



Aanvraag voor een Klein DataProject (KDP)  
Elektronisch verzenden aan  
Data Archiving and Networked Services  
t.a.v. Dr. J.H. Harmsen  
e-mail: [henk.harmsen@dans.knaw.nl](mailto:henk.harmsen@dans.knaw.nl)

Kleine dataprojecten kunnen zich richten op het beschrijven en toegankelijk maken van één of meer belangrijke datasets. Het kunnen ook haalbaarheidsstudies zijn, of verkenningen op het gebied van data-archivering op een nieuw terrein, discipline of data-formaat. Het kunnen ook pilots zijn voor vernieuwde manieren van archiveren of toegankelijk maken van onderzoeksdata. Na afloop van het project zijn de data die bij dit KDP betrokken zijn Open Access<sup>1</sup> beschikbaar via DANS.

Titel project:	<i>Up and Running: the recovery of NLGis</i>
Datum aanvraag:	31 januari 2014
Discipline:	Sociale en Economische Geschiedenis, Demografie, Sociologie
Naam aanvrager:	Dr. R. L. Zijdeman Prof. Dr. K. Mandemakers Prof. Dr. K. Hofmeester
Naam instelling:	Internationaal Instituut voor Sociale Geschiedenis
Adres:	Cruquiusweg 31, 1019 AT, Amsterdam
Telefoon:	+31 20 688 5 886, +31 6 420 80 665
e-Mail:	<a href="mailto:richard.zijdeman@iisg.nl">richard.zijdeman@iisg.nl</a>

---

<sup>1</sup> Indien de wet dit toelaat.

**Samenvatting (15-25 woorden):**

Dit project integreert databases op gemeenteniveau met NLGis shapefiles en ontsluit deze binnen 6 maanden, via een website, API en R package.

**Projectplan (400-600 woorden)***Introductie*

Met de niet langer voor handen zijnde applicatie NLGis<sup>i</sup> had Nederland een koppositie in het historisch onderzoeksveld van de demografie, economie en sociologie. Waar andere landen op dit moment nog bezig zijn shapefiles van historische kaarten te maken, bestond voor onderzoekers van Nederland de mogelijkheid om data van de afgelopen twee eeuwen op gemeenteniveau in kaart te brengen met behulp van aan de kaarten gekoppelde databases en bovendien de mogelijkheid om eigen onderzoeksdata op de kaarten te plotten.

Dit project zal Nederland de pionierspositie op het gebied van historisch onderzoek teruggeven, door de kaarten en data in NLGis opnieuw aan te bieden. Naast de mogelijkheid zoals deze in NLGis bestond (het online plotten van kaarten met bestaande of geüploade data) biedt dit project een veel betere integratie van de kaarten en databases in de onderzoeksworkflow door de databases via een API beschikbaar te maken en de package "hismap" te maken in open source statistiek programma R<sup>ii</sup>. Specifiek biedt dit project:

- het garant staan voor interactieve ontsluiting van de bestaande databases: Historische Ecologische Databank (HED) en de Historische Databank Nederlandse Gemeenten (HDNG), voor een periode van ten minste 5 jaar én de mogelijkheid om deze databanken en data van onderzoekers zelf te plotten op de kaarten van NLGis op de volgende manieren:
  - o via een website (zoals NLGis)
  - o via een API (garandeert toegang tot diverse (commerciële) pakketten)
  - o via package "hismap", een nieuw te schrijven package voor het open source statistiekpakket R.

*Uitvoering*

De datasets zijn gestandaardiseerd en kunnen gelinkt worden via de Amsterdamse Code en CBS Code voor gemeenten. Bovendien zouden uitbreidingen van deze datasets en verbeteringen aan de kaarten, zoals voorgesteld in het KDP project "*Webservices voor voormalige gemeentes*" kunnen worden gekoppeld aan de bestaande datasets. De data zullen interactief worden ontsloten door het Internationaal Instituut voor Sociale Geschiedenis. Het instituut biedt onderzoekers wereldwijd en reeds jarenlang dergelijke services aan, zoals de Historische Steekproef Nederland (HSN)<sup>iii</sup>, CLIO-INFRA<sup>iv</sup>, en diverse Collaboratories<sup>v</sup>.

R is open-source statistiek software en inmiddels uitgegroeid tot een serieuze tegenhanger van commerciële pakketten als IBM Statistics (SPSS) en Stata. Wereldwijd wordt dit pakket onderwezen op universiteiten zoals Cambridge University, en in Massive Open Online Courses (MOOCs), zoals Coursera<sup>vi</sup>. Kwaliteit van packages in R wordt gegarandeerd door de CRAN Repository Policy<sup>vii</sup>, waar nieuwe packages aan dienen te voldoen eer ze in CRAN (Comprehensive R Archive Network<sup>viii</sup>) opgenomen worden. Functionaliteit zoals voorgesteld bestaat reeds in R voor contemporaine kaarten (onder meer in "rworldmap" package<sup>ix</sup>). De techniek is dus gegarandeerd. Ook de techniek om een API te maken voor de databases is gegarandeerd, dergelijke API's worden onder meer gebruikt door de WorldBank<sup>x</sup> en het Open Archives Initiative<sup>xi</sup> en kunnen leiden tot belangrijke uitbreidingen, zoals het modelleren van geografische relaties tussen gemeenten in een vorm die aansluit bij het Semantisch Web, zoals voorgesteld in het KDP voorstel "Webservices voor voormalige gemeentes".

#### *Afronding*

Het project is geslaagd indien de drie producten (website, API en R package "hismap") binnen 6 maanden na aanvang van het project toegankelijk zijn. Voor details zie "Begroting en doorlooptijd".

#### **Begroting en doorlooptijd (max. 1 A4):**

Voor de te ontwerpen elementen wordt veelal gebruik gemaakt van bestaande open access software (Drupal 7 database API<sup>xii</sup> voor de API, diverse map packages in R voor de "hismap" package, en IISG in-house software voor de website). Derhalve kunnen de elementen in een vrij kort tijdsbestek gemaakt worden. Voor een betrekkelijk kleine opdracht als deze is het echter niet mogelijk iemand full-time aan te stellen. Daarom zal gebruik gemaakt worden van een software-ontwikkelaar die reeds onder contract staat en die voor 2 dagen in de week (0,4 fte) aan het project zal werken. De looptijd van het project duurt dan ook 6 maanden (10 \* 1/0,4 weken). Bij aanvang op 1 juni zou het project dus 1 december 2014 zijn afgerond. Een ontwikkelaar met enige jaren ervaring wordt ingeschaald in schaal 10.3 (uurtarief 31,39 euro), waarbij de totale kosten op 10K euro uitkomen.

Element	Duur bij 1,0 fte	Duur bij 0,4 fte
API	1,0 weken	2,5 weken
R Library	4,0 weken	10,0 weken
Website	5,0 weken	12,5 weken
<i>Totaal</i>	<i>10,0 weken</i>	<i>25,0 weken</i>

URL's gecontroleerd op 30 januari 2014.

---

- <sup>i</sup> <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:44426>
- <sup>ii</sup> <http://www.r-project.org>
- <sup>iii</sup> <http://www.iisg.nl/hsn/indexnl.html>
- <sup>iv</sup> <http://www.clio-infra.eu>
- <sup>v</sup> <http://collab.iisg.nl>
- <sup>vi</sup> <https://www.coursera.org>
- <sup>vii</sup> <http://cran.r-project.org/web/packages/policies.html>
- <sup>viii</sup> <http://cran.r-project.org>
- <sup>ix</sup> <http://cran.r-project.org/web/packages/rworldmap/index.html>
- <sup>x</sup> <http://data.worldbank.org/node/9>
- <sup>xi</sup> <http://www.openarchives.org>
- <sup>xii</sup> <https://api.drupal.org/api/drupal>