



Centraal Bureau
voor de Statistiek

Handleiding

CBS Open data overstap van OData v3 naar OData v4

April 2020

Inhoudsopgave

1	Inleiding.....	3
2	OData 4 uitleg	4
3	Model verschil tussen OData versie 4 en OData versie 3	6
4	Overstappen van OData 3 naar OData 4	7
5	Inhoudelijke verschillen tussen OData 3 en OData 4	9

1. Inleiding

Het CBS publiceert veel cijfers in de tabellen op de databank StatLine, die te vinden is op de website van het CBS. Sinds 2014 zijn de datasets van StatLine ook te benaderen als open data via het portaal voor open data van het CBS en via het portaal van de overheid (data.overheid.nl).

Deze handleiding heeft betrekking op de nieuwste versie van open data, versie 4. Deze is als bètaversie beschikbaar.

OData 4 levert in beginsel precies dezelfde data en metadata als OData 3. Het verschil zit in een andere structuring van de data en metadata. Ook zullen er via OData 4 meer datasets beschikbaar komen dan alleen die van StatLine. Een belangrijk verschil is dat in OData 4 de data per cel aangeleverd worden, in plaats van per rij.

De opbouw van deze handleiding is als volgt. In hoofdstuk 2 wordt uitleg gegeven over de OData 4 service. Hoofdstuk 3 gaat in op de verschillen tussen de nieuwe en de oude service. In hoofdstuk 4 wordt een voorbeeld gegeven hoe je een oude link kunt omzetten. Tot slot zijn in hoofdstuk 5 de inhoudelijke verschillen beschreven.

De inhoud van deze handleiding is bedoeld voor data-analisten en app ontwikkelaars, die OData 3 query's willen omzetten naar OData 4 query's. Er wordt niet ingegaan op de specifieke functies van open data 4 query's. Voor ontwikkelaars die met deze functies query's willen maken met bijvoorbeeld Python of R zijn er handleidingen en voorbeelden in het portaal voor open data, onder de menu-optie 'Informatie voor ontwikkelaars'.

2. OData4 uitleg

De OData4 service is als volgt te benaderen:

<https://beta-OData4.cbs.nl/>

```
▼ @odata.context: "https://beta-odata4.apps.prod02.ap.cbsp.nl/$metadata"
▼ value:
  ▼ 0:
    name: "Catalogs"
    kind: "EntitySet"
    url: "Catalogs"
  ▼ 1:
    name: "Datasets"
    kind: "EntitySet"
    url: "Datasets"
```

Het endpoint bevat de toegang tot alle catalogs en tot alle datasets die via deze service worden aangeboden.

<https://beta-OData4.cbs.nl/Catalogs>

<https://beta-OData4.cbs.nl/Datasets>

De afzonderlijke catalogs zijn te vinden door de Catalog identifier op te nemen in de url:

<https://beta-OData4.cbs.nl/CBS/>

Via het eindpunt van de catalog is catalog informatie te vinden en een overzicht van alle datasets die bij deze catalog horen.

<https://beta-OData4.cbs.nl/CBS/Catalogs>

<https://beta-OData4.cbs.nl/CBS/Datasets>

De endpoints van de afzonderlijke datasets zijn te vinden via de identifier van de catalog en de identifier van de dataset.

<https://beta-OData4.cbs.nl/CBS/3723oned>

Let op: de identifiers zijn case sensitive.

Binnen de dataset heb je een aantal subsets. Het aantal subsets kan per dataset verschillen; het is afhankelijk van de structuur van de dataset.

Link	Uitleg
https://beta-odata4.cbs.nl/CBS/3723oned/MeasureGroups	Informatie over de onderwerp groep
https://beta-odata4.cbs.nl/CBS/3723oned/MeasureCodes	Informatie over onderwerpen en eenheid
https://beta-odata4.cbs.nl/CBS/3723oned/Dimensions	Informatie over de dimensies
https://beta-odata4.cbs.nl/CBS/3723oned/RegioSGroups	Informatie over de afzonderlijke dimensies met groep (parent) informatie en de afzonderlijke codelijsten per dimensie
https://beta-odata4.cbs.nl/CBS/3723oned/RegioSCodes	
https://beta-odata4.cbs.nl/CBS/3723oned/PeriodenGroups	
https://beta-odata4.cbs.nl/CBS/3723oned/PeriodenCodes	
https://beta-odata4.cbs.nl/CBS/3723oned/Observations	De waarden met code verwijzingen naar de dimensies en de measure
https://beta-odata4.cbs.nl/CBS/3723oned/Properties	Informatie over de dataset.

3. Model verschil tussen OData versie 4 en OData versie 3

De nieuwe op OData versie 4 gebaseerde service voor open data wijkt in een aantal opzichten af van de vorige versie. In onderstaande tabel is een voorbeeld van dataset 81298ned uitgewerkt.

OData3	overeenkomst	OData4	Toelichting
TableInfos	=	Properties	Algemene informatie dataset
UntypedDataSet	>	Observations	De daadwerkelijke cijfers
TypedDataSet			
DataProperties	<	Dimensions	Informatie over dimensies
		MeasureCodes	Onderwerpen
		MeasureGroups	Hiërarchie van onderwerpen
CategoryGroups	<		Hiërarchie van dimensies
WerldEUnNietEU	=	WerldEUnNietEUCodes	Codelijst
Perioden	=	PeriodenCodes	Codelijst
		PeriodenGroups	Hiërarchie informatie

Toelichting overeenkomsten en verschillen:

- De algemene informatie van een dataset (was TableInfos, is Properties) is uitgebreid met enkele velden.
- Het grote verschil tussen de Odata 3 implementatie en de Odata 4 implementatie van de open data services van het CBS is de overgang van record oriëntatie naar cel oriëntatie. In Odata 3 zijn de gegevens (waarden) per dataset opgeslagen in record vorm. Alle onderwerpen van een dataset staan naast elkaar in het record. Daarbij zijn er twee subsets, te weten UntypedDataSet (de waarden inclusief statistische symbolen) en TypedDataSet (de waarden zonder symbolen, direct geschikt voor gebruik in grafische toepassingen. In Odata 4 zijn de gegevens (waarden) per dataset opgeslagen in cel vorm. Er is sprake van één subset. Met het veld "ValueAttribute" kun je de statistische symbolen filteren. De onderwerpen staan onder elkaar en zijn op dezelfde manier te gebruiken als dimensies (via de optie filter).
- In OData3 kenden we de DataProperties met informatie over het record. In OData4 is dit opgedeeld in 3 subsets, waarbij de Measure (onderwerpen) zijn verdeeld in twee subsets, één subset met codes (en gegevens over eenheid en formaat) en één subset met hiërarchie informatie. De laatste hoeft niet altijd voor te komen.
- De afzonderlijke dimensies kennen een subset met code lijsten en een subset met hiërarchie informatie. Beide zijn niet verplicht.
- In OData4 hebben de dimensies van type Geodimension, Gedetail en TimeDimenson een subset Groups met hiërarchie informatie erbij.

4. Overstappen van OData 3 naar OData 4

In OData3 zijn de gegevens (waarden) per dataset opgeslagen in record vorm. Alle onderwerpen van een dataset staan naast elkaar in het record. Daarbij zijn er twee subsets, te weten UntypedDataSet (waarden inclusief statistische symbolen) en TypedDataSet (waarden zonder symbolen, direct geschikt voor gebruik in grafische toepassingen).

In OData4 zijn de gegevens (waarden) per dataset opgeslagen in cel vorm. Er is sprake van één subset. Met het veld "ValueAttribute" kun je de statistische symbolen filteren. De onderwerpen staan onder elkaar en zijn op dezelfde manier te gebruiken als dimensies (via de optie filter).

Voorbeeld OData3 met UntypedDataSet:

[https://opendata.cbs.nl/ODataapi/OData/81298ned/UntypedDataSet?\\$filter=WerldEUnNietEU eq 'T001047' and \(Perioden eq '2008MM01' or Perioden eq '2008MM02'\)&\\$select=WerldEUnNietEU, Perioden, Invoerwaarde_1, Uitvoerwaarde_2](https://opendata.cbs.nl/ODataapi/OData/81298ned/UntypedDataSet?$filter=WerldEUnNietEU eq 'T001047' and (Perioden eq '2008MM01' or Perioden eq '2008MM02')&$select=WerldEUnNietEU, Perioden, Invoerwaarde_1, Uitvoerwaarde_2)

levert:

```
{
  "odata.metadata": "https://opendata.cbs.nl/ODataapi/OData/81298ned/$metadata#Cbs.OData.WebAPI.UntypedDataSet&$select rde_2", "value": [
    {
      "WerldEUnNietEU": "T001047", "Perioden": "2008MM01", "Invoerwaarde_1": "28099547", "Uitvoerwaarde_2": "31024727"
    }, {
      "WerldEUnNietEU": "T001047", "Perioden": "2008MM02", "Invoerwaarde_1": "27899694", "Uitvoerwaarde_2": "31068191"
    }
  ]
}
```

Vergelijkbaar voorbeeld OData4:

[https://beta-OData4.cbs.nl/CBS/81298ned/Observations/?\\$filter=Measure in \('Moo1607', 'Moo1636'\) and WerldEUnNietEU eq 'T001047' and Perioden in \('2008MM01', '2008MM02'\)](https://beta-OData4.cbs.nl/CBS/81298ned/Observations/?$filter=Measure in ('Moo1607', 'Moo1636') and WerldEUnNietEU eq 'T001047' and Perioden in ('2008MM01', '2008MM02'))

```
{"@odata.context": "https://odata4-a.cbs.nl/CBS/81298ned/$metadata#Observations", "value":
  [{"Id": 0, "Measure": "M001607", "ValueAttribute": "None", "Value": 28099547.0, "WerldEUnNietEU": "T001047", "Perioden": "2008MM01"},
  {"Id": 2, "Measure": "M001607", "ValueAttribute": "None", "Value": 27899694.0, "WerldEUnNietEU": "T001047", "Perioden": "2008MM02"},
  {"Id": 1, "Measure": "M001636", "ValueAttribute": "None", "Value": 31024727.0, "WerldEUnNietEU": "T001047", "Perioden": "2008MM01"},
  {"Id": 3, "Measure": "M001636", "ValueAttribute": "None", "Value": 31068191.0, "WerldEUnNietEU": "T001047", "Perioden": "2008MM02"}]}
```

De doorgevoerde wijziging is:

Oud : \$select=WerldEUnNietEU, Perioden, Invoerwaarde_1, Uitvoerwaarde_2

Nieuw: \$filter=Measure in ('Moo1607', 'Moo1636')

Toelichting:

- De afzonderlijke dimensies (WerldEUnNietEU, Perioden) hoeft je niet op te geven.
- De identifiers 'Moo1607', 'Moo1636' zijn nieuw en te vinden in dataset 900002NED.

Voorbeeld om deze identifiers op te halen

[https://beta-OData4.cbs.nl/CBS-asd/900002NED/Observations/?\\$filter=DatasetIdentifier eq '81298ned' and OData3Identifier in \('Invoerwaarde_1', 'Uitvoerwaarde_2'\)&\\$select=DatasetIdentifier, OData4Identifier, OData3Identifier](https://beta-OData4.cbs.nl/CBS-asd/900002NED/Observations/?$filter=DatasetIdentifier eq '81298ned' and OData3Identifier in ('Invoerwaarde_1', 'Uitvoerwaarde_2')&$select=DatasetIdentifier, OData4Identifier, OData3Identifier)

```
{"@odata.context": "https://beta-OData4.cbs.nl/CBS-asd/900002NED/$metadata#Observations(DatasetIdentifier,OData4Identifier,OData3Identifier)", "value": [{"DatasetIdentifier": "81298ned", "OData4Identifier": "Moo1636", "OData3Identifier": "Uitvoerwaarde_2"}, {"DatasetIdentifier": "81298ned", "OData4Identifier": "Moo1607", "OData3Identifier": "Invoerwaarde_1"}]}
```

Gebruik Value attribute.

Het CBS kent statistische symbolen. Bij alle recente tabellen is het aantal symbolen beperkt. De huidige set aan symbolen omvat:

Symbol	Beschrijving	Waar te vinden
niets (blanco)	het cijfer kan op logische gronden niet voorkomen	In Observations, ValueAttribute = 'Impossible'
.	het cijfer is onbekend, onvoldoende betrouwbaar of geheim	Dit gegeven wordt weggelaten. Ontbrekende cijfers zitten niet in Observations.
*	voorlopige cijfers	Veld Status, onderdeel van de TimeDimension
**	nader voorlopige cijfers	Veld Status, onderdeel van de TimeDimension

Als je de waarden logisch onmogelijk wilt uitsluiten, is de eenvoudigste manier om het \$filter statement uit te breiden met ValueAttribute eq 'None'.

In het voorbeeld betekent dit:

[https://beta-OData4.cbs.nl/CBS/81298ned/Observations/?\\$filter=Measure in \('Moo1607', 'Moo1636'\) and WereldEUnNietEU eq 'T001047' and Perioden in \('2008MM01', '2008MM02'\) and ValueAttribute eq 'None'](https://beta-OData4.cbs.nl/CBS/81298ned/Observations/?$filter=Measure in ('Moo1607', 'Moo1636') and WereldEUnNietEU eq 'T001047' and Perioden in ('2008MM01', '2008MM02') and ValueAttribute eq 'None')

In de OData3 versie bevatte de UnTypedDataset de symbolen blanco en "." (punt). In de TypedDataset was "." (punt) vervangen door een blanco.

In het verleden (datasets voor 2015) hanteerde het CBS meer statistische symbolen.

Symbol	Beschrijving	Waar te vinden
niets (blanco)	het cijfer kan op logische gronden niet voorkomen	In Observations, ValueAttribute = 'Impossible'
.	het cijfer is onbekend, onvoldoende betrouwbaar of geheim	Dit gegeven wordt weggelaten. Ontbrekende cijfers zitten niet in Observations.
x	Geheim	In Observations, ValueAttribute = 'Secret'
-	nihil	In Observations, ValueAttribute = 'Zero'. In TypedDataset had dit de waarde o.
*	voorlopige cijfers	Veld Status, onderdeel van de TimeDimension
**	nader voorlopige cijfers	Veld Status, onderdeel van de TimeDimension

Als je bij oude datasets ook de waarde nihil wilt meenemen. Kun je de \$filter uitbreiden met ValueAttribute in ('None', 'Zero'). Voor recent gepubliceerde of bijgewerkte datasets speelt dit niet; het komt alleen voor in oudere datasets die na 2015 niet meer vernieuwd zijn..

5. Inhoudelijke verschillen tussen OData3 en OData4

Identifiers bij MeasureCodes

In Odata 4 zijn een aantal inhoudelijke veranderingen door gevoerd. De identifier van Measure (voorheen Topic) is gewijzigd. Bij het toekennen van de identifiers wordt de volgende systematiek gebruikt:

DatasetIdentifier	OData4Identifier	OData3Identifier	Title
70072ned	M0000352	TotaleBevolking_1	Totale bevolking
70072ned	3000	Mannen_2	Mannen
70072ned	4000	Vrouwen_3	Vrouwen
70072ned	70200_1	k_5Tot10Jaar_5	5 tot 10 jaar
70072ned	70200_2	k_5Tot10Jaar_14	5 tot 10 jaar
7379vt	X000222	TeZamenWonende_4	Te zamen wonende

Uitleg:

- Identifiers die beginnen met een **M** zijn variabelen. Over datasets heen worden zoveel mogelijk dezelfde identifiers gebruikt.
- Identifiers beginnend met een **cijfer, A, B of T** zijn classificaties. Deze codes kun je ook tegenkomen bij dimensies.
- Identifiers beginnend met **X** staan in gearcheverde tabellen. Bij gearcheverde tabellen zijn de identifiers generiek toegekend; deze identifiers zijn niet over verschillende datasets te combineren.
- Als er een **_1** (of een ander volgnummer) voorkomt, betekent het dat het item vaker voorkomt in de dataset. Via de bijbehorende MeasureGroup kun je nadere duiding vinden.

Met het toekennen van identifiers hopen we dat het makkelijker wordt om informatie te linken.

Identifiers bij dimensies

In OData3 was het mogelijk dat identifiers spaties bevatten. Voorbeeld:

[https://opendata.cbs.nl/ODataapi/OData/70072ned/RegioS?\\$top=1](https://opendata.cbs.nl/ODataapi/OData/70072ned/RegioS?$top=1)

Key "NL01 "

En in OData4 zijn de spaties verwijderd uit de identifiers:

[https://beta-OData4.cbs.nl/CBS/70072ned/RegioSCodes?\\$top=1](https://beta-OData4.cbs.nl/CBS/70072ned/RegioSCodes?$top=1)

Identifier "NL01"

Dataset met GeoDimension of met GeoDetail informatie

Bij datasets met een GeoDimension dimensie (regio-indelingen tot het niveau van gemeenten) zijn de measures (voorheen topics) van het formaat "String" uit de dataset gehaald en in één nieuwe dataset (900003NED) samengebracht. Ditzelfde geldt voor datasets met GeoDetail dimensie (regio-indelingen wijken en buurten). Ook deze informatie is in dataset 900003NED ondergebracht. Alle measures zijn van het type *numeric*. Deze informatie is te vinden in de subset Properties in het veld DatasetType.

Naast *numeric* kennen we het DatasetType *string* als er sprake is van alleen measures van het formaat string. Tot slot kennen we het DatasetType *Hybride*. Hier kunnen zowel string als numeric measures in zitten. Dit laatste type datasets is heel lastig te gebruiken in Excel en andere Odata 4 tools.

Lege Cellen

Zoals eerder ook is aangegeven, worden lege cellen niet langer meegegeven in de service

DCAT-AP-NL 1.1

Het CBS heeft zich geconformeerd aan het DCAT-AP-NL 1.1 model. Op sommige plaatsen is de informatie uitgebreid. Ook zijn namen aangepast.

Voorbeeld OData4:

[https://beta-OData4.cbs.nl/CBS/83131NED/\\$metadata](https://beta-OData4.cbs.nl/CBS/83131NED/$metadata)

Voorbeeld OData3:

[https://opendata.cbs.nl/ODataApi/OData/83131NED/\\$metadata](https://opendata.cbs.nl/ODataApi/OData/83131NED/$metadata)