie je een aantal datalogische concepten terug.

Ik ga hierin overigens slechts in op twee aspecten van de datalogie, met name gegevenskwaliteit en in mindere mate het dataplatform.

Bij data voor een digital twin gaat het niet alleen meer om data in databases, maar om alle mogelijke data, die overal vandaan kunnen komen, denk aan linked data en big data. Dit is voor wat betreft het soort eisen geen verandering, het blijft gaan om zaken als nauwkeurigheid, actualiteit, consistentie, enz. , maar de manier waarop daarmee moet worden omgegaan is wel veranderd. Immers voor het world wide web geldt: AAA: Anyone can say Anything about Anything! Dus niet ieder gegeven is zomaar te gebruiken.

Een digital twin bevat de componenten data, rekenmodellen en visualisaties. We richten ons hier verder op de data en de eisen die daaraan gesteld moeten of kunnen worden. Die eisen zijn af te leiden van de functies.



Digital Twin: functies



- Ontwerp
- Prototyping
- Visualisatie/ VR/ AR
- Simulatie
- Registratie
- Analyse
- Monitoring
- Evaluatie
- Testen
- Voorspellen
- Optimalisatie van processen
- Configuratie van producten
- Training
- Verkoop en marketing c.q. burgerparticipatie

23-05-2021

GeoBuzz

63

De functies van digital twins stellen hoge eisen aan de datakwaliteit. Niet dat digital twins altijd aan de hoogste graad van nauwkeurigheid moeten voldoen, maar die moet wel in overeenstemming zijn met de beoogde functies van de betreffende digital twin. VOORBEELD je wilt de ontwikkeling van de bevolking van een stad voorspellen naar verdeling van de leeftijden over de wijken. Dan heb je de samenstelling van de woningvoorraad nodig naar een aantal kenmerken, zoals grootte, aantal kamers, hoogteligging, bouwjaar, renovatiejaar, typering van de wijk, e.d., maar geen details over de indeling, het materiaalgebruik en de inrichting van de woningen. Niettemin als je diezelfde digital twin gebruikt voor voorspelling van de belastingopbrengst, het treffen van energiebesparende maatregelen en watermanagement, worden de eisen een stuk hoger.

Eisen aan de data



De 9 kwaliteitsdimensies van NORA



Kwaliteitsdimensie	Definitie
Juistheid	De mate waarin gegevens de echte waarde goed weergeven.
Compleetheid	De mate waarin gegevens aanwezig zijn.
Validiteit	De mate waarin gegevens voldoen aan de verwachte structuur en opslagvorm.
Consistentie	De mate waarin gegevens vrij van tegenspraak zijn en samenhang vertonen met andere gegevens.
Actualiteit	De mate waarin gegevens recent genoeg zijn.
Precisie	De mate waarin gegevens exact of onderscheidend genoeg zijn.
Plausibiliteit	De mate waarin gegevens worden beschouwd als waar en geloofwaardig door gebruikers.
Traceerbaarheid	De mate waarin de totstandkoming en het gebruik van gegevens zijn vastgelegd.
Begrijpelijkheid	De mate waarin gegevens eenvoudig gelezen en geïnterpreteerd kunnen worden door gebruikers.

23-05-2021

GeoBuzz

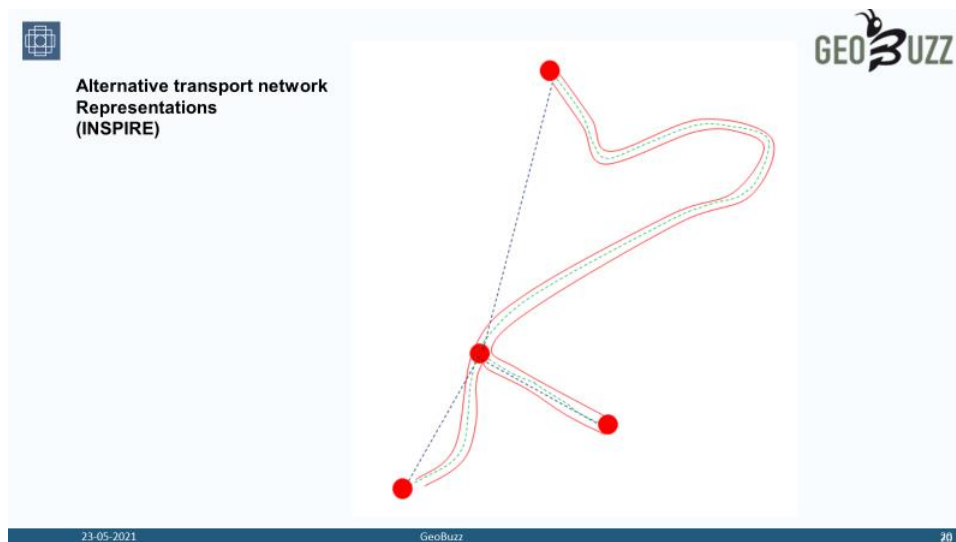
12

NORA, de Nederlandse overheidsreferentiearchitectuur, heeft een Raamwerk gegevenskwaliteit gepubliceerd. Laten we dat eens toepassen op de data, waarop een digital twin gebaseerd is. Het raamwerk kent een negental dimensies.

Precisie geeft de marge, waarbinnen de waarde van een gegeven mag liggen +/- 10 cm is minder precies dan +/- 1 mm. (Ook wel aangegeven met 'spreiding' of 'standaardafwijking'). Een digital twin moet behoorlijk precies zijn als deze voor bijvoorbeeld voor watermanagement gebruikt wordt.

Geometrische precisie is de mate van detail waarmee ruimtelijke gegevens worden ingewonnen. Dit gaat over onder meer de gridgrootte en aantal decimalen van coördinaten. Er wordt in de context van rasterdata ook wel gesproken over **resolutie**.

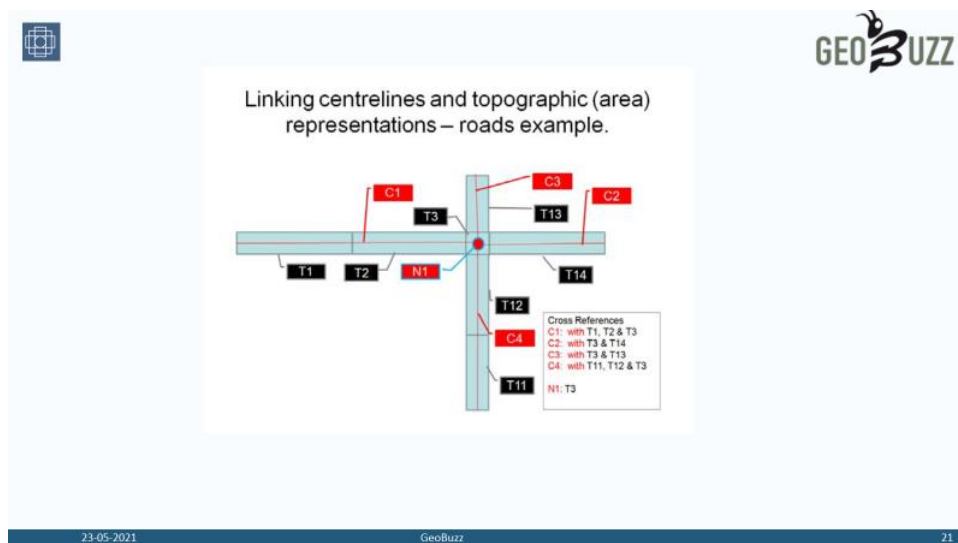
Een voorbeeld met een vectorbestand (uit INSPIRE DataSpecification TN v 3.0 Annex C):



Je ziet hier een weg op drie manieren gerepresenteerd:

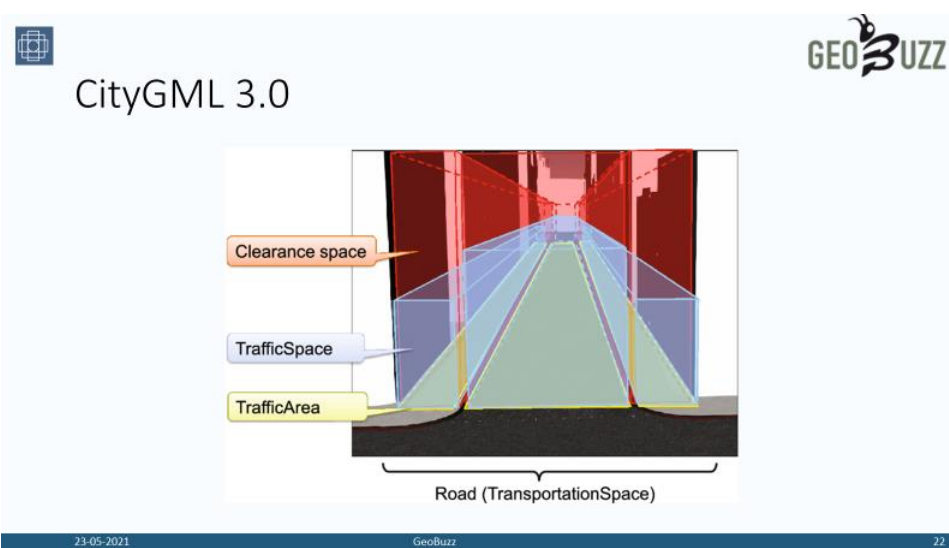
1. Fysiek (topografisch) model: de rode lijnen begrenzen vlakobjecten, die het oppervlak van een snelweg representeren (of de oevers van een rivier op een bepaald moment in het jaar of op winterpeil).
2. Middellijnmodel: de groene stippellijn. Dit model kan topologie ondersteunen en gebruikt worden voor analyses. Alle ruimtelijke objecten zijn geogerefererd.

Middellijnen zijn vaak verbonden met de onderliggende topografie.



3. Topologisch model: de blauwe stippellijn, een set verbindingen tussen knopen (verbindingpunten en eindpunten), al dan niet met de coördinaten van die knopen. Wordt vooral als rekenmodel gebruikt.

Door afstemming - door het laten verwijzen van knopen uit het ene naar het andere model - van topologisch model en het middellijnmodel kennen gegevens tussen beide gedeeld worden.



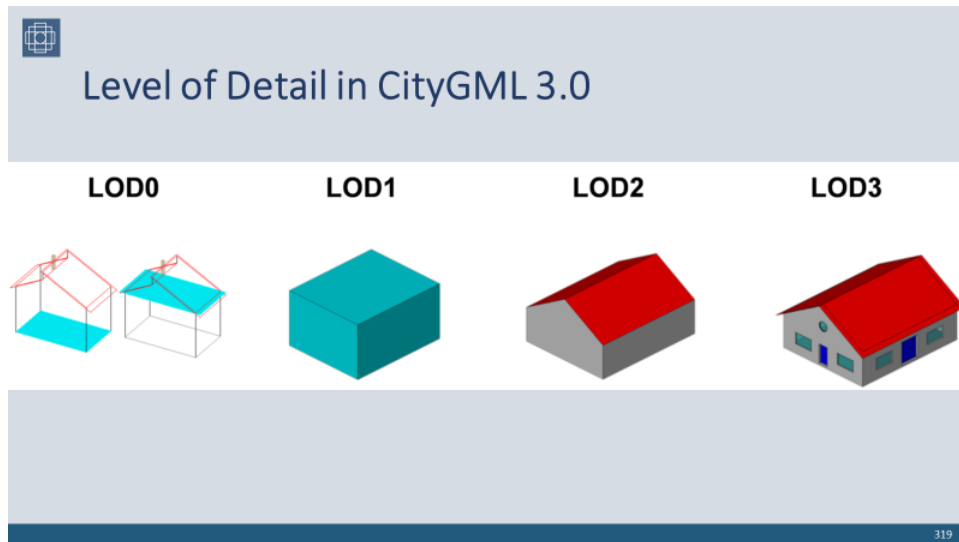
bron: CityGML 3.0: Tatjana Kutzner, Kanishk Chaturvedi, Thoma Kolbe, 'New Functions Open Up New Applications', 2020

Eigenlijk kun je zo een model maken, waarin een verbinding in een topologische netwerk als het ware wordt opgeblazen tot een gedifferentieerd 3D-model van de weg, met daarin onderscheiden:

- verkeersruimte en vrije ruimte (functioneel object)
- fysieke weg (fysiek object)

Je kunt nog een object toevoegen: de registratieve ruimte, meestal een vlakobject dat de weg administratief afbakt.

Ander voorbeeld:



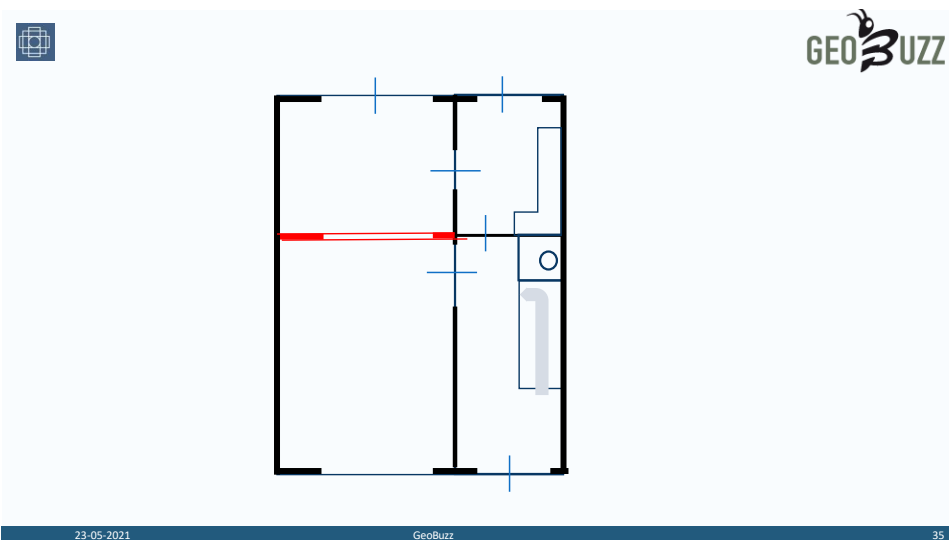
Juistheid is de kans dat een gegeven overeenstemt met de werkelijkheid.

Juistheid en precisie werken als communicerende vaten: als ik de mogelijke waarde van een gegeven ruimer neem, is de kans dat het juist is, natuurlijk groter (in dezelfde context), als ik het preciezer wil hebben en ik gebruik hetzelfde gegeven, is de kans dat het fout is natuurlijk groter.

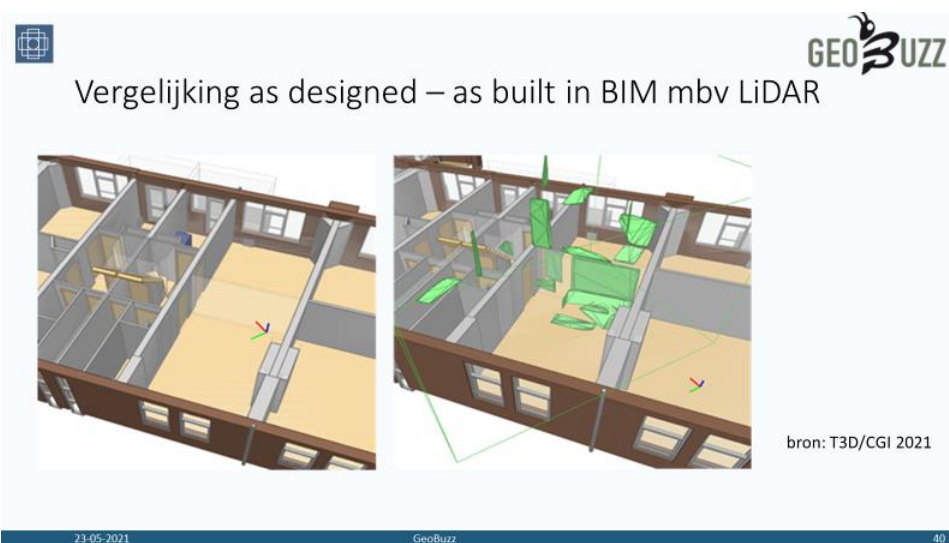
De combi van precisie en juistheid is **Nauwkeurigheid**: een precies gegeven met een hoge kans op juistheid is een nauwkeurig gegeven.

(Hoge nauwkeurigheid: weinig toevallige fouten en weinig systematische fouten ofwel 'in de roos').

(anders gezegd: een gegeven vertelt ons dat de breedte van een bepaalde groenstrook 20m is en het betrouwbaarheidsinterval 2m bij een kans van 95%, dus ligt de breedte van die groenstrook met 95% kans tussen 19 en 21 m. Een betrouwbaarheidsinterval van 98% en 50 cm is natuurlijk vele malen nauwkeuriger).



Je kunt je voorstellen dat er een digital twin gemaakt wordt, waarin woningplattegronden een rol spelen, bijvoorbeeld voor de bepaling van de WOZ-waarde of voor woningbehoefteberekening (verhuisprocessen). In een bepaald type woning blijkt de woonkamer vaak doorgebroken. Ik zou kunnen besluiten bij bepaalde woningcomplexen niet iedere woning daarop te inspecteren, maar een schatting te doen op basis van een steekproef. Door de steekproef te vergroten kan ik de nauwkeurigheid vergroten. Op basis van de gestelde eisen kan ik de steekproef bepalen en afwegen tegen de kosten.



Dit doet me overigens denken aan een voorbeeld uit T3D, betreffende de toepassing van LiDAR-scanning. In een experiment werden een opgeleverde woning vergeleken met het ontwerp. Dit werd mogelijk door de toepassing van diverse algoritmen op de puntenwolk.

Validiteit

Het gegevens moet weergeven wat bedoeld is om weer te geven. In een tabel met afstanden zijn andere regels van toepassing dan in een tabel met oppervlakten.

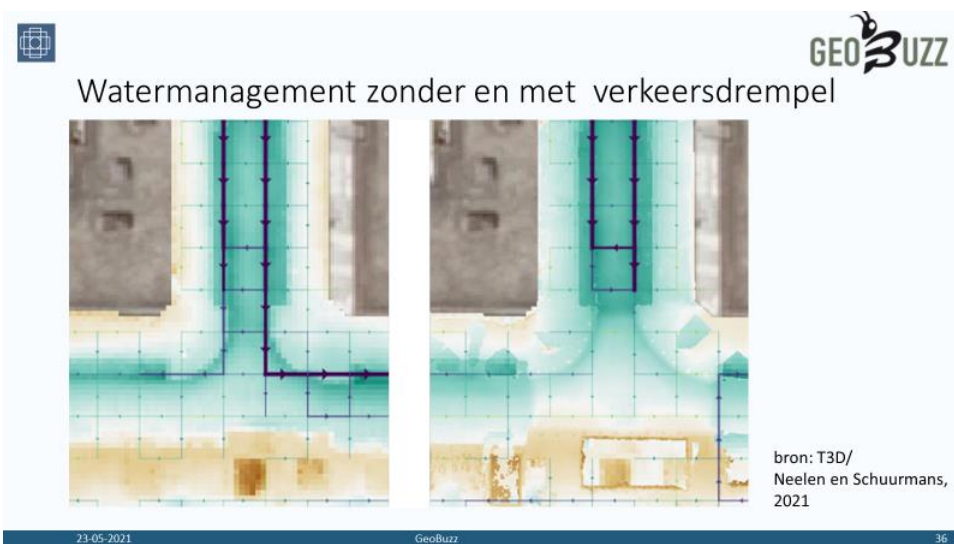
NORA: beperker: 'de mate waarin gegevens voldoen aan de verwachte structuur en opslagvorm':

Zo heb je bijvoorbeeld formaatvaliditeit (syntactische correctheid), bijvoorbeeld heeft een attribuutwaarde een minimale en maximale waarde.



Geometrische validiteit is de mate waarin iedere geometrie voldoet aan geometrische regels

De punten die onderdeel uitmaken van deze geometrie moeten aan specifieke voorwaarden voldoen. Zo mogen zij bijvoorbeeld niet in een incorrecte ringvolgorde zitten, geen dubbele punten bevatten, het juiste aantal punten bevatten en geen onnodig korte lijnen hebben.

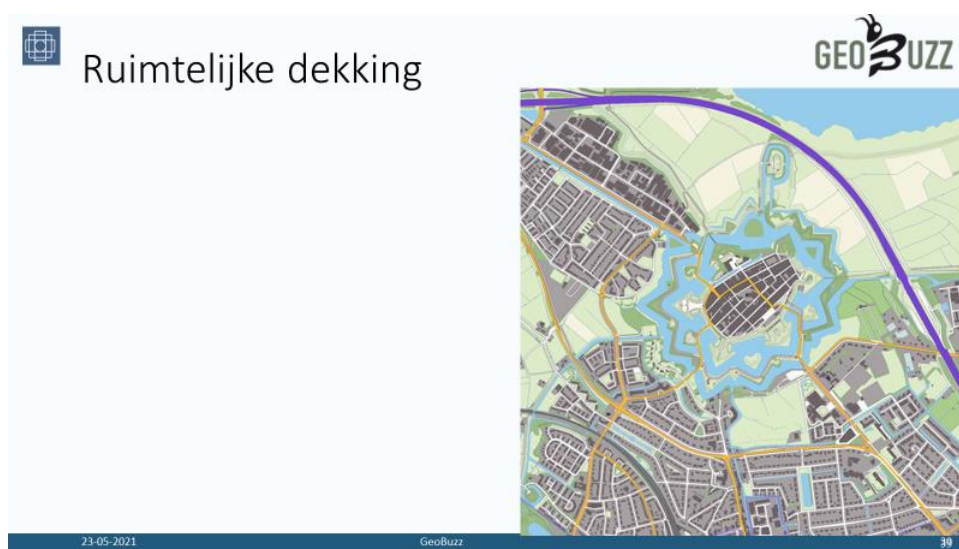


Volledigheid.

Neem bijvoorbeeld watermanagement. Als je de gevolgen van een regenbui gaat simuleren en analyseren schijnen verkeersdrempels belangrijk te zijn om de stroming van het water te bepalen. Maar je moet ze dan wel **allemaal** meenemen natuurlijk.



Iets gelijksoortigs geldt voor obstakels als het gaat om crowdmanagement. Hier een voorbeeld van een POSITIEVE werking van zo'n obstakel: het vergemakkelijkt de doorstroming van de menigte en zorgt ervoor dat de ruimte eerder leeg is. Dus hier is het ONTBREKEN van een obstakel een belangrijk gegeven!



Een ander voorbeeld van onvolledigheid: ruimtelijke dekking, d.w.z. géén witte plekken in de kaart!

Compleetheid

De mate waarin gegevens aanwezig zijn.

- attributocompleteheid, bijvoorbeeld bouwjaar BAG-pand niet ingevuld
- datasetcompleteheid, ontbreken van een openbare ruimte in BAG op basis van straten in NWB (Nationaal wegenbestand); ontbreken van stroken op een snelweg (dan wordt autonavigatie of het voorspellen van verkeersstromen lastig)
- ihb: ruimtelijke dekking
 - afdekking door een objecttype van een gebied zonder gaten
- historiecompleteheid
- identificeerbaarheid, bijvoorbeeld BAG ID in vestiging Handelsregister
- metadatacompleteheid

- overcompleteheid



Nog een voorbeeld: COMEX (netwerk voor Corridormanagement op Europees niveau) heeft informatiebehoefte om de maximale toegestane snelheid op de vaarweg in kaart te brengen. Deze informatie is nu (nog) niet opgenomen als kernregistratie en niet opgenomen in het NWB-V, maar inmiddels wel opgenomen in VNDS. Dit gegeven is onmisbaar voor simulaties, analyses en besturing.

Actualiteit.

Gegevens moeten tijdig bijgewerkt worden (updatefrequentie, periodieke controle).

Actualiteit kan volledigheid ondersteunen. Als de aanwezigheid van verkeersdrempels niet bijgehouden wordt, is het resultaat onvolledig! Het kan uiteraard ook zijn dat de drempels vervangen zijn door een ander type, bijvoorbeeld minder steil.

Consistentie, een heel belangrijke. Dit kan samenhangen met actualiteit; als ik bijvoorbeeld twee registraties combineer van twee verschillende tijdstippen en er is intussen ergens een tunnel aangelegd, dan kan ik die combi waarschijnlijk moeilijk gebruiken voor verkeerssimulatie. En zeker ook niet voor watermanagement.

Voorbeeld van een logische inconsistentie:

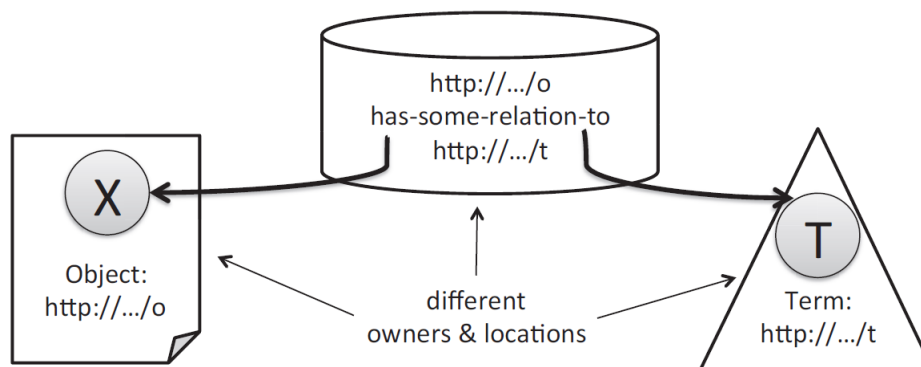
- De Dorpstraat en de Nieuwstraat in gemeente G hebben geen gemeenschappelijke panden of verblijfsobjecten
- Het adres van Jan is Dorpsstraat 33
- Het adres van Marietje is Nieuwstraat 92
 - Jan en Marietje wonen samen

Topologische consistentie bijvoorbeeld: rijstrook en groenvlak die over elkaar heen liggen (als afgesproken is dat op één plek er slechts één verhardingstype mag bestaan).

Een ander voorbeeld dat ik ken: verschillende definities van 'bouwwerk' in verschillende regelingen.

Hier ligt ook een aandachtspunt voor data op het web. Als ik over hetzelfde object onderling tegenstrijdige gegevens vind moet ik hetzij gezond verstand hetzij geavanceerde technieken gebruiken om het juiste gegeven vast te stellen. (Statistische methoden, raadplegen van de metagegevens, bepaling van de context waarin het gegeven vastgelegd is, analysetools, AI, enz.). (**Ontologieën** zijn hierbij een belangrijk hulpmiddel; Je kunt om te beginnen willekeurige concepten van wie dan ook van het web gebruiken en onder een eigen term gebruiken in de ontologie. Met behulp van die ontologie kun je dan gaan redeneren over dat concept en het relateren aan andere concepten in je ontologie

Je kunt om te beginnen willekeurige concepten van wie dan ook van het web gebruiken en onder een eigen term gebruiken in de ontologie. Met behulp van die ontologie kun je dan gaan redeneren over dat concept en het relateren aan andere concepten in je ontologie.



Webarchitectuur voor linked data (bron: A semantic web primer)

}




GO FAIR

- richtlijnen voor de manier van beschrijven, opslag en publicatie van wetenschappelijke data; ook breder toegepast;
- community (GO FAIR Initiative); FAIRification workgroups
- FAIR principles:
 - Findable - vindbaar
 - Accessible - toegankelijk
 - Interoperable - uitwisselbaar
 - Reusable - herbruikbaar
- uitgewerkt in 15 principes

F_{indable}



A_{ccessible}



I_{nteroperable}



R_{eusable}



23-05-2021

GeoBuzz

48

Deze ontwikkeling vraagt om speciale aandacht voor enkele aspecten van de data, die ik niet direct terugzien in het NORA-raamwerk: de zogenaamde FAIR principles: gegevens moeten voldoen aan de criteria Findable, Accessible, Interoperable en Reusable.

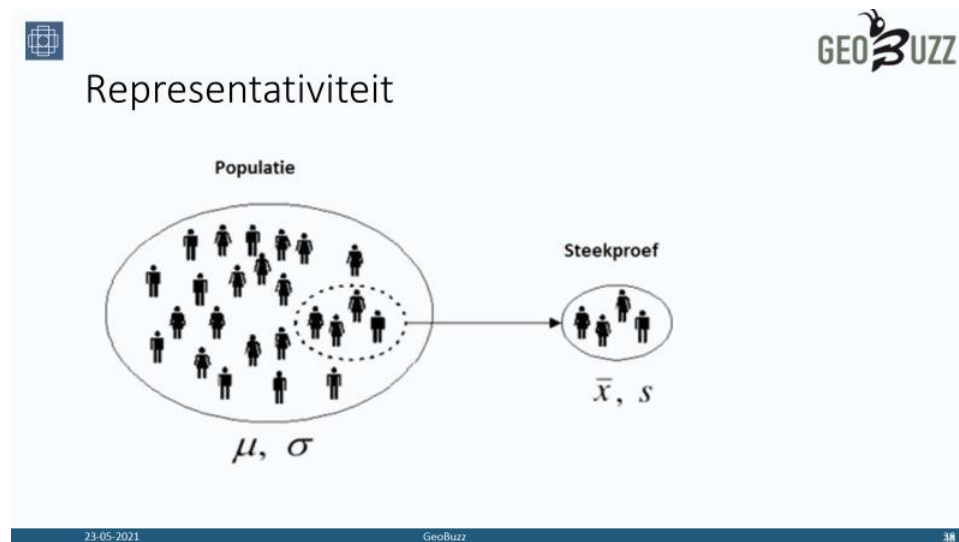
Traceerbaarheid

De mate waarin totstandkoming en gebruik van het gegeven is vastgelegd.

- Herleidbaarheid
- Data lineage
- reproduceerbaarheid

Plausibiliteit valt uiteen in een aantal subdimensies: authenticiteit, bewijsbaarheid, reputatie ('de gegevens komen van het RIVM'), waarschijnlijkheid, representativiteit.

M.b.t. waarschijnlijkheid: als je ziet dat ergens de afstand Amsterdam-Maastricht is vastgelegd als 2150 km, dan moet er een belletje gaan rinkelen. Je kunt dit vastleggen als signaalregels. Een onwaarschijnlijk gegeven is een gegeven dat een signaalregel overschrijdt. Het gegeven is daarmee niet automatisch fout. Er is slechts een signaal dat het gegeven mogelijk fout is.



Representativiteit is de mate waarin een dataset een goede weergave geeft van het geheel. Een dataset is representatief als de inhoud ervan statistisch gezien lijkt op die van andere datasets in het geheel.

(- Authenticiteit: de mate waarin de identiteit van de bron van de gegevens aantoonbaar is. Dit gaat erover of gegevens echt afkomstig zijn van de bron die je verwacht. Authenticiteit kan bijvoorbeeld worden aangetoond door de bewerker of verzender een digitale handtekening toe te laten voegen aan een gegeven.

- Bewijsbaarheid: de mate waarin de juistheid van een gegeven kan worden aangetoond. Dit gaat over de objectiviteit van een waarde; zijn er bewijzen aanwezig die onderbouwen dat deze waarde echt klopt? De kans is anders groot dat de waarde vooral gebaseerd is op wat één persoon denkt. Voorbeeld: inmeting.
- Reputatie, bijvoorbeeld: 'deze data komen van het RIVM'.
- Waarschijnlijkheid. Voorbeeld: de A2 van Amsterdam naar Maastricht is 2150 km lang. Je kunt dit vastleggen als signaalregels. Een onwaarschijnlijk gegeven is een gegeven dat een signaalregel overschrijdt. Het gegeven is daarmee niet automatisch fout. Er is slechts een signaal dat het gegeven mogelijk fout is.)
- Representativiteit is de mate waarin een dataset een goede weergave geeft van het geheel. Een dataset is representatief als de inhoud ervan statistisch gezien lijkt op die van andere datasets in het geheel.

Begrijpelijkheid

Bijvoorbeeld B1-taalniveau

én: verstrek **Duidelijkheid**: verstrek bij de data voldoende metadata!

Toevoeging aan NORA:

FAIR:

Findable

Accessible

Interoperable

Reusable

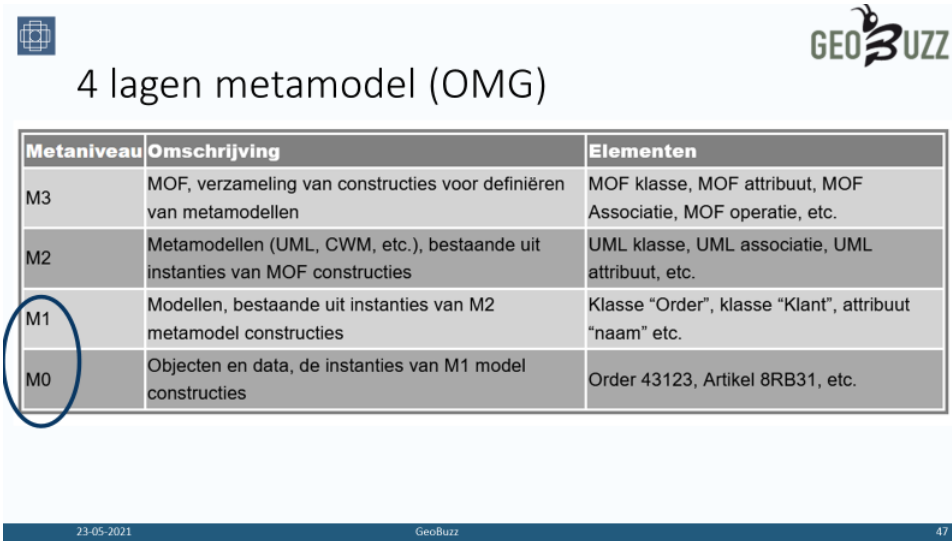
M1-niveau

Kennen jullie dit plaatje?

(van Magritte)



Het betekent: je kijkt niet naar een pijp, maar naar de afbeelding van een pijp.



Metaniveau	Omschrijving	Elementen
M3	MOF, verzameling van constructies voor definiëren van metamodellen	MOF klasse, MOF attribuut, MOF Associatie, MOF operatie, etc.
M2	Metamodellen (UML, CWM, etc.), bestaande uit instanties van MOF constructies	UML klasse, UML associatie, UML attribuut, etc.
M1	Modellen, bestaande uit instanties van M2 metamodel constructies	Klasse "Order", klasse "Klant", attribuut "naam" etc.
M0	Objecten en data, de instanties van M1 model constructies	Order 43123, Artikel 8RB31, etc.

Deze manier van kijken is geformaliseerd in een standaard: Metamodel Informatiemodellering (Geonovum). Je ziet 4 lagen, waarin per laag standaarden staan voor de onderliggende laag. We beperken ons hier tot de onderste twee.

Tot nu toe bekeken we de dingen op M0-niveau: een digital twin is een model van een systeem. Maar de digital twin kunnen we zelf ook als een systeem beschouwen. Dat moet aan gelijksoortige criteria voldoen als die voor het M0-niveau.

In concreto gaat het hier om eisen aan het **gegevensmodel** van de Digital Twin Eventueel te onderscheiden naar technisch, logisch en conceptueel. We lopen kort de criteria langs.

Validiteit

Beschrijft het gegevensmodel echt de digital twin.

(Kan ik een dt-model voor een gebouw gebreken voor een dt-model voor een stad? Er zijn wel veel overeenkomsten maar ook veel verschillen)

Juistheid

De structuur van de data moet isomorf zijn met de structuur van het gerepresenteerde systeem.

Precisie

Dit vergt precieze naamgevingen en definities.

Bijvoorbeeld 'kamer' is meer precies dan 'ruimte'.

Definities moeten, uitgaande van een bepaalde korrelgrootte, axiomatisch zijn opgebouwd.

Een X is een Y, waarvoor geldt:

Volledigheid

Alle relevante klassen moeten benoemd zijn.

Actualiteit

Laatste update van een standaard gebruiken, bijvoorbeeld CityGML 3.0

Consistentie

Voorbeeld: CityGML kan niet zonder de relatie tussen panden en verblijfsobjecten opgenomen worden, uitgaande van een mapping van Panden naar Buildingparts en Verblijfsobjecten naar Units (de relatie moet n:m zijn en is in CityGML 1:nn).

Denk verder aan homoniemen

Plausibiliteit

- Voorbeeld: door elkaar halen van componenten en subtypen: als ik zou tegenkomen: een bankgebouw is een onderdeel van een gebouw moet er een belletje gaan rinkelen: een bankgebouw is een subtype of een eventueel een rol van een gebouw, maar geen onderdeel.

Traceerbaarheid

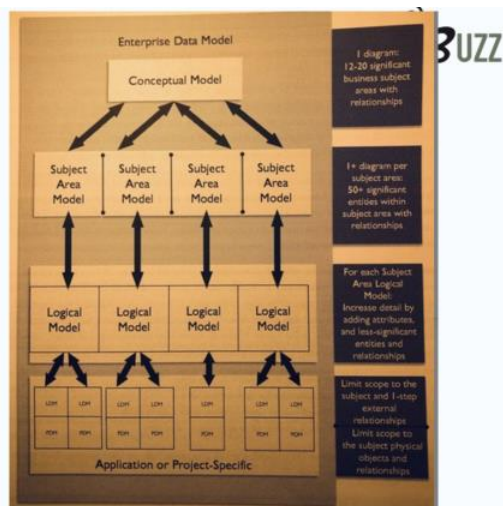
- ...

Begrijpelijkheid

- begrijpelijke namen voor klassen en attributen gebruiken
- model begrijpelijk beschrijven en ook FAIR!
- hiertegen wordt ZEER VAAK gezondigd: je krijgt een databaseschema, veelal ad hoc in elkaar gezet en daar moet je het mee doen!

II. Gegevensplatforms en digital twins/ enkele implementatievereisten

Enterprise Data Model



23-05-2021

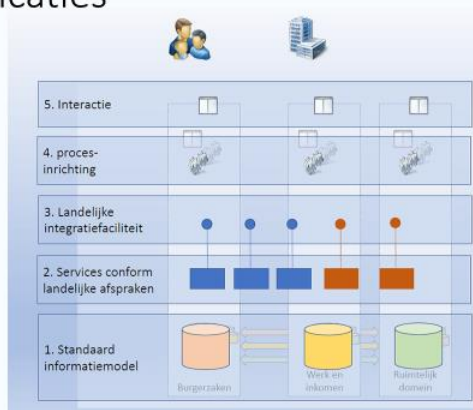
GeoBuzz

51



Scheid data van applicaties

Voorbeeld:
Common ground
gemeenten



23-05-2021

GeoBuzz

52



Metadata repository



- Gegevenswoordenboek
- Gegevenscatalogus
- Ontologie

23-05-2021

GeoBuzz

53



Regel de rechten



- Data-eigenaarschap
- POD (Personal Online Datastore)(ook hier: data scheiden van diensten)

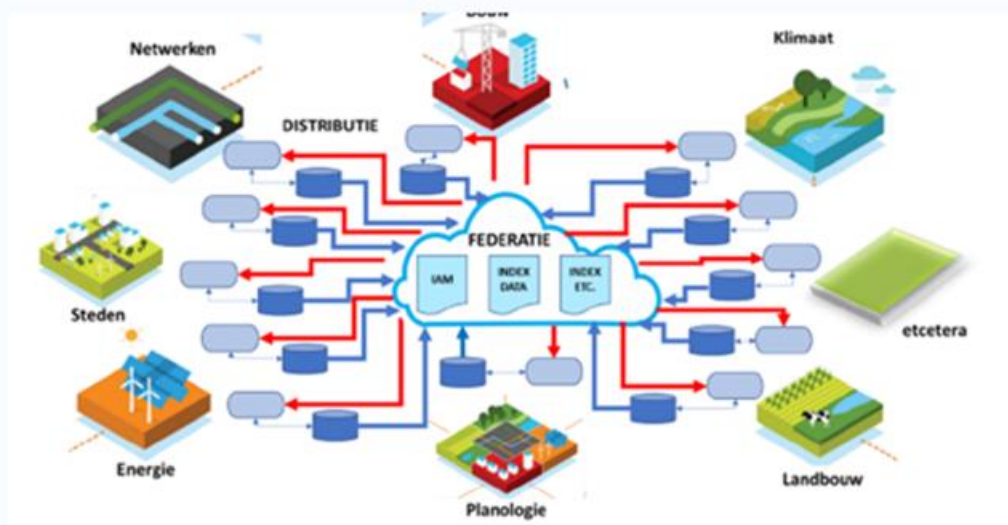
23-05-2021

GeoBuzz

54



Federatief stelsel/ stelselarchitectuur DTFL (Geonovum 2022)



23-05-2021

GeoBuzz

55

Om gegevens efficiënt en effectief te kunnen delen dient bij voorkeur een goed uitwisselingsstelsel tot stand te komen. Digitale tweelingen zijn tegelijk 'consumenten' en 'producenten' van data. Het geheel van afspraken en voorzieningen die hiervoor nodig zijn noemen we het uitwisselingsstelsel of kortweg stelsel. Deze referentie architectuur bevat ontwerpprincipes, standaarden en afspraken voor het stelsel voor dit uitwisselingsstelsel. We spreken ook wel over het stelsel voor digital twinning.

Kenmerkend voor dit stelsel is een decentrale verwerving, opslag en bewerking van gegevens in combinatie met een lichte, federatieve structuur voor de uitwisseling van gegevens en modellen.

Figuur 1 geeft een gestileerde weergave van dit uitgangspunt. Diverse deelnemers aan het stelsel verzamelen, beheren en bewerken gegevens en deze worden in de vorm van informatieproducten via netwerken (internet) uitgewisseld. Enkele federatieve voorzieningen, zoals indexen voor het snel vinden van de gewenste gegevens en het zorgen voor een identiteit- en toegangscontrole, maken het uitwisselen van gegevens mogelijk.

III. Conclusies



Conclusies

1. Digital twin: oude wijn in nieuwe zakken? (productmodel en BIM revisited)
2. Essentiële eigenschappen:
 - Een digital twin is een **model**
 - Een digital twin is dynamisch
 - Een digital twin realiseert terugkoppeling (een 'closed loop')
 - Een digital twin is integraal
3. Een digital twin stelt hoge eisen aan datakwaliteit en datamodelkwaliteit
4. Een digital twin vergt een architectuur die uitgaat van scheiding van data van applicaties en een adequate regeling van datarechten.