



Downloaden Iv3 open data met R

23 maart 2016

Inhoud

1. Inleiding	3
2. Programma R	4
2.1 Installatie	4
2.2 Programma starten	4
2.2.1 De console	4
2.2.2 Opdrachten via een script	5
3. Iv3 voorbeeldscript	6
3.1 Packages installeren	6
3.2 Package laden	7
3.3 Locatievariabelen definiëren en vullen	8
3.4 Metadata ophalen	9
3.4.1 Ophalen lijst beschikbare open data-tabellen	9
3.4.2 Ophalen dimensies van een tabel	10
3.5 Iv3 data downloaden	13
3.5.1 Functie download_table	13
3.5.2 Voorbeeld 1: per dimensie één filterwaarde	13
3.5.3 Voorbeeld 2: download met meer filterwaardes per dimensie	14

1. Inleiding

Deze handleiding is bedoeld voor gebruikers die 'een gehele lv3 open dataset' willen downloaden, bijvoorbeeld alle jaarrekeningen 2014 van gemeenten. Het betreft dan een grote dataset die niet via het wat gebruiksvriendelijkere 'lv3 open data portaal' (<http://iv3statline.cbs.nl/portal.html>) of de lv3-app (<http://iv3statline.cbs.nl/index.html>) kan worden gedownload.

CBS heeft voor het downloaden van grotere datasets voor het programma R een package ontwikkeld. In deze handleiding wordt beschreven hoe het programma R is te installeren. Vervolgens wordt met een voorbeeld beschreven hoe dit package gebruikt kan worden voor het downloaden van de lv3 open data. Hierbij wordt met name ook aandacht besteed aan het opgeven van selectiecriteria voor het kunnen selecteren van de gewenste data.

De stappen die in de handleiding worden genomen, vereisen geen voorkennis van het programma R. Eventueel benodigde vaardigheden worden in deze handleiding uiteengezet. Deze handleiding geeft echter geen volledige beschrijving van het programma R en haar mogelijkheden. Voor een beschrijving van of cursus voor R zijn de volgende links/documenten handig:

- <https://cran.r-project.org/> (website programma R)
- <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf> (introductie van het programma R)
- <https://cran.r-project.org/doc/contrib/Short-refcard.pdf> (een cheat-sheet met commando's en opdrachten in R)

De lv3 data wordt door CBS via open data in een structuur beschikbaar gesteld waarin CBS al haar open data beschikbaar stelt. Voor meer technische achtergrond over deze structuur en de technische wijze waarop deze door CBS beschikbaar wordt gesteld, wordt verwezen naar het volgende document:

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

Tot slot voor meer informatie over lv3 open data wordt verwezen naar de informatiepagina op de CBS-site over lv3 open data. Deze is te vinden via:

- www.cbs.nl met als zoekterm lv3+open+data

Voor meer informatie over de lv3, waaronder de lv3 modellen wordt verwezen naar:

- www.cbs.nl/kredo

2. Programma R

Het programma R is een programmeertaal en software omgeving voor onder meer het maken van statistische bewerkingen en data-analyse. Voor het samenstellen van deze handleiding is gewerkt met versie 3.2.2 voor Windows op een Windows 10 64 bits versie.

2.1 Programma installeren

Het programma R is te downloaden via de website <https://cran.rstudio.com/bin/windows/base/>. De installatie wijst zich daarbij vanzelf; volg de aanwijzingen van de installer. Indien de installatie niet mocht lukken, dan biedt de volgende site wellicht een oplossing: <https://cran.rstudio.com/bin/windows/base/README.R-3.2.2>.

STAP 1. Download en installeer het programma R

2.2 Programma starten

Na installatie is R te openen via de geïnstalleerde apps/programma's (Windows toets, links onder in het scherm → Alle apps → R of onder Windows 7/8 via (startknop → Alle programma's → R) op de computer of via de snelkoppeling op het bureaublad (indien hiervoor is gekozen tijdens de installatie). Er zijn twee opties waarvan de selectie afhankelijk is van het eigen platform:

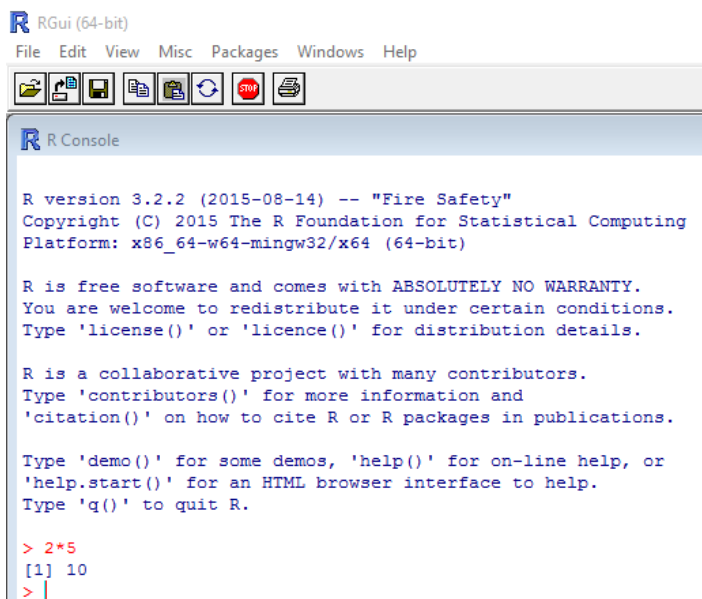
- a) R x i386 3.2.2 bij een Windows versie 32 bits
- b) R x 64 3.2.2. bits bij een Windows versie 64 bits

STAP 2. Open het programma R

2.2.1 De console

Standaard opent het programma met een scherm (venster) dat wordt genoemd de Console. In dit scherm kunnen opdrachten worden gegeven op de commandline achter het teken '>'. Bijvoorbeeld als achter het teken '>' de opdracht "2*5" wordt ingetoetst (wat verschijnt in een rode tekstkleur) en daarna een Enter wordt gegeven (= opdracht uitvoeren), dan geeft het programma als uitkomst "10" (in een blauwe tekstkleur). Zie onderstaande schermafbeelding voor een voorbeeld (zie Figuur 1).

Figuur 1 Schermafbeelding voorbeeld weergave R Console



```
R RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console

R version 3.2.2 (2015-08-14) -- "Fire Safety"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

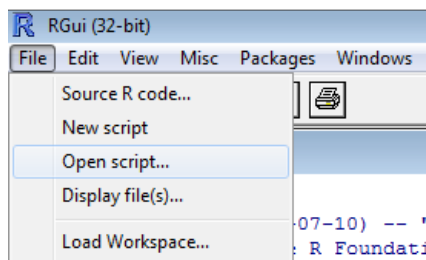
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 2*5
[1] 10
> |
```

2.2.2 Opdrachten via een script

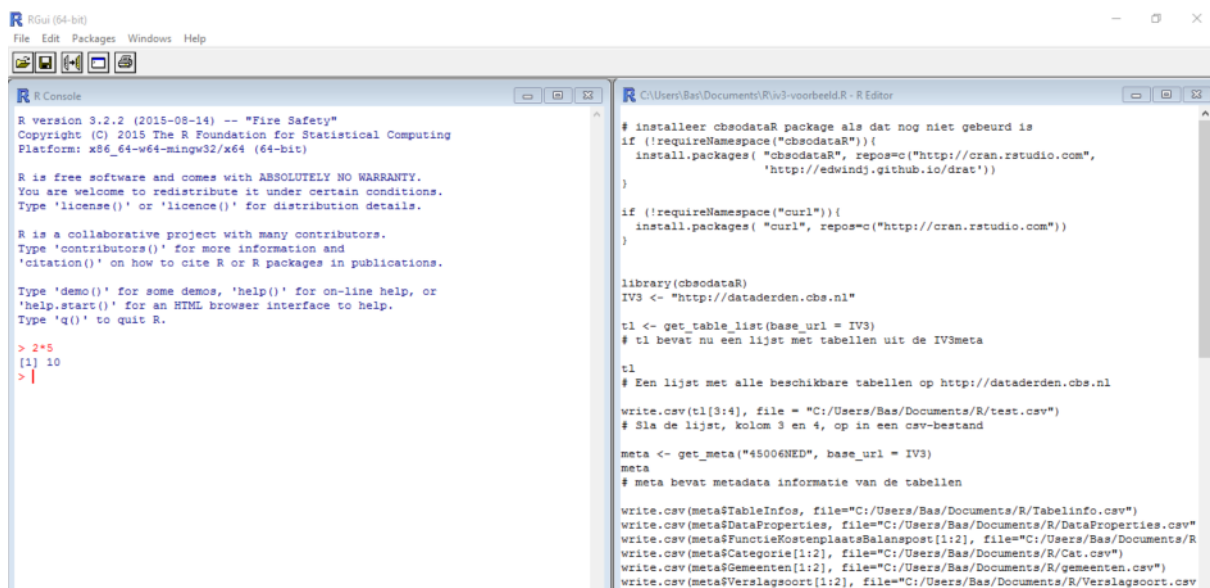
Een andere manier om in R opdrachten uit te voeren, is via een script. In een script kan de gebruiker verschillende opdrachten opgeven die het programma R achter elkaar moet uitvoeren. Voordelen van het gebruik van een script is dat de databewerking exact terug te vinden is en dat de bewerkingen gemakkelijk nog een keer kunnen worden uitgevoerd. In het script kan daarnaast commentaar/toelichting opgenomen worden door de regel te laten beginnen met het teken #. Om in R een script te openen, kies in de menubalk File → Open script (zie Figuur 2) en vervolgens het te openen script (door naar de map te navigeren waar het script is opgeslagen).

Figuur 2 Schermafdruk R openen script via menubalk



Het script is nu geladen in de R editor. In de onderstaande figuur staat het script aan de rechterkant van het scherm (zie Figuur 3).

Figuur 3 Schermafdruk R - Console (links) en geopend script (rechts)



Via de menu-optie Edit → Run all kan een script in één keer worden uitgevoerd. Ook kan een gedeelte van het script worden uitgevoerd. Selecteer daarvoor in het script het gedeelte dat moet worden uitgevoerd (tekst wordt dan blauw gearceerd) en kies vervolgens via de menubalk Edit → Run line or selection om het uit te voeren. In de volgende paragraaf 'Iv3 voorbeeld script' zal deze laatste optie (een gedeelte van het script uitvoeren) veelvuldig worden toegepast. NB het uitvoeren kan ook via de sneltoets Ctrl+r.

Belangrijk! Als een selectie in een script (blauw gearceerde tekst) wordt uitgevoerd, dan kopieert R deze tekst naar de console, voert de commando's hierin uit en laat de output zien in de console in rode tekst (indien van toepassing).

3. Iv3 voorbeeldscript

In deze handleiding worden voorbeelden gegeven voor het downloaden van Iv3 data, waarbij gebruik wordt gemaakt van een script. Dit script heeft als naam 'voorbeeld downloaden iv3 open data met R.R' en wordt beschikbaar gesteld op de informatiepagina 'Iv3 open data' die te vinden is via zoekterm Iv3+open+data op www.cbs.nl. In de hieronder beschreven voorbeelden moet het script worden aangepast. Het script moet daarom worden gedownload op de eigen computer.

STAP 3. Download het voorbeeldscript naar de eigen computer

STAP 4. Open het script 'iv3-voorbeeld.R' in R via de menuoptie File → Open script.

3.1 Packages installeren

Het programma R maakt gebruik van packages. Packages zijn onder andere verzamelingen van R-functies, data en gecompileerde code. Het CBS heeft een package ontwikkeld om Iv3 data zonder al te veel benodigde technische kennis en/of programmeerervaring te kunnen downloaden. Dit package 'cbsodataR' en het package 'Curl' waarvan het gebruik maakt, zitten niet bij de standaard installatie van R. Zij moeten daarom apart worden geïnstalleerd. In het voorbeeldscript 'iv3-voorbeeld.R' zijn voorbeeldregels opgenomen hoe beide packages kunnen worden geïnstalleerd (zie hieronder Figuur 4).

Figuur 4 Opdrachtregels - installeren packages

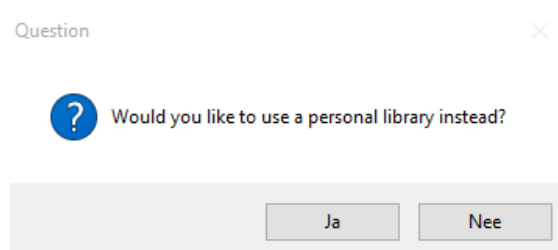
```
J:\08_Persoonlijke_mappen\BTNF\R\voorbeeld downloaden iv3 open data met R v20160114.R - R Editor
# In dit script worden twee voorbeelden gegeven hoe Iv3 data is te downloaden
# uit de Iv3 open data-tabellen met behulp van het package cbsodataR
#
#-----
# installeren packages
#-----
# installeer cbsodataR en Curl package als dat nog niet gebeurd is
# dit hoeft maar éénmalig te worden gedaan
#
if (!requireNamespace("cbsodataR")) {
  install.packages("cbsodataR", repos=c("http://cran.rstudio.com"))
}
if (!requireNamespace("curl")) {
  install.packages("curl", repos=c("http://cran.rstudio.com"))
}
```

STAP 5. Selecteer in het script de bovenstaande regels.

STAP 6. Druk op de toets combinatie Ctrl+r om de geselecteerde opdrachten uit te voeren.

R gaat nu de twee packages downloaden en installeren. R kan de te installeren packages echter niet wegschrijven op de standaard voorgestelde locatie. Het programma vraagt daarom waar het de 'personal library' moet plaatsen (zie hieronder Figuur 5).

Figuur 5 Vraag gebruik 'personal library' bij installeren packages



STAP 7. Druk op de knop “Ja” bij de vraag “Would you like to use a personal library instead?”.

Wanneer op “Ja” is gedrukt, draagt het programma een locatie aan waar de library moet worden geplaatst.

STAP 8. Selecteer eventueel eerst een alternatieve opslagplaats

STAP 9. Druk weer op “Ja”

NB het installeren van de packages hoeft maar één keer te worden uitgevoerd; bij een volgende sessie zijn de packages al beschikbaar. Wel moet in elke nieuwe R-sessie het package cbsodataR worden geladen. Hoe dit moet, staat beschreven in de volgende paragraaf.

3.2 Package laden

Wanneer je in een nieuwe R-sessie gebruik wilt maken van het package cbsodataR dan moet het package eerst worden geladen. De regels hiervoor zijn opgenomen in het voorbeeldscript (zie hieronder Figuur 6).

Figuur 6 Opdrachtregels - laden package

```
R J:\08_Persoonlijke_mappen\BTNF\R\voorbeeld downloaden iv3 open data met R v20160114.R - R Editor
#-----
# laden package cbsodataR
#-----
# Het package cbsodataR moet eerst worden geladen alvorens er in een nieuwe
# R-sessie gebruik van worden gemaakt
library(cbsodataR)
```

STAP 10. Selecteer in het script de bovenstaande regels

STAP 11. Voer de tekst uit (Ctrl+r).

Het package cbsodataR wordt nu geladen. Het package bevat een aantal functies zoals ‘get_table_list’, ‘get_meta’ en ‘download_table’ die in het verdere verloop van dit voorbeeld zullen worden aangeroepen. Het is dan ook belangrijk dat het package goed is geladen. Als het package niet (goed) is geladen, dan geeft R een foutmelding bij de aanroep van deze functies. Zie hiervan een voorbeeld in Figuur 7 waarin functie ‘get_table_list’ is aangeroepen terwijl de package niet (goed) is geladen.

Belangrijk! Treden dit soort melding op, dan is het advies om de hele uitvoering van het script opnieuw te beginnen waarbij het package goed (geïnstalleerd en/of) geladen moet worden.

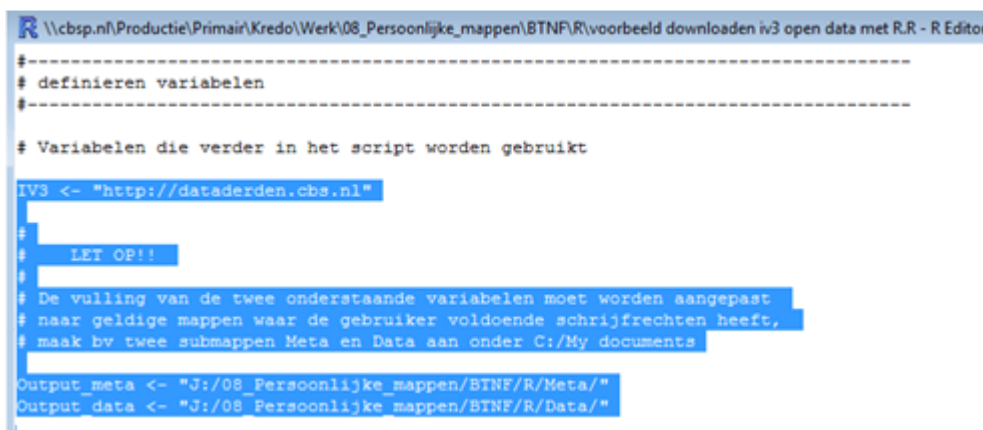
Figuur 7 Schermoutput console - aanroep `get_table_list` bij niet geladen package `cbsodataR`

```
> IV3 <- "http://dataderden.cbs.nl"
> Output_meta <- "C:/Users/Bas/Documents/R/Meta/"
> Output_data <- "J:/08_Persoonlijke_mappen/BTNF/R/Data/"
> Tbl_lijst <- get_table_list (base_url = IV3, Catalog = "IV3")
Error: could not find function "get_table_list"
> # Tbl_lijst bevat nu een lijst met tabellen uit de Iv3
> setwd (Output_meta) #de 'workdirectory' instellen naar een opgegeven
> write.csv(Tbl_lijst[3:4], file = "lijst.csv")
Error in is.data.frame(x) : object 'Tbl_lijst' not found
```

3.3 Locatievariabelen definiëren en vullen

In de afbeelding hieronder (zie Figuur 8) worden drie variabelen gedefinieerd met als namen `IV3`, `Output_meta` en `Output_data`, waarbij aan de variabelen een waarde wordt toegekend. Deze variabelen worden verderop in het script gebruikt.

Figuur 8 Opdrachtregels - variabelen definiëren en vullen met locatie



```
R \\cbsp.nl\Productie\Primair\Kredo\Werk\08_Persoonlijke_mappen\BTNF\R\voorbeeld downloaden iv3 open data met R.R - R Editor
#-----
# definiëren variabelen
#-----
# Variabelen die verder in het script worden gebruikt
IV3 <- "http://dataderden.cbs.nl"
#
# LET OP!!
#
# De vulling van de twee onderstaande variabelen moet worden aangepast
# naar geldige mappen waar de gebruiker voldoende schrijfrechten heeft,
# maak bv twee submappen Meta en Data aan onder C:/My documents
Output_meta <- "J:/08_Persoonlijke_mappen/BTNF/R/Meta/"
Output_data <- "J:/08_Persoonlijke_mappen/BTNF/R/Data/"
```

De waarde van de variabele `IV3` is een URL en hoeft niet te worden aangepast. **De waarden van de laatste twee variabelen, `Output_meta` en `Output_data` moeten wel worden aangepast.** Deze variabelen bevatten verwijzingen naar locaties (mappen) op de computer waarnaar de diverse databestanden moeten worden geschreven. Ze moeten dan ook verwijzen naar een **geldige locatie** op de computer van de gebruiker en waar je als gebruiker **voldoende schrijfrechten** hebt. Bepaal waar de output naartoe moet worden geschreven (maak bijvoorbeeld een map `Meta` en een map `Data` (in een submap) onder bv `C:\My documents`) en geef deze twee variabelen in de R editor de goede waarden.

Let daarbij op dat in R in namen van paden een slash ("/") moet worden gebruikt en niet een backslash ("\").

STAP 12. Wijzig de waarden van variabelen `Output_meta` en `Output_data` in het voorbeeldscript naar geldige locaties op de eigen computer

Wanneer de waarden van de variabelen `Output_meta` en `Output_data` in de R editor zijn aangepast naar bestaande locaties op de computer, dan kun je het script selecteren zoals weergegeven in bovenstaand voorbeeld en uitvoeren.

STAP 13. Selecteer in het script de (aangepaste) tekst uit de bovenstaande Figuur 8 (definiëren en vullen 'locatievariabelen')

STAP 14. Voer de tekst uit (Ctrl+r)

3.4 Metadata ophalen

Metadata zijn gegevens die de karakteristieken van de data zelf beschrijven. Het zijn dus eigenlijk data over data.

3.4.1 Ophalen lijst beschikbare open data-tabellen

Alle Iv3 data zijn per verslagjaar en per overheidslaag in een aparte tabel opgeslagen. Om te weten te komen wat de naam is van de tabel waaruit data gedownload moeten gaan worden, is het nodig een overzicht te maken van alle beschikbare tabellen. Dit overzicht van beschikbare tabellen is te downloaden/bekijken via de onderstaande commando's (zie Figuur 9).

Figuur 9 Opdrachtregels - tonen overzicht beschikbare tabellen

```
R J:\08_Persoonlijke_mappen\BTNFR\voorbeeld downloaden iv3 open data met R v20160114.R - R Editor
#-----
# Ophalen lijst beschikbare open data-tabellen
#-----
Tbl_lijst <- get_table_list (base_url = IV3, Catalog = "IV3")
# Tbl_lijst bevat nu een lijst met tabellen uit de Iv3
Tbl_lijst
# Een lijst met alle beschikbare tabellen op http://dataderden.cbs.nl
```

De werking van deze regels is als volgt. Ten eerste wordt met behulp van functie `get_table_list` alle metadata van de tabellen betreffende Iv3 geplaatst in een variabele met de naam `Tbl_lijst`. Vervolgens wordt de inhoud van deze variabele (de metadata) getoond op het scherm (in de console).

STAP 15. Selecteer in het script de bovenstaande tekst.

STAP 16. Voer de tekst uit (Ctrl+r).

Per regel wordt nu getoond de meta-gegevens van elke open data tabel¹. Hierin zijn vooral de derde en de vierde kolom belangrijk. Dit zijn de kolommen met de naam `Identifier` en `Title`, zijnde de verkorte tabelnaam en de titel van de tabel. Het is handig om alleen deze twee kolommen te bekijken, wat kan door deze kolommen te exporteren naar een CSV-bestand. Dit kan via de commando's zoals opgenomen in het voorbeeldscript (zie Figuur 10).

Figuur 10 Opdrachtregels - schrijf inhoud variabele `Tbl_lijst` naar csv-bestand

```
R J:\08_Persoonlijke_mappen\BTNFR\voorbeeld downloaden iv3 open data met R v20160114.R - R Editor
# Sla de lijst met beschikbare tabellen, alleen kolom 3 en 4, op in een csv-bestand
setwd (Output_meta) #de 'workdirectory' instellen naar een opgegeven variabele
write.csv(Tbl_lijst[3:4], file = "lijst.csv")
```

STAP 17. Selecteer in het script de bovenstaande regels.

STAP 18. Voer de tekst uit (Ctrl+r).

Er wordt nu een CSV-bestand gecreëerd in de map zoals eerder is opgegeven bij de variabele `Output_meta`.

Als in de console het bericht `'cannot open file 'lijst.csv' : Permission denied'` verschijnt, dan heeft R geen rechten om het bestand in de opgegeven map (waarde variabele `Output_meta`) weg te schrijven. Als dit optreedt, wijzig dan in het script de parameters `Output_meta` en/of `Output_data` naar een map waar wel voldoende schrijfrechten voor zijn **en voer de betreffende opdrachtregels**

¹ Afhankelijk van de schermgrootte breekt R de tabel af en kunnen kolommen onder elkaar komen te staan.

voor het maken en vullen van deze variabelen nog een keer uit (dat wil zeggen de opdrachtregels zoals opgenomen in Figuur 8).

Het CSV-bestand kan worden geopend via de Verkenner. Indien het bestand wordt geopend in bijvoorbeeld Excel, dan kan de tekst leesbaarder worden gemaakt via de optie “Tekst naar kolommen”. Het gedownloade bestand ziet er dan in Excel uit zoals in onderstaande figuur is opgenomen (zie Figuur 11). De waarde van de kolom ‘**Identifier**’ is de (verkorte) naam van de tabel en wordt verderop in het script gebruikt om aan te geven **uit welke tabel data moet worden gedownload**.

Figuur 11 Schermafdruk van in Excel geopend overzicht met beschikbare tabellen

	A	B	C	D	E	F	G	H
1		Identifier	Title					
2		1 45024NED	Gemeenschappelijke regelingen 2010 onbewerkte lv3-data					
3		2 45025NED	Gemeenschappelijke regelingen 2011 onbewerkte lv3-data					
4		3 45026NED	Gemeenschappelijke regelingen 2012 onbewerkte lv3-data					
5		4 45027NED	Gemeenschappelijke regelingen 2013 onbewerkte lv3-data					
6		5 45028NED	Gemeenschappelijke regelingen 2014 onbewerkte lv3-data					
7		6 45029NED	Gemeenschappelijke regelingen 2015 onbewerkte lv3-data					
8		7 45007NED	Gemeenten 2010 onbewerkte lv3-data					
9		8 45008NED	Gemeenten 2011 onbewerkte lv3-data					
10		9 45001NED	Gemeenten 2012 onbewerkte lv3-data					
11		10 45004NED	Gemeenten 2013 onbewerkte lv3-data					
12		11 45005NED	Gemeenten 2014 onbewerkte lv3-data					

STAP 19. Zoek de gewenste tabel en noteer de naam van de tabel (kolom Identifier)

3.4.2 Ophalen dimensies van een tabel

Alle lv3 open data-tabellen hebben een gelijke structuur. Deze bestaat uit:

1. Onderwerpen

Is datgene waar het cijfer betrekking op heeft, het onderwerp.

In elke lv3 open data-tabel zijn er twee onderwerpen:

- a. eerste plaatsing, wat staat voor een cijfer gepubliceerd 10 dagen na uiterste inzenddatum;
- b. tweede plaatsing, wat staat voor een cijfer gepubliceerd 70 dagen na uiterste inzenddatum;

2. Dimensies

De dimensies geven meer informatie over van wie en waarover van het onderwerp. In de lv3-tabellen zijn er altijd vier dimensies, namelijk:

a. FunctieKostenplaatsBalanspost

in de lv3-matrix de y-as, met de daarop de functies (beleidsterreinen) en balansposten;

b. Categorie

in de lv3-matrix de x-as met de kosten- en opbrengstsoorten en de soorten balansstanden;

c. Gemeenten

In de tabellen betreffende gemeenten komt de dimensie Gemeenten voor, waarin de meta van de gemeenten wordt beschreven waarop het onderwerp betrekking heeft. NB in bijvoorbeeld de tabellen met data van de provincies heet deze dimensie Provincies.

d. Verslagsoort

Dit kan zijn begroting, jaarrekening of kwartaalrekening.

De Iv3 open data-tabellen voor met name gemeenten kunnen door hun grote aantal berichtgevers (ca. 400) zeer groot van omvang zijn. Het is daarom aan te raden om van te voren goed te bepalen in welke data men is geïnteresseerd, bijvoorbeeld van verslagjaar 2014 alleen de jaarrekeningen. Dit is wenselijk om vervolgens via het opgeven van een filter de download te kunnen beperken. Om een filter te maken is het wel van belang om te weten waarop is te filteren. Daartoe moet de metadata van de dimensies (variabelen) van de tabel worden opgehaald. Hoe dit gedaan kan worden, wordt toegelicht met een voorbeeld. In dit voorbeeld worden de meta-data van de dimensies van de tabel voor gemeenten 2014 gedownload. In het eerder gedownload overzicht met welke tabellen er zijn, is te zien dat deze tabel als verkorte naam (Identifier) "45005NED" heeft (zie ook Figuur 11). Om te kijken uit welke dimensies (variabelen) deze tabel is opgebouwd en welke waardes hierin voorkomen, selecteer dan in het voorbeeldscript de onderstaande tekst, waarbij eerst het script zonnodig moet worden aangepast als het om andere tabel gaat:

Figuur 12 Opdrachtregels - metadata downloaden en tonen

```

R J:\08_Persoonlijke_mappen\BTNF\R\voorbeeld downloaden iv3 open data met R v201601
# plaats in variabele meta de metadata van tabel 45005NED
meta <- get meta("45005NED", base_url = IV3)
# meta bevat metadata informatie van de tabellen
# toon de inhoud van meta
meta

```

STAP 20. Pas eventueel in het script in de bovenstaande tekst de tabelnaam aan.

STAP 21. Selecteer de tekst.

STAP 22. Voer de tekst uit (Ctrl+r).

Met behulp van de functie `get_meta` is de meta-informatie van tabel 45005NED in variabele `meta` geplaatst, en is vervolgens de inhoud van variabele getoond in de console. In de console is daarbij nu het volgende zien (zie Figuur 13):

Figuur 13 Schermoutput console - downloaden en tonen meta

```

> meta <- get_meta("45005NED", base_url = IV3)
Retrieving meta data for table '45005NED' http://dataderden.cbs.nl/ODataApi/oda$
Retrieving http://dataderden.cbs.nl/ODataApi/odata/45005NED http://dataderden.c$
Retrieving http://dataderden.cbs.nl/ODataApi/odata/45005NED http://dataderden.c$
Retrieving http://dataderden.cbs.nl/ODataApi/odata/45005NED http://dataderden.c$
Retrieving http://dataderden.cbs.nl/ODataApi/odata/45005NED http://dataderden.c$
Retrieving http://dataderden.cbs.nl/ODataApi/odata/45005NED http://dataderden.c$
Retrieving http://dataderden.cbs.nl/ODataApi/odata/45005NED http://dataderden.c$
> meta
45005NED: 'Gemeenten 2014 onbewerkte Iv3-data', 2014
  FunctieKostenplaatsBalanspost: 'FunctieKostenplaatsBalanspost'
  Categorie: 'Categorie'
  Gemeenten: 'Gemeenten'
  Verslagsoort: 'Verslagsoort'
> # meta bevat metadata informatie van de tabellen

```

In woorden betekent dit dat de tabel 45005NED de volgende dimensies (soort van variabelen) bevat:

- FunctieKostenplaatsBalanspost (een lijst met Iv3 functies, balansposten en kostenplaatsen in deze tabel 45005NED)
- Categorie (een lijst met Iv3 categorieën)
- Gemeenten (een lijst met beschikbare gemeenten)

- Verslagsoort (een lijst met verslagsoorten: begroting, jaarrekening en kwartaalrekeningen)

De lijsten met beschikbare waarden voor respectievelijk de dimensies 'FunctieKostenplaatsBalanspost', 'Categorie', 'Gemeenten' en 'Verslagsoort' kunnen het beste worden gedownload naar CSV-bestanden om ze daarna via de Verkenner te openen en te bekijken. Dit downloaden gaat via de onderstaande regels uit het script (zie Figuur 14).

Figuur 14 Opdrachtregels - lijst met waardes van dimensies exporteren naar CSV-bestand

```
J:\08_Persoonlijke_mappen\BTNFR\voorbeeld downloaden iv3 open data met R v20160114.R - R Editor
# Sla alle ingeladen meta-informatie op
setwd (Output_meta) #de 'workdirectory' instellen naar de waarde van variabe Output_meta
write.csv (meta$FunctieKostenplaatsBalanspost[1:2], file="Lijst FunctieKostenplaatsBalanspost.csv")
write.csv (meta$Categorie[1:2], file="Lijst Categorie.csv")
write.csv (meta$Gemeenten[1:2], file="Lijst Gemeenten.csv")
write.csv (meta$Verslagsoort[1:2], file="Lijst Verslagsoort.csv")
```

STAP 23. Selecteer in het script de bovenstaande tekst.

STAP 24. Voer de tekst uit (Ctrl+r).

In de map Output_meta staan nu vier nieuwe CSV-bestanden met daarin per dimensie welke voorkomens hierin zijn. Elk voorkomen bestaat uit drie kolommen, namelijk:

- Key;
- Title (omschrijving);
- Description (toelichting).

De waarden in de kolom Key worden later gebruikt bij het opstellen van een filter voor het ophalen van een selectie van Iv3 data. Zo kan worden opgezocht dat (in dit voorbeeld) in de dimensie Verslagsoort de waarde "2014X005" voor Key voorkomt, wat staat (volgens de Title) voor "Jaarrekening". Zo ook bijvoorbeeld in dimensie FunctieKostenplaatsBalanspost de Key-waarde "001 " als Key-waarde, wat volgens de Description staat voor "Functie - 001 - Bestuursorganen". Het is belangrijk om te weten dat de waarde van Key altijd een vast aantal tekens heeft. De lengtes van de variabele Key in de diverse dimensies zijn als volgt:

- FunctieKostenplaatsBalanspost bestaat uit 6 tekens;
- Categorie bestaat uit 8 tekens;
- Gemeenten bestaat uit 9 tekens;
- Verslagsoort bestaat uit 8 tekens.

Het komt regelmatig voor dat een **waarde van Key uit een aantal zichtbare tekens bestaat en voor de rest uit spaties**. Bijvoorbeeld bij de dimensie FunctieKostenplaatsBalanspost kan Key de waarde "001 " hebben, oftewel drie zichtbare tekens en drie spaties. Hiermee moet rekening worden gehouden bij het opstellen van het filter voor het ophalen van een selectie van data; **de lengte van een filterwaarde voor Key moet altijd zijn zoals hierboven aangegeven**.

STAP 25. Open de csv-bestanden met daarin per dimensie de voorkomende Key's.

STAP 26. Noteer de Key's van dimensies waarop u later gaat filteren, b.v. de key van de gewenste verslagperiode.

3.5 Iv3 data downloaden

3.5.1 Functie download_table

Aanroep van de functie

Het package cbsodataR bevat de functie (commando) “download_table” om Iv3 data te downloaden. De functie “download_table” bevat zeven parameters, die in willekeurige volgorde kunnen worden aangeroepen:

1. Id
betreft uit welke open data-tabel gegevens worden gedownload. Dit is de waarde van Identifier van de tabel, die is verkregen in de stap [3.4.1](#) Ophalen lijst beschikbare open data-tabellen. Er kan maar voor één tabel tegelijk worden gedownload.
2. Base_url
Voor Iv3 data moet deze altijd de waarde `http://dataderden.cbs.nl` krijgen
3. Dir
de naam van de map waar het programma de output moet schrijven
4. FunctieKostenplaatsBalanspost
5. Categorie
6. Gemeenten
7. Verslagsoort

De laatste vier parameters, ‘FunctieKostenplaatsBalanspost’, ‘Categorie’, ‘Gemeenten’ en ‘Verslagsoort’, komen overeen met de dimensies van de tabellen. Elk van deze parameters kan één of meerdere waarden van de Keys bevatten, zoals die verkregen zijn in het voorbeeld in paragraaf [3.4.2](#) Ophalen dimensies van een tabel. Het is niet verplicht om deze parameters op te geven; ze mogen weggelaten worden. Als een parameter van deze vier wordt weggelaten, dan is er geen filter op de betreffende dimensie en worden dus de data voor alle Keys van die dimensie opgehaald. Dus wanneer bijvoorbeeld er voor geen een dimensie Key's worden opgegeven, dat wordt alle data gedownload voor alle functies, categorieën, gemeenten en verslagsoorten. Naar mate er minder gefilterd wordt, duurt de opdracht ook langer. Wanneer bijvoorbeeld uit een gemeentetabel alle data voor de jaarrekeningen worden gedownload, dan zal de uitvoering van de opdracht zeker langer dan 30 minuten zijn.

Zie verder de onderstaande voorbeelden voor de precieze aanroep.

Output van de functie

De output van de functie download_table bestaat uit zeven CSV-bestanden. Het bestand ‘data.csv’ betreft de data. De andere bestanden betreffen metadata (vier dimensies, tabelinformatie en data-eigenschappen).

3.5.2 Voorbeeld 1: per dimensie één filterwaarde

In het onderstaande voorbeeld (zie Figuur 15) worden de Iv3-gegevens uit tabel 45005NED (gemeenten 2014) opgehaald. Met behulp van de metadata (verkregen bij [3.4.2](#) Ophalen dimensies van een tabel) worden alleen de gegevens uit de jaarrekening van de gemeente Alkmaar voor de functie 610 met lastencategorie 4.2.2 opgehaald.

Figuur 15 Opdrachtregels - downloaden selectie van jaargegevens 2014 voor Alkmaar

```
J:\08_Persoonlijke_mappen\BTNF\R\voorbeeld downloaden iv3 open data met R v20160114.R - R Editor
#-----
# Voorbeeld 1: download met één filterwaarde voor elke dimensie
#-----
download_table(id = "45005NED" # tabel nummer (uit 2014)
, Verslagsoort = "2014X005" # pak het jaarverslag
, base_url = IV3 # haal deze tabel uit IV3
, Gemeenten = "GM0361 "
, FunctieKostenplaatsBalanspost = "610 "
, Categorie = "L4.2.2 "
, dir = Output_data
)
```

STAP 27. Pas eventueel parameters aan om de eigen gewenste selectie te maken.

STAP 28. Selecteer de bovenstaande tekst in het script.

STAP 29. Voer de tekst uit (Ctrl+r) om de Iv3 data te downloaden.

NB omdat voor elke dimensie één waarde is opgegeven, zal er in dit voorbeeld wat betreft de gedownloade IV3-data, voor elk onderwerp (1ste plaatsing en 2de plaatsing) elk één waarde worden opgehaald. De gedownloade data wordt weggeschreven in zeven CSV-bestanden, in de map die eerder is opgegeven bij de variabele Output_data. Bij een succesvolle uitvoering verschijnt de onderstaande tekst in de console (zie Figuur 16):

Figuur 16 Schermoutput console - succesvolle download

```
Retrieving meta data for table '45005NED' http://dataderden.cbs.nl/ODataApi/odata/45005NED
Retrieving http://dataderden.cbs.nl/ODataApi/odata/45005NED http://dataderden.cbs.nl/ODataApi/odata/45005NED/TableInfos
Retrieving http://dataderden.cbs.nl/ODataApi/odata/45005NED http://dataderden.cbs.nl/ODataApi/odata/45005NED/DataProperties
Retrieving http://dataderden.cbs.nl/ODataApi/odata/45005NED http://dataderden.cbs.nl/ODataApi/odata/45005NED/FunctieKostenplaatsBalanspost
Retrieving http://dataderden.cbs.nl/ODataApi/odata/45005NED http://dataderden.cbs.nl/ODataApi/odata/45005NED/Categorie
Retrieving http://dataderden.cbs.nl/ODataApi/odata/45005NED http://dataderden.cbs.nl/ODataApi/odata/45005NED/Gemeenten
Retrieving http://dataderden.cbs.nl/ODataApi/odata/45005NED http://dataderden.cbs.nl/ODataApi/odata/45005NED/Verslagsoort
Writing TableInfos.csv...
Writing DataProperties.csv...
Writing FunctieKostenplaatsBalanspost.csv...
Writing Categorie.csv...
Writing Gemeenten.csv...
Writing Verslagsoort.csv...
Retrieving data from table '45005NED'
Reading...
Writing...
Done!
```

3.5.3 Voorbeeld 2: download met meer filterwaardes per dimensie

In dit voorbeeld worden zowel de gegevens voor de gemeente Alkmaar als Amsterdam opgehaald. De Key-waardes worden daarbij als volgt opgegeven:

Gemeenten = c("GM0361 ", "GM0363 ")

Meer dan twee waardes voor de parameter opgeven kan ook, bv.

Gemeenten = c("GM0361 ", "GM0363 ", "GM0599 ")

In het onderstaande voorbeeld worden verder weer de gegevens opgehaald voor de jaarrekening 2014 voor functie 610.

De parameter Categorie wordt weggelaten in dit voorbeeld. Dit betekent dat alle data worden opgehaald voor alle beschikbare 'Keys' van dimensie Categorie (i.e. alle categorieën, zowel lasten als baten) voor functie 610 voor de twee geselecteerde gemeenten.

NB als een dimensie (Verslagsoort, ..., Categorie) weggelaten wordt, dan worden alle data voor alle beschikbare Keys in deze dimensie opgehaald. Let hierbij op dat wanneer dit bij de dimensie Verslagperiode wordt gedaan, de opdracht, zeker bij de tabellen van gemeenten, lang kan duren (tot meer dan 30 minuten).

STAP 30. Selecteer de onderstaande tekst in het script

STAP 31. Voer de tekst uit (Ctrl+r) om de Iv3-data te downloaden

Figuur 17 Opdrachtregels - downloaden Iv3-data van twee gemeenten en exporteren naar CSV-bestand

```
R J:\08_Persoonlijke_mappen\BTNF\R\voorbeeld downloaden iv3 open data met R v20160114.R - R Editor
#-----
# Voorbeeld 2: download met meer filterwaarden voor een dimensie
#-----
download_table(id = "45005NED" # tabel nummer (uit 2014)
, Verslagsoort = "2014X005" # pak het jaarverslag
, base_url = IV3 # haal deze tabel uit IV3
, Gemeenten = c("GM0361 ", "GM0363 ") #voor gemeenten Alkmaar en Amsterdam
, FunctieKostenplaatsBalanspost = "610 "
, dir = Output data
)
```

NB de gedownloade IV3-data zal bestaan uit (voor beide onderwerpen) alle waarden voor Alkmaar en Amsterdam uit de jaarrekening voor alle categorieën voor alleen functie 610.