

Guidelines voor de logische laag

Collibra Data Catalog

Juni 2023 – Versie 1.0



Inhoudsopgave

Inleiding

Opbouw PostNL Data Catalogus

Logische termen – basis principes, toevoegen, status

Vast te leggen metadata

Relaties

Inleiding

De PostNL Data Catalogus

Introductie

PostNL gebruikt het platform Collibra Data Governance Center (DGC) ter ondersteuning van haar data organisatie op het gebied van data kwaliteit, data governance en data analyse. Het helpt bij het vinden van data, biedt een inventarisatie van metadata en verstrekt informatie die nodig is om vast te stellen of de data geschikt is voor het beoogd gebruik.

De PostNL Data Catalogus is een instrument waarmee de data organisatie van PostNL de toegankelijkheid, juistheid en relevantie van de data bedrijfsbreed kan verbeteren. Hiermee wordt cruciale ondersteuning geboden bij:

- **Data gebruik:** metadata verhoogt de kennis over data. Hoe meer gebruikers weten over data hoe beter ze de bruikbaarheid en de grenzen daarvan kunnen bepalen.
- **Datamanagement:** de data catalogus biedt inzicht en een beter begrip van de data die PostNL bezit. Dit maakt de data bekend en beheersbaar en is een randvoorwaarde voor professionalisering van data management capabilities zoals data governance en data kwaliteit management.

De PostNL Data Catalogus is een van de solution building blocks binnen de data capability 'Meta Data Management'.

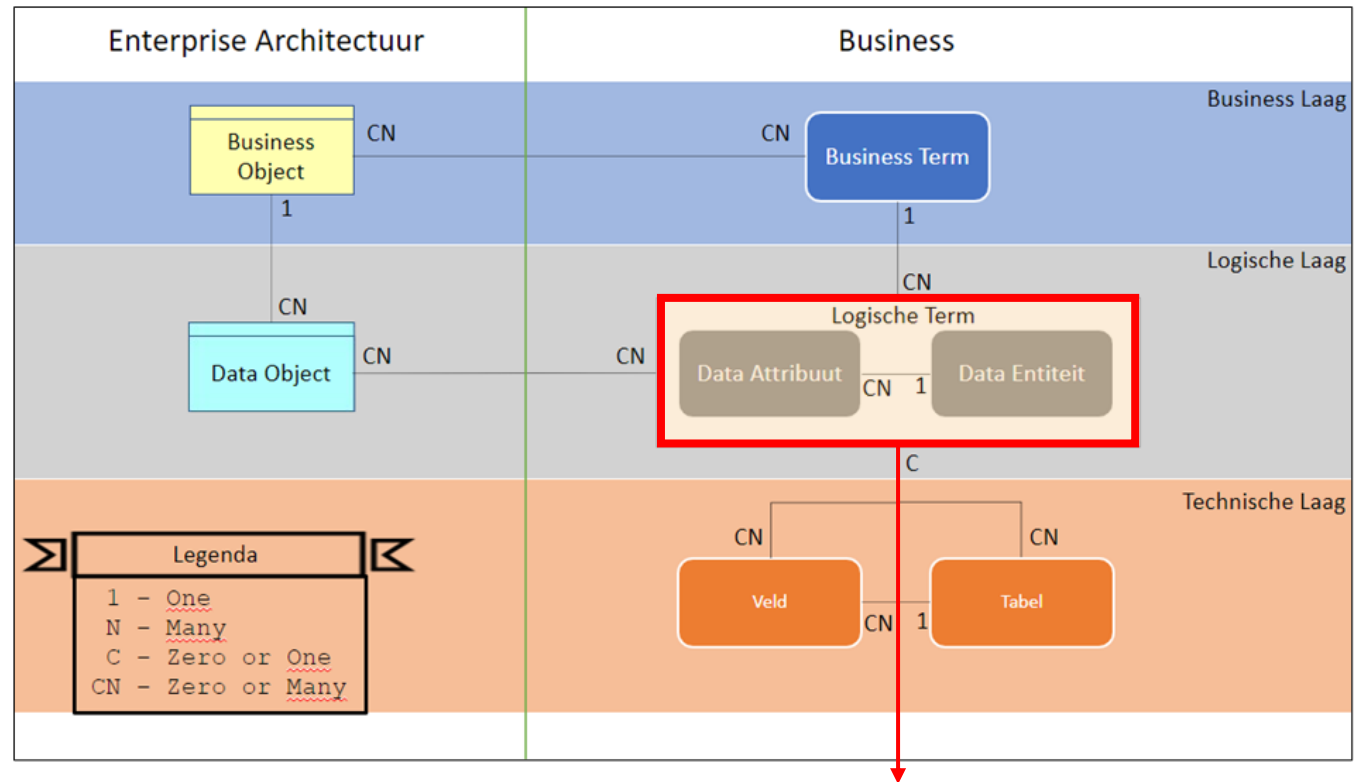
*“**Metadata Management** is de discipline die gaat over het verzamelen, onderhouden en standaardiseren van metadata (inclusief toegang en distributie)”.*

Opbouw PostNL Data Catalogus

Drie lagen: business, logisch en technisch

De PostNL Data Catalogus is opgebouwd uit drie verschillende lagen:

- **Business laag**
De termen die voor 'de business' gemeengoed zijn en in het dagelijks gebruik worden gehanteerd.
- **Logische laag**
De benodigde termen (entiteiten en attributen) om de business termen te koppelen aan de technische laag.
- **Technische laag**
De velden en tabellen zoals die in de systemen en applicaties daadwerkelijk voorkomen.



Focus in dit document ligt specifiek op de 'Logische laag'. Oftewel, wat is een logische term en wat is de bijbehorende metadata?

Logische termen

Basis principes (1/3)

- *“Een term die wordt gebruikt om een relatie te leggen tussen de Business terms en de tabellen en velden zoals ze in de systemen staan”.*
- Het gaat hier om de woorden en termen die gebruikt kunnen worden om technische tabel- en veldnamen te versimpelen.
- In de Data Catalog, kan een logische term ontstaan doordat er een behoefte is om een business term te koppelen aan de technische laag (of enkel alleen de logische laag). Wanneer er geen relatie bestaat tussen de business laag en de andere lagen, hoeft er dus ook geen logische term aangemaakt te worden.
- Indien een business term voldoende overeenkomt met de technische laag, kan er voor gekozen worden om de business term en de logische termen hetzelfde te laten zijn. Dit geldt dan voor de naam, en eventueel ook voor de bijbehorende metadata.



Logische termen

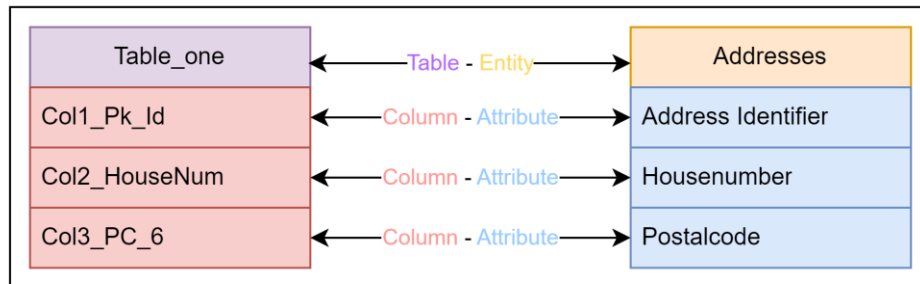
Basis principes (2/3)

- Een logische term kan twee verschijningsvormen hebben: in de vorm van een entiteit of in de vorm van een attribuut.
- Logische termen worden gebruikt in het logische datamodel. Dit model beschrijft de structuur en de relaties tussen de logische termen. Logische datamodellen worden vastgelegd in ARIS en worden gemodelleerd volgen het Entity Relation Diagram (ERD).
- Het logisch datamodel kan worden gezien als een versimpeling/verduidelijking van het reeds bestaande fysieke datamodel, maar is wel technologie onafhankelijk. Wel kan het LDM als leidraad werken van hoe het daadwerkelijke fysieke datamodel er uit komt te zien.
- Omdat fysieke datamodellen veelal systeembenamingen gebruikt voor tabellen en kolommen die te technisch zijn voor de 'gewone' lezers, is hiermee gelijk de noodzaak voor de logische laag aangeduid.
- **Let op:** de entiteiten en attributen worden los van elkaar uit ARIS ingeladen. Hierin is *niet* de onderlinge relatie tussen entiteit en attribuut in meegenomen. ARIS blijft de SoR voor de modellen.

Logische termen

Basis principes (3/3)

- Een logische term kan een data entiteit of een data attribuut zijn.
 - Een data entiteit is de logische tegenhanger van een tabel op de technische laag
 - Een data attribuut is de logische tegenhanger van een veld op de technische laag
- Net als dat een tabel uit een of meerdere velden bestaat, bestaat een entiteit uit een of meerdere attributen.

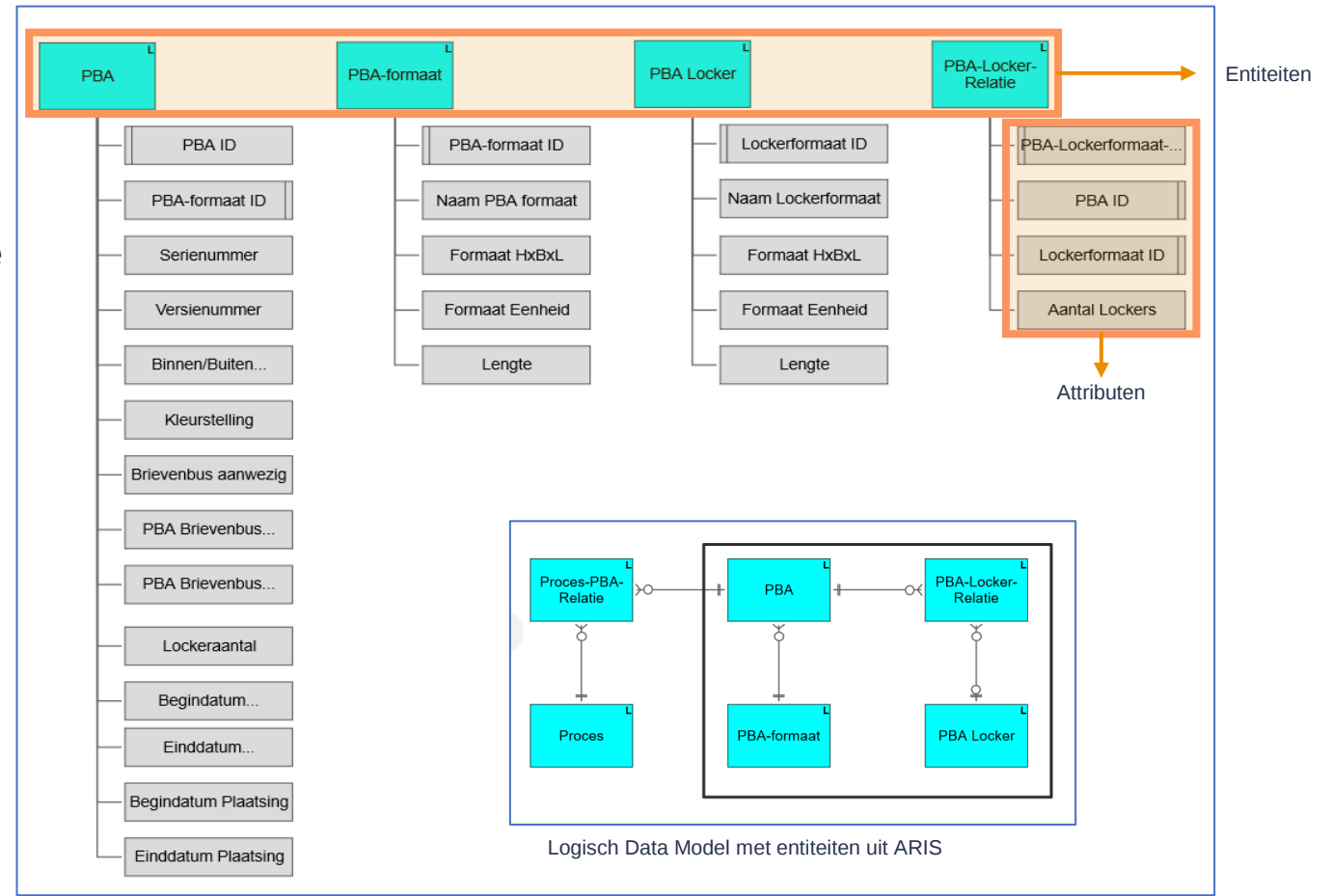


- Doordat een data entiteit uit minimaal een of meer data attributen bestaat, is er vanuit de business laag enkel een relatie te leggen met een entiteit. Vanuit daar worden de bijbehorende attributen gelinkt.
- Indien een entiteit uit één enkel attribuut bestaat, kunnen deze beide dezelfde naam krijgen.

Logische termen

Toevoegen

- Een eerste batch van logische termen wordt geïmporteerd vanuit ARIS. Deze termen zijn in het verleden opgesteld voor verschillende datamodellen, systemen en applicaties.
- Deze eerste batch van logische termen zal **onvoldoende** zijn om de volledige logische laag te vullen, want:
 - Niet elke logische term staat in ARIS.
 - Systemen kunnen verder ontwikkeld of aangepast zijn, maar het kan zo zijn dat aanpassing nog niet in ARIS is doorgevoerd.
- Om die reden wordt het ook mogelijk om handmatig logische termen toe te voegen aan de logische laag.

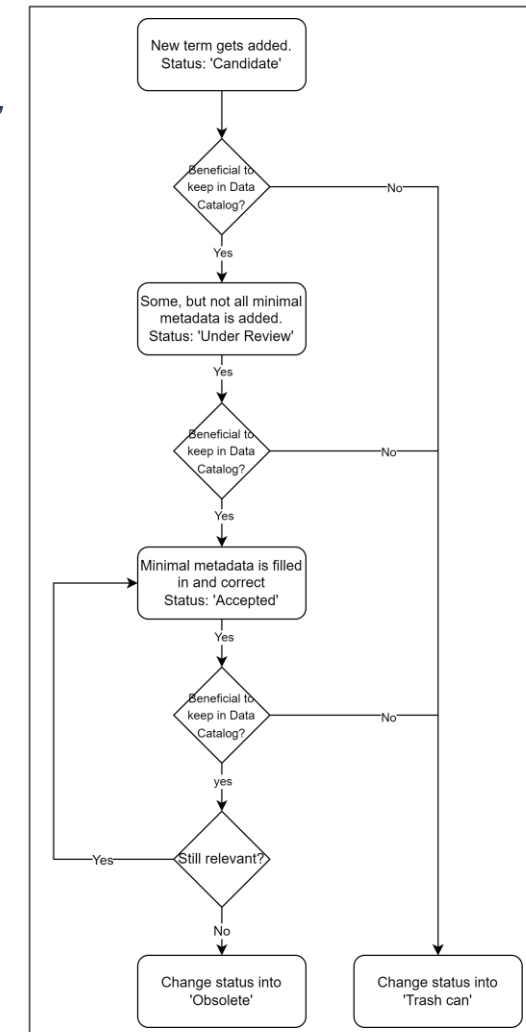


Logische termen

Status

Een term kan in verschillende fases verkeren. Met de status wordt aangegeven of het bijvoorbeeld een geaccepteerde term met metadata is of dat het een term is die 'Under Review' is.

Status	Betekenis
Candidate	Initiële status van een term. Dit betekent dat de term is aangemaakt of geïmporteerd. Er wordt hier nog niet gekeken naar eventuele metadata.
Under Review	Stakeholders reviewen de Asset. Dit betekent dat er metadata is toegevoegd, maar dat nog niet alle verplichte velden zijn ingevuld of kloppend zijn.
Accepted	De term én de verplichte metadata zijn volledig en correct ingevuld. De betekenis wordt gedragen door de belangrijkste stakeholders.
Obsolete	De term is achterhaald. De metadata van deze term blijft in dit geval wel beschikbaar voor informatie. Na bepaalde periodes wordt er gekeken of hierop een schoning nodig is.
Trash can	Wanneer een term foutief is opgevoerd krijgt deze de status van Trash can. Termen met deze status worden periodiek, met goedkeuring van de eigenaar, verwijderd
Invalid	Buiten scope
In Progress	Buiten scope
Approval Pending	Buiten scope



Vast te leggen metadata

Afgestemd met de Data Management Organisaties

1	Definition
2	Data Manager
3	Data Domain
4	System Of Record
5	System of Entry
6	Data Labelling
7	Personally Identifiable Information
8	PII Type
9	CIA Classification
10	Confidentiality
11	Integrity
12	Availability
13	Business Ruling
14	Datatype
15	Language
16	Standard Value Format
17	Date Created
18	Date Changed
19	Made/Changed by
20	Status
21	Context
22	Source (origin metadata)
23	<i>Rel Business Layer – Terms</i>
24	<i>Rel Architecture - Objects</i>
25	<i>Rel Technical Layer – Table</i>
26	<i>Rel Technical Layer – Field</i>
27	<i>Synonym</i>

In samenspraak met de Data Management Organisaties binnen PostNL is vastgesteld dat de hier links genoemde zgn. metadata velden in de PostNL Data Catalogus dienen te worden vastgelegd.

- De rood gekleurde velden zijn verplicht.
- De zwart gekleurde velden zijn optioneel.
- De blauw gekleurde velden laten de relatie zien tussen de drie lagen en/of business terms onderling.
- De velden 10 t/m 12 zijn vooralsnog buiten scope.

Op de volgende slides wordt elk van deze velden nader toegelicht.

Vast te leggen metadata

Toelichting op de velden (1/5)

Metadata	Description	Guideline	Mandatory
Definition	De definitie die duidelijk de term in één à twee zinnen beschrijft.	• Tekstveld • Definitie bevat niet de opgevoerde term zelf • Volgt de meest gebruikelijke spelling en leestekens • Uniek voor de context waarin de term beschreven wordt	Verplicht
Data Manager	Verantwoordelijke Data Manager voor deze specifieke term.	• Selectielijst met daarin de specifieke Data Manager die verantwoordelijk is voor deze specifieke term • Actueel overzicht	Verplicht
Data Domain	Verantwoordelijke domein voor deze specifieke term.	• Selectielijst met daarin het specifieke Data Domein dat verantwoordelijk is voor deze specifieke term • Actueel overzicht	Verplicht
System of Record	Het systeem waarin de waarheid en correcte data in is opgeslagen; de Golden Records.	• Selectielijst met daarin alle systemen en applicaties die gebruikt worden binnen PostNL die als Golden Record zijn gedefinieerd • Kan dezelfde waarde bevatten als de SoE. • Als de term een lineage heeft tot en met de technische laag, dan is dit veld verplicht	Optioneel
System of Entry	Het systeem waarin de data wordt gegenereerd.	• Selectielijst met daarin alle systemen en applicaties die gebruikt worden binnen PostNL voor data opvoer • Kan dezelfde waarde bevatten als de SoR. • Als de term een lineage heeft tot en met de technische laag, dan is dit veld verplicht	Optioneel

Vast te leggen metadata

Toelichting op de velden (2/5)

Metadata	Description	Guideline	Mandatory
Data Labelling	Duidt het label aan dat van toepassing is op gegevensbescherming	- Selectielijst met waarden: Public: Komt overeen met gegevens met de vertrouwelijkheidswaarde 'None'. Internal: Komt overeen met gegevens met de vertrouwelijkheidswaarde 'Low'. Confidential**: Komt overeen met gegevens met de vertrouwelijkheidswaarde 'Medium'. Secret: Komt overeen met gegevens met de vertrouwelijkheidswaarde 'High'. - Moet afgestemd zijn met en voldoen aan de regels en principes van PostNL Cybersecurity Office. ** Als de kolom PII-gegevens bevat, dan dient deze minimaal het label van "Confidential" te krijgen.	Optioneel
Personally Identifiable Information	Geeft aan of het persoonsgegevens bevat.	- Selectielijst met daarin de waarden: - Yes, does contain Personally Identifiable Information - No, does not contain Personally Identifiable Information - Uncertain if it contains Personally Identifiable Information - Unknown	Verplicht
PII Type	Geeft aan wat voor soort persoonsgegevens het bevat.	- Meerkeuze selectielijst met daarin de waarden: - Business Partner - Consumer - Employee - Kan in de toekomst verder uitgebreid worden. - Indien 'Personal Identifiable Information' is ingevuld, dan is dit veld verplicht	Optioneel

Vast te leggen metadata

Toelichting op de velden (3/5)

Metadata	Description	Guideline	Mandatory
CIA Classification	De classificatie van data en de gerelateerde systemen is bepaald door wat de impact voor PostNL is wanneer de gestelde eisen aan Confidentiality, Integrity en Availability niet gehaald worden.	- Selectielijst met daarin de waarden Baseline en Above Baseline - Wanneer de impact van een van de drie categorieën (Confidentiality, Integrity or Availability) hoog is, dan is de CIA Classificatie Above Baseline. In alle andere gevallen is Baseline voldoende. - Dient afgestemd te zijn en overeen te komen met de regels en principes van PostNL Security Office.	Optioneel
Confidentiality	Wat is de impact wanneer er sprake is van onbevoegde verstrekking van informatie.	- Selectielijst met daarin de waardes: low, medium en high: Low: Het verlies van Confidentiality, Integrity of Availability heeft naar verwachting een <u>beperkt nadelig effect</u> op bedrijfsprocessen, bedrijfsmiddelen of individuen. Medium: Het verlies van Confidentiality, Integrity of Availability heeft naar verwachting een <u>nadelig effect</u> op bedrijfsprocessen, bedrijfsmiddelen of individuen. High: Het verlies van Confidentiality, Integrity of Availability heeft naar verwachting een <u>ernstig of catastrofaal nadelig effect</u> op bedrijfsprocessen, bedrijfsmiddelen of individuen. - Dient afgestemd te zijn en overeen te komen met de regels en principes van PostNL Security Office.	Optioneel
Integrity	Wat is de impact wanneer er sprake is van onbevoegde aanpassing of vernietiging van informatie.		Optioneel
Availability	Wat is de impact wanneer er sprake is van een verstoring in de toegang voor het gebruik van informatie of informatiesystemen.		Optioneel

Vast te leggen metadata

Toelichting op de velden (4/5)

Metadata	Description	Guideline	Mandatory
Business Ruling	De business regels die gelden voor een term.	- Tekstveld - Voorbeeld: Nederlandse Postcode: Bestaat vier cijfers en twee letters. Het eerste getal mag niet met 0 beginnen. De lettercombinaties 'SS', 'SD' en 'SA' worden niet gebruikt.	Optioneel
Data Type (bij attribuut)	Geeft aan wat voor type data er in bij het data attribuut verwacht wordt.	Selectielijst met daarin de waarden: Text, Date, DateTime, Numeric, Code, Measure, Amount, Identifier, Indicator, Quantity, Value, Percent.	Optioneel
Language	De taal waarin een term betekenis heeft.	Selectielijst met daarin de waarden: NL, EN, DE and FR	Verplicht
Standard Value Format	De standaard waarde opmaak die geldt voor een term.	- Tekstveld - Geeft aan volgens welk format de term ingevuld dient te worden. - Kan ingevuld worden als reguliere expressie - Voorbeeld: Nederlandse Postcode: /^[1-9][0-9][0-9][0-9][A-Z][A-Z]\$/gm	Optioneel
Source	Geeft aan wat of wie (functie/rol) de bron is van de ingevulde metadata.	- Tekstveld - Kan een of meerdere bronnen bevatten - Kan verwijzen naar functies, rollen en/of systemen.	Optioneel

Vast te leggen metadata

Toelichting op de velden (5/5)

Metadata	Description	Guideline	Mandatory
Synonym	De relatie met terms die eenzelfde of soortgelijke definitie hebben binnen hetzelfde Data Domein.	• Heeft een relatie met andere term in Collibra. Dit vereist dat de term waarmee een relatie moet worden gelegd bestaat in Collibra. • Kan alleen bestaan met andere terms op dezelfde laag • Relatie dient zowel bij term A als term B aangegeven te worden	Optioneel
Relaties naar andere lagen	De relaties naar andere termen worden op aparte dia's uitgelegd.	• Heeft een relatie met andere term in Collibra. Dit vereist dat de term waarmee een relatie moet worden gelegd bestaat in Collibra.	Optioneel

Vast te leggen metadata

Een voorbeeld

- Een aantal voorbeelden van ingevulde metadata van de logische term Adres

The screenshot shows the 'Logical Layer Glossary' interface for the 'Adres' Data Entity. The interface includes a sidebar with icons for Definition, Data Manager, Data Domain, Context, and Entity type. The main content area displays the following information:

- Definition:** De verzameling van een (Huis)nummer (optioneel een Huisnummer-toevoeging) en Postcode; OF Een Huisnummer (optioneel een Huisnummer-toevoeging), Straatnaam en Woonplaatsnaam.
- Data Manager:** No value has been given yet. Double click or use the edit button.
- Data Domain:** No value has been given yet. Double click or use the edit button.
- Context:** Masterdata
- Entity type:** Logical

- Een term kan in verschillende verschijningsvormen een andere context hebben

☐	Adres	Het adres van het gebouw waar PostN...	Masterdata
☐	Adres	Dit object bevat alle unieke adressen ...	Business Partner
☐	Adres	De verzameling van een (Huis)numme...	Masterdata
☐	Adres	De aanduiding van een bepaalde locat...	Commercieel Product
☐	Adres	Een adres is de aanduiding van een be...	Masterdata
☐	Adres	Adres volgens Basis Administratie Geb...	Masterdata
☐	Adres	De verzameling van gegevens voor de ...	Business Partner

Logische term

Relatie *vanuit* de business laag

- Wanneer een Business term ook voorkomt op applicatie en systeem niveau (technische laag) dient dit aangegeven te worden met een relatie.
 - Voorbeeld: De zakelijke klant die te vinden is in Salesforce of de Retaillocatie die in BLS te vinden is.
- Het is niet mogelijk om een Business term direct te koppelen aan een Technische term (tabellen en velden). Hiervoor is de logische laag nodig. Hierin staan logische terms die als tussenstap dienen om de relatie te laten zien. Daarnaast kunnen er op de technische laag systeemvelden voorkomen die moeilijk te snappen zijn. In deze gevallen biedt de logische term de uitkomst.
- Een logische term kan hetzelfde zijn als een Business Term, Business Object of Data Object. Hiermee kan het zelfs zo zijn dat definities en andere metadata een-op-een gelijk zijn aan elkaar. Toch blijft, in ieder geval Architectuur eigenaar en verantwoordelijk voor Business- en Data objecten. Dit betreft dus ook de bijbehorende metadata.

Logische term

Relatie met technische termen op de technische laag

- Zoals op de vorige dia beschreven is het niet mogelijk om een Business term direct te koppelen aan een Technische term (tabellen en velden) (of andersom).
- Wanneer er een behoefte is om de technische implementatie op applicatie en systeem niveau (technische laag) inzichtelijk te maken dient er een relatie gecreëerd te worden op de logische laag.
 - Voorbeeld: De technische implementatie van een proceslocatie.
- Een logische term kan op de logische laag gekoppeld worden aan een technische velden en technische tabellen.
 - Data Entiteit is de logische tegenhanger van een tabel
 - Data Attribuut is de logische tegenhanger van een veld
- Pas wanneer er een systeem is ingeladen op de technische laag, kan deze in de data catalog aan de gerelateerde term (Entiteit of Attribuut) gekoppeld worden.

Relaties

Relatie met Enterprise Objecten

- Net als de business terms kunnen logische termen ook relatie hebben met een business object of een data object.
 - In deze gevallen krijgt de logische term dezelfde naam en/of definitie, maar blijft de verantwoordelijkheid ervoor bij de business.
 - In deze gevallen kan er vanuit de logische term gerefereerd worden naar het object waar deze op gebaseerd is.
 - Dit doe je door een relatie te leggen tussen de logische term en het object.
 - Ook wanneer de logische terms niet exact overeenkomen met de objecten, dan kan er alsnog gerefereerd worden naar een object.