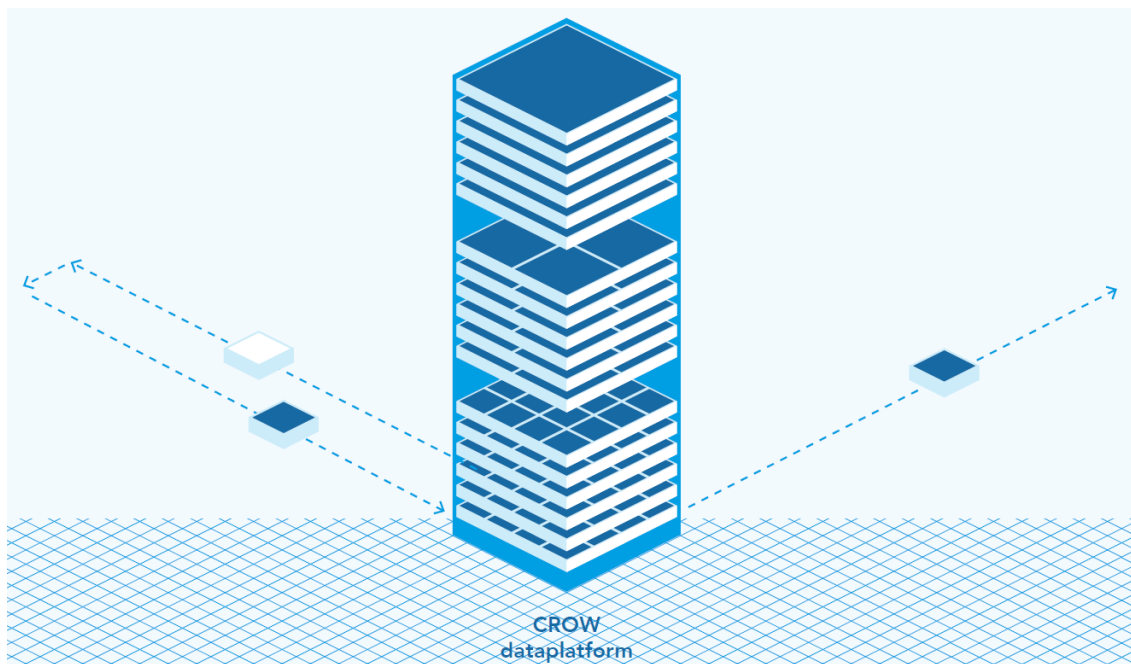


# Aanleverspecificaties

## CROW – Data Platform



Datum 29-11-2018  
Versie 2.0

**CROW**

Galvanistraat 1, 6716 AE Ede

Postbus 37, 6710 BA Ede

T 0318 69 53 00

E [crow@crow.nl](mailto:crow@crow.nl)

[www.crow.nl](http://www.crow.nl)

IBAN NL63INGB0667211446

BIC INGBNL2A

BTW NL823697733B04

HANDELSREGISTER 41150929

## Inhoudsopgave

1	Inleiding .....	3
1.1	Het publicatieproces .....	3
2	Definities .....	4
2.1	Algemeen .....	4
2.2	Definities van dit document .....	4
3	Wijze van aanlevering .....	5
4	Eisen aan de dataset .....	6
4.1	Eisen aan de inhoud .....	6
4.2	Eisen aan het model .....	6
4.3	Eisen aan de vorm .....	7
4.3.1	Optioneel: Eisen ten behoeve van een publicatiestructuur in de CROW Kennisbank .....	8

# 1 Inleiding

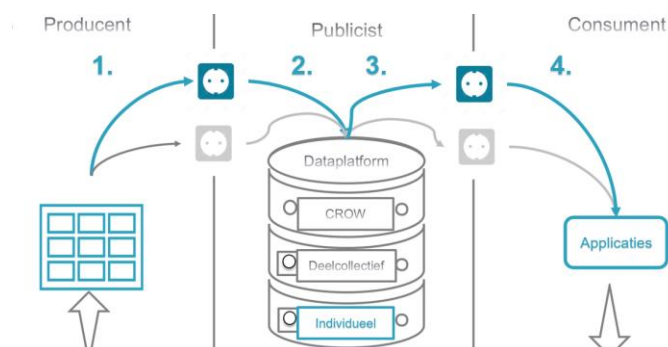
Dit document bevat de aanleverspecificaties voor het Dataplatform van CROW. Het maakt deel uit van de complete set aan documentatie die CROW beschikbaar heeft omtrent het CROW Linked Data Platform (verder: LDP). De documentatie omtrent het Data Platform bestaat uit een aantal documenten, waarbij elk document gericht is op een bepaalde stap in de publicatieproces.

*[ Aanleverspecificaties ] : ~ Voorwaarden waaraan een data(set) van een producent moet voldoen waaronder CROW de data(set) kan accepteren voor opname in het Data Platform*

Binnen het publicatieproces wordt er onderscheid gemaakt tussen drie rollen: (1) Producent, (2) Publicist en (3) Consument. Elke rol heeft zijn eigen taken/verantwoordelijkheden. In onderstaande tabel worden deze nader toegelicht.

## 1.1 Het publicatieproces

Op hoofdlijnen onderscheiden we 4 stappen in het publicatieproces:



1. **De producent levert een dataset aan bij de publicist.**
2. De publicist publiceert de dataset in het LDP.
3. De publicist zorgt ervoor dat de dataset beschikbaar is in het CROW Kennisbank en voor de relevante webservice(s)/endpoint(s).
4. De consument kan een eindgebruiker zijn of een software systeem die data uit het Data Platform haalt voor eigen gebruik. Dit kan bijvoorbeeld door het ophalen van data met zijn/haar software door middel van een webservice.

Dit document beschrijft de eisen die CROW stelt aan de kwaliteit van aan te leveren datasets bij CROW (=Stap 1). Informatie over andere stappen of onderdelen van het publicatieproces is opgenomen in andere delen van de documentatie. Hiervoor kunt u contact opnemen met CROW. Na publicatie van een dataset wordt door CROW de URI van de dataset gedeeld met de Producent.

## 2 Definities

### 2.1 Algemeen

Term	Definitie
<b>Class diagram</b>	Diagram dat de informatiestructuur weergeeft door middel van klassen, attributen en relaties.
<b>Semantiek</b>	De structuur van elementen en de betekenis van ieder element in die structuur.
<b>SKOS</b>	SKOS (Simple Knowledge Organization System) is een vakgebied dat zich bezig houdt met het ontwikkelen van specificaties en standaarden om het gebruik van kennisorganisatiesystemen (KOS) te ondersteunen, zoals thesauri, classificatie schema's, vakgroep systemen en taxonomieën binnen het framework van het semantisch web. SKOS biedt een standaard manier om kennisorganisatiesystemen te representeren waarbij gebruik wordt gemaakt van het Resource Description Framework (RDF).
<b>CSPEC</b>	CROW Specificatieschema, volgens dit schema worden de specificaties gestructureerd
<b>CDOC</b>	CROW Documentschema. Dit model beschrijft de publicatiestructuur voor de CROW Basisspecificaties en de resultaatsbeschrijvingen
<b>Structuur</b>	De opbouw van informatie in het bestand.
<b>XML</b>	Extensible Markup Language, een standaard van W3C om gestructureerde gegevens in platte tekst weer te kunnen geven

### 2.2 Definities van dit document

Term	Definitie
<b>LDP</b>	Linked Data Platform. Online database van CROW waarin datasets van zowel CROW, collectieven als individuele organisaties opgeslagen kunnen worden. De datasets in deze database kunnen beschikbaar worden gesteld aan externe gebruikers zodat zij tijd- en plaats onafhankelijk beschikking hebben over de laatste versie van die dataset.
<b>Rol Producent</b>	Eigenaar van een dataset: de organisatie die de dataset heeft ontwikkeld en aanlevert cq heeft aangeleverd bij de publicist. <b>Verantwoordelijk voor de juistheid van de data: kwaliteit, volledigheid, betrouwbaarheid .</b>
<b>Rol Publicist</b>	De organisatie die de dataset opslaat in het LDP en vanuit beschikbaar stelt aan consumenten. Verantwoordelijk voor (1) het publicatieproces en (2) voor de beschikbaarheid van de dataset.
<b>Rol Consument</b>	Eindgebruiker van de data: persoon (cq de software die hij/zij gebruikt) die gebruik maakt van een gepubliceerde dataset. Bijvoorbeeld door deze te bekijken of toe te passen bij het opstellen van een contractdocument. <b>Verantwoordelijk voor het verkrijgen en bezitten van juiste verificatiegegevens en/of CROW producten voor het ophalen/bevragen van de data.</b>

Tabel 1 Rollen & Verantwoordelijkheden publicatie derden

### 3 Wijze van aanlevering

CROW biedt op dit moment één methode om datasets aan te leveren. Dit is een aanlevering van een dataset **per e-mail**. In de toekomst zal dit via een webformulier worden geregeld via de website van CROW.

Elke dataset dient als los bestand bij de begeleidende e-mail gevoegd te worden en bestaat uit het volgende formaat:

- .ttl (Terse RDF Triple Language) Syntax om de dataset in RDF framework te beschrijven. Verdere beschrijving over deze manier van aanleveren staat in hoofdstuk 4.

In dezelfde e-mail moeten de volgende zaken worden benoemd:

- Contactgegevens van de volgende personen:
  - o Eigenaar dataset: algemeen aanspreekpunt voor CROW
  - o Technisch aanspreekpunt: voor vragen over de semantische structuur, modellering etc.
  - o Aanspreekpunt voor klantvragen: voor het doorsturen van vragen van eindgebruikers van die dataset.
- Indien het een nieuwe versie betreft van een eerder gepubliceerde dataset:
  - o Verwijzing naar de vorige versie van de dataset. **Let op:** ten behoeve van versiebeheer is het noodzakelijk om te vermelden wat de verschillen zijn tussen de versies.
- Indien de dataset zichtbaar gemaakt moet worden in de Kennisbank van CROW, dan zijn de volgende zaken van belang:
  - o Tekstuele beschrijving van de boomstructuur in de dataset.
  - o Naam van de verzameling welke gebruikt moet worden als bovenste tak in de boomstructuur (top-node). Zie de toelichting in Paragraaf 4.3.3.
  - o Beschrijving van de relaties die gebruikt zijn binnen de boomstructuur.

## 4 Eisen aan de dataset

### 4.1 Eisen aan de inhoud

De Producent van de data is verantwoordelijk voor de juistheid van de data (kwaliteit, volledigheid, betrouwbaarheid).

CROW stelt op dit moment enkel eisen aan de semantiek (4.2 Eisen aan het model) en structuur (4.3 Eisen ten behoeve het publicatiestructuur.)

### 4.2 Eisen aan het model

**Eis 1: Door CROW gepubliceerde model, CSPEC, dient als basis te worden gebruikt voor het definiëren van classes en relaties. Indien in de dataset classes en/of relaties voorkomen die afwijken, dan wordt de dataset niet meegenomen ter publicatie.**

Door CROW gepubliceerde model, CDOC, is optioneel om te gebruiken.

Voorbeeld:

In CSPEC en CDOC staan restricties m.b.t. de domain en range van de relaties die nauwkeurig gevold moet worden.<sup>1</sup>

De relaties en klassen in CSPEC en CDOC zijn beschreven in het volgende document: “Toelichting bij classes en relaties in CROW-datamodellen CSPEC en CDOC v1.0, dd 08-05-2018” (de actuele versie van dit document is op te vragen bij CROW).

In figuur 1 is het *class diagram* van CSPEC weergegeven. Dit *class diagram* beschrijft welke entiteiten bekend zijn in de CROW-datamodellen en beschikbaar voor instantiëring. De betekenis van de data is beschreven d.m.v. de relaties tussen de entiteiten.

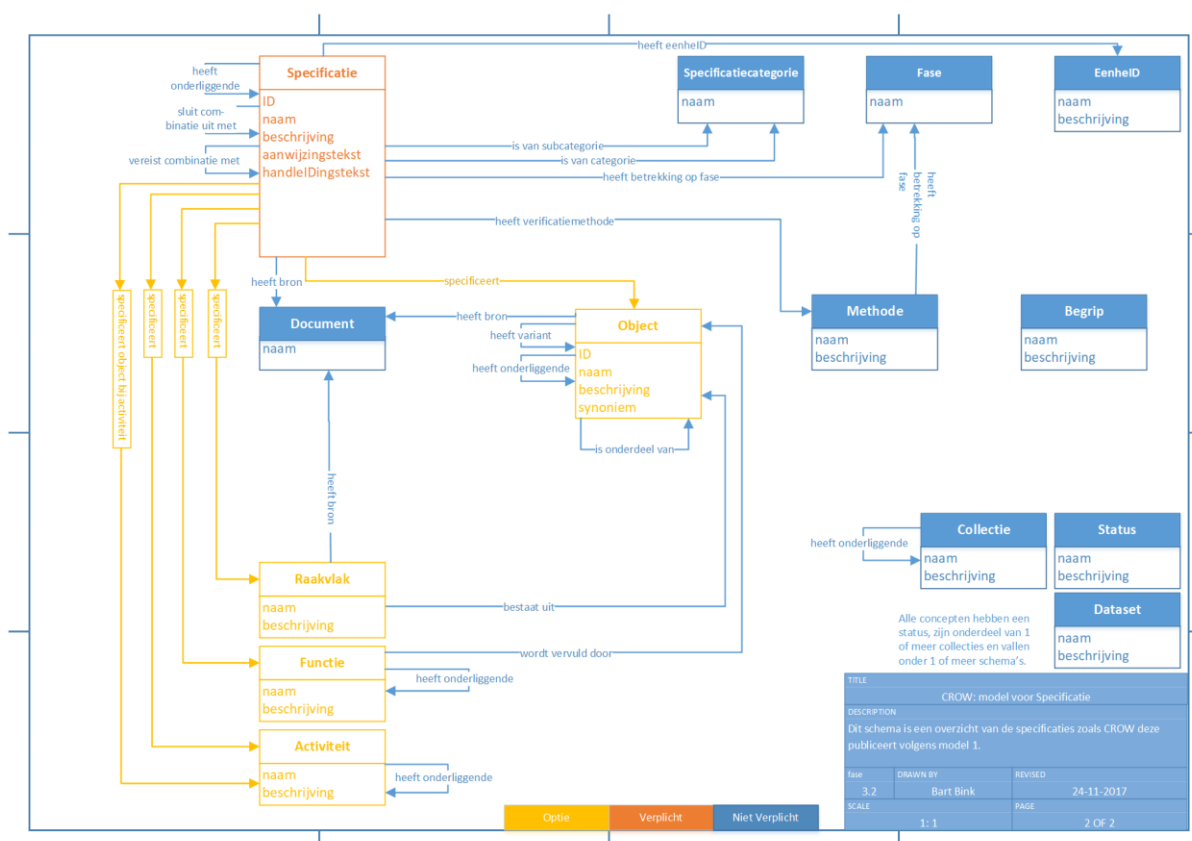
In het *class diagram* ligt tevens vast welke entiteiten en relaties **verplicht** zijn, welke **optioneel** zijn en welke **niet verplicht** zijn. De optionele klassen of relaties houden in dat er in ieder geval 1 of meer entiteiten **moeten** aangeleverd zijn.

---

<sup>1</sup> Domain, Ranges en restricties: Domain is wat er links staat van de relatie. En rechts staat de range van de relatie. De restrictie is de regel die geldt voor deze. De domain en ranges met daarbij behorende restrictie zijn terug te vinden in het CROW datamodel.

Eis 2: De volgende semantische modellering dienen minimaal in de dataset aanwezig te zijn:

- Alle specificaties c.q. eisen dienen geclassificeerd te zijn als **<cspec:Specificatie>**
- Elke instantie van **<cspec:Specificatie>** dient d.m.v. de relatie **<cspec:specificeert>** gekoppeld te zijn aan minimaal één van de volgende concepten:
  - **Cspec:Object**
  - **Cspec:Activiteit**
  - **Cspec:Functie**
  - **Cspec:Raakvlak**



Figuur 1 CROW Specificatieschema

### 4.3 Eisen aan de vorm

De dataset moet als separaat Turtle-bestand (.ttl) worden aangeleverd bij CROW (zie hoofdstuk 3 Wijze van aanlevering)

Het Turtle-bestand dient in lijn te zijn met de W3C Recommendation RDF 1.1 Turtle, gepubliceerd op <http://www.w3.org/TR/turtle/>.

#### 4.3.1 Optioneel: Eisen ten behoeve van een publicatiestructuur in de CROW Kennisbank

CROW biedt de mogelijkheid om een (in het LDP) gepubliceerde dataset op te nemen in de CROW Kennisbank zodat eindgebruikers deze dataset eenvoudig kunnen bekijken en kunnen raadplegen. Dit is een aanvullende dienstverlening vanuit CROW en valt buiten het publicatieproces van CROW. Echter, omdat veel producenten hier gebruik van willen maken en de voorbereidingen hiervoor al tijdens de productie van de dataset getroffen moeten worden, wordt hier in dit document toch aandacht aan besteed.

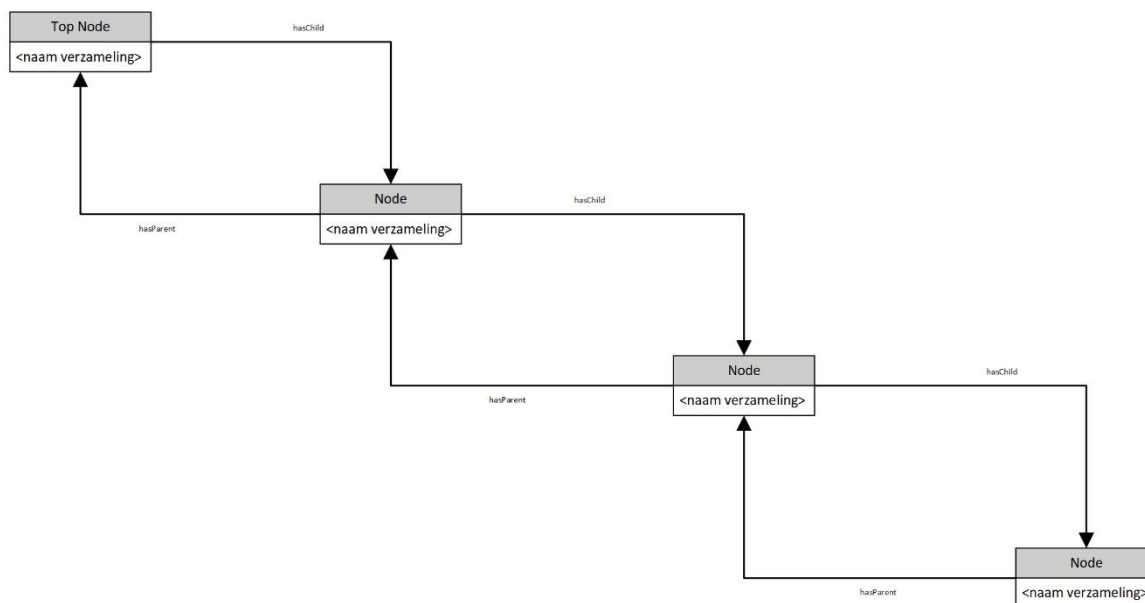
**De eisen in deze paragraaf zijn dus OPTIONEEL: deze zijn enkel relevant indien de dataset opgenomen moet worden in de CROW Kennisbank.**

Voor elke dataset die wordt gepubliceerd in het LDP en getoond moet worden in de CROW Kennisbank, is er door CROW een publicatiestructuur ontwikkeld. Dit is een hiërarchische boomstructuur waarmee de gebruiker doorklikt in de boom om tot een specificatie te komen. Deze publicatiestructuur wordt volledig gebaseerd op de informatie in de dataset en daarmee is de opbouw van de boomstructuur in de CROW Kennisbank ook volledig bepaald door de producent van de dataset.

Zoals eerder vermeld zijn de volgende zaken van belang:

- Tekstuele beschrijving van de boomstructuur in de dataset.
- Naam van de verzameling welke gebruikt moet worden als top-node (bovenste tak) in de boomstructuur en de onderliggende nodes.
- Beschrijving van de relaties die gebruikt zijn binnen de boomstructuur.

Onderstaande illustratie geeft het bovengenoemde schematisch weer.



*Figuur 2 Boomstructuur publicatie CROW Kennisbank*