



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

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 Access¹ beschikbaar via DANS.

Titel project:	DMD – Diepere Maritieme Data
Datum aanvraag:	31 januari 2014
Discipline:	geschiedenis
Naam aanvragers:	Jurjen R. Leinenga en Victor de Boer
Naam instelling:	zelfstandig onderzoeker, resp. VU, Amsterdam
Adressen:	Houtweg 337, 7823 PW Emmen

¹ Indien de wet dit toelaat.

VU, De Boelelaan 1105, 1081 HV Amsterdam

Telefoons: 0591-620021

06-47418031

e-Mailadressen: leinenga@xs4all.nl
vdeboer@gmail.com

Samenvatting:

Het project beoogt bestaande datasets te verrijken door analoge archiefbronnen te ontsluiten voor maritiem-historisch onderzoek met digitale foto's van de bronnen.

Projectplan (400-600 woorden)

Binnen het CLARIN project "Dutch Ships and Sailors" (DSS) wordt een aantal maritiem-historische archiefbronnen gedigitaliseerd en als *Linked Data* beschikbaar gemaakt voor hergebruik door onderzoekers. Waar DSS zich alleen richt op metadata is een verdere verrijking met scans of foto's van de originele archiefbronnen van grote waarde voor historisch onderzoekers. Op deze manier kunnen analyses op de metadata direct worden getoetst aan de originele bronnen.

Het project wil een dataset met informatie uit scheeps- en bemanningslijsten in Noord-Nederlandse archieven, maritieme musea en privébezit verrijken door de hierbij gebruikte analoge bronnen digitaal te fotograferen en te documenteren en aan de dataset te linken. Hiermee wordt de dataset zelf verifiëerbaar en worden de originelen – deels in kwetsbare toestand – ontlast. De gebruiker kan voor het eerst ook niet in de database opgenomen informatie ontsluiten, indien dat voor eigen onderzoek relevant is. Door dit linken aan de bronnen zal de levensduur van de dataset worden verlengd en de waarde ervan worden verhoogd.

Het gaat totaal om naar schatting zo'n 25.000 foto's. Door het zelf fotograferen in overleg met de archieven is er geen conflict met copyright van derden.

Het aantal instellingen dat dient te worden bezocht, is : 13 archieven, 2 maritieme musea en 2 particulieren in de provincies Groningen, Friesland, Drenthe en Overijssel. Een gedetailleerde lijst hiervan is toegevoegd.

Het project gaat uit van twee wetenschappelijke projectmedewerkers, met voor Jurjen Leinenga een inzetbaarheid van tweeëneenhalve werkdag per week voor een periode van acht weken en voor Victor de Boer van twee werkdagen per week voor een periode van acht weken. Jurjen Leinenga, historicus, zal het contact met de archieven, musea en particulieren leggen, het fotograferen op zich nemen en dit documenteren. Victor de Boer, Linked Data specialist, zal de links tussen de dataset en de fotocollectie aanmaken en de RDF metadata maken. Het tijdpad en de uitvoering kunnen door Leinenga worden gecontroleerd en indien nodig bijgestuurd.

Geproduceerde (Linked) Data

De data zoals geproduceerd binnen het DSS project zullen via Linked Data technologieën beschikbaar worden gemaakt via webserver, gehost bij de VU Amsterdam danwel Huygens ING, conform de DSS projectbeschrijving. De foto's en metadata, zoals die binnen het hier voorgestelde project worden geproduceerd, zullen worden aangeboden via een webserver van VU Amsterdam danwel DANS. Linked Data technologie biedt een uitgelezen kans om heterogene en gedistribueerde bronnen en metadata te koppelen door middel van RDF links tussen URIs. Hierdoor is het niet noodzakelijk dat de foto's op dezelfde webserver gehost worden als de metadata. Deze aanpak is al eerder succesvol gebleken in het CLARIN Verrijkt Koninkrijk project, waar de Linked metadata en de textuele content van de historiografische tekst van Loe de Jong op aparte servers gehost zijn (zie ook de Boer et al. 2013).

Waarde voor maritiem-historisch onderzoek

In eerder onderzoek is gebleken dat het digitaal aanbieden van metadata onderzoekers kan helpen om snel onderzoeksvragen te formuleren, hypotheses te toetsen en overzichten te genereren. Daarbij is steeds expliciet naar voren gekomen dat een heldere *provenance* (herkomst) van de data essentieel is voor correct gebruik van deze data (zie ook Ockeloën et al.). Door het toevoegen van de scans van de originele analoge bronnen zal de *provenance-chain* expliciet digitaal opvraagbaar zijn.

Illustratie van gebruik

Simon is als onderzoeker geïnteresseerd in overeenkomsten en verschillen tussen zeelieden die eind 19^e eeuw aanmonsterden in Delfzijl en Harlingen. Een SPARQL query op de DSS data resulteert in het inzicht dat mensen die in Delfzijl aanmonsterden vaker naar Groot-Brittannië afreisden. Simon vertrouwt dit niet en breidt zijn query uit, waarna hij een lijst met scans van originele bronnen te zien krijgt. Na het bekijken van een aantal hiervan ziet Simon dat dit daadwerkelijk in de originele monsterrollen staat. Simon besluit hier verder onderzoek naar te doen.

Literatuur

Victor de Boer, Johan van Doornik, Lars Buitinck, Maarten Marx, Tim Veken, Kees Ribbens. Linking the Kingdom: Enriched Access To A Historiographical Text. Proceedings of KCAP 2013, Banff, Canada, 23-26 June 2013.

Niels Ockeloën, Antske Fokkens, Serge ter Braake, Piek Vossen, Victor de Boer, Guus Schreiber, and Susan Legene. BiographyNet: Managing Provenance at multiple levels and from different perspectives. 3rd International Workshop on Linked Science 2013—Supporting Reproducibility, Scientific Investigations and Experiments (LISC2013)

Begroting en doorlooptijd (max. 1 A4):

[...]

[De begroting is uit de publieksversie van de projectbeschrijving gelaten. Binnen het project zullen twee onderzoekers werkzaam zijn: dr. Jur Leinenga en dr. Victor de Boer.]