

Guidelines voor de technische laag

Collibra Data Catalog

Juni 2023 – Versie 0.9

Inhoudsopgave

Inleiding

Opbouw PostNL Data Catalogus

Technische termen – basis principes, 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'.

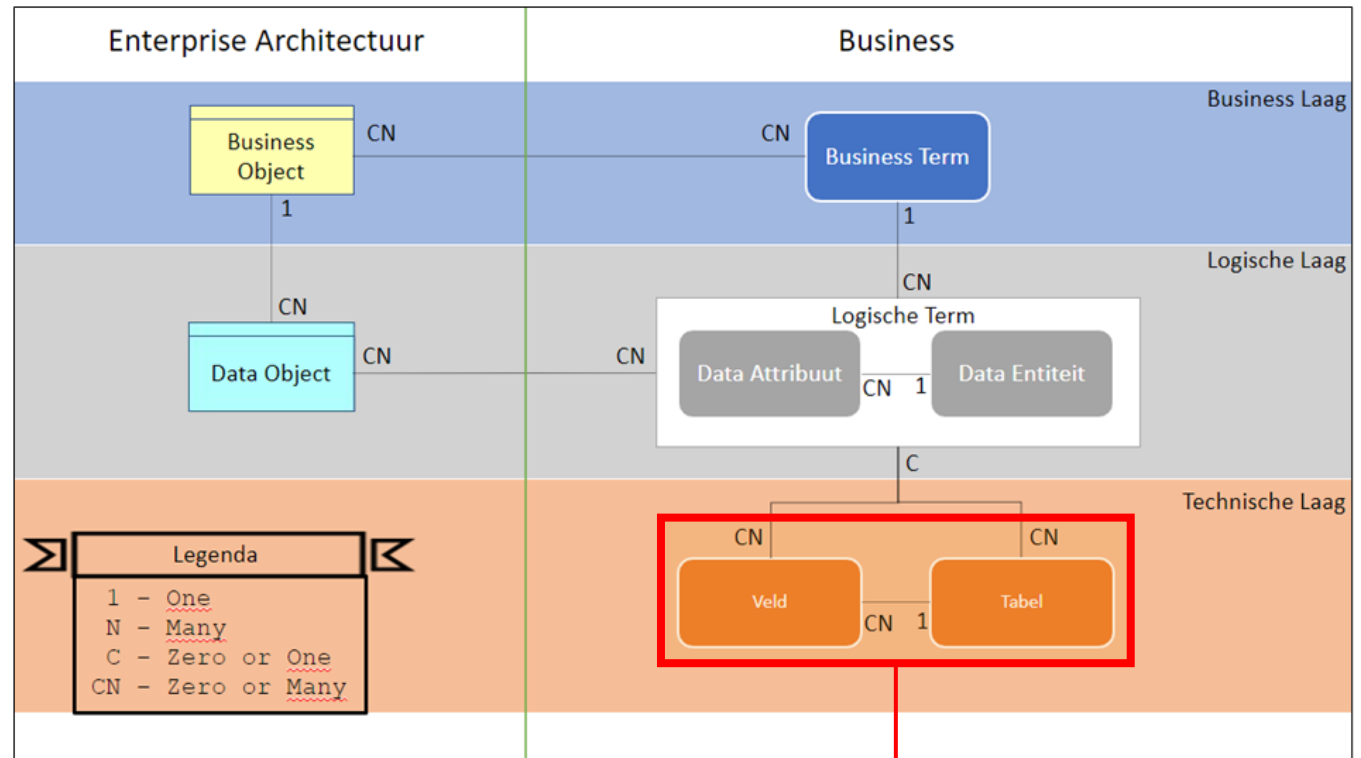
*“**Metadata Management** is de discipline die gaat over het verzamelen, onderhouden en standaardiseren van metadata (inclusief toegang en distributie)”.*

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 'Technische laag'. Oftewel, wat is een technische term en wat is de bijbehorende metadata?

Technische termen

Basis principes (1/2)

- De technische laag wordt ingeladen via het datalake. Het gaat hierbij om de tabellen en velden die op de fysieke data laag te vinden zijn.
- Ondanks dat de data via het datalake wordt ingeladen, is het een vereiste dat de structuur zoals deze op het datalake staat, overeenkomt met hoe deze in het System of Record staat. Op deze manier wordt de verbinding tussen de Data Catalog en het SoR gesimuleerd. Dit betekent bijvoorbeeld ook dat er geen tabellen en velden in voorkomen die naderhand zijn toegevoegd in het Datalake.
- De vulling van de technische laag (termen + metadata) komt voor het merendeel tot stand bij het technisch inladen, overerven of bij bulkmutaties. Ondanks dat, zal een gedeelte handmatig moeten gebeuren (o.a. het aangeven van de juiste privacy classificaties).
- Een technische term kan twee verschijningsvormen hebben: in de vorm van een tabel of in de vorm van een kolom.
- Samen vormen de technische termen de fysieke data die in een systeem of applicatie te vinden zijn. Het fysieke datamodel beschrijft de manier waarop de data in de database achter de applicatie wordt opgeslagen. Vaak is hier ook vanaf te leiden op welke manier de tabellen gekoppeld zijn aan elkaar of wat voor datatype er in de kolommen wordt verwacht.

Technische Metadata

Technische Metadata uit SoE/SoR van de Business term wordt vastgelegd. Dit betreffen de technische velden en tabellen, business ruling en technische definities.

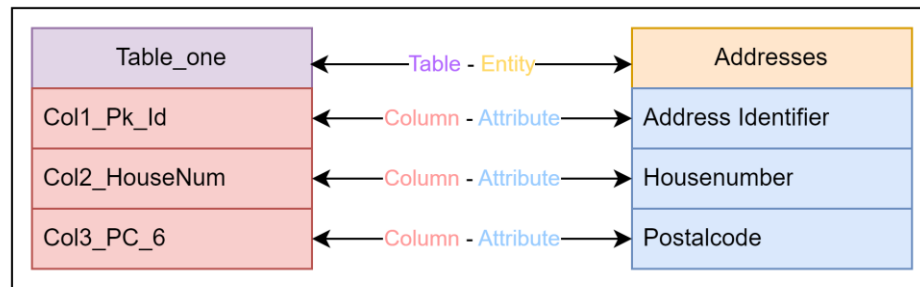
- Platform PIA
- Data steward
- Solution Consultant

Veld / Tabel

Technische termen

Basis principes (2/2)

- Technische termen krijgen systeembenamingen mee die vaak (te) technisch zijn voor de 'gewone' lezers. Om dat te versimpelen worden deze gekoppeld aan de logische termen van de logische laag.



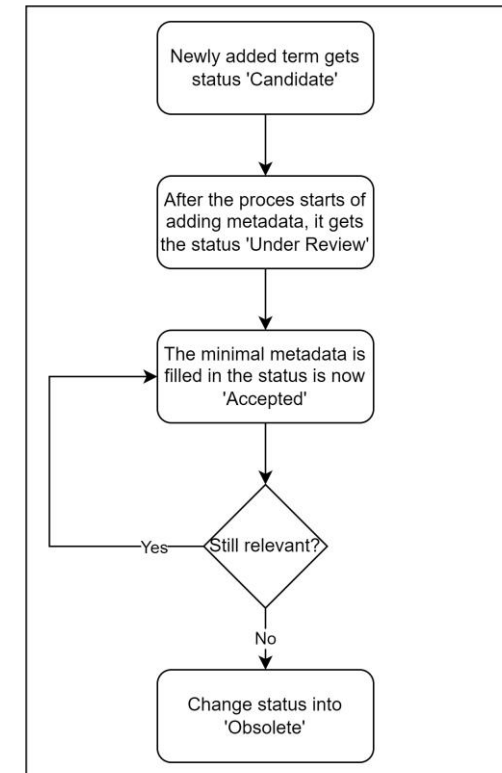
Bronnen kunnen handmatig of volgens een vast tijdschema geladen of gesynchroniseerd worden in de Data Catalog. Wanneer deze bronnen opnieuw worden gesynchroniseerd, wordt de reeds ingevuld metadata *niet* overschreven.

Technische 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.
Invalid	Buiten scope
In Progress	Buiten scope
Approval Pending	Buiten scope



Vast te leggen metadata

Afgestemd met de Data Management Organisaties

1	Definition
2	Explanation
3	Example
4	Data Domain
5	Platform
6	Platform Supplier
7	Cloud Service
8	Cloud Service Supplier
9	System Of Record
10	System of Entry
11	Data Labelling
12	Personally Identifiable Information
13	PII Type
14	CIA Classification
15	Confidentiality
16	Integrity
17	Availability
18	Business Ruling
19	Database Type
20	Datatype
21	Validation
22	Language
23	Standard Value Format
24	Date Created
25	Date Changed
26	Made/Changed by

27	Status
28	Source (origin metadata)
29	<i>Rel Logical Layer – Data Entity</i>
30	<i>Rel Logical Layer – Data Attribute</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 met de logische laag.
- De velden 15 t/m 17 zijn vooralsnog buiten scope.
- De velden 24 t/m 26 worden automatisch door Colibra zelf gegenereerd (standaard functionaliteit).

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

Vast te leggen metadata

Toelichting op de velden (1/6)

Metadata	Description	Guideline	Mandatory
Definition	De definitie die duidelijk de term in één à twee zinnen beschrijft.	• Tekstveld • Definitie bevat niet de term zelf • Volgt de meest gebruikelijke spelling en leestekens • Uniek voor de context waarin de term beschreven wordt.	Verplicht
Explanation	Een aanvulling op de definitie die duidelijk de term beschrijft.	• Tekstveld • Uitleg bevat niet de term zelf • Volgt de meest gebruikelijke spelling en leestekens • Uniek voor de context waarin de term beschreven wordt	Optioneel
Example	Voorbeelden van de term.	• Tekstveld • Voorbeeld bevat niet de term zelf • Volgt de meest gebruikelijke spelling en leestekens	Optioneel
Data Domain	Verantwoordelijke <i>domein</i> voor deze specifieke term	• Selectielijst met daarin het specifieke Data Domein die verantwoordelijk is voor deze specifieke term • Actueel overzicht • Indien in één en hetzelfde veld een waarde kan worden opgenomen die in meerdere data domeinen kan zitten (bijv. het e-mail adres van een consument en het e-mail adres van een business partner) geldt dat het data domein wordt ingevuld dat vanuit privacy perspectief als het 'meest risicovol' wordt gezien. In dit voorbeeld zou het veld email adres het data domein 'Commercie - Klant (consument)' krijgen.	Verplicht

Vast te leggen metadata

Toelichting op de velden (2/6)

Metadata	Description	Guideline	Mandatory
Provided by Platform	Geeft aan op welk platform het systeem ontwikkeld is.	Selectielijst met daarin alle systemen en applicaties die gebruikt worden binnen PostNL waarop ontwikkeld kan worden zijn gedefinieerd	Verplicht
Provided by Supplier	Geeft aan wie de ontwikkelaar van het platform is.	Selectielijst met daarin alle softwareleveranciers waarmee PostNL samenwerkt of heeft samengewerkt	Optioneel
Hosted on Platform	Geeft aan in welke cloud omgeving het systeem gehost wordt.	Selectielijst met daarin alle systemen en applicaties die gebruikt worden binnen PostNL waarmee gehost kan worden zijn gedefinieerd	Verplicht
Provided by Cloud service supplier	Geeft aan wie de leverancier is van de cloud omgeving.	Selectielijst met daarin alle cloud oplossingen waarmee PostNL samenwerkt of heeft samengewerkt	Optioneel
System of Record	Applicatie waarin dé waarheid wordt vastgelegd en van waaruit distributie naar ontvangende systemen en applicaties plaatsvindt.	- 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	Verplicht
System of Entry	Applicatie waarin de gegevens initieel worden opgevoerd c.q. 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	Verplicht

Vast te leggen metadata

Toelichting op de velden (3/6)

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.	Verplicht
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	Verplicht

Vast te leggen metadata

Toelichting op de velden (4/6)

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 (5/6)

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
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
Database type	Geeft aan wat voor soort structuurtype de database heeft.	• Selectielijst met daarin de waarden: Network Database, Object-oriented Database of Relational Database	
Data type	Geeft aan wat voor type of soort gegevens in een veld zijn toegestaan	• Selectielijst met daarin de waarden: Identifier, Tekst, Integer, String, Date, Indicator of Code	

Vast te leggen metadata

Toelichting op de velden (6/6)

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
Validatie	Geeft aan of een veld in de bron gevalideerd wordt tegen de waarheid.	• Boolean met de optie ja of nee.	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
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 technische term (column) 'waarnemingssoortcode'.

The screenshot shows a web interface for a 'Technical Layer Glossary'. The main entry is for 'waarnemingssoortcode', which is marked as a 'Column' and a 'Candidate'. It has a rating of five stars and zero reviews. A progress bar indicates 5% completion. The breadcrumb trail is: SYS Datalake > AwsDataCatalog > prod_ebx_spr > waarnemingssoort > waarnemingssoortcode. The entry includes three sections: 'Definition' (The unique combination of the WaarnemingSoort and the WaarnemingsoortReden), 'Explanation' (This unique combination represents a unique context from which the observation follows), and 'Example' (J01). A left sidebar contains navigation icons for home, search, and other functions.

Technical Layer > Technical Layer Glossary

waarnemingssoortcode

Column Candidate ☆☆☆☆☆ (0) 0 0 5%

SYS Datalake > AwsDataCatalog > prod_ebx_spr > waarnemingssoort > waarnemingssoortcode

Definition ⓘ

The unique combination of the WaarnemingSoort and the WaarnemingsoortReden

Explanation ⓘ

This unique combination represents a unique context from which the observation follows

Example ⓘ

J01

Relaties

Relatie met logische termen

- Het is 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 technische term kan op de logische laag gekoppeld worden aan een logische term (data entiteit of een data attribuut).
 - 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 logische term gekoppeld worden.