



Downloaden Iv3 open data via het dataportaal

23 maart 2016

Inhoudsopgave

1. Inleiding	3
2. Opzoeken Iv3 open datatabel via het portaal	4
3. Downloaden data uit de Iv3 open datatabel	7
3.1 Downloaden metadata	7
3.2 Maken selectie	8
3.3 Downloaden data	10

1. Inleiding

Deze handleiding is bedoeld voor het downloaden van lv3 data via het lv3 open dataportaal. In deze handleiding wordt aan de hand van een voorbeeld, het downloaden van alle data van een verslagjaar van één berichtgever, de werking van het dataportaal uitgelegd. Het voorbeeld resulteert in het downloaden van de data naar een bestand in CSV-formaat.

Alle lv3 data zijn per verslagjaar per overheidslaag gegroepeerd in een tabel. De tabellen voor gemeenten, gemeenschappelijke regelingen, provincies en waterschappen bevatten echter meer datacellen dan het maximale aantal van één miljoen dat via het dataportaal kan worden gedownload. Via het dataportaal alle data downloaden van alle berichtgevers voor een verslagjaar (met uitzondering van de provincies) is dan ook niet mogelijk. NB voor bijvoorbeeld het downloaden van de jaarrekeningen van een bepaald jaar van alle gemeenten biedt CBS ondersteuning via een script voor het statistisch programma R met daarbij een handleiding. Meer informatie hierover is te vinden via de 'lv3 open data'-pagina, die te vinden is via de zoekterm lv3+open+data op www.cbs.nl.

Door te filteren is het wel mogelijk om via het dataportaal een selectie van data te downloaden, bijvoorbeeld voor maximaal 7 berichtgevers alle data van een verslagjaar, of voor 20 berichtgevers alle data van de jaarrekeningen, of voor alle berichtgevers alle data voor 15 functies voor de jaarrekeningen.

Deze handleiding gaat niet in op de betekenis van de te downloaden variabelen oftewel de variabelen in de lv3-modellen, op basis waarvan CBS de data van berichtgevers krijgt aangeleverd. Meer informatie over deze modellen is te vinden via www.cbs.nl/kredo.

Tot slot, de lv3 data is via open data beschikbaar in een technische structuur, die gelijk is aan de structuur waarin CBS al haar data via open data beschikbaar stelt. Voor meer technische achtergrond over deze structuur en de wijze waarop CBS open data beschikbaar stelt, wordt verwezen naar de algemene handleiding over CBS open data services:

- <http://www.cbs.nl/NR/rdonlyres/2561A2B7-CE51-47B9-A838-9968EF67FFB0/0/2014handleidingcbsopendataservices.pdf>

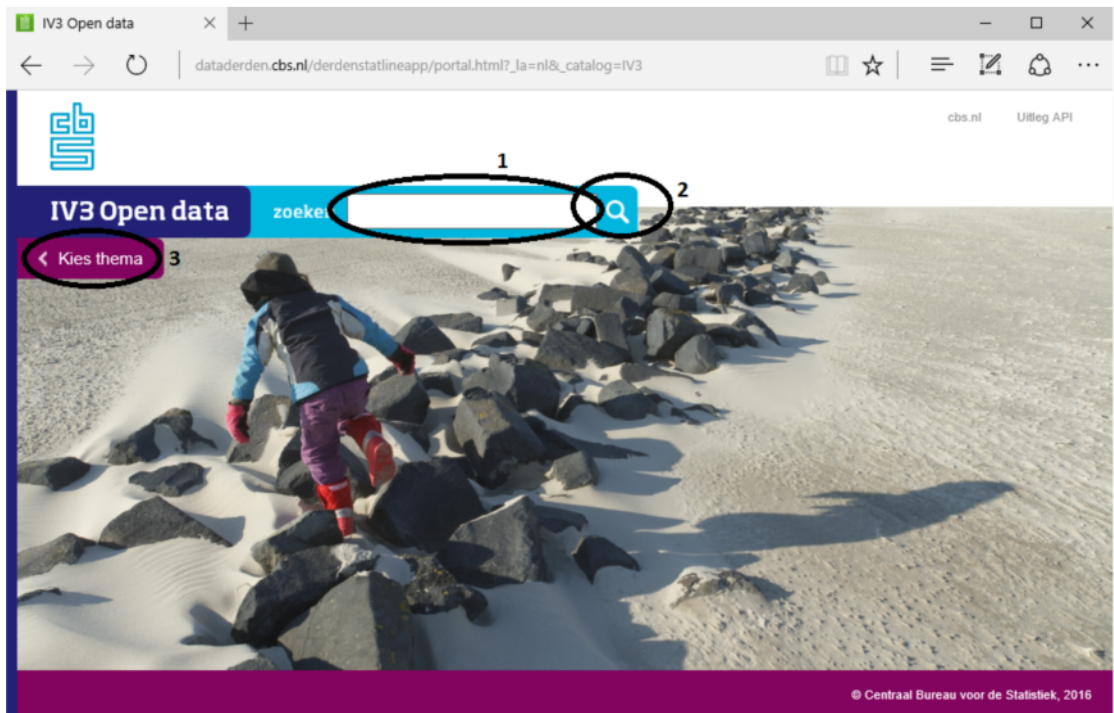
2. Opzoeken Iv3 open datatabel via het portaal

Het webadres van het dataportaal is: <http://iv3statline.cbs.nl/portal.html>

STAP 1. Klik op de bovenstaande link voor het openen van het Iv3 open dataportaal

Hierna opent zich (in de Internetbrowser) onderstaand scherm.

Figuur 1 Schermafdruck Iv3 open data-portaal - Kies thema



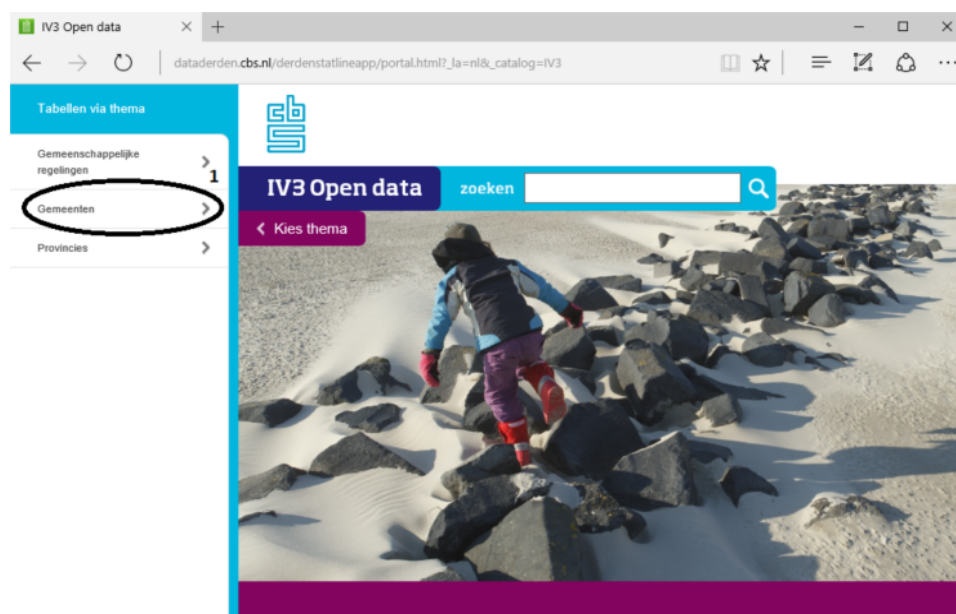
In het bovenstaande scherm kan op twee manieren worden gezocht naar een datatabel. De eerste manier is door het opgeven van een zoekterm bv. gemeenten 2012 (zie bij 1 in bovenstaande schermafdruck) en te klikken op het vergrootglas voor het uitvoeren van de zoekopdracht (zie bij 2).

De tweede wijze waarop de beschikbare tabellen kunnen worden gevonden, is via het kiezen van een thema (zie bij 3). In dit voorbeeld wordt voor deze laatste mogelijkheid gekozen.

STAP 2. Klik op 'Kies thema' in bovenstaand scherm (zie bij 3)

Er opent zich nu een scherm vergelijkbaar met onderstaande schermafdruck.

Figuur 2 Schermafdruk Iv3 open data-portaal - Tabellen via thema

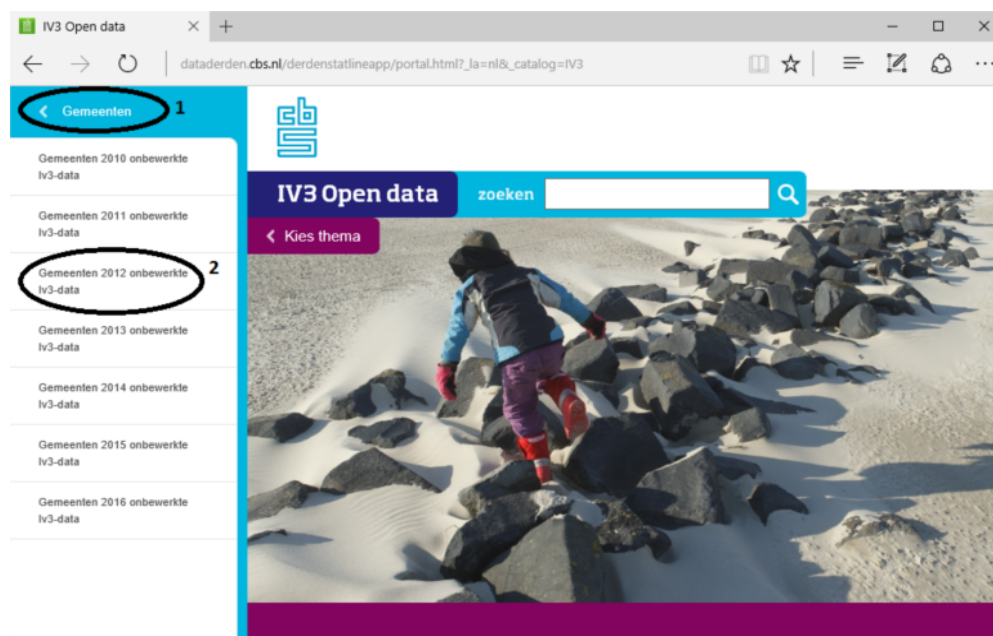


In het bovenstaande scherm is aan de linkerkant een overzicht te zien van de soorten overheden (thema's) waarover data beschikbaar zijn. Selecteer de soort overheid waarvan data wordt gezocht, bv. Gemeenten (zie bij 1 in bovenstaande schermafdruk).

STAP 3. Klik op een soort overheid (thema) in bovenstaand scherm

Hierna opent een scherm vergelijkbaar met het scherm in onderstaande schermafdruk.

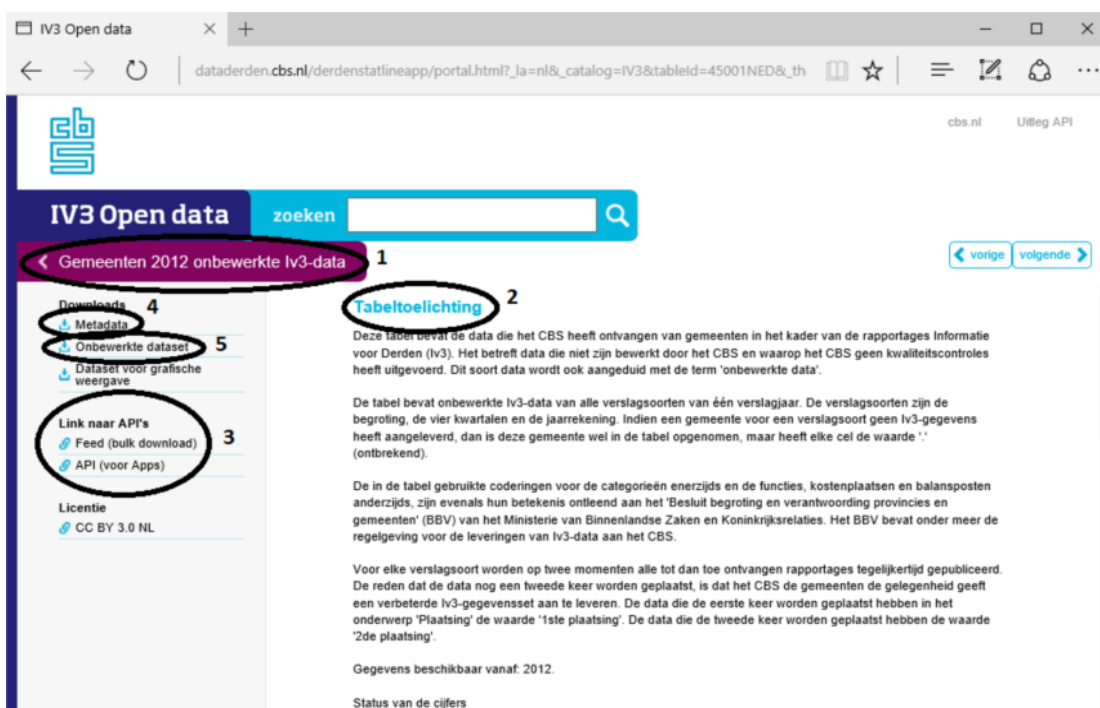
Figuur 3 Schermafdruk Iv3 open data-portaal - Selecteer tabel van thema



In bovenstaand scherm staat links de geselecteerde soort overheid (thema, zie bij 1 in bovenstaand schermafdruk) en de datatabellen die voor deze overheid beschikbaar zijn. Uit de beschikbare tabellen kan een tabel worden geselecteerd door deze aan te klikken, in dit voorbeeld de tabel 'Gemeenten 2012 onbewerkte Iv3 data' (zie bij 2). Hierna opent dan een scherm vergelijkbaar met onderstaande schermafdruk.

STAP 4. Selecteer een datatabel binnen de soort overheid (thema)

Figuur 4 Schermafdruk Iv3 open data-portaal - geselecteerde tabel



In dit scherm is van de betreffende tabel (zie bij 1 in bovenstaande schermafdruk) een toelichting op de tabel opgenomen, de zogeheten tabeltoelichting (zie bij 2). Daarnaast zijn op dit scherm opgenomen een tweetal mogelijkheden voor het benaderen/downloaden van de data.

Ten eerste zijn beschikbaar de 'Links naar API's' die kunnen worden gebruikt voor het maken van een verbinding met de open data in een programmeeromgeving (zie bij 3). Deze mogelijkheid wordt verder niet behandeld in deze handleiding.

Ten tweede is opgenomen functionaliteit voor het downloaden van de metadata van de betreffende tabel (zie bij 4) evenals de data zelf (zie bij 5) naar bijvoorbeeld een bestand in CSV-formaat. Deze functionaliteit wordt in het volgende hoofdstuk verder toegelicht.

3. Downloaden data uit de Iv3 open datatabel

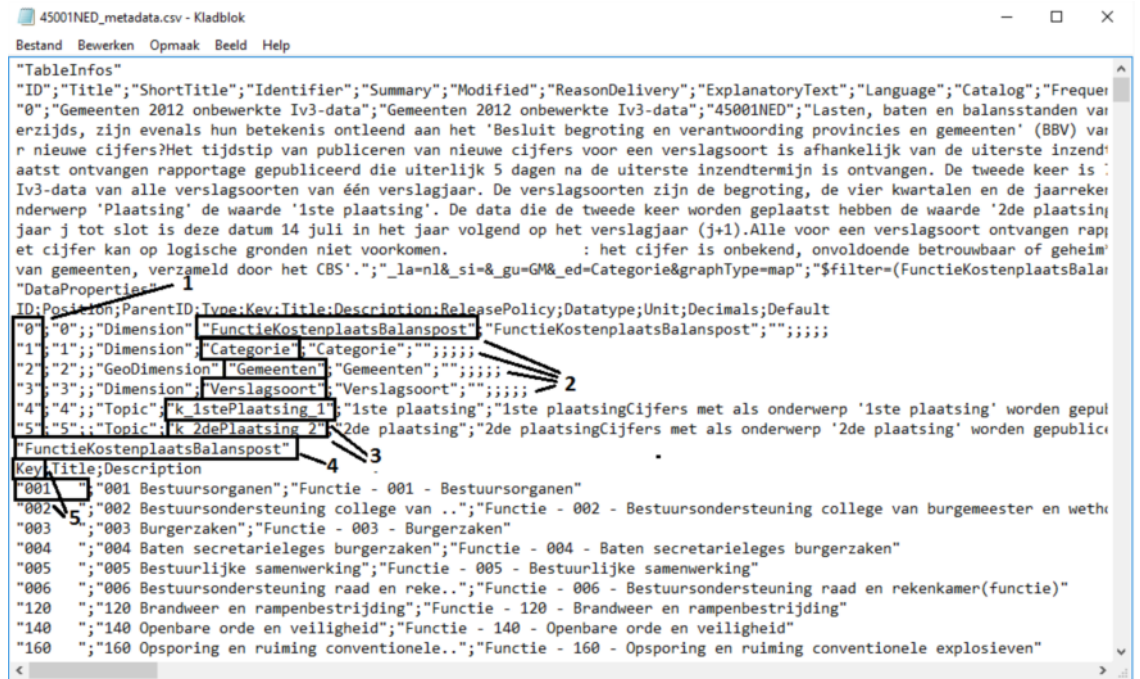
3.1 Downloaden metadata

Metadata zijn gegevens die de karakteristieken van de data zelf beschrijven. Het zijn dus eigenlijk data over data. De metadata van een tabel kunnen in een CSV-bestand worden gedownload door te klikken op ‘Metadata’ (zie bij 4 in bovenstaand scherm).

STAP 5. Download het bestand met de metadata

Er is nu een CSV-bestand met metadata gedownload. De inhoud van dit bestand is vergelijkbaar met onderstaande schermafdruck.

Figuur 5 Voorbeeld bestand met metadata in CSV-formaat



Te zien (zie bij 1 in bovenstaande schermafdruck) is dat de open data-tabel zes variabelen bevat (genummerd van 0 t/m 5). De eerst vier variabelen (FunctieKostenplaatsBalanspost, Categorie, Gemeenten en Verslagsoort, zie bij 2) zijn zogeheten dimensies (dimensions). De dimensies zijn de variabelen die betekenis geven aan de eigenlijke datavariabelen, zijnde de onderwerpen of topics, die als namen 1ste plaatsing en 2de plaatsing hebben (zie 3)¹.

Welke waarden per dimensie voorkomen is te zien in het verdere verloop van dit CSV-bestand.

In dit voorbeeld is de eerste dimensie die daarbij wordt weergegeven

FunctieKostenplaatsBalanspost (zie bij 4). Alle mogelijke voorkomens van deze dimensie worden in het CSV-bestand weergegeven. Van belang daarbij is het veld Key. De waarden van de Key's (dat in dit voorbeeld als eerste waarde "001" heeft) komen voor in het eigenlijke databestand (zie 3.3 Downloaden data). In dit metabestand kunt u de betekenis van de Key's vinden. Bijvoorbeeld van Key "001" is de betekenis (Title) "001 Bestuursorganen" met als toelichting (Description) "Functie - 001 - Bestuursorganen".

Ook van alle andere dimensies kunt u zien welke voorkomens er zijn van Keys en wat hun 'Title' (betekenis) en 'Description' (toelichting) is.

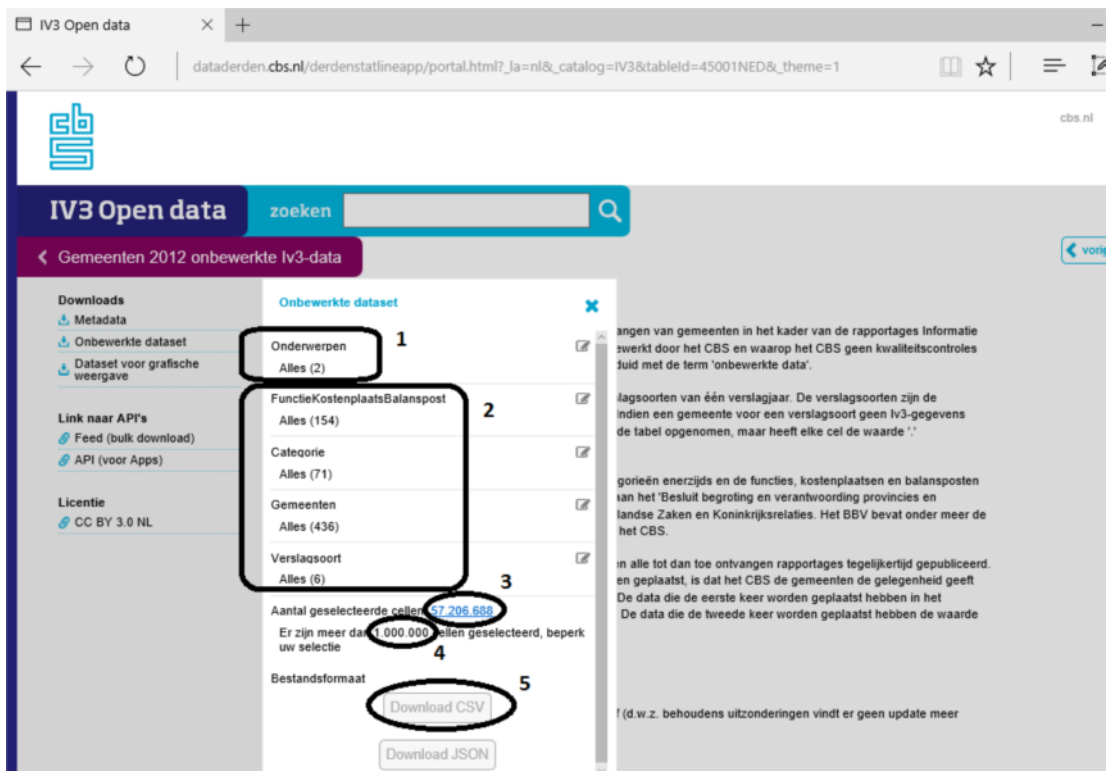
¹ De dimensies FunctieKostenplaatsBalanspost en Categorie zijn daarbij de variabelen uit het vragenformulier, Iv3-model, waarmee CBS de data uitvraagt. Zie voor meer informatie de modellen zelf; zie www.cbs.nl/kredo.

Tot slot, het CSV-bestand is een enigszins lastig te lezen bestand. Het bestand kunt u leesbaarder maken door bijvoorbeeld de inhoud van het bestand te kopiëren en te plakken in een rekenblad(programma), dat wil zeggen te importeren als tekst, waarbij aangegeven moet worden dat de velden zijn 'gescheiden' door een puntkomma.

3.2 Maken selectie

Het downloaden van data uit een tabel gaat door te klikken onder het kopje 'Downloads' op 'Onbewerkte dataset' (zie bij 5 in Figuur 4 Schermafbeelding Iv3 open data-portaal - geselecteerde tabel). Hierna opent zich een scherm vergelijkbaar met onderstaande schermafbeelding.

Figuur 6 Schermafbeelding dataportaal – filter op tabel



In bovenstaand scherm is te zien hoeveel onderwerpen zijn geselecteerd, namelijk twee (zie bij 1 in bovenstaande schermafbeelding). Zoals eerder uit het metadatabestand, en ook uit de tabeltoelichting, was af te leiden zijn dit de onderwerpen '1^{ste} plaatsing' en '2^{de} plaatsing', ook te zien door op het kopje Onderwerpen te klikken (zie bij 1).

Ook zijn de (in dit voorbeeld) vier dimensies te zien, alsmede hoeveel items in elke dimensie zijn geselecteerd (zie bij 2). De aantallen geselecteerde items in de dimensies en onderwerpen bepalen hoeveel datacellen uit de tabel zijn geselecteerd. In het voorbeeld zijn dit er 57.206.688 (zie bij 3), wat veel meer is dan het maximaal aantal cellen dat kan worden gedownload (zie bij 4). Is het aantal geselecteerde cellen wel binnen het toegestane aantal, dan kan voor een bestandsformaat worden gekozen waarnaar de data moet worden gedownload, bv het CSV-formaat (zie bij 5). Echter het aantal geselecteerde cellen is in dit voorbeeld meer dan toegestaan. Dit is ook te zien doordat de knoppen voor het downloaden van de data naar een bepaald bestandsformaat niet aanklikbaar zijn.

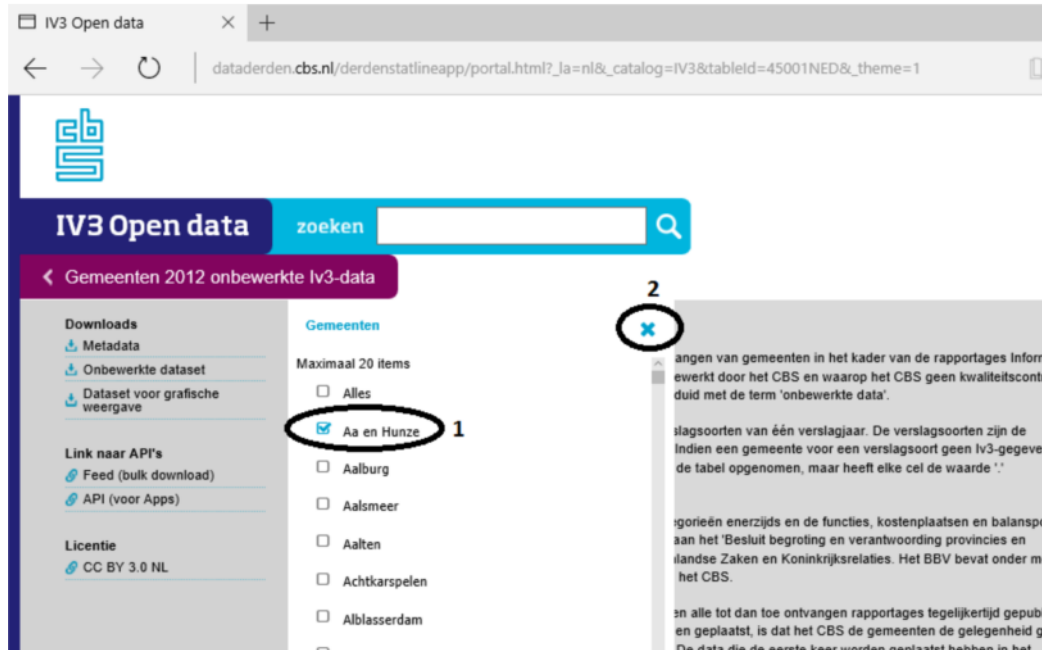
Om in dit voorbeeld te kunnen downloaden moet er worden gefilterd, bijvoorbeeld niet alle berichtgevers en/of niet alle verslagperiodes en/of niet alle onderwerpen. U kunt daarbij per dimensie (FunctieKostenplaatsBalanspost, Categorie, Gemeenten en Verslagsoort) of alle

voorkomens of maximaal 20 voorkomens selecteren. Dit filteren start u door te klikken op een dimensie of de onderwerpen.

STAP 6. Klik op een dimensie of de onderwerpen voor het aanbrengen van een filter

Er opent zich een scherm vergelijkbaar met het onderstaande waarin de dimensie Gemeenten is geselecteerd (zie Figuur 7).

Figuur 7 Schermafdruk dataportaal - selecteren voorkomens (items) in dimensie



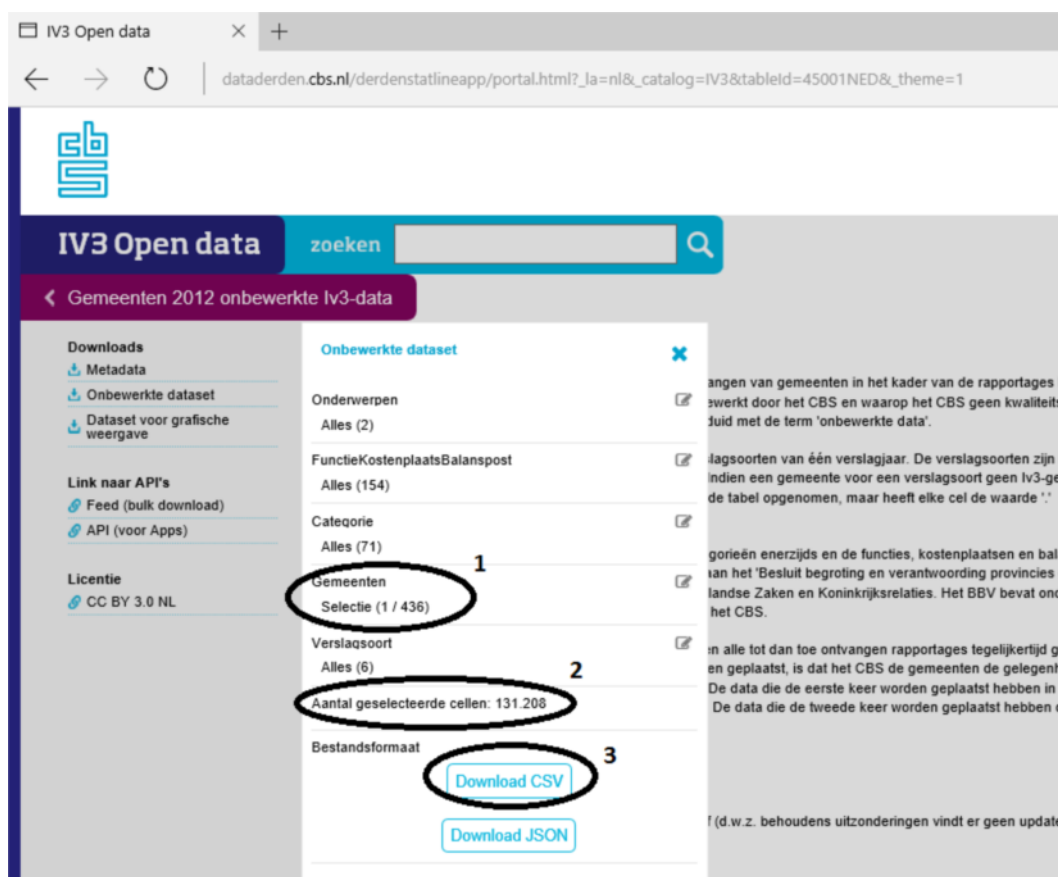
In dit scherm selecteert u een gemeente door op de naam van deze gemeente te klikken (zie bij 1 in bovenstaande schermafdruk). Het selecteren van de gemeente sluit u vervolgens af door te klikken op het kruisje (zie bij 2)

STAP 7. Selecteer binnen de dimensie of de onderwerpen de gewenste voorkomens (alles of maximaal 20)

STAP 8. Sluit het filteren binnen de dimensie of onderwerpen af door te klikken op het kruisje

Hierna wordt teruggekeerd in het overzicht scherm met de aantallen geselecteerde items per dimensies en onderwerpen, alsmede het aantal geselecteerde cellen, zie onderstaande schermafdruk.

Figuur 8 Schermafdruk dataportaal - gefilterde tabel



Ten zien is nu dat het aantal geselecteerde gemeente 1 is van de 436^2 gemeenten die voorkomen (zie bij 1 in bovenstaande schermafdruk). Ook is te zien dat het aantal geselecteerde cellen is gedaald naar 131.208 en dat de melding over het beperken van de selectie is verdwenen (zie bij 2). NB dat het aantal geselecteerde cellen niet groter is dan het maximaal aantal te downloaden cellen, is ook te zien doordat de knoppen voor het downloaden nu aanklikbaar zijn (zie bij 3).

Tot slot kan nog worden opgemerkt dat nu alle data voor één geselecteerde gemeente voor het verslagjaar 2012 zijn geselecteerd en dat dit dus 131.208 cellen (cijfers) zijn. Aangezien er **binnen** elke Iv3 open data tabel voor elke berichtgever evenveel datacellen zijn, en het per keer maximum aantal te downloaden gegevens op 1 miljoen ligt, kunnen uit deze tabel per keer van maximaal zeven berichtgevers (gemeenten) alle data worden gedownload.

3.3 Downloaden data

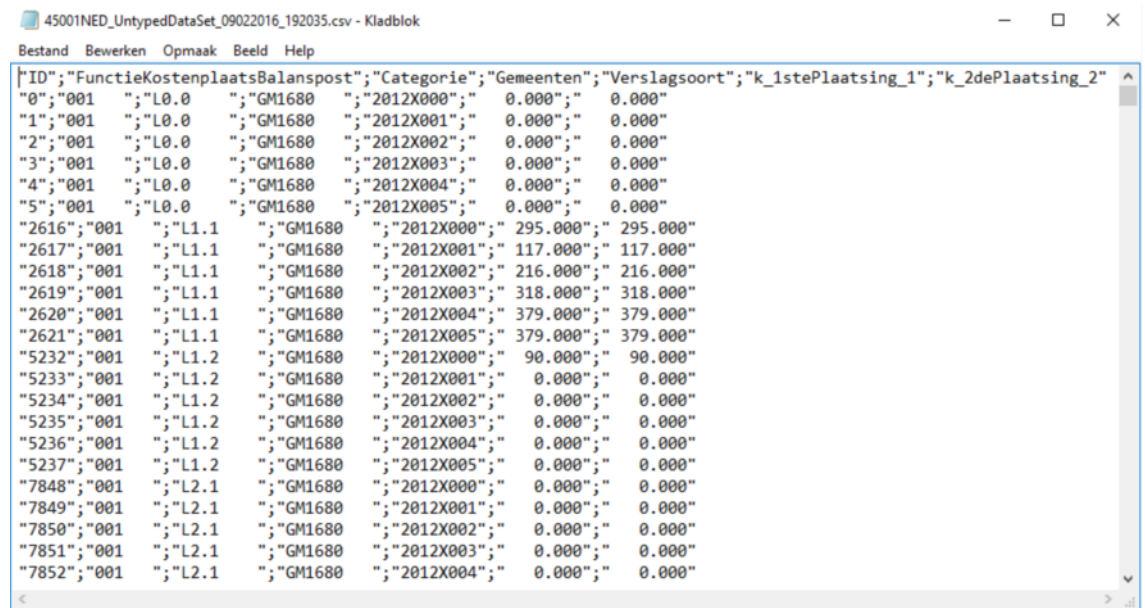
Om de data te downloaden klikt u in bovenstaand scherm op één van de knoppen voor het downloaden naar een bepaald formaat (CSV of JSON). In dit voorbeeld wordt de data gedownload naar een bestand met het CSV-formaat. Dit moet worden gedaan door te klikken op de knop Download CSV (zie bij 3). Hierbij kan worden opgemerkt dat in dit voorbeeld het downloaden van alle data voor een gemeente voor een verslagjaar circa 1 minuut duurt en resulteert in een CSV-bestand van circa 5 MB.

STAP 9. Klik op de knop Download CSV

² Nb het feitelijke aantal gemeenten in Nederland in 2012 was 415; echter de stadsdelen van Amsterdam en de deelgemeenten van Rotterdam leverden in 2012 ook Iv3 data aan en worden binnen deze tabel beschouwd als gemeente.

In onderstaand scherm is een weergave te zien van de inhoud van het bestand met daarin de gedownloade data.

Figuur 9 Voorbeeld bestand met Iv3-data in CSV-formaat



In bovenstaande schermafdruk is te zien dat het databestand is opgebouwd uit zeven velden. Het eerste veld ID is een unieke verwijzing naar de cel in de datatabel waaruit de data zijn gedownload en heeft verder geen inhoudelijke betekenis. Vervolgens komen de vier dimensies, FunctieKostenplaatsBalanspost, Categorie, Gemeenten en Verslagsoort. De waarden die deze velden kunnen aannemen zijn de waarden van de Key's zoals die eerder gezien zijn in het (gedownloade) CSV-bestand met de metadata. De laatste twee velden zijn de onderwerpen: 1ste plaatsing en 2de plaatsing.

Tot slot: ook dit CSV-bestand met data, kunt u leesbaarder maken door bijvoorbeeld de inhoud van het bestand te kopiëren en te plakken in een rekenblad(programma) (bijvoorbeeld Excel), dat wil zeggen te importeren als tekst, waarbij u aangeeft dat de velden zijn 'gescheiden' door een puntkomma.