

Data Archiving and Networked Services



Data Archiving and Networked Services (DANS)

**JAARVERSLAG
2009-2010**

Door data gedreven

1. Inleiding

Missie en organisatie

De missie van DANS is om de blijvende toegankelijkheid van digitale onderzoeksbestanden voor de alfa- en gammawetenschappen te garanderen. DANS levert databestanden en daaraan gerelateerde diensten aan onderzoekers en bevordert samenwerkingsprojecten met en tussen onderzoeksgroepen. DANS is partner in Europese en andere internationale dataorganisaties, zodat gegevensbestanden uit het buitenland gemakkelijk te verkrijgen zijn voor Nederlandse onderzoekers. DANS streeft naar een zo open mogelijke toegang tot wetenschappelijke gegevensbestanden, maar beschermt gegevens waar dat nodig is. Op deze wijze biedt DANS garanties omtrent datakwaliteit, duidelijke voorwaarden voor datagebruik en heldere criteria voor digitale databewaarplaatsen.

Inmiddels werken er zo'n veertig mensen bij DANS, waarvan ruim een derde gefinancierd op projectbasis. Uiteraard vergde die groei regelmatig aanpassingen van de organisatie- en overlegstructuur. De organisatiestructuur is de afgelopen vijf jaar flexibel aangepast aan de eisen van een groeiende organisatie die bezig was zijn plaats te bepalen in het wetenschappelijke informatie landschap. De vier coördinatiegroepen binnen DANS zijn geëvolueerd tot Archief, ICT-ontwikkeling, Infrastructuur en Secretariaat. Het merendeel van medewerkers heeft aangegeven een direct leidinggevende binnen deze groepen, in de vorm van een coördinator, op prijs stellen. In de eigen organisatie worden vier bekwame coördinatoren gevonden om de nieuwe vier secties aan te sturen. Deze organisatorische verandering wordt in goed overleg met de OC uitgevoerd en behoeft daarom geen reorganisatie. Begin 2010 draait de situatie naar behoren en wordt nog een enkele onderdeel (ICT-beheer) van infrastructuur overgeheveld naar ICT-diensten.

Behalve de basiswerkzaamheden heeft DANS een groot aantal projecten uitgevoerd, vaak in een ondersteunende rol van wetenschappelijke communities. Ook internationaal heeft DANS zich stevig geroerd. In vrijwel alle grootschalige infrastructuurprojecten op alfa- en gammagebied, nationaal en internationaal, speelt DANS een aanzienlijke, soms leidende, rol.

2. Ontwikkelingen 2009 en 2010

In 2009 werkte DANS met onderzoeksgroepen in binnen- en buitenland aan tientallen landelijke en internationale dataprojecten. Het onderzoek aan die data zelf wordt doorgaans door medewerkers van universiteiten en wetenschappelijke instituten verricht. DANS speelt daarbij veeleer een dienstverlenende en ondersteunende rol. De eigen R&D is toepassingsgericht of draagt bij aan de kennis op het gebied van digitale duurzaamheid, zoals het MIXED-project, waarin software wordt ontwikkeld voor het duurzaam opslaan van data in een intermediair XML-formaat.

e-depot Nederlandse archeologie

Eén van de belangrijkste resultaten van DANS in 2009 was de spectaculaire groei van het E-Depot Nederlandse Archeologie (EDNA), dat in korte tijd uitgroeide tot het grootste digitale archief binnen DANS. EDNA begon nog maar vijf jaar geleden als pilot project, maar heeft in korte tijd het veel oudere sociaalwetenschappelijk archief (oorspronkelijk Steinmetz-archief, 45 jaar geleden opgericht) en het historisch data-archief (uit 1989) in volume en aantal datasets gepasseerd. Van de ruim 12.000 datasets in het DANS-archief zijn er eind 2009 zo'n 10.000 archeologisch van aard.

De archeologie is academisch gezien een betrekkelijk kleine discipline. Dat het vakgebied desalniettemin zoveel data produceert, komt vooral omdat het sinds het "Verdrag van Malta" verplicht is om overal in de Europese Unie waar de schop de grond in gaat voor een nieuwbouwwijk, een industrieterrein of een spoorlijn, te controleren of zich iets van cultuurhistorische waarde in de bodem bevindt. Al deze projecten leveren informatie op over dat stukje van het grondgebied, en al deze informatie wordt (wat Nederland betreft) verzameld in EDNA. Dat gebeurde in een reeks van projecten, in samenwerking met de Rijksdienst Cultureel Erfgoed (voorheen: Rijksdienst Oudheidkundig Bodemonderzoek) en de Universiteit Leiden. Ook oudere informatie, in de vorm van archeologische rapporten, werd gedigitaliseerd.

De aard van de gegevensbestanden in de archeologie is anders dan in de sociale wetenschappen of in de geschiedenis. In de sociale wetenschappen hebben we vooral van doen met survey-onderzoek onder de bevolking. Uit zo'n enquête, die over duizenden respondenten gaat en doorgaans enkele tonnen kost, komt doorgaans maar één grote datatabel. Maar bij een archeologische opgraving wordt de plattegrond weergegeven in een geografisch informatiesysteem, worden de scherven getekend in software voor computer-aided design (CAD), zijn er databases van de vondsten, digitale foto's van de opgraving en bodemdoorsneden, rapporten in de vorm van tekstbestanden, et cetera. Eén opgraving kan op deze manier duizenden bestanden omvatten, waarbij één afbeelding in hoge resolutie groter kan zijn dan alle cijfers over de Nederlandse bevolking in een sociaalwetenschappelijke datamatrix. Maar ook geteld in aantal projecten is het archeologisch archief spectaculair gegroeid. En hoewel de top-tien van meest geraadpleegde bestanden in 2009 wordt aangevoerd en gedomineerd door een sociaalwetenschappelijk bestand (het Nationaal Kiezersonderzoek staat op 1), behoren ook data over de Steentijd van Nederland en over de opgravingen op de tracé's van de Hoge Snelheidslijn en de Betuwelijn tot DANS' "*Greatest Hits*". Als buitencategorie kan overigens niets op tegen de gedigitaliseerde Volkstellingen 1795-1971, die ook door veel amateurhistorici worden geraadpleegd.

Door data gedreven: 45-jarig jubileum na 4 jaar DANS



In december vierde DANS met een feestelijk symposium het feit dat in Nederland al meer dan 45 jaar wetenschappelijke data worden gearchiveerd en beschikbaar gesteld. In 1964 werd het eerste wetenschappelijke data-archief in Nederland opgericht: het Steinmetzarchief voor de sociale wetenschappen. Maar dat is niet het enige jubileum op datagebied. In 1989, twintig jaar geleden dus, werd het Nederlands Historisch Data Archief opgericht en vijf jaar later het Wetenschappelijk Statistisch Agentschap. Vijf jaar geleden tenslotte ging het elektronisch Depot voor de Nederlandse archeologie van start. Al die instituten zijn intussen onderdeel geworden van DANS, dat in de verschillende lustra de aanleiding zag voor dit symposium.

Tijdens de bijeenkomst in Den Haag heeft de geschiedenis van de diverse archieven maar een zijdelingse rol gespeeld. De opzet was juist om daar alle aandacht te geven aan de toekomst, die er voor de Nederlandse alfa- en gammawetenschappen zonnig uitziet.

Niet alleen kunnen er nog enorme hoeveelheden data beschikbaar komen die bijdragen aan beter, goedkoper en dus meer goed onderzoek. Ook leidt de groeiende beschikbaarheid van data voor hergebruik tot steeds meer nieuwe methodes en tot het beantwoorden van andere onderzoeksvragen dan tot nu toe. Op dit symposium kwamen onderzoekers uit zeer uiteenlopende disciplines binnen de alfa- en gammawetenschappen aan het woord over nieuwe vragen, methodes en antwoorden die zij met behulp van innovatief datagebruik hebben verkend.

Een miljoen bestanden

Dat het miljoenste bestand dat in 2010 bij DANS werd gearchiveerd een digitale foto was, was geen toeval. 2010 was een jaar waarin beeld-, geluid- en videodata een belangrijke rol speelden. Het meest opmerkelijk daarbij was de digitale archivering en toegankelijk stelling van een reeks collecties met kwalitatieve interviews, die op het grensgebied liggen van de "oral history" en de kwalitatieve sociale wetenschappen.

De mogelijkheid om kwalitatieve onderzoeksdata toegankelijk te maken voor heranalyse is door onderzoekers lang in twijfel getrokken. De nuances van een interview in de vorm van een lopend gesprek konden lastig worden gecodificeerd, zo meenden de meeste specialisten. De insteek van de vragen en antwoorden is ook vaak zo persoonlijk, dat buiten de context van dat specifieke onderzoek de interviewgegevens zo goed als onbruikbaar zouden zijn. Ook de noodzaak van anonimiseren (niet alleen van de respondenten, maar ook van de personen die zij noemen) zou bij dit type onderzoeksbronnen onmogelijk zijn.

Al bij de start van DANS werd een wens van de Sociaal Wetenschappelijke Raad meegegeven om aandacht te besteden aan de mogelijkheden van het archiveren van kwalitatieve onderzoeksdata. In 2010 heeft dit tot vernieuwende resultaten geleid. Samen met een enthousiaste community van historici en sociale wetenschappers heeft DANS in 2010 aangetoond dat het hergebruik van kwalitatief onderzoeksmateriaal niet alleen heel goed mogelijk is, maar ook tot nieuwe inzichten leidt.

De mogelijkheden die Verrijkte Publicaties (*Enhanced Publications*) bieden, gaven hiertoe de aanstoot. In dergelijke hybride publicaties worden verschillende media in

onderlinge samenhang aangeboden. Het gaat in dit geval om video- of spraakbestanden van interviews, die zijn gekoppeld aan al dan niet gecodeerde en geannoteerde interviewteksten. Samen met een goede documentatie van de interviewcontext en een analyse van de interviewinhoud, biedt de multidisciplinaire e-publicatie "Wat veteranen vertellen" een combinatie van boek en website¹. Interviewfragmenten kunnen door de lezer van het boek online worden afgeluisterd, en desgewenst kunnen ook de volledige interviews in het archief worden opgezocht en gedownload.

DANS heeft inmiddels een aanzienlijke collectie van enkele duizenden kwalitatieve interviews verzameld, die vooral betrekking hebben op oorlog en traumaverwerking. Deze collectie zal de komende jaren nog verder uitgroeien en geleidelijk aan ook andere thema's gaan bestrijken.

Toekomstvisie

Eind 2009 presenteerde DANS een visienota, waarin het instituut voorstelt om zijn werkterrein de komende jaren uit te breiden. Tot nu toe beperkte het instituut zich tot de alfa- en gammawetenschappen. Een uitbreiding naar de bètawetenschappen brengt veel voordelen mee, mede omdat een groot deel van diensten die DANS heeft ontwikkeld toepasbaar zijn op elk vakgebied. Een bredere activiteit van DANS is goed voor het bevorderen van digitale duurzaamheid, waarvoor niet alleen de Nationale Coalitie Digitale Duurzaamheid (NCDD) pleit maar ook organisaties als KNAW en NWO. Beide laatste hebben zich ook verbonden aan het streven naar Open Access. DANS zal daar als disciplineonafhankelijke organisatie een sterke impuls aan kunnen geven. Overige voordelen van de nieuwe strategie en aanpassingen in de dienstverlening zijn na te lezen in de visienota voor de DANS strategie 2011-2015.

¹ <http://www.watveteranenvertellen.nl/index.html>

² Schatting i.v.m. EDNA cijfers

3. Projecten

3.1 Historische Databank Nederlandse Gemeenten (HDNG)

Als onderdeel van het door NWO gefinancierde project *Hub for Aggregated Social History* (HASH) bouwt DANS een webportaal waar verschillende historische databronnen over Nederlandse gemeenten (zoals de volkstellingen) eenvoudig kunnen worden doorzocht. Dergelijk onderzoek is nu een tijdrovende bezigheid omdat de bronnen zijn verspreid over diverse bestanden en locaties en omdat deze beperkt zijn gestructureerd. HASH verzamelt deze bronnen en maakt het mogelijk de ontbrekende structuur toe te voegen. Vervolgens kan een onderzoeker via het webportaal alle bronnen doorzoeken op gemeente, jaartal en onderwerp. Het resultaat kan vervolgens eenvoudig worden gedownload in Excel voor verdere analyse.

3.2 DRIVER II

Het Europese project DRIVER II (Digital Repository Infrastructure Vision for European Research), gefinancierd vanuit het zevende Kaderprogramma (Framework Program, FP7) van de Europese Unie, bouwt een netwerk van interdisciplinaire Open Access repositories. DANS is bij dit project betrokken voor de ontwikkeling van verrijkte publicaties: publicaties die direct zijn gelinkt met de bijbehorende onderzoeksdata. Hiervoor heeft DANS samen met KNAW Onderzoekinformatie een *demonstrator* gebouwd. Op diverse internationale conferenties is deze gepresenteerd



om de mogelijkheden en uitdagingen te presenteren aan zowel onderzoekers als beleidsmakers. Op basis hiervan zijn onder andere in Nederland meerdere vervolgprojecten gestart om mogelijkheden en uitdagingen van verrijkte publicaties verder te onderzoeken.

3.3 Alfalab

Zes instituten van de KNAW hebben het initiatief genomen om de toepassing van digitale methoden en technieken in de geesteswetenschappen een impuls te geven. Onder de noemer Alfalab werken Huygens Instituut, Meertens Instituut, Fryske Akademy, de Virtual Knowledge Studio, IISG en DANS samen aan een eerste versie van een online laboratorium waarin geesteswetenschappelijke data en digitale gereedschappen kunnen worden ondergebracht en gebruikt. In de eerste fase van het project, die loopt van 2009 tot en met 2011, worden de bestaande technieken en methoden en de beschikbare data en gereedschappen onderzocht. Met behulp van surveys, interviews en workshops worden ook de behoeftes en verwachtingen van onderzoekers in diverse disciplines in kaart gebracht. Type en omvang van het huidige gebruik en de mogelijke succesfactoren voor toekomstige impact worden daarbij in beeld gebracht en er wordt een eerste online werkomgeving gebouwd waarin digitale tekstuele, geografische en demografische data en analysegereedschappen worden gedemonstreerd. DANS is verantwoordelijk voor het werkpakket 'Portaal' waarin zal worden gedemonstreerd hoe verschillende data en gereedschappen op verschillende niveaus kunnen worden geïntegreerd. In 2010 is gestart met de bouw van een webportaal, een Linked Data demonstrator waarin verschillende soorten annotaties kunnen worden verzameld en onderzocht en een Rich Internet Publication waarin een publicatie worden geïntegreerd met data en tools.

<http://alfalablog.huygensinstituut.nl/wp-content/themes/alfalab/images/alfalab.gif>

3.4 Persistent Identifiers

DANS voorziet sinds 2008 in een nationale *persistent-identificer*-infrastructuur waarmee onderzoekers duurzaam aan gedeponeerde onderzoeksresultaten kunnen refereren. Sindsdien is het gebruik en de behoefte aan meer diensten toegenomen. Hierdoor is de behoefte ontstaan aan internationale samenwerking. In 2010 nam DANS deel aan het Knowledge Exchange project PersID, waar zij met andere Europese partners heeft gewerkt aan het opbouwen van een gezamenlijke Europese persistent identifier infrastructuur. Binnen dit project werken de nationale infrastructuren van Denemarken, Duitsland, Finland, Italië, Nederland en Zweden samen op het gebied van techniek, beleid en communicatie.

Image 2010: <http://www.knowledge-exchange.info/files/billeder/PersId/persid-3.gif>

3.5 OpenAire

OpenAIRE wordt gesubsidieerd uit het Zevende Kaderprogramma van de Europese Commissie. Doel is om in Europa Open Access op praktijkniveau ingevoerd te krijgen. Bij het project zijn 38 partners uit 27 Europese landen betrokken. In augustus 2008 introduceerde de Europese Commissie een 'Open Access-pilot', een verkenning waarvoor rond een vijfde van het budget van het Zevende Kaderprogramma (voor wetenschappelijk onderzoek) werd gereserveerd. Onderzoekers in zeven richtingen, waarbij de sociaal-economische & geesteswetenschappen als één worden beschouwd, worden in dat kader verplicht om hun onderzoekpublicaties in een Open Access-repository te deponeren. Dat moet dan wel een repository zijn van een onderzoekinstelling of een overkoepelende repository voor een specifiek vakgebied. OpenAIRE moet hen helpen te voldoen aan de gestelde eisen. Daarvoor wordt onder andere gewerkt aan een Europese Helpdesk en het opzetten van een OpenAIRE-portal en e-infrastructuur voor de repository-netwerken. Ook worden in samenwerking met vijf vakdisciplines de mogelijkheden onderzocht van het beheer van wetenschappelijke gegevens.

In het projectconsortium is de Europese topexpertise op het gebied van Open Access- en repository-infrastructuur verenigd. Er zal een netwerkstructuur worden opgezet met partners in alle lidstaten van de Europese Unie, plus Noorwegen. Ieder land heeft zijn consortiumpartners, of – in het geval van Luxemburg – heeft zijn steun toegezegd aan het project, waarbij wordt voortgebouwd op resultaten uit het DRIVER II project. Gecoördineerd door SURFfoundation zal een Europese helpdesk worden opzetten, waarvoor Universiteit Utrecht de Nederlandse liaison zal zijn. DANS gaat werken aan een 'subject specific roadmap report', waarin het de specifieke eisen zal beschrijven die gelden voor repositories voor sociaal-economische en geesteswetenschappen.



3.6 Dutch Atlantic Connections

Bij DANS is in 2009 in de vorm van een stage door Carien Meijerman van 1 februari 2009 tot 1 augustus 2009 gewerkt aan de creatie van een database waarin een aantal databestanden over de Nederlandse handel van en naar het Atlantisch gebied

in de zeventiende en achttiende eeuw geïntegreerd worden. In 2009 zijn de meeste van deze bestanden, die in de afgelopen twintig jaar door verschillende historische onderzoekers gecreëerd zijn, geconverteerd naar up-to-date formaten. Daarnaast zijn zij geharmoniseerd en zijn speciaal persoonsnamen, scheepsnamen en geografische aanduidingen zoveel mogelijk geüniformeerd. Deze databestanden, zijn op dit moment al afzonderlijk raadpleegbaar bij DANS via EASY. Dit project is uitgevoerd in samenwerking met dr. H.J. den Heijer van de Leidse Universiteit (maritieme geschiedenis).

Bij DANS is in de vorm van een stage door Erik Odegard van 1 januari tot 30 juni 2010 verder gewerkt aan de creatie van een geïntegreerde database met betrekking tot een aantal databestanden over de Nederlandse handel van en naar het Atlantisch gebied in de zeventiende en achttiende eeuw. Deze databestanden waren begin 2010 grotendeels al afzonderlijk raadpleegbaar bij DANS via EASY. In deze periode is het ontwerp gemaakt voor de geïntegreerde relationele database en daadwerkelijk uitgevoerd. Deze relationele database is zo uitgevoerd dat hier ook nieuwe gegevensbestanden aan toegevoegd kunnen worden. Dit project is begeleid door dr. H.J. den Heijer van de Leidse Universiteit (maritieme geschiedenis).

3.7 MeMO

In mei 2009 is begonnen aan het project Medieval Memoria On-Line (MeMO). Dit door NWO gefinancierde project onder leiding van dr. Truus van Bueren (Universiteit Utrecht) beoogt verschillende medievistische databronnen integral doorzoekbaar en toegankelijk te maken. DANS heeft in deze eerste fase meegewerkt aan het opstellen van de MeMO Description Standard voor het beschrijven van middeleeuws materiaal, zoals grafzerken, schilderijen en klooster-registers. Verder is gewerkt aan het vertalen van de description standard naar database implementaties, en voorbereidend werk geleverd voor het invoerwerk. MeMO loopt tot medio 2012.

3.8 EASY II

Eind 2008 is besloten om niet meer te werken met eSciDoc als laag tussen Fedora en EASY. Alhoewel eSciDoc veel van de functionaliteit biedt die DANS nodig heeft, kan het EASY team niet rekenen op de gewenste stabiliteit en performance. Aangezien eSciDoc nog in ontwikkeling is wordt gekozen om het heroverwegen van integratie van eSciDoc uit te stellen totdat het een meer volwassen status heeft. Voor EASY II betekent dit dat het systeem direct op Fedora gebouwd zal gaan worden, en er dus veel functionaliteit zelf door DANS ontwikkeld zal moeten worden. Gevolg hiervan is dat 2009 voornamelijk in het licht stond van het ontwikkelen van basis-architectuur en -functionaliteit. Er is een stevig fundament neergelegd dat zowel wat betreft schaalbaarheid en performance voldoet aan de eisen van DANS. Er is een overzicht gemaakt van de functionaliteit van EASY I om te bepalen wat uiteindelijk opgeleverd zou moeten gaan worden, en in welke volgorde. Op basis van dit overzicht werd de oplevering van de eerste release in 2010 verwacht.

3.9 NCDD

In 2009 heeft de Nationale Coalitie Digitale Duurzaamheid (NCDD) een onderzoek uitgevoerd naar de duurzame toegankelijkheid van digitale data. DANS heeft de inventarisatie uitgevoerd van de sector Wetenschap (De inputnotitie sector Wetenschap is te vinden op: <http://www.ncdd.nl/activiteiten-natverkenning.php>). Ten behoeve van de inventarisatie zijn een groot aantal betrokkenen geïnterviewd. De uitkomsten van de verkenning zijn gepubliceerd in het eindrapport "Toekomst voor ons digitaal geheugen. Strategische agenda voor duurzame toegankelijkheid".

De verkenning heeft naar voren gebracht dat er meer aandacht dient te komen voor de langetermijnarchivering van digitale wetenschappelijke objecten. Hierbij is het van belang in samenwerking te komen tot een gemeenschappelijke basisinfrastructuur voor digitale archivering en voor de manier waarop digitale wetenschappelijke objecten kunnen worden hergebruikt.



3.10 Kwalitatieve data

Er is toenemende behoefte aan diensten op het gebied van archivering en beschikbaarstelling van kwalitatieve data. Het gaat dan met name om interviews. Er wordt begonnen met het inrichten van een infrastructuur om grote hoeveelheden interviews duurzaam op te slaan en toegankelijk te maken. Aspecten die hierbij een rol spelen zijn: het opstellen van gebruikslicenties, het ontwikkelen van documentatiestandaarden, en het aanbieden van instrumenten om de interviews toegankelijk te maken. DANS ontwikkelt een infrastructuur om zogenaamde "verrijkte publicaties" te kunnen maken. Deze infrastructuur bestaat uit bouwstenen om een verbinding aan te brengen tussen een wetenschappelijke publicatie met de onderliggende kwalitatieve onderzoeksdata.

3.11 Geodata

In het voorjaar verscheen het boek "Tijd en Ruimte, waarin DANS een coördinerende rol vervulde. Nieuwe toepassingen van GIS in de alfawetenschappen" waarin veel aandacht is voor uitkomsten van het door DANS geleide project "Alfa-Geo". "Alf-Geo" had als doel het gebruik van GIS en ruimtelijke data binnen de alfa wetenschappen te stimuleren.

3.12 Kleine Dataprojecten (KDP's)

In 2009 zijn verschillende kleine dataprojecten (KDP's) afgerond en nieuwe begonnen. Kleine dataprojecten zijn bedoeld voor het beschrijven en toegankelijk maken van één of meer belangrijke datasets. De maximaal toe te kennen subsidie vanuit DANS is tienduizend euro. Eén van de voorwaarden is het Open Access toegankelijk maken van de data via EASY. In 2009 werd onder meer 'Soundbites uit vervlogen tijden' afgesloten, ter ontsluiting van geluidsfragmenten van Nederlandse dialect sprekers in samenwerking met het Meertens instituut. Verder werden de volgende projecten gesubsidieerd in het kader van het KDP programma: Harmonisatie intermediair beste werkgevers data 2001-07, Veteran Tapes en Memoria. In 2010 werd onder meer 'Visualisatie beroepentelling volkstelling' afgesloten, met als doel te experimenteren met data visualisatie op één dataset om daarmee ervaring op te doen voor toekomstige ontwikkelingen. Als onderwerp van studie is de beroepentelling 1899 genomen. Verder werden de volgende projecten gesubsidieerd in het kader van het KDP programma: Beschermheiligen in Nederland, Soldatenkinderen NL en Prototype kennissysteem Risicofactoren onderwijsdeelname sub-Sahara Afrika.

4. Verkenningen

4.1 Onderzoek kosten van digitale archivering

Van september 2008 tot juni 2009 heeft bij DANS een onderzoek plaatsgevonden naar de kosten van digitaal archiveren. Dit onderzoek is in het kader van een internationale stage uitgevoerd door Anna Palaiologk, economiestudente aan de Universiteit van Macedonië in Thessaloniki, Griekenland. Bij DANS werd zij begeleid door Laurents Sesink en Heiko Tjalsma.

Hoe de kosten te berekenen van het archiveren van digitale bestanden is een onderwerp dat in een snel stijgende belangstelling staat bij archieven, bibliotheken en andere (erfgoed)instellingen. Wat de prijs is van het permanent beschikbaar houden van digitale informatie is niet een simpel te beantwoorden vraag, die echter wel relevant is en vooral van de kant van beleidsmakers en financiers steeds meer gesteld wordt.

Anna Palaiologk heeft op grond van de bestaande literatuur twee modellen toegepast op DANS. Het ene is het ABC- Activity Based Costing model, waarin de kosten van afzonderlijke activiteiten gerelateerd worden aan kostobjecten als producten of diensten. Op deze manier kunnen speciaal de meest arbeidsintensieve (en daarom kostbaarste) werkzaamheden van een data-archief geïdentificeerd worden. Het andere is het BSC – Balanced Scorecard model. Dit is een methode om de verschillende uitvoerende werkzaamheden van een organisatie af te meten aan de strategische doelen daarvan. Deze twee modellen zijn getest door alle activiteiten van DANS zoveel mogelijk in te delen volgens het OAIIS model (Open Archival Information System Reference Model), standaard referentiemodel voor digitale bibliotheken en archieven. In een vervolgstage zal vooral aandacht besteed moeten worden aan de materiële kosten zoals opslag en een nog verdergaande specificering van de ABC en BSC-modellen.

4.2 Taal- en Tekstdata in Nederland.

In 2009 is verder gewerkt aan dit inventariserende en verkennende onderzoek naar de data-infrastructuur van het taal- en tekstwetenschappelijk onderzoek in Nederland. Het betreft hier een oriëntatie op disciplines waar DANS zelf nog niet actief in is. Het onderzoek richt zich vooral op de vraag hoe de situatie ten aanzien van de toegang tot en de langetermijn-bewaring van deze databanken is.

Om erachter te komen welke gebruikerswensen leven bij de Nederlandse taal- en letterkundigen is in 2009 met een beperkt aantal voor het onderzoeksveld representatieve personen gesproken, zowel onderzoekers als vertegenwoordigers van de op dit terrein werkzame instituten. Daarnaast werd een inventarisatie van de in Nederland existierende databanken op dit terrein noodzakelijk geacht. Op deze manier kan een beter inzicht in het hele veld verkregen worden. Daartoe is in het najaar van 2008 door DANS een inventarisatie opgezet in samenwerking met het Europese CLARIN-project. DANS trad daarbij op als de Nederlandse national contact partner voor de algemene, Europese, CLARIN inventarisatie van taalkundige bestanden. Deze inventarisatie (resources database) is in 2009 gevuld en via de website van CLARIN.EU ter beschikking gesteld.

In 2010 is het inventariserende en verkennende onderzoek naar de data-infrastructuur van het taal- en tekstwetenschappelijk onderzoek in Nederland afgerond.

De laatste gesprekken zijn gevoerd met voor het onderzoeksveld representatieve personen gesproken, zowel onderzoekers als vertegenwoordigers van de op dit terrein werkzame instituten. Ook is intensief overleg gevoerd met CLARIN.NL, hetgeen geleid heeft tot onderlinge afstemming. In 2010 was CLARIN nog in ontwikkeling. Het Nederlandse taal- en tekstlandschap zal er, door CLARIN-NL, over enige jaren ongetwijfeld anders uitzien. In het eindrapport wordt hier rekening mee gehouden. In 2010 is de eindrapportage voltooid die begin 2011 gepubliceerd zal worden.

5. Publicaties & symposia



DANS Symposium Publications 5

Nederland in vergelijkend perspectief. Proceedings Tweede Nederlandse Workshop European Social Survey - 13 november 2009. Harry Ganzeboom, Marion Wittenberg (redactie). Amsterdam, Aksant Academic Publishers, 2009 - DANS Symposium Publications 5. ISBN 978-90-5260-364-3 - 200 p. - € 19,90



DANS Data Guide 7

Dutch Parliamentary Election Studies, Data Source Book 1971-2006. Bojan Todosijevic, Kees Aarts, Harry van der Kaap. Den Haag, DATA Archiving and Networked Services - DANS, 2010. DANS Data Guide 7. ISBN 978 94 90531 04 1. Prijs €21,50, excl. verzendkosten.



Tijd en ruimte

Tijd en ruimte. Nieuwe toepassingen van GIS in de alfawetenschappen. O. Boonstra, A. Schuurman (red.) Uitgave in samenwerking met Data Archiving and Networked Services - DANS/KNAW. Utrecht, Matrijs, 2010. ISBN 978-90-5345-388-9. Prijs: € 29,95 (tot 1 september 2010); daarna € 34,95.

ICT in de mediëvistiek: Het memoria-onderzoek in Nederland

Utrecht, 9 januari 2009

Het project Medieval memoria Online heeft een subsidie van bijna 550 duizend euro toegewezen gekregen van de Nederlandse organisatie voor Wetenschappelijk Onderzoek NWO. Dat gebeurde in het kader van de subsidieronde NWO-Middelgroot.

Digital Preservation Workshop

Den Haag, 30 januari 2009

Het door DANS ontwikkelde 'Data Seal of Approval (DSA)' of Datakeurmerk kreeg ruime aandacht op de workshop over 'Trusted Digital Archives' die het instituut op 30 januari in Den Haag organiseerde. Dat kwam vooral doordat de richtlijnen van het Datakeurmerk ruim zijn geformuleerd en daardoor breed toepasbaar.

Symposium Het Biografisch Portaal Nederland

Amsterdam, 27 maart 2009

Op deze bijeenkomst wordt gediscussieerd over het Biografisch Portaal als nieuw onderzoeksinstrument. In de ochtend wordt de Nederlandse situatie geplaatst in een internationale context. In het middagedeelte volgt een discussie over de mogelijkheden en perspectieven van een biografisch portaal. De centrale vraag is wat Nederlandse onderzoekers verwachten van dit nieuwe onderzoeksinstrument.

Symposium 'Jeugddata'

Den Haag, 25 november 2009

Er wordt steeds meer jeugdonderzoek gedaan in Nederland. Aan universiteiten en in onderzoeksinstituten is de aandacht voor jeugd en vooral voor problemen met opgroeien en opvoeden sterk toegenomen. DANS heeft in 2009 een Verkenning Jeugddata afgerond. Uit de Verkenning blijkt dat er behoefte is aan een centrale plaats waar gegevens uit jeugdonderzoek te vinden zijn.

Jubileumsymposium 'Door data gedreven'

Den Haag, 2 december 2009

In 2009 was het vijfenveertig jaar geleden dat het eerste wetenschappelijke data-archief in Nederland werd opgericht: het Steinmetzarchief voor de sociale wetenschappen. Maar dat is niet het enige jubileum op datagebied. In 1989, twintig jaar geleden dus, werd het Nederlands Historisch Data Archief opgericht en vijf jaar later het Wetenschappelijk Statistisch Agentschap. Vijf jaar geleden tenslotte ging het Electronisch Depot voor de Nederlandse archeologie van start. Al die instituten zijn intussen onderdeel geworden van DANS, dat in de verschillende lustra de aanleiding zag voor dit symposium.

Presentatie boek 'Tijd en Ruimte' en lancering NLGis web

Den Haag, 11 februari 2010

NLGis website en het boek Tijd en Ruimte zijn beide voorbeelden van het toenemend belang van ruimtelijke informatie in wetenschapsgebieden als geschiedenis, archeologie, kunstgeschiedenis en taal- en letterkunde.

Mini symposium 'Verkiezingen - stem van de wetenschap' en presentatie van het boek Dutch Parliamentary Election Studies 1971-2006

Den Haag, 21 mei 2010

Op 21 mei jl. organiseerde Data Archiving and Networked Services (DANS) in samenwerking met prof. dr. Kees Aarts, Universiteit Twente en bestuurslid Stichting Nationaal Kiezersonderzoek Nederland (SKON) een mini symposium in perscentrum Nieuwspoort in Den Haag. Tijdens deze bijeenkomst werd het boek Dutch Parliamentary Election Studies 1971-2006 officieel gepresenteerd aan drs. Ank Bijleveld, staatssecretaris van Binnenlandse Zaken en Koninkrijksrelaties.

DANS Dataprijs 2010

Amsterdam, 18 juni 2010

De eerste DANS Dataprijs is op 18 juni 2010 in Amsterdam uitgereikt door de president van de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW), prof. dr. Robbert Dijkgraaf. De achterliggende gedachte van de prijs is om dataverzameling en -beschikbaarstelling in het zonnetje te zetten door expliciete waardering te geven aan de vaak slecht zichtbare onderzoekers die de noeste en niet altijd dankbare arbeid verrichten om data bij elkaar te brengen, te documenteren en toegankelijk te maken. Data aan de hand waarvan baanbrekend onderzoek wordt verricht en nieuwe inzichten ontstaan. De DANS Dataprijs kan worden toegekend aan elke onderzoeker of elke onderzoeksinstelling in Nederland die werkt op het gebied van de alfa- en gammawetenschappen. Prof. dr. C.A. Mandemakers, initiator en projectleider van de Historische Steekproef Nederland bij het Internationaal Instituut voor Sociale Geschiedenis, heeft de DANS Dataprijs 2010 gewonnen.

6. Jaarcijfers

	2010	2009	2008	2007	2006	2005
Tijdsinzet onderzoek, onderwijs en collecties						
Onderzoeksinzet in fte's	0,0	0,0	0,1	1,9	1,4	1,2
<i>vast wetenschappelijk personeel</i>		0,0	0,0	1,7	0,4	1,2
<i>tijdelijk wetenschappelijk personeel, excl. promovendi</i>		0,0	0,1	0,2	1,0	
<i>promovendi</i>		0,0	0,0	0,0		
<i>overig wetenschappelijk personeel (incl. analisten en onderzoeksondersteunend personeel)</i>				0,0		
Publicaties						
Artikelen in tijdschriften, gerefereerd		0	1	4	2	0
Artikelen in tijdschriften, niet gerefereerd	1	2	0	2	0	0
Hoofdstukken in boeken, bundels en proceedings	8	2	7	4	7	0
Boeken/monografieën	4	5	0	0	2	0
Dissertaties		0	1	0	0	1
Boekredactie		1	2	4	1	3
Artikelen in tijdschriften, vakpublicaties (t/m 2007 incl. externe rapporten, software, datasets, ontwerpen)	1					
Artikelen in tijdschriften, populariserend		0	2	1	0	2
Rapporten	3	3	1			
Overige producten (computerprogramma's, software, websites, portals; aparte categorie sinds 2008)			3			
Wetenschappelijke voordrachten						
Voordrachten/lezingen als keynote speaker	1	3	3	8	10	11
Overige voordrachten/lezingen	21	41	8	7	7	6
Wetenschappelijke bijeenkomsten						
Wetenschappelijke bijeenkomsten georganiseerd of medegeorganiseerd door instituut (congressen, symposia, workshops, etc.)	8	12	5	4	4	2
Gastonderzoek						
Aantal gastonderzoekers/fellows in het instituut (verblijf > 1 maand)	1					

Onderwijsinzet						
Onderwijsinzet in fte's	0,2	0,1	0	0,0		0,2
DANS-kengetallen						
Personele inzet consultancy (totaal)	4	3	2,3	2,1		
<i>waarvan vast</i>	4	3	2,3	2,1		
<i>waarvan tijdelijk</i>				0,0		
Bezoekers website (unieke bezoekers)	151.408²	209.093	237.266	162.899	154.664	61.010
DANS	48.951	49.463	55.956	70.011	80.650	8.500
EDNA	8.000 ³	32.920	39.962	33.476	32.254	1.729
Nederlandse Volkstellingen	34.452	41.141	54.055	42.844	31.367	37.848
Amsterdamse Kooplieden	907	1.965	2.896	2.469	2.070	1.758
Nationale rekeningen	2.850	40.125	68.087	8.550	8.323	11.175
EASY	56.248	43.479	16.310	5.549		
Datasets gedownload⁴	16.340	8.323	3.585	1.214	1.017	742
EASY	13.823	7.624	3.059			
Overig	2.517	699	526	1.214	1.017	742
EASY - files gedownload	365.479	407.083	111.298	17.198		
<i>waarvan Humanities</i>	306.750	217.240	71.098	2.468		
<i>waarvan Social sciences</i>	20.453	27.974	14.592	4.895		
<i>waarvan Behavioural sciences</i>	6.908	11.609	3.836	1.628		
<i>waarvan Socio-cultural sciences</i>	19.105	14.419	20.639	7.658		
<i>waarvan Life sciences and medicine</i>	566	402	1.110	245		
<i>waarvan Geospatial</i>	19.052 ⁵	135.439	23	304		
Geleverd via DVD (Kadasterbestanden)	37.500	12.500				
Datasets gedocumenteerd	7.716	6.630	3.497	274	308	2.650
Indeling EASY hoofddisciplines (sommige datasets vallen onder meerdere disciplines)						
Social sciences	304	36	511			
Socio-cultural sciences	335	156	985			
Humanities	6.854	6.597	1.398			

² Schatting i.v.m. EDNA cijfers

³ Sinds 1-1-2010 kan er geen data meer gedownload worden uit EASY via de EDNA website. Het getal voor 2010 is een schatting i.v.m. de overgang naar nieuwe website.

⁴ Geen optelling in verband met dubbelstellingen

⁵ Teruggang tov 2009 heeft waarschijnlijk te maken met het Top10NL bestand

Geospatial sciences	7	1	4			
Life sciences and medicine	13	6	88			
Behavioural sciences	203	9	379			
Projecten						
Europese en Nationale samenwerking	16	17				
Verbeteren data infrastructuur	30	20				
Ondersteuning data projecten onderzoekers	25	26				
Exploitatie en consolidatie wetenschapsgebieden	8	7				
Invoeren standaarden en richtlijnen	17	18				
Leidinggevend of belangrijke partner in infrastructurele projecten	12	13				
Consultancy	6	13				
Persistent Identifier						
Uitgegeven nummers (nulmeting 31- 12-2009)	328.347	259.06 1				