# Reharvest en updaten NARCIS-websitegrafieken

Versie 6-2-’18, laatst bewerkt door Emil Bode

### Achtergrond

We hebben nu op de NARCIS-website een aantal grafieken (zie <https://www.narcis.nl/metrics/Language/en>). Twee hiervan worden vanuit een script in NARCIS continu ververst, vier zijn statisch, en moeten door middel van een script ververst worden.

De betreffende vier zijn:

* Open access voor verschillende publicatietypen
* Open access voor proefschriften
* Open access voor artikelen
* Open access van artikelen, voor verschillende universiteiten

We willen deze één keer per jaar verversen, hier staat kort omschreven hoe. Uitgebreidere, technische instructies staan in een apart document, maar dit is de algemene opzet, om te plannen.

Het plan is om elk jaar de data van 1 juli op 1 augustus te presenteren, maar dit kan aangepast worden.

Algemene stappen om de grafieken te vernieuwen:

1. Eventueel installatie en opzetten bestanden
2. We kijken of er nog zaken in de scripts aangepast moeten worden
3. De nieuwe data wordt binnengehaald door een eerste script te runnen
4. De nieuwe grafieken worden gemaakt door een tweede script te runnen
5. Controle
6. De grafieken worden op de website geplaatst.
7. De nieuwe dataset wordt in EASY gedeponeerd (of DataVerse)

Hieronder zijn ze verder toegelicht:

1. Installatie en opzetten bestanden
   1. Als het proces verder vanaf de MacBook van Emil gebeurt kan dit overgeslagen, maar anders moet RStudio geinstalleerd worden, en moeten scripts gedownload, dit document doorgelezen.
   2. Wie: Wie verder gaat met stappen 2, 3 en 4.
   3. Hoe lang: Uur
2. Aanpassingen in de scripts en bestanden klaarzetten
   1. Wat aangepast moet worden zijn 2 hoofdzaken:
      1. Als we een andere selectie van grafieken willen, of bijv. extra lijnen erbij, een andere presentatie moet het grafiekgenereerscript aangepast worden
      2. Sommige labels zullen aangepast moeten worden:
         1. Sommige universiteiten zullen van systeem wisselen, dus de omzetting van de url naar universiteitslabel moet aangevuld worden
         2. De indeling in types verandert nog wel eens, dit moet aangevuld worden
         3. In overleg met Chris: kijken of sommige inhoud nog aangepast moet worden vanwege fouten in bronsystemen. Bijvoorbeeld: In 2017 gaf de UvA aan hun proefschriften het label ‘Other’ mee. Dit moet in het script dan verbeterd worden
   2. Wat altijd moet:
      1. de oude webpagina opslaan en klaarzetten, het script combineert dit met de nieuwe grafieken.
      2. Tenzij je alles opnieuw binnen wilt halen: de oude dataset op de goede plek zetten. Het script haalt dan alleen vernieuwingen op.
   3. Wie dit doet: waarschijnlijk Emil, overleg met Chris
   4. Tijdsduur: Paar uur. Waarschijnlijk in eerste instantie tot half uur, maar als er zaken anders lopen dan gedacht moet je hier terugkomen.
3. Nieuwe data binnenhalen
   1. In feite gewoon script aanzetten, en wachten tot hij klaar is.
      1. Alleen ondertussen kijken of en waar foutmeldingen komen, zo nodig kijken waar ze vandaan komen en aanpassen. In dat geval terug naar stap 2
   2. Wie: Waarschijnlijk Emil
   3. Hoe lang: als alles goed gaat binnen een dag, anders terug naar stap 2. Als het tegenzit moet deze stap dus een paar keer opnieuw, dus ruim de tijd voor nemen
4. De nieuwe grafieken worden gemaakt door een tweede script
   1. Dit werkt eigenlijk hetzelfde als de vorige stap, maar gaat sneller. Aanzetten en wachten tot hij klaar is, tot een kwartier.
      1. Ook hier ondertussen kijken of en waar foutmeldingen zijn, eventueel stap 2 weer doorlopen. Stap 3 hoeft niet opnieuw
   2. Wie: Waarschijnlijk Emil
   3. Hoe lang: Snel, < 1 uur
5. Controle
   1. Nagaan of er rare dingen zijn gebeurd.
      1. Om te beginnen natuurlijk kijken of de grafieken voor je gevoel kloppen (als een universiteit helemaal mist is er iets fout gegaan)
      2. Verder wordt in stap 3 ook een overzicht gemaakt van dingen die het script er vreemd uit vond zien. Dit nalopen en kijken of het problemen oplevert.
      3. Zie verder uitgebreide instructie
   2. Wie: Emil
   3. Hoe lang: tot een paar uur bezig met controle, maar het kan betekenen dat we weer terug moeten naar stap 2. Dit moet dus wel ruim vooraf gebeuren.
6. Grafieken op website
   1. Na afloop van stap 5 heb je een paar html-bestanden, die op de site gezet kunnen worden
   2. Wie: Wilko
   3. Hoe lang: Onbekend, paar uur? Wilko wil het wel graag vooraf weten wanneer dit moet gebeuren
7. DataSet-deposit
   1. Nieuwe dataset mag naar Easy of DataVerse, eventueel nog documentatie aanpassen
   2. Wie: Zelfde als bij stap 2 t/m 5
   3. Hoe lang: Uur

Totaalplaatje:

Vooraf weet je niet precies waar moeilijkheden zitten, dus het is handig om ruim te rekenen. In het gunstigste geval is alles na 2 dagen klaar (met een halve dag er echt mee bezig zijn), maar het kan ook tegenzitten, alles moet dan een aantal keer opnieuw.