



# Royal Netherlands Academy of Arts and Sciences (KNAW) KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

## Eindelijk een demografisch walhalla

Guéret, Christophe

2014

### **document version**

Early version, also known as pre-print

[Link to publication in KNAW Research Portal](#)

### **citation for published version (APA)**

Guéret, C. (2014). *Eindelijk een demografisch walhalla*. Hack de Overheid.  
<http://www.hackdeoverheid.nl/eindelijk-een-demografisch-walhalla-3/>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[pure@knaw.nl](mailto:pure@knaw.nl)

# Eindelijk een demografisch walhalla

Gepubliceerd op 21 maart 2014

door Christophe Guéret



Wil je weten hoe groot het gemiddelde Amsterdamse huishouden in 1930 was? Of hoe de verhouding tussen sociale status en beroep was gedurende de 19e eeuw? Heb je altijd al willen weten hoeveel procent van de alleenstaande moeders bij hun ouders woonden in 1971? De antwoorden op deze vragen zijn te vinden in de Nederlandse historische demografische gegevens.

## Een schat aan census data

Deze historische data is vrij verkrijgbaar op de volkstellingen website en bestrijkt de periode van 1795 tot 1971 met drie type gegevens (volkstellingen, beroepstellingen en woningtellingen). Op deze website kun je ongeveer 2000 Excel bestanden vinden waarin alle data uit historische demografische onderzoeken te vinden is. Verder zijn er scans te vinden van de stukken die nog niet gedigitaliseerd zijn. Al deze data is zorgvuldig gearchiveerd bij DANS.

## 2000 bestanden in één demografische database

Het probleem is dat deze data tot voor kort lastig bruikbaar was. De data bevatte namelijk spelfouten, was niet geharmoniseerd en lastig doorzoekbaar. De gefrustreerde gebruiker van de data moest veel werk handmatig uitvoeren en uitgebreide kennis van de data was geen overbodige luxe. Dit alles zorgde ervoor dat de data voor veel mensen niet toegankelijk was. Dit is zonde omdat demografische data zeer rijk is en inzichten kan verschaffen over een land. Dit is al helemaal het geval wanneer deze gecombineerd kan worden met andere databronnen, zoals oude kranten en foto's.

Het CEDAR-project maakt het echter mogelijk deze demografische data toegankelijk te maken voor iedereen die er mee aan de haal wil. Data Archiving and Networked Services (DANS) heeft samen met de Vrij Universiteit, Erasmus Universiteit, eHumanities KNAW en het IISH gewerkt aan het produceren van een dataset die alle demografische gegevens samenbrengt. In een door ons ontwikkelde API is deze data verder op een gemakkelijke wijze benaderbaar.

## Perfekte match voor Linked Open Data

De eerste versie van de volledige database is onlangs uitgebracht. Voor het datamodel hebben we gekozen om een graph met een lay out te gebruiken, die lijkt op de modellen (SDMX) die statische bureaus gebruiken om data te delen. Dit zorgt voor een accurate weergave en maakt het mogelijk om de database te koppelen met andere data binnen onze database en aan externe datasets. De daadwerkelijke implementatie wordt gedaan met behulp van Linked Open Data technologieën (SPARQL, RDF, JSON-LD) en de gratis en open source software tools die te vinden zijn op Github.

## Symposium om aan de slag te gaan met deze data!

Omdat de eerste versie van de database recent is gelanceerd, wordt er op 31 maart en 1 april een symposium georganiseerd. Op de eerste dag zullen alle ins- en outs van de huidige versie besproken worden en horen we ervaringen van experts uit Zwitserland en Frankrijk. De tweede dag staat in het teken van de praktijk en kunnen de developers aan de slag met de data. Er zijn ook historici aanwezig die graag samen met de developers kunnen nadenken over nuttige toepassingen van deze data.

De blog van de organisatie van Hack de Overheid is op de taart maken, de ontwikkelaars met de meest interessante ideeën kans op budget om hun prototype te ontwikkelen. Iedere geïnteresseerde is welkom en kan zich hier aanmelden.

Biografie

Laatste berichten



## Christophe Guéret

Christophe is a research associate at Data Archiving and Networked Services and the eHumanities group of the KNAW. His research activities are around the design of decentralised interconnected knowledge systems and their social implications. Christophe is a guest blogger for HackdeOverheid.nl

## Gerelateerde berichten:

1. Datablog: POMS De NPO heeft een API voor het systeem POMS (Publieke...
2. Datablog: Incontxt Incontxt is een aparte portal voor het beter toegankelijk maken...

0 Comments

Hack de Overheid

Login ▾

Sort by Best ▾

Share ↗ Favorite ★



Start the discussion...

Be the first to comment.

ALSO ON HACK DE OVERHEID

WHAT'S THIS?

### Geodata, hoe het zoveel beter kan!

1 comment • a year ago



**Bruno Scheele** — Interessant om te lezen! Maar de blog kan wel wat linkjes gebruiken. Ik kan de issue-tracker op Github bijvoorbeeld nergens

### Datablog: Incontxt

1 comment • a year ago



**me** — wat the hek>?

### Challenge: €5.000 voor allerbeste cultuurapp

5 comments • 8 months ago



**Breyten Ernsting** — Dirk, ga gewoon bouwen. Je weet best hoe werkt :)

### Culture Harvests 2014

2 comments • 9 months ago



**Lex Slaghuis** — Hallo Gerben, De huidige code base met Docker.io is gebaseerd op Python ETL. Om een harvester toe te voegen of daar in zaken