

Bevindingen pilot attribuutbeschrijving & DCAT

Data.Overheid
Data.Eindhoven
2021



EINDHOVEN

Data@eindhoven.nl

Vullen attribuutbeschrijvingen



Voorkom open deuren!

Attribuut: status
Omschrijving: ~~status van de data~~

De volgende aspecten kunnen van belang zijn om mee te nemen bij het opstellen van attribuutbeschrijvingen.

- ⚡ Afkortingen [1], [2], [4]
- ⚡ Algemene context [1], [2], [3], [4], [5], [6]
- ⚡ Eenheden en format [2], [3], [4], [6]
- ⚡ Externe (informatie)bronnen [2], [3], [6]
- ⚡ Geometrie
- ⚡ Ontbrekende en standaard waardes [2], [3], [5], [6]
- ⚡ Schrijfstijl [1]

Let op: niet alle bovenstaande aspecten zijn voor elk type attribuut en/of elke beoogde gebruiker relevant. Maatwerk op basis van behoefte is noodzakelijk. Dit document kan hierbij een startpunt zijn.

(Volgorde van de aspecten is alfabetisch, niet op basis van prioriteit)

Afkortingen:

Schrijf afkortingen, codes, symbolen en acroniemen volledig uit en licht de betekenis en bron waar mogelijk toe.

Enkele veel voorkomende voorbeelden, inclusief toelichting:

- ⚡ **ID:** Uniek identificatie nummer voor <vul hier dataobject in>.
- ⚡ **BOR:** BOR staat voor Beheer Openbare Ruimte en heeft betrekking op het type gemeentelijke beheerder van <vul hier dataobject in>.
- ⚡ **Kadastrale gemeente codes:** In de dataverzameling <vul hier titel dataverzameling in> komen de volgende codes voor: EHV00, GTL01, STM00, STR03, TGR00, WSL01. Deze verwijzen naar de verschillende kadastrale gemeentes binnen Eindhoven. EHV00: Eindhoven, GTL01: Gestel, etc.

Algemene contex:

Het opnemen van context in een attribuutbeschrijving vergroot de vindbaarheid van data, en helpt gebruikers de betreffende dataverzameling inhoudelijk beter te begrijpen en dus gebruiken.

De volgende aspecten maken onderdeel uit van de context:

- ⚡ **Wat?** Wat is er gemeten / wordt er beschreven? Gebruik hiervoor relevante synoniemen.
- ⚡ **Waarvoor/Waarom?** Wat was het oorspronkelijke gebruiksdoel? Dit geeft inzicht in de oorspronkelijke scope en biedt inspiratie voor hergebruik.
- ⚡ **Hoe?** Hoe is het attribuut tot stand gekomen (wat is de bron / wijze van inwinning, en verdere bewerkingen), welke standaarden / definities / aannames zijn hierbij gevolgd, zijn de waardes onderworpen aan validatie of kwaliteitscontrole, en wanneer heeft dit plaatsgevonden? (Bijvoorbeeld: Vanaf welke hoogte is de breedte van de boomstam gemeten en met welke nauwkeurigheid? Is het bedrag inclusief of exclusief BTW? Is de waarde automatisch berekend of niet? Etc.).

Context is vaak lastig te achterhalen door de beheerder van het portaal, de datahouder dient deze aan te leveren (bijvoorbeeld via een data inventarisatie catalogus).

Eenheden en format:

Wanneer sprake is van getallen, kun je de volgende structuur gebruiken om deze inhoudelijk toe te lichten: <Onderwerp getal, bijvoorbeeld oppervlakte> van <onderwerp> in <eenheid, bijvoorbeeld vierkante meters [m²]>. Het betreft een <format van het veld, bijvoorbeeld integer / string / date-time veld>.

Verder kan je de volgende informatie vermelden (wanneer van toepassing):

- ⚡ Zijn de waardes afgerond?
- ⚡ Is er sprake van een bepaalde standaard ordening of sortering?
- ⚡ Hoe zit het format van het veld eruit. Bijvoorbeeld ISO 6801 format: YYYY-MM-DDTHH:MM:SS+HH:MM.
- ⚡ Voor velden met datum en tijd is het handig om de betreffende tijdszone te vermelden.

Ook is hier de **context** van belang: hoe zijn de getallen tot stand gekomen, en waarom in deze vorm?

Externe (informatie)bronnen:

Om gebruikers nog verder te ondersteunen bij het gebruik van de attributen, kunnen verwijzingen (idealiter middels URI's) naar externe (informatie)bronnen worden opgenomen. Denk hierbij aan:

- ⚡ Verwijzing naar externe bron van data.
- ⚡ Verwijzing naar scripts / code waarmee bewerkingen gedaan zijn.
- ⚡ Verwijzing naar informatiebronnen waar definities en referentiewaarden verder worden toegelicht.

Geometrie:

Wanneer sprake is van geometrie, kun je de volgende structuur gebruiken om deze inhoudelijk toe te lichten: Geometrie (<type geometrie, bijvoorbeeld punt of polygoon>) die de ligging van <onderwerp> beschrijft in <coördinaat referentiesysteem inclusief EPSG code>.

Als de geometrie in verschillende coördinaat referentiesystemen beschikbaar is, kan je hier naar verwijzen. Verder is hier de **context** van belang: hoe is de geometrie tot stand gekomen, en waarom in deze vorm?

Ontbrekende en standaard waarden:

Om onverwachts gedrag bij gebruik te verklaren (denk aan hoge gemiddeldes als 9999 is ingevuld wanneer waarden ontbreken), en om gebruikers te helpen bij het valideren van de waarden, is het wenselijk om toe te lichten hoe er om gegaan wordt met ontbrekende waarden, en of er bepaalde standaard waarden zijn.

Ontbrekende waarden: Kunnen velden leeg zijn of zijn bepaalde attributen juist verplicht?

Veel voorkomende waarden om ontbrekende waarden te duiden wanneer velden níét leeg gelaten worden zijn: 9999, 0, -1, of null.

Standaard waarden: Zijn er restricties aan de waarden?

- ⚡ Betreft het een vrije tekst veld of gaat het om een beperkt aantal toegestane waarden? Voeg waardenlijsten toe wanneer van toepassing.
- ⚡ Betreft het een veld met alle mogelijke getallen, of enkele toegestane ranges?
- ⚡ Zijn er maximale of minimale waarden?
- ⚡ Wanneer er sprake is van een telling, bij welke waarde is deze dan gestart (begonnen bij 0 of bij 1 bijvoorbeeld)?

Schrijfstijl:

De volgende algemene wenselijke aspecten aan schrijfstijl ten behoeve van instructies zijn ook voor attriboot beschrijvingen van toepassing:

- Actief schrijven met korte zinnen.
- Zowel inhoudelijk als qua structuur zoveel mogelijk consistentie waarborgen binnen een beschrijving, maar ook tussen verschillende beschrijvingen.
- Goed de gebruiker in beeld houden. Gaat het om een professional, dan is een andere toon passend dan voor een inwoner.

Signaleer je een gebruikersvraag?
Licht deze dan waar dat relevant is
toe in de attribuutbeschrijving.

Bronnen

1. [2018, T. Craven: Thesaurus Construction](#)
2. [2018, Smithsonian Data Management Best Practices](#)
3. [2017, P. Kononow: Captain Obvious' Guide to Column Descriptions - Data Dictionary Best Practices](#)
4. [2017, Axiom Data Science: Resource Content — documentation \(axiomdatascience.com\)](#)
5. [Data description in practice | Universiteit Utrecht \(uu.nl\)](#)
6. [How to Make a Data Dictionary – OSF Guides](#)

Voorbeelden:

Parkeerautomaten

Veldnaam:	Automaatnummer	Tariefcode	Mobiliteitstoeslag	GSM code	Betaaldagen	Koopavond
Beschrijving:	<p>Het automaatnummer is het unieke nummer van de parkeerautomaat. Deze staat op de automaat vermeld (op het beeldscherm), op de parkeertickets die door de automaat worden uitgegeven, én op het rekeningafschrift van de klant in geval van een pin-transactie.</p>	<p>De tariefcode is opgebouwd uit één letter en één cijfer. Deze geven aan welk tarief van toepassing is (cf. de tarieventabel uit de Verordening Parkeerbelastingen).</p>	<p>Starttarief, als er geen starttarief van toepassing is, is "nvt" ingevuld.</p>	<p>Dit is de code die parkeerders dienen te selecteren wanneer ze in de nabijheid van deze parkeerautomaat willen parkeren. Deze code is opgebouwd als "51"+ de eerste 2 cijfers van het automaatnummer.</p>	<p>Dagen dat parkeertarief geldig is.</p>	<p>Alleen ingevuld indien sprake is van een afwijkende tijdsperiode ten opzichte van de standaard betaaltijd.</p>
Type veld:	Integer	Tekst	Tekst	Integer	Tekst	Tekst
Voorbeeld inhoud:	9106	C2	0,30	5191	Maandag t/m vrijdag	Koopavond (vrijdagavond)

Hoogtebouten of NAP-bouten

Veldnaam:	Peilmerknnummer	Korte omschrijving	Datum meting	NAP hoogte [m]	Geo_point_2d
Beschrijving:	Dit geeft het nummer aan van de betreffende hoogtebout. NAP-merken die door de Adviesdienst Geo-Informatie en ICT van Rijkswaterstaat zijn gepubliceerd, bestaan uit een letter (B,D,E,G) plus cijfers. Als een nummer bestaat uit alleen cijfers, dan is het peilmerk door de Gemeente Eindhoven zelf geplaatst om het net te verdichten. Uiteraard is er bij het verdichten aangesloten op de hoogtemerken van Rijkswaterstaat	Hier is het object omschreven waarin het peilmerk zich bevindt. Zoveel mogelijk is bij huizen en gebouwen de straat genoemd met het huisnummer. In combinatie met de kaart en de korte omschrijving is het meestal niet moeilijk om het hoogtemerk terug te vinden in het veld	Deze kolom geeft het jaar en soms ook de datum van een waterpassing behorende bij de hoogte	Hier wordt de hoogte van de bovenkant van de bout weergegeven in meters. Ofwel, het hoogteverschil tussen het peilmerk en het NAP-vlak. Wanneer geen teken wordt vermeld, is dit verschil positief, het peilmerk ligt dan boven het NAP-vlak. Een peilmerk met een negatieve hoogte ligt onder het NAP vlak. Indien het gaat om een ingemetselde NAP steen, representeert het hart van de horizontale sleuf de weergegeven hoogte	Geometrie (punt) die de ligging van de hoogtebout beschrijft binnen de viewer van het portaal in WGS 1984, EPSG:4326. Ook op te halen als de oorspronkelijke RD-nieuw coördinaten
Type veld:	Tekst	Tekst	Datum	Decimaal	geo_point_2d
Voorbeeld inhoud:	G153	TRAFO PASTORIESTR 108	2012-05-13	19.618	[51.4187505012,5.47356754264]

Bonus: DCAT op een Opendatasoft (ODS) portaal

De reguliere DCAT velden op een ODS portaal zijn niet toereikend om de benodigde DCAT DO-NL velden op Data.Overheid te vullen. ODS biedt wel de mogelijkheid om middels een API een rechtstreekse mapping naar bepaalde applicatieprofielen te leveren op basis van de reguliere metadata vanaf het ODS portaal. Het mappen hoeft dan niet bij Data.Overheid te gebeuren.

Aangemaakt	<input type="text" value="01-01-2004"/>
Uitgegeven	<input type="text" value="24-04-2018"/>
Aanmaker	<input type="text" value="Gemeente Eindhoven"/>
Bijdrager	<input type="text" value="Gemeente Eindhoven"/>
Naam contactpersoon	<input type="text" value="Gemeente Eindhoven"/>
E-mail contactpersoon	<input type="text" value="data@eindhoven.nl"/>
Periodiciteit van de aangroei	<input type="text" value="Jaarlijks"/>
Ruimtelijk	<input type="text" value="Eindhoven"/>
Tijdelijk	<input type="text"/>
Granulariteit	<input type="text" value="1:5000-1:12500"/>
Kwaliteit van de data	<input type="text"/>
Type publisher	<input type="text" value="Gemeente Eindhoven"/>
Is in overeenstemming met	<input type="text"/>

ODS DCAT vullen

Bronnen: <https://resources.data.gov/resources/dcat-us/>,
<https://dcat-ap-donl.readthedocs.io/en/latest/classes-dataset.html>

- ⚡ **Aangemaakt:** Als onbekend, oudste datum uit dataverzameling genomen.
- ⚡ **Uitgegeven:** Eerste verschijnen van data op het portaal.
- ⚡ **Aanmaker, Bijdrager, Naam contactpersoon:** Gemeente Eindhoven.
- ⚡ **E-mail contactpersoon:** Algemeen adres om vanuit daar, waar nodig, vakafdelingen aan te kunnen schrijven.
- ⚡ **Periodiciteit van aangroei:** Verversingsfrequentie op basis van https://waardelijsten.dcat-ap-donl.nl/overheid_frequency.json.
- ⚡ **Ruimtelijk:** Gebied op basis van https://waardelijsten.dcat-ap-donl.nl/overheid_spatial_gemeente.json.
- ⚡ **Tijdelijk:** *“This field should contain an interval of time defined by the start and end dates for which the dataset is applicable.”* Vaak niet van toepassing, in dat geval niet gevuld.
- ⚡ **Granulariteit:** Resolutie / precisie: schaal of voor data op adresniveau “PostcodeHuisnummer”.
- ⚡ **Kwaliteit van de data:** *“Indicates whether a dataset conforms to the agency’s information quality guidelines.”* Voor nu leeg, eventueel te vullen op basis van BIV classificatie of indicatie geven voor systematische controle.
- ⚡ **Type publisher:** Type organisatie, op basis van https://waardelijsten.dcat-ap-donl.nl/overheid_spatial_scheme.json
- ⚡ **In overeenstemming met:** *“An established standard to which the described resource conforms. / URI that identifies the version of the Project Open Data schema being used.”* Ook gebruikt om te verwijzen naar aanverwante bronnen (documentatie, rapportages). Desondanks vaak niet van toepassing, in dat geval leeg gelaten.