



Data Expert Team

Grand Design Landen

Datum	5/10/2019
Auteur	Data Expert Team
Rapport	Data Expert Team Grand Design Landen
Versie	0.6



Versiebeheer

Versie	Datum	Auteur	Bijzonderheden
0.1	05-03-2019	Hans Bottenberg	Initiële versie
0.2	15-03-2019	Hans Bottenberg	
0.3	18-03-2019	Hans Bottenberg	
0.4	21-03-2019	Hans Bottenberg	
0.5	09-04-2019	Hans Bottenberg	
0.6	10-05-2019	Mustafa Susam	Canonieke Export toegevoegd

Versie	Datum	Akkoord EDM	Akkoord Architect

Distributielijst

Persoon	Functie

Copyright PostNL BV. © te Den Haag

Niets uit deze uitgave mag verveelvoudigd en/of openbaar worden gemaakt (voor willekeurig welke doeleinden) door middel van druk, fotokopie, microfilm, geluidsband, elektronisch of op welke andere wijze dan ook zonder voorafgaande schriftelijke toestemming van PostNL BV. Dit rapport is enkel en alleen bedoeld voor intern gebruik voor hierboven genoemd(e) bedrijf/bedrijven.

Inhoudsopgave

1. Inleiding	4
1.1. Aanleiding	4
1.2. Huidige Situatie	4
1.3. Beoogde Situatie	5
1.4. Betrokken Partijen	5
2. Voorstel Data Expert Team	6
2.1. Uitgangspunten & Scope	6
2.2. Data	7
2.3. Functionaliteit (CRUD)	11
2.4. Databeheer Processen	13
2.5. Data Eigenaarschap	14
2.6. Datakwaliteitsmaatregelen	14

1. Inleiding

Vanuit verschillende business units is de behoefte voor een generieke referentietabel met feestdagen uitgesproken. Deze wens is ontstaan omdat het huidige systeem waarin de feestdagen worden onderhouden binnenkort uitgezet wordt.

Het uitwerken van het design voor landen is randvoorwaardelijk voor het realiseren van de tabel met feestdagen, en komt dus initieel uit die wens voort.

Landen zijn referentiegegevens. Deze kunnen gezien worden als corporate data en vallen onder de gedelegeerde verantwoordelijkheid van Enterprise Data Management.

Landen worden naast voor feestdagen ook gebruikt voor de tariefbepaling van bepaalde producten van PostNL en worden daartoe gegroepeerd in Zones en Zone Groepen. Daarmee is de scope van dit document breder dan alleen de landen zelf.

1.1. Aanleiding

De applicaties SAP ECC en SAP PI worden op dit moment uitgefaseerd. Het gevolg hiervan is dat er t.z.t. geen centrale feestdagenservice meer is binnen het PostNL landschap. Omdat feestdagen per land en regio zijn gemodelleerd, bracht dit de noodzaak tot het modelleren van landen en regio's.

Daarnaast is het beleid om referentietabellen centraal beschikbaar te stellen op het MDM platform Informatica voor algemeen gebruik. Dit is een aanvullende reden om het grand design Landen op zichzelf uit te werken en als referentiedata ter beschikking te stellen.

1.2. Huidige Situatie

De huidige implementatie is dat de landentabel in MDM 'decentraal' beheerd wordt. Dat wil zeggen: de tabel landen is opgeslagen in verschillende repositories en daar vindt separaat onderhoud op plaats. Het gaat om de volgende repositories:

- PBPR (PostNL Business Partner Repository)
- CMP (Centraal MDM Producten)
- BLS (BedrijfsLocatieSysteem)

Er is geen controle op consistentie tussen de verschillende landentabellen.

Afbeelding: tabellen in repository CMP

Landen	Zones	ZoneLandengroepSet
<i>Rowid Object</i>	<i>Rowid Object</i>	<i>Rowid Object</i>
Naam NL	Zone ID	ZonelandengroepsetID
Naam EN	ZoneOmschrijving NL	Zonelandengroepset Omschrijving NL
Naam DE	ZoneOmschrijving EN	Zonelandengroepset Omschrijving EN
Naam FR	ZoneOmschrijving DE	Zonelandengroepset Omschrijving DE
ISO-code	ZoneOmschrijving FR	Zonelandengroepset Omschrijving FR
ISO-code 2	Externe omschrijving NL	
Geldig van	Externe omschrijving EN	
Geldig tot	Externe omschrijving DE	
Wijzigingsindicator	Externe omschrijving FR	
	Status	

1.3. Beoogde Situatie

Architectuur stelt dat PostNL “in control” moet zijn en blijven van de eigen data en dus niet (volledig) afhankelijk mag zijn van externe partijen. Dat betekent o.a. dat PostNL wil niet afhankelijk mag zijn van extern geleverde API's.

Ons voorstel richt zich op het realiseren van een “universele” referentietabel die binnen deze kaders past. Voor de vulling van die tabel zullen wij ons richten op de volledige ISO landentabel. Indien mogelijk wordt deze tabel geautomatiseerd opgehaald uit een externe service, waarbij het te allen tijde mogelijk moet zijn om aan te sluiten op een andere externe bron die dezelfde gegevens kan leveren. Tevens moet PostNL op elk moment zelf de tabel kunnen aanpassen.

Bovenstaande geldt ook voor regio's. Een aanvullende eis voor de regiotabel is dat regio's die niet afkomstig zijn van ISO ook op elk moment opgevoerd moeten kunnen worden door PostNL.

Tot slot zijn ook Zones (verzameling landen) en Zone Groepen (verzameling zones) in scope om als referentietabellen aan te bieden. Zones en Zone groepen zijn PostNL specifieke objecten die niet ontleend worden aan externe bronnen.

Indien mogelijk zal een canoniek bericht ter beschikking gesteld worden door Informatica voor afnemers die er gebruik van willen maken.

Het resultaat zal zijn:

- Een generieke 'landenservice' met een tabel die alle door ISO erkende landen bevat, aangevuld met referentiedata over regio's, zones en zone groepen.

1.4. Betrokken Partijen

Data Expert Team EDM

Data Architect EDM (Nelleke van der Meer)

Data Engineer (Mustafa Susam)

Data Engineer (Seray Basgol)

Data Quality Analyst (Alwine van Aarle)

Data Architect EDM (Hans Bottenberg)

Domein Architect (Iddo Bakker)

Verskillende AC'ers (ten behoeve van in gebruik name)

Marinus Baan als Solution Consultant van beoogde oplossing.

Hot Item (ten behoeve van ontwikkeling en realisatie)

2. Voorstel Data Expert Team

2.1. Uitgangspunten & Scope

2.1.1. Uitgangspunten

Voor het ontwerp, de realisatie en het beheer van de referentietabel voor Landen nemen wij de volgende uitgangspunten:

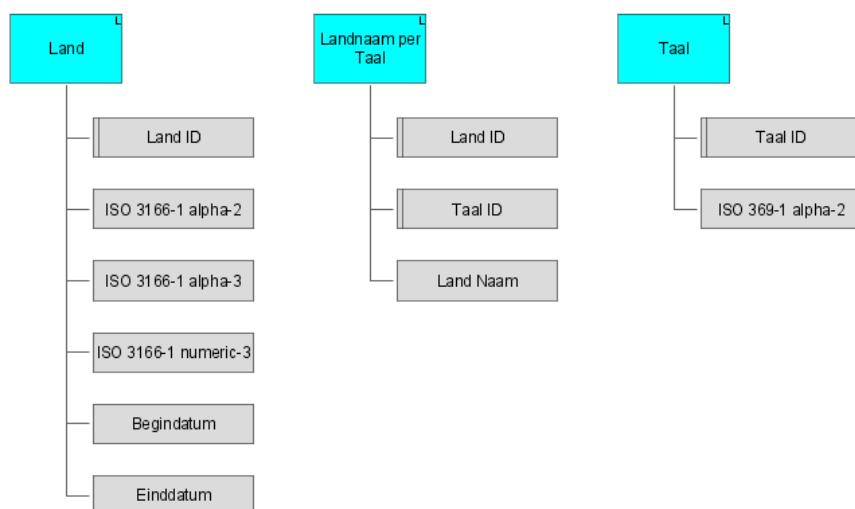
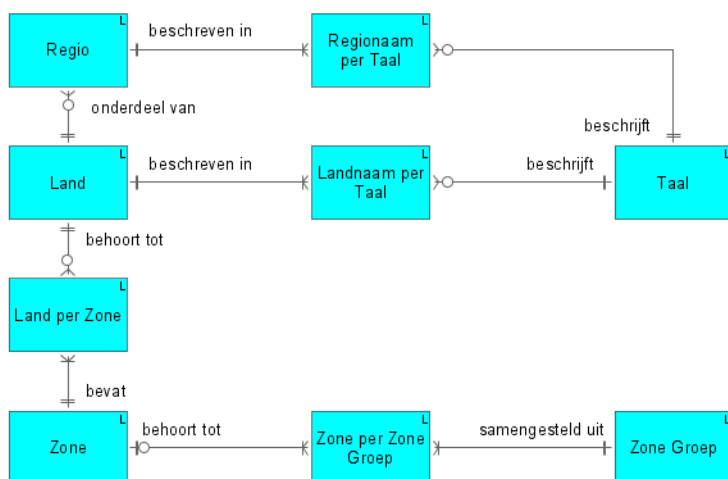
- PostNL steunt op de landentabel die vastgesteld is door de ISO in ISO standaard 3166-1 en voor de regiotabel op de ISO standaard 3166-2
- De gekozen solution is Informatica omdat dit in de architectuur het centrale masterdata platform is
- De landentabel en de regiotabel zijn corporate masterdata en het eigenaarschap valt daarom bij de Directeur Insights (gedelegeerd: Manager Enterprise Data Management)
- De referentietabellen worden ter beschikking gesteld aan afnemers middels een canoniek bericht
- PostNL is op elk moment in staat om een eventuele geautomatiseerde invoer van externe brongegevens te overrulen en op basis van eigen criteria de registratie aan te passen
- Zones zijn herbruikbaar over de business lines heen; dat wil zeggen dat Mail NL dezelfde zone zal gebruiken als Pakketten & Logistiek indien de gewenste samenstelling hetzelfde is.

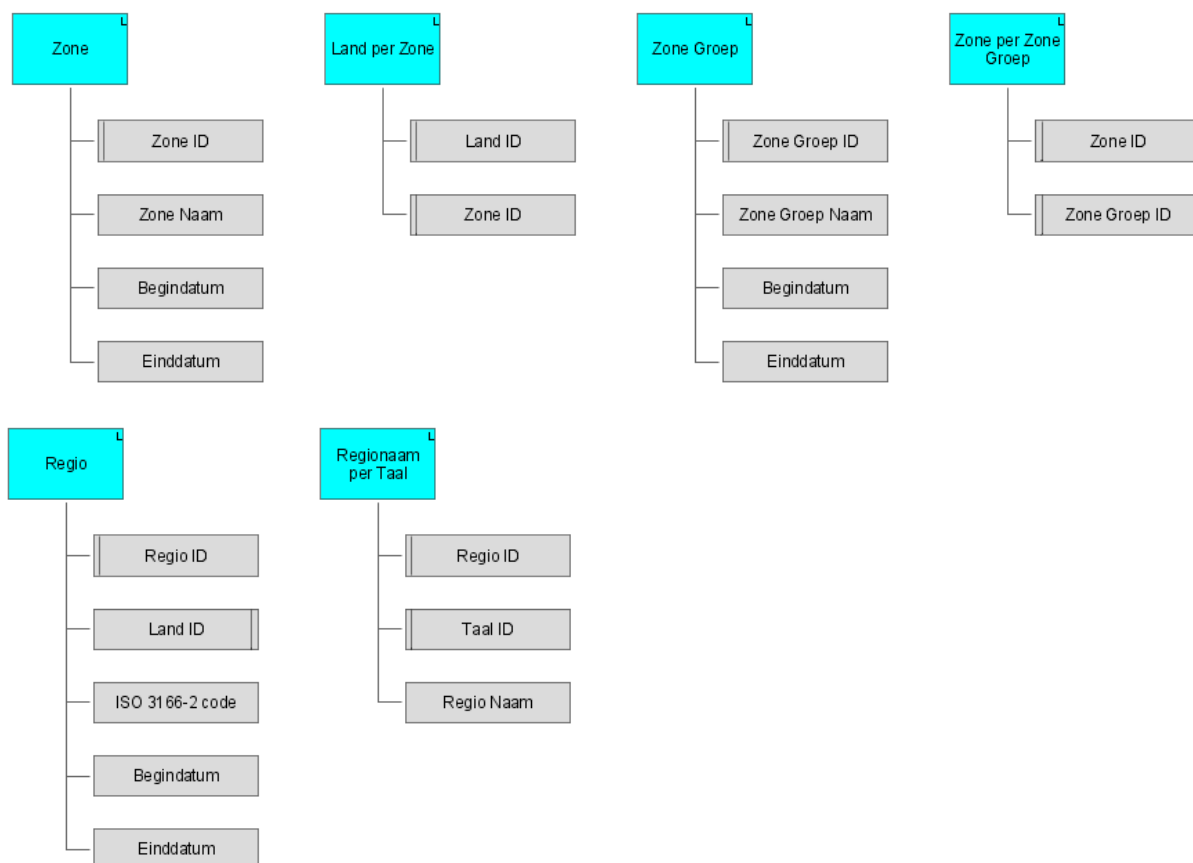
2.1.2. Scope

- Het koppelen van afnemende partijen is buiten scope
- De referentietabellen moeten universeel toepasbaar zijn
- Alle landen die erkend zijn door ISO worden opgenomen in de referentietabel
- Regio's, Zones en Zone Groepen zijn in scope van het LDM Landen
- Procedures rondom gegevensbeheer van de tabel zijn beschreven
- Procedures omtrent en toetsing van datakwaliteit zijn beschreven

2.2. Data

2.2.1. LDM & Attributen





Hieronder worden de definities van de attributen getoond:

Regio

Regiotabel op basis van de door ISO erkende 'onderverdelingen van landen ' in standaard 3166-2

Attributen:

ISO 3166-2 code	
Einddatum	
Regio ID	
Begindatum	
Land ID	Sleutelwaarde behorende bij het land

Land

Landentabel op basis van de door ISO erkende landen in standaard 3166-1

Attributen:

ISO 3166-1 alpha-3	ISO 3166-1 alpha-3 code behorende bij het land
ISO 3166-1 alpha-2	ISO 3166-1 alpha-2 code behorende bij het land

Begindatum	Beschrijft de datum vanaf wanneer het land
Land ID	Sleutelwaarde behorende bij het land
ISO 3166-1 numeric-3	ISO 3166-1 numeric-3 code behorende bij het land
Einddatum	

Land per Zone

Koppeltabel met alle geldige combinaties van land-zone

Attributen:

Land ID	Sleutelwaarde behorende bij het land
Zone ID	

Regionaam per Taal

Koppeltabel met regionamen per taal

Attributen:

Regio ID	
Taal ID	Sleutelwaarde behorende bij de taal
Regio Naam	

Zone

Tabel met zones die door PostNL erkend zijn

Attributen:

Begindatum	
Zone ID	
Einddatum	
Zone Naam	

Landnaam per Taal

Koppeltabel met land namen per taal

Attributen:

Taal ID	Sleutelwaarde behorende bij de taal
Land ID	Sleutelwaarde behorende bij het land
Land Naam	Naam van het land uitgedrukt in de betreffende taal

Zone per Zone Groep

Attributen:

Zone Groep ID	
---------------	--

Zone ID	
---------	--

Taal

Talen op basis van de door ISO erkende talen (ISO standaard 369-1)

Attributen:

Taal ID	Sleutelwaarde behorende bij de taal
ISO 369-1 alpha-2	

Zone Groep

Attributen:

Einddatum	
Zone Groep ID	
Begindatum	
Zone Groep Naam	

2.2.2. Gemaakte keuzes in LDM

- Hoewel de ISO drie soorten landcodes aanbiedt, is ervoor gekozen om een eigen sleutel te genereren. De ISO hanteert coderingsnormen die in de tijd kunnen veranderen (dat is in het verleden wel eens gebeurd), waardoor bijvoorbeeld de situatie kan ontstaan dat een landcode wijzigt. Het overnemen van de ISO landcode als sleutel is daardoor onwenselijk. Door wel alle drie landcodes op te nemen als attribuut kunnen wijzigingen in de tijd bijgehouden worden en kan PostNL daar indien nodig gericht op acteren.
- De regiotabel ontstaat uit ISO 3166-2, formeel genoemd *Codes for the representation of names of countries and their subdivisions*. Strikt genomen het betreft het hier 'delen' van een land, waarbij niet op voorhand gezegd kan worden of het provincies, staten of regio's zijn. Dat komt omdat elk land eigen soorten onderverdelingen kent. In sommige gevallen betekent het ook dat er meerdere soorten *subdivisions* in deze tabel voorkomen, die kunnen overlappen met elkaar. Zo is bijvoorbeeld België onderverdeeld in zowel regio's als provincies, waarbij provincies in dit geval een onderdeel van een regio zijn. Deze hiërarchie wordt door ISO niet verder gerepresenteerd in subtabellen.
- De ISO standaard 639 biedt verschillende varianten en codes. In het LDM is gekozen voor de ISO 639-1 code, die het meest gebruikt wordt voor 'algemene doeleinden', het meest herkenbaar is voor gebruikers, en meer dan 200 van de 'meest ontwikkelde' talen bevat. De ISO 639-1 is niet volledig, maar kent voldoende talen om bijvoorbeeld ook Fries en Catalaans te benoemen.
- De koppeltabellen kennen geen eigen primary keys; de combinatie van de foreign keys vormen tezamen een unieke sleutel.

2.3. Functionaliteit (CRUD)

2.3.1. Handmatige opvoer nieuwe objecten: Create

Bestaande objecten mogen niet opnieuw aangemaakt worden. Om dat te voorkomen dient onderstaande logica gevolgd te worden. In alle gevallen, met uitzondering van de koppeltabellen, genereert het systeem een eigen oplopende, numerieke sleutel (0001, 0002 etc.).

Land

- Bepaal aan de hand van de 'ISO 3166-1 alpha-2' code of het land voorkomt in de tabel Land
 - o Zo nee, creëer nieuw Land en vul alle attributen met uitzondering van de einddatum
 - o Zo ja, einde proces

Landnaam per Taal

- Bepaal aan de hand van de unieke combinatie van Land ID en Taal ID of het land al bekend is in de gewenste taal.
 - o Zo nee, voer de nieuwe unieke combinatie op en voer de naam van het land in de betreffende taal op
 - o Zo ja, einde proces

Regio

- Bepaal aan de hand van de 'ISO 3166-2 code' of de regio voorkomt in de tabel Regio
 - o Zo nee, creëer nieuwe Regio en vul alle attributen met uitzondering van de einddatum
 - o Zo ja, einde proces

Regionaam per Taal

- Bepaal aan de hand van de unieke combinatie van Regio ID en Taal ID of de regio al bekend is in de gewenste taal.
 - o Zo nee, voer de nieuwe unieke combinatie op en voer de naam van de regio in de betreffende taal op
 - o Zo ja, einde proces

Zone

- Bepaal of de Zone al bestaat a.d.h.v. de samenstelling van landen die de nieuwe zone moeten gaan vormen. Dit is vastgelegd in Land per Zone. Indien de samenstelling gelijk is aan die van een bestaande zone, dan is er sprake van dezelfde zone.
 - o Zo nee, creëer een nieuwe Zone
 - o Zo ja, einde proces

Land per Zone

- Bepaal aan de hand van de unieke combinatie van Land ID en Zone ID of het land al gekoppeld is aan de betreffende Zone.
 - o Zo nee, voer de nieuwe unieke combinatie op
 - o Zo ja, einde proces

Zone Groep

- Bepaal of de Zone Groep al bestaat a.d.h.v. de samenstelling van zones die de nieuwe Zone Groep moeten gaan vormen. Indien de samenstelling gelijk is aan die van een bestaande Zone Groep, dan is er sprake van dezelfde Zone Groep.
 - o Zo nee, creëer een nieuwe Zone Groep
 - o Zo ja, einde proces

Zone per Zone Groep

- Bepaal aan de hand van de unieke combinatie van Zone ID en Zonegroep ID of de zone al gekoppeld is aan de betreffende zonegroep.
 - o Zo nee, voer de nieuwe unieke combinatie op
 - o Zo ja, einde proces

Opvoer mapping

Er wordt voor landen gebruikgemaakt van de gratis te gebruiken resource voor landen op <https://restcountries.eu/>. De te gebruiken velden kunnen als volgt worden gemapt.

Veld in LDM	Veld in data bron
ISO 3166-1 alpha-2	alpha2Code
ISO 3166-1 alpha-3	alpha3Code
ISO 3166-1 numeric-3	numericCode
Land Naam	name

2.3.2. Handmatig updaten objecten: Update

Onderstaande attributen mogen handmatig aangepast worden door de data administrator en de data steward.

Entiteit	Attribuut	Verplicht/optioneel
Land	ISO 3166-1 alpha-2	V
	ISO 3166-1 alpha-3	V
	ISO 3166-1 numeric-3	V
	Einddatum	O
Landnaam per Taal	Naam Land	V
Regio	ISO 3166-2 alpha-2	V
	Einddatum	O
Regionaam per Taal	Naam Regio	V
Zone	Zone Naam	V
Zone Groep	Zone Groep Naam	V

2.3.3. Handmatig verwijderen objecten

Masterdata objecten mogen in regel nooit verwijderd worden; er mag alleen een einddatum toegekend worden. Voor die gevallen waarin er als gevolg van een fout een object in de tabel ontstaat die niet had moeten ontstaan, dient het mogelijk te zijn om het object volledig te verwijderen uit de tabel.

2.3.4. Geautomatiseerd exporteren gegevens

Alle records waarvan de begindatum vandaag of in de toekomst ligt, en de einddatum niet in het verleden ligt, dienen geëxporteerd te kunnen worden.

Om de export te kunnen maken zijn de volgende gegevens nodig:

Uit de Landen tabel: "ISO 3166-1 alpha-2", "ISO 3166-1 alpha-3" en "ISO 3166-1 numeric-3"

Uit de Landnaam per taal tabel: "Land Naam"

Uit de Regio tabel: "ISO 3166-2 code"

Uit de Regionaam per taal tabel: "Regio Naam"

Uit de Zone tabel: "Zone Naam"

Uit de Zone Groep tabel: "Zone Groep Naam"

Uit de Taal tabel: "ISO 369-1 alpha-2"

Het export bestand wat gemaakt moet worden is een XML bestand opgebouwd volgens de aangeleverde Canonieke XSD. De XSD geeft de structuur aan (en dient niet voor de validatie). De XSD is leidend voor het export bericht van deze repository. Begindatum/Einddatum in xsd's staan in formaat: YYYY-MM-DD

<<<Referentie naar canoniek XSD >>>

De naam van het XML export bestand heeft een bepaalde formaat. Het formaat is als volgt:

Landen_ jjjjmmdd_versienummer.xml

-Het eerste deel: Landen_ is een vaste tekst, en zal hetzelfde blijven voor alle bestanden.

-Het tweede deel is een datum om de export bestand uniek per export maken

-Laatste deel is een versienummer bestaande uit 2 cijfers beginnend met 01 en oplopen per export/dag indien er meerdere exports gemaakt worden per datum.

2.3.5. Datakwaliteit controleren

Het dient mogelijk te zijn om de volledige datasets in scope van dit design op geautomatiseerde of handmatige wijze uit te lezen of te exporteren, ten behoeve van datakwaliteitscontrole.

2.3.6. Autorisatiematrix

Functionaliteit	Data Administrator	Data Steward	Data Raadpleger
2.3.1. Handmatige opvoer	Muteren	Muteren	N.v.t.
2.3.2. Handmatig updaten	Muteren	Muteren	N.v.t.
2.3.3. Handmatig verwijderen	Muteren	Muteren	N.v.t.
2.3.4. Exporteren	Muteren	Muteren	N.v.t.
Overig: raadplegen tabellen	n.v.t.	n.v.t.	Raadplegen

2.4. Databeheer Processen

Per proces wordt de trigger, frequentie, rol, input, output en de gebruikte functionaliteit genoemd.

Land + Land per Taal

- Trigger: keuze om handmatig de lijst bij te werken
- Frequentie: ad hoc
- Input: ISO landcode tabel 3166-1 van <https://restcountries.eu/>
- Data Administrator bepaalt aan de hand van de ISO 3166-1 alpha-2 code of alle landen voorkomen in de tabel Land
 - o Zo nee, creëer voor elke constatering van een nieuwe ISO 3166-1 alpha-2 code een nieuw Land en vul alle attributen. Voer het land op in de tabel Land per Taal met de ISO naam.
 - o Zo ja, vergelijk per Land alle attributen met de ISO tabel en werk gegevens bij indien nodig.
- Indien gewenst: controle Data Steward

Zone + Land per Zone

- Trigger: de business wil een nieuwe zone laten opnemen.
- Frequentie: ad hoc

-
- Input: Zone Naam; ten minste 1 Land per Zone
 - Data Administrator voert Zone op en daarna ten minste 1 regel in Land per Zone
 - *Indien gewenst: controle Data Steward*

Zone Groep + Zone per Zone Groep

- Trigger: de business wil een nieuwe zone groep laten opnemen.
- Frequentie: ad hoc
- Input: Zone Groep Naam; ten minste 1 Zone per Zone Groep
- Data Administrator voert Zone Groep op en daarna ten minste 1 regel in Zone per Zone Groep
- *Indien gewenst: controle Data Steward*

Regio + Regio per Taal

- Trigger: keuze om handmatig de lijst bij te werken
- Frequentie: ad hoc
- Input: ISO tabel 3166-2 van <https://restcountries.eu/>
- Data Administrator bepaalt aan de hand van de ISO 3166-2 code of alle regio's voorkomen in de tabel Regio
 - o Zo nee, creëer voor elke constatering van een nieuwe ISO 3166-2 code een nieuwe Regio en vul alle attributen. Voer de regio op in de tabel Regio per Taal met de ISO naam.
 - o Zo ja, vergelijk per Regio alle attributen met de ISO tabel en werk gegevens bij indien nodig.
- *Indien gewenst: controle Data Steward*

2.5. Data Eigenaarschap

Landen zijn gegevens die voor PostNL als geheel van belang zijn als referentiedata. Landen vallen dus onder de categorie Corporate Master Data. De Directeur Insights is eigenaar van corporate data; EDM acteert als gedelegeerd eigenaar.

2.6. Datakwaliteitsmaatregelen

In het design zijn de volgende maatregelen genomen om de datakwaliteit te borgen:

- De benoemde databeheerprocessen beschrijven alle noodzakelijke stappen die genomen dienen te worden om te komen tot de juiste data invoer.
- Het logisch datamodel beschrijft de volgende constraints: beschrijving object, optionaliteit, cardinaliteit, datatype, primaire- en foreign keys.
- De functiescheiding tussen Data Administrator en Data Steward zoals beschreven in de autorisatiematrix maakt onafhankelijke controle op de data invoer mogelijk.
- Het beschikbaar stellen van een data export maakt het mogelijk om controles te doen op de datakwaliteit.