

Data Archiving and Networked Services (DANS)

Jaarverslag 2005

Den Haag, 2008 (elektronische versie)

Data Archiving and Networked Services

DANS

Inhoud

Woord Vooraf 3

1 Missie en werkgebied van DANS 6

2 Hoofdactiviteiten van DANS 7

- 2.1 Archiveren 7
- 2.2 Academisch Digitaal Archief 8
- 2.3 Thematische Ontwikkel Programma's 8
- 2.4 Research & Development 9
- 2.5 Datakeurmerk 10

3 DANS in 2005 13

- 3.1 Personeel en inrichting van de organisatie 13
- 3.2 Bestuurlijke organisatie 15
- 3.3 Integratie data-archieven 15
- 3.4 Archivering en beschikbaarstelling 16
- 3.5 Thematische Ontwikkelprogramma's (TOP's) 17
- 3.6 Archieftechnische infrastructuur 17
- 3.7 Communicatie 18
- 3.8 Internationaal 18

Bijlagen 19

Colofon 22

Woord Vooraf

Na een lange periode van voorbereidingen ging DANS op 1 juli 2005 feitelijk van start met twaalf mensen in een nieuw gebouw. Drie maanden later werd de website openbaar en op 1 juni 2006 werd de oprichting een feit met de officiële opening door de minister van Onderwijs, Cultuur en Wetenschap, Maria van der Hoeven. Eigenlijk is het jaarverslag van DANS over 2005 dus maar een half jaarverslag: het gaat over een periode van zes maanden.

Een aantal projecten waarbij DANS in het verslagjaar betrokken was, was al eerder van start gegaan. Voordat DANS bestond waren er de data-archieven van het NIWI, het sociaal-wetenschappelijke Steinmetzarchief en het Nederlands Historisch Data Archief (NHDA) als onderdelen van de KNAW. Zij maken nu deel uit van DANS. De taken van het Wetenschappelijk Statistisch Agentschap (WSA) van NWO werden eveneens overgedragen aan DANS.

De kennisproductie in de alfa- en gammawetenschappen verandert onder invloed van de ontwikkelingen in de informatie- en communicatietechnologie. De behoefte aan databestanden en de manier waarop die bestanden worden aangeboden is steeds weer anders. Wat gisteren nog revolutionair leek is vandaag al heel gewoon en kan morgen al weer hopeloos achterhaald zijn. De behoeften van onderzoekers verschuiven onophoudelijk. 'Onmiddellijke toegang tot alle relevante informatie zonder belemmeringen, en dat vanuit één toegangspunt', zo formuleerde een onderzoeker zijn datahonger op een symposium waarop werd gesproken over *repositories* voor de wetenschap.

Hoewel deze wens utopisch lijkt en er door informatiespecialisten allerlei bezwaren kunnen worden geuit tegen de haalbaarheid daarvan, zou dit toch het doel kunnen zijn waarnaar wetenschappelijke data-aanbieders streven.

Een aantal jaren geleden berekende ik op grond van wat kencijfers van de British Library en de snelheid van de toenmalige scanners, dat er ca. 10.000 scanjaren nodig zouden zijn om alle boeken ter wereld te digitaliseren. Nu denk ik dat het nog maar enkele jaren zal duren of (bijna) ieder boek zal digitaal beschikbaar zijn. Hoewel het met data anders ligt geeft dit voorbeeld wel aan dat digitale dagdromen sneller werkelijkheid kunnen worden dan je denkt.

DANS streeft er in elk geval naar om wetenschappelijke data optimaal toegankelijk te maken voor onderzoekers in de alfa- en gammawetenschappen. Dit houdt in dat belemmeringen in de toegang zoveel mogelijk uit de wereld worden geholpen. *Open access* is daarbij een belangrijke steunpilaar. Met de opstellers van de Berlin Declaration on Open Access (20-23 oktober 2003; <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>) zijn wij van mening:

'The Internet has fundamentally changed the practical and economic realities of distributing scientific knowledge and cultural heritage. For the first time ever, the Internet now offers the chance to constitute a global and interactive representation of human knowledge, including cultural heritage and the guarantee of worldwide access.'

Waar het kan maken we data gratis en voor iedereen beschikbaar. Dit geldt in de eerste plaats voor data die met publieke middelen zijn gefinancierd.

Open Access is geen religie of dogma. Evenals voor publicaties kan voor databestanden het auteursrecht een beperking op vrije verspreiding vormen. Ook de privacy van respondenten en de wetgeving met betrekking tot de bescherming van persoonsgegevens moet worden gerespecteerd. Maar binnen die kaders draagt DANS bij aan een zo groot mogelijke toegankelijkheid, waarbij *some rights reserved* kunnen zijn.

Peter Doorn, directeur DANS



Het moment van de officiële opening van DANS op 1 juni 2006 werd geaccentueerd door de uitvoering van een speciaal geproduceerde dans, 'If you forget me', op een choreografie van Neel Verdoorn en uitgevoerd door studentes van de Artez hogeschool voor de Kunsten in Arnhem (Foto John Thuring)

1 Missie en werkgebied van DANS

In juli 2005 is, op initiatief van KNAW en NWO, DANS van start gegaan. Dat gebeurde met als belangrijkste houvast het advies van de daarvoor door beide organisaties ingestelde *Taskforce*, dat in december 2004 verscheen. Voorheen bestaande data-instituten op deelterreinen, zoals het Steinmetzinstituut, het Nederlands Historisch Data Archief en het Wetenschappelijk Statistisch Agentschap, werden onderdeel van DANS.

De missie van DANS is om de blijvende toegankelijkheid van digitale onderzoeksbestanden voor de alfa- en gammawetenschappen te garanderen. DANS doet dit door eigen initiatieven: het aanbieden van diensten en voorzieningen, maar ook door de samenwerking met en tussen onderzoekers te bevorderen. De organisatie krijgt de vorm van een netwerk met een centrum dat verantwoordelijk is voor de data-infrastructuur. DANS is ook partner in Europese en andere internationale dataorganisaties, zodat gegevensbestanden uit het buitenland gemakkelijk te verkrijgen zijn voor Nederlandse onderzoekers. In oktober is de DANS-website www.dans.knaw.nl online gegaan.

DANS streeft uit hoofde van haar opdracht naar een zo open mogelijke toegang tot wetenschappelijke gegevensbestanden, in overeenstemming met de Verklaring van Berlijn voor 'Open Access' van 22 Oktober 2003. Tevens bestaat er in het wetenschappelijke veld een dringende behoefte aan garanties omtrent datakwaliteit, duidelijke voorwaarden voor datagebruik (vooral van privacygevoelige data; tevens correcte naamsvermelding van de oorspronkelijke eigenaar van een databestand) en heldere criteria voor de digitale databewaarplaatsen – allemaal voorwaarden die op hun beurt bijdragen aan een maximale toegankelijkheid van de bestanden die zich daarvoor lenen. Centraal hierbij staat dat data altijd vindbaar, toegankelijk en bruikbaar moeten zijn.

Het DANS-centrum voert vijf groepen van activiteiten uit: bewaring en beschikbaarstelling van data in de digitale archieven; het opzetten van thematische ontwikkelprogramma's (TOP's) binnen deelgebieden van wetenschappelijk onderzoek; academische digitale archiveringsprojecten (ADA-projecten) om bestaande maar moeilijk toegankelijk geworden databestanden te saneren, research & development en tenslotte het ontwikkelen en de toepassing bevorderen van het DATA-keurmerk.

2 Hoofdactiviteiten van DANS

2.1 Archiveren

Sinds de start van DANS bevat het DANS data-archief de databestanden van het voormalige Nederlands Historisch Data Archief (NHDA) en het sociaal-wetenschappelijke Steinmetzarchief. Deze bestanden worden gratis ter beschikking gesteld aan onderzoek en onderwijs. In de komende jaren zullen ook databestanden uit andere disciplines in het DANS-archief worden opgenomen.

De on-line DANS-catalogus biedt beschrijvingen van alle aanwezige datasets. De bijbehorende documentatie, bestaande uit bijvoorbeeld vragenlijsten of codeboeken, is voor het grootste deel te downloaden.

Het *Nederlands Historisch Data Archief* bevat vooral historische datasets over Nederland en de voormalige koloniën. Ze hebben betrekking op alle historische tijdvakken en zijn vrijwel altijd gebaseerd op archiefbronnen. Veel bestanden zijn direct via de catalogus te downloaden.

Het *Steinmetzarchief* bevat ruim 1700 Nederlandse en internationale datasets op het gebied van de sociale wetenschappen, van 1960 tot heden. De bestanden kunnen met een orderformulier worden besteld. Een deel van de sociaal-wetenschappelijke informatie is toegankelijk via de DANS NESSTAR catalogus. Behalve studiebeschrijvingen is daar informatie te vinden over de variabelen in het bestand en kunnen de data worden bekeken. De databestanden in NESSTAR zijn direct te downloaden.

Naast bovenstaande collecties bevat de DANS catalogus ook informatie over de bestanden waarvoor het voormalige *Wetenschappelijk Statistisch Agentschap* als intermediair optrad. DANS biedt dus toegang tot onder meer de bestanden van het CBS en TD/Kadaster.

DANS ontwikkelt een systeem waarmee onderzoekers of databasebeheerders zelf hun data en bijbehorende bestanden on-line kunnen deponeren in het DANS archief. DANS zorgt ervoor dat de bestanden direct permanent en duurzaam worden opgeslagen en onder tevoren afgesproken voorwaarden veilig beschikbaar worden gesteld aan andere wetenschappers.

Wanneer onderzoekers hun data in eigen data-archief opslaan en zelf de toegang regelen, kan DANS als depot fungeren. De onderzoekers hebben op deze manier een extra zekerheid dat hun data blijvend bewaard worden. Ook kan DANS door middel van een portaal zorgen voor een extra ingang tot het archief.

2.2 Academisch Digitaal Archief

Er bevinden zich bij Nederlandse onderzoeksinstituten veel oude databestanden. Vaak zijn ze ontoegankelijk, soms weet bijna niemand van het bestaan ervan. Sommige van die bestanden zijn echter wel degelijk van wetenschappelijke waarde. Ze kunnen worden hergebruikt of aangewend voor controle van onderzoek.

DANS biedt een vorm van archivering achteraf aan onder de naam ADA: Academisch Digitaal Archief. In zo'n project worden de waardevolle bestanden geïdentificeerd en voor de toekomst veiliggesteld, terwijl de bestanden zonder waarde of betekenis definitief kunnen worden opgeruimd. Voor deze activiteit heeft DANS zelfs een eigen methode ontwikkeld, de ADA-methode, gebaseerd op haar eigen ervaring met wetenschappelijke data-archieven en nieuwe inzichten in digitale duurzaamheid.

In een ADA-project wordt de wetenschappelijke dataproductie van een onderzoeksgroep of instelling geheel of gedeeltelijk geïnventariseerd, geselecteerd, gereconstrueerd en gearchiveerd. Daarbij worden metadata toegevoegd om de bestanden toegankelijk te maken en te houden. De inhoudelijke selectie van te bewaren databestanden blijft in handen van de onderzoekers.

De ADA-methode is modulair van opzet en daardoor uiterst flexibel. De gearchiveerde databestanden kunnen aan de wetenschap beschikbaar worden gesteld.

DANS levert in een ADA-project advisering, begeleiding en ondersteuning. Het project wordt nooit alleen door DANS uitgevoerd: betrokkenheid van de opdrachtgever is altijd nodig. Deze verstrekt informatie over de in het verleden gemaakte bestanden en de bijbehorende IT-infrastructuur. De opdrachtgever maakt ook de inhoudelijke selecties van databestanden en draagt bij aan de documentatie ervan.

De inzet van DANS-personeel kan sterk variëren, zodat ook de uitvoeringskosten nogal uiteen zullen lopen. Deze worden daarom van project tot project bepaald na vooronderzoek en in overleg met de opdrachtgever. Advisering en begeleiding zijn in principe gratis.

2.3 Thematische Ontwikkel Programma's

Thematische Ontwikkel Programma's (TOP's) zijn bedoeld om op een bepaald deel-terrein de data-infrastructuur te versterken. In TOP's worden met onderzoekers *networked services* gebouwd voor het opslaan en hergebruiken van gegevens. TOP's maken databestanden met nationale en/of internationale reikwijdte beter toegankelijk en verrijken deze door combinatie en harmonisatie met andere bestanden. In TOP's worden nieuwe terreinen van digitale archivering en toegang tot onderzoeksdata ontgonnen of bestaande verbeterd en verbonden. Ze sluiten

aan bij actuele ontwikkelingen in de verschillende wetenschapsgebieden. In TOP's kan ook worden samengewerkt met al bestaande datacentra op een bepaald onderzoeksterrein.

DANS kan in een thematisch ontwikkelprogramma expertise, menskracht en faciliteiten inbrengen. DANS fungeert als regiepartner of als mede-aanvrager bij het verwerven van financiering bij bestuurlijke partijen. DANS kan ook een adviserende of bestuurlijke rol vervullen door deel uit te maken van de raad van advies of stuurgroep van de TOP.

Thematische ontwikkelprogramma's zijn tijdelijk. De omvang en tijdsduur zijn afhankelijk van het doel van het programma. Bij afronding van een TOP moet de continuïteit van de toegang tot de data altijd zijn gewaarborgd in de vorm van een Thematisch Data Archief, dat decentraal in een trusted repository of centraal in het DANS-archief is opgeslagen.

Een TOP staat ten dienste van innovatief wetenschappelijk onderzoek door het verbeteren van de toegang, harmonisatie of verrijking van dataverzamelingen. Een TOP moet daarnaast internationaal meetellen. De initiatiefnemers zijn vooraanstaande wetenschappers. Een TOP maakt deel uit van het DANS-netwerk. Daarnaast moet een TOP brede steun genieten in het betreffende gebied. De verbetering van de data-infrastructuur waar de TOP toe leidt, biedt mogelijkheden voor vernieuwend wetenschappelijk onderzoek. Aan een TOP is een onderzoeksagenda gekoppeld, waarin beschreven wordt welk nieuw onderzoek erdoor ontstaat.

Een TOP is innovatief, bijvoorbeeld door het toepassen van nieuwe ICT onderzoeks- of archiveringsmethoden, of levert een grote bijdrage aan de vernieuwing op een wetenschapsgebied.

2.4 Research & Development

Hoe kunnen we zeker weten dat datasets over vijf, vijftig of honderd jaar nog te gebruiken zijn? Als een onderzoeker in een publicatie verwijst naar een dataset, zal de URL daarvan dan over een paar jaar nog wel betrekking hebben op dezelfde dataset? Hoe bouw je een *repository* die grote hoeveelheden data digitaal duurzaam kan opslaan, die eenvoudig en snel doorzoekbaar is en die ook over vijftig jaar nog toegankelijk is? Deze vragen zijn richtinggevend voor de activiteiten van DANS op het gebied van *research & development*. Problemen waarmee data-archieven te maken hebben, vereisen oplossingen die langer meegaan dan doorsnee producten kunnen garanderen. Een archiefsysteem moet eenvoudig en efficiënt te gebruiken zijn en het moet in staat zijn de toenemende aantallen datasets moeiteloos te verwerken. De datasets moeten digitaal duurzaam opgeslagen worden, ze dienen continu beschikbaar te zijn en 'links' naar de datasets vanuit publicaties moeten ook na tientallen jaren nog werken. Het archiefsysteem

dient zo flexibel te zijn dat het nieuw te ontwikkelen ontsluitingsmethoden probleemloos kan ondersteunen. De zo verkregen duurzaamheid is nog niet vanzelfsprekend. Daarvoor is nog heel wat onderzoek en ontwikkeling nodig.

Research & Development is een van de speerpunten van DANS. Om fundamenteel onderzoek toe te passen in systemen voor archivering van en toegang tot data beschikt DANS over een eigen ontwikkelteam. Dit team van academisch geschoolde informatici volgt, in nauwe samenwerking met de DANS data-archivisten, ontwikkelingen in het veld op de voet en draagt daaraan zelf ook bij door middel van papers, presentaties en onderzoek.

Waar mogelijk worden bij R&D projecten van DANS Open Standaarden gebruikt en wordt er aangehaakt bij Open Source Software.

Op dit moment ontplooit DANS activiteiten op de volgende terreinen:

- *GRID Computing*: hoe kunnen GRIDs helpen bij de toegang tot gedistribueerd opgeslagen data?
- *Persistent Identifiers*: hoe gaan we er voor zorgen dat resources permanent beschikbaar blijven via dezelfde URL?

Digitale Duurzaamheid: kunnen we in de toekomst de datasets van nu nog wel gebruiken?

2.5 Datakeurmerk

Digitale archieven hebben als doel om digitale informatie toegankelijk te houden voor toekomstig gebruik. Hoewel er duizenden digitale archieven in allerlei vormen bestaan, is er geen eenduidige strategie om de kwaliteit van digitale data te garanderen en deze voor de lange termijn beschikbaar te houden. DANS heeft de opdracht aanbevelingen te doen voor criteria waaraan digitale archieven horen te voldoen. Onderzoek naar archiveringsmethoden en -technieken en evaluatie van internationale richtlijnen vormen de basis voor die richtlijnen. Samen zullen die een keurmerk ondersteunen dat betrouwbaarheid garandeert *van de digitale data, van de data-bewaarplaatsen en bij het gebruik van de data*: het DATA-keurmerk

In de eerste plaats zullen de data-archieven die DANS zelf beheert (Steinmetzarchief/WSA en NHDA) aan deze criteria voldoen. Maar ook andere organisaties die zich aan willen sluiten bij het netwerk van thematische data-archieven (TDA) dat DANS vormt, zullen hun archieven hiermee in overeenstemming moeten brengen, evenals de partners die gezamenlijk met DANS een Thematisch Ontwikkelprogramma (TOP) opzetten.

De eisen zijn zo generiek geformuleerd dat ze toepasbaar zijn voor alle vakgebieden van de geesteswetenschappen en de gedrags- en maatschappijwetenschappen en voor alle soorten data. De criteria voor toekenning van het DATA-keurmerk zullen in overeenstemming zijn met gangbare nationale en internationale richtlijnen voor digitale data-archivering.

Om de acceptatie van het DATA-keurmerk zo breed mogelijk te laten zijn, wordt ervoor gekozen om zo weinig mogelijk regels dwingend voor te schrijven en zoveel mogelijk gebruik te maken van aanbevolen richtlijnen, *preferred standards* en *good practices*.

Een veelgehoorde klacht over de waarde van informatie op het Internet voor wetenschappelijk onderzoek is dat de kwaliteit van de informatie niet goed kan worden vastgesteld. Het DATA-keurmerk is dan ook bedoeld om juist op dat gebied garanties te verschaffen. Een gebruiker van data met het keurmerk wordt door middel van reviews van collega-onderzoekers of links naar relevante informatie (publicaties en onderzoekinformatie) in staat gesteld om zich een oordeel te vormen over de volgende aspecten:

- zijn de data nuttig voor ander onderzoek of hergebruik?
- zijn de data gebaseerd op eigen werk van de *depositor* (onderzoeker of instelling die het onderzoek beschikbaar stelt)?
- is dataverzameling of digitalisering verricht volgens de op het vakgebied geldende criteria?
- staat de *depositor* (onderzoeker of instelling) te goeder naam en faam bekend?

Niet alleen de wetenschappelijk verantwoorde kwaliteit moet gegarandeerd zijn; hetzelfde geldt voor de vindbaarheid, toegankelijkheid en bruikbaarheid. De *toegankelijkheid* van gegevensbestanden heeft betrekking op de licentievoorwaarden. DANS streeft in het algemeen naar *open access*, maar het blijft wel mogelijk om voorwaarden aan de toegang te stellen. De licentievoorwaarden voor gebruik van de gegevensbestanden moeten duidelijk zijn omschreven en voldoen aan de vigerende wet- en regelgeving. Dit houdt bijvoorbeeld in dat de privacy van personen is gegarandeerd. De depositor (onderzoeker of instelling) is verantwoordelijk voor de inhoud van zijn gegevensbestanden, ook als die via een dataarchief beschikbaar worden gesteld.

Uitgangspunten voor de toegang tot en gebruik van de data zijn *Gedragsscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek* (uitgave SWR) en *de Notitie Wetenschappelijke integriteit* (uitgave VSNU, KNAW, NWO). Op technisch vlak zal dit door middel van authenticiteit- en autorisatiemechanismen gecontroleerd kunnen worden. Privacygevoelige persoonsdata zullen in zogenaamde *safe rooms* op universiteiten en bij DANS onderzocht kunnen worden. DANS beveelt aan de licentievoorwaarden voor open toegang te hanteren zoals ontwikkeld en geformuleerd door Creative Commons en Science Commons. Deze zijn algemeen toepasbaar voor op het internet beschikbare hulpbronnen – dus ook voor wetenschappelijke gegevensbestanden die niet privacygevoelig zijn – en in overeenstemming met nationale en internationale wetten en regels.

Er worden in diverse landen pogingen in het werk gesteld om criteria voor certificering van *trusted digital repositories* op te stellen. Deze criteria hebben betrekking op de taken van een repository om de toegang tot data op de lange

termijn te kunnen garanderen. DANS zal zoveel mogelijk bij deze internationale ontwikkelingen aansluiten.

De bij DANS ondergebrachte data-archieven en bij het DANS-netwerk aangesloten Thematische Data Archieven (TDA's) hebben tot taak om:

- de beoordeling te organiseren van de kwaliteit van de beschikbaar gestelde gegevensbestanden en maatregelen te treffen om fouten en onzorgvuldigheden in de gegevens die achteraf worden geconstateerd te (laten) herstellen:
- gegevens te allen tijde vindbaar, toegankelijk en bruikbaar te maken
- gegevensbestanden die beschikbaar worden gesteld te beheren en te onderhouden (dat wil zeggen: zodanig te archiveren dat de gegevens ook op de lange termijn toegankelijk blijven).

3 DANS in 2005

3.1 Personeel en inrichting van de organisatie

Op 1 juli 2005 is DANS van start gegaan met tien medewerkers. Daaraan was een procedure vooraf gegaan waarbij aan medewerkers van de oude NIWI-instituten en het WSA de mogelijkheid werd geboden te reflecteren op een aantal functies in het nieuwe instituut. Het resultaat was dat van het NHDA vier medewerkers overstapten, terwijl twee collega's van het Steinmetzarchief en één van het WSA een nieuwe start bij DANS maakten. Buiten de genoemde instituten leverde het NIWI nog drie medewerkers. Per saldo werden alle tien nieuwe formatieplaatsen bezet door medewerkers van de voorgaande instituten.

In het formatieplan zoals door de *Task Force* van KNAW en NWO opgesteld waren twintig formatieplaatsen voorzien. Omdat het om een organisatie in opbouw ging, waarbij maar een deel van de activiteiten overgenomen diensten van voorgaande instituten behelst, koos de directie voor een gefaseerde realisering van het plan: nieuwe formatieplaatsen werden pas dan ingevuld als de organisatie er klaar voor was om ze goed te laten functioneren en de resultaten van hun werk tot hun recht te laten komen. Zo werden in de loop van 2005 nog twee formatieplaatsen ingevuld met als resultaat een bezetting van twaalf van de twintig plaatsen aan het begin van 2006.

Tabel 3.1 Invulling van de DANS-personeelsbezetting naar herkomst medewerkers

Herkomst	1 juli 2005	1 januari 2006
Steinmetzarchief	2	2
NHDA	4	4
NIWI	3	4
WSA	1	1
Extern	-	1
Totaal	10	12



Voor de locatie van het centrale DANS-kantoor is het gebouw Borneo van NWO gekozen (foto), gelegen in Den Haag net achter de Laan van Nieuw Oost Indië. De gunstige ligging en het moderne karakter van het gebouw, de daarbij behorende goede voorzieningen en de nabijheid van de verwante NWO-organisatie waren overwegingen die hierbij een belangrijke rol speelden. Het DANS-kantoor is ingericht met flexplekken omdat voor de meeste DANS-medewerkers een dynamisch bestaan wordt voorzien. De data-archivering zal steeds meer op locatie plaatsvinden bij de instituten en faculteiten, en ook de medewerkers die hun krachten geven aan het opzetten en onderhouden van Thematische Ontwikkelprogramma's zullen dat grotendeels in decentrale projecten doen.

Over de bedrijfsvoering zijn afspraken gemaakt met twee andere Haagse KNAW-instituten, het Rathenau Instituut en het Huygens Instituut, waarmee gezamenlijk een clusterbureau Den Haag is gevormd. Dit bedrijfsbureau verleent DANS diensten op het gebied van personeelszaken en financieel beheer. Het ICT-beheer is grotendeels uitbesteed aan de nieuwe afdeling Informatisering & Automatisering van het Bureau van de KNAW.

3.2 Bestuurlijke organisatie

Op basis van een bestuurlijk convenant van KNAW en NWO werd voor DANS een Stuurgroep samengesteld onder voorzitterschap van mw. mr. Y. Van Rooij. De stuurgroep bestond in 2005 uit vijf leden, waarvan er twee door NWO werden afgevaardigd en twee door de KNAW, terwijl de voorzitter uit het universitair bestuur afkomstig was. Met het functioneren van de Stuurgroep is beantwoord aan het streven in het bedrijfsplan naar een bestuurlijke structuur die het langdurig commitment van beide moederorganisaties aan DANS zou weerspiegelen, en tegelijk het kanaal zou bieden voor de verantwoording van de DANS-directie aan beide organisaties.

Naast de Stuurgroep zou er volgens het Bedrijfsplan een Wetenschappelijke Adviesraad moeten komen met vertegenwoordigers van onderzoekers en data-producenten. Deze zou de DANS-directie begeleiden bij de opbouw van de organisatie. De instelling van deze Adviesraad werd in het verslagjaar voorbereid.

3.3 Integratie data-archieven

Het sociaal-wetenschappelijk Steinmetzarchief en het Nederlands Historisch Data Archief (NHDA), die van 1997 tot 2005 deel uitmaakten van het NIWI, zijn onderdeel van DANS geworden. DANS nam ook de taken over van het Wetenschappelijk Statistisch Agentschap (WSA, voorheen een onderdeel van NWO) om als intermediair op te treden bij de levering aan onderzoekers van grote databestanden van organisaties als het CBS, het SCP en de Topografische Dienst/Kadaster.

In de loop van 2005 is de integratie van Steinmetzarchief, NHDA en WSA voltooid, nagenoeg zonder merkbare nadelen voor de gebruikers van de data-archieven. Onderzoekers kunnen hun databestanden bij DANS laten archiveren en omgekeerd ook rekenen op de levering van bestanden door DANS. De procedures voor archivering en toegang zijn in 2005 herzien en worden in 2006 juridisch goetst. Pas als de vernieuwing van de archieftechnische infrastructuur is gerealiseerd, zal efficiencywinst worden geboekt door een grondige verandering van de primaire werkprocessen (zie ook 3.6).

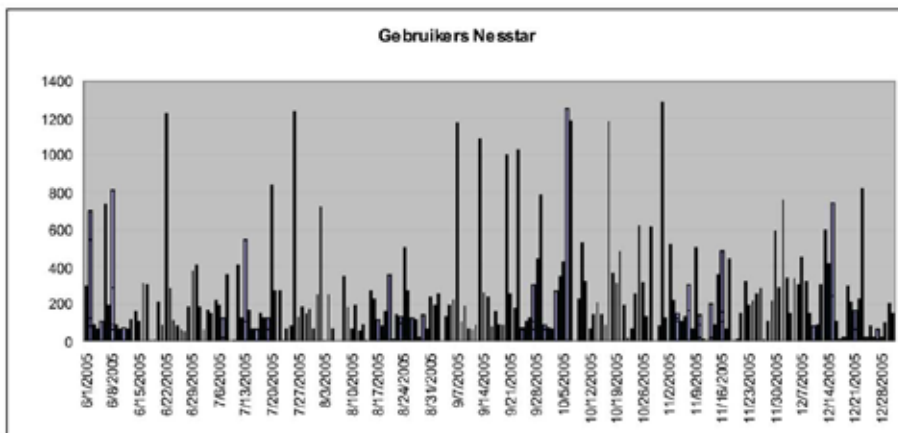
In overleg met de grote dataleveranciers van het WSA zijn de datacontracten waar mogelijk herzien. Bijna alle data kunnen nu gratis worden geleverd, maar nog niet met alle partijen zijn de onderhandelingen afgerond. Met het CBS wordt gestreefd naar een overkoepelende overeenkomst voor toegang tot de data van dat instituut. Het gaat daarbij niet alleen om de microbestanden uit het voormalige WSA-contract, maar ook om de toegang tot historische gegevens en het zogenaamde Sociaal Statistisch Bestand, een conglomeraat van registratiebestanden van de overheid en steekproefonderzoeken, waaruit onder andere de Volkstelling van 2001 is samengesteld.

3.4 Archivering en beschikbaarstelling

De archivering en beschikbaarstelling van databestanden is een kernactiviteit van DANS. Het verkeer op de DANS-website, waarin het digitale archief een centrale plaats inneemt, is tussen de opening in oktober 2005 en het eind van dat jaar gegroeid naar rond de 250 individuele bezoekers per dag. In totaal werd de website sinds de opening in oktober tot aan het eind van 2005 door meer dan acht duizend mensen bezocht.

In het verslagjaar is ook begonnen met het verkennen van de behoefte aan nieuwe data-archieven binnen de alfa- en gammawetenschappen. Dat bracht in elk geval de dringende behoefte onder archeologen aan een data-archief aan het licht, en de behoefte bij de taal- en letterkundigen aan het virtueel bijebrengen van verspreid opgeslagen corpora en databanken. De behoefte aan een archief voor digitaal audio-, video- en stilstaand beeldmateriaal wordt nader onderzocht.

Aan het eind van 2005 werden ook voorbereidingen getroffen voor een meer systematische, uitvoerige ronde van gesprekken met bestuurders en onderzoekers in de wereld van het universitair onderzoek om een betrouwbaar beeld te krijgen van de behoefte aan nieuwe data-archieven en Thematische Ontwikkelprogramma's (zie 3.5). Deze ronde is in de eerste maanden van 2006 van start gegaan.



Figuur 2.2 De gebruikersaantallen van de databank Nesstar tonen geen spoor van de overgang in DANS-beheer in juli 2005 of de start van de DANS-website in oktober 2005.

3.5 Thematische Ontwikkelprogramma's (TOP's)

Terwijl de dienstverlening aan onderzoekers werd opgebouwd, had DANS ook de opdracht om aan de vernieuwing en verbetering van de data-infrastructuur te werken: de 'Networked Services' uit het DANS-acronym.

In 2005 is direct begonnen met de voorbereiding van Thematische Ontwikkelprogramma's (TOP's): programma's om op een bepaald deelterrein van onderzoek de data-infrastructuur te versterken en databestanden met nationale en/of internationale reikwijdte beter toegankelijk te maken en te verrijken (zie 2.3).

Daarbij is een groot aantal suggesties en wensen verzameld, er zijn selectiecriteria en –voorwaarden geformuleerd, er is met mogelijke financiers overlegd en er is met diverse onderzoeksgroepen gewerkt aan voorstellen en aanvragen voor financiering. Dit laatste gebeurde aanvankelijk vooral *ad hoc* en op verzoek van de onderzoekers. Zo werd met groepen taal- en letterkundigen, archeologen, sociale wetenschappers en historici gewerkt aan de voorbereiding van aanvragen. TOP's worden in samenwerking met onderzoeksgroepen opgezet. Het initiatief ervoor moet vooral uit de wetenschappelijke wereld komen, maar DANS ziet voor zichzelf wel een stimulerende en begeleidende rol weggelegd. DANS heeft zelf geen middelen om TOP's te financieren. De initiatiefnemers zijn dus op dat punt afhankelijk van externe financiering. Ook daarbij kunnen kennis en deskundigheid van DANS worden ingezet.

3.6 Archieftechnische infrastructuur

Een projectplan voor de vernieuwing van de archieftechnische infrastructuur is in 2005 opgesteld en er is een begin gemaakt met de uitvoering daarvan. Deze vernieuwing is pas in de loop van 2006 gerealiseerd onder de naam DANS EASY. Daarbij komt een grote omslag in de werkprocessen kijken. De archivering zal uiteindelijk op een nieuwe manier gaan plaatsvinden. Onderzoekers zullen met behulp van DANS EASY op eenvoudige wijze hun eigen bestanden in een data repository kunnen plaatsen.

Zo beschikken ze altijd over een betrouwbare back-up van hun eigen bestanden, terwijl ze toch hun gegevens kunnen delen met andere onderzoekers. Ook worden Academische Digitale Archief-projecten opgezet (zie 2.2). In deze ADA-projecten wordt samengewerkt met onderzoeksinstituten en faculteiten om hun digitale dataproductie van de afgelopen jaren snel te kunnen inventariseren, ordenen, archiveren en ontsluiten, waarbij ook weer een beroep kan worden gedaan op de mogelijkheden van de DANS-infrastructuur.

3.7 Communicatie

Kort na de zomer van 2005 is een communicatieplan opgesteld met een beperkte reikwijdte: hoofddoel was het bekend maken van de naam en activiteiten van DANS bij zowel de relaties van de drie geïncorporeerde instituten (Steinmetz, NHDA en WSA) als het alfa- en gammagedeelte van de academische wereld in zijn volle breedte. Er is daarna op diverse wijzen gewerkt aan de bekendmaking van DANS. Er werden lezingen gehouden, informatiebladen (*factsheets*) uitgegeven, presentaties verzorgd op congressen, bijeenkomsten georganiseerd rondom data-gebruik en mailings verstuurd. Ook de website vormt een belangrijk communicatiemedium, maar kon in die rol pas tot z'n recht komen nadat aan het bestaan en het adres door middel van twee mailings brede bekendheid was gegeven.

In het verslagjaar is een opzet gemaakt voor een nieuw blad over elektronische data en onderzoek, *e-data&research*, gezamenlijke uit te geven met enkele andere instituten. De publicatiereeks *DANS studies in digital archiving* is gestart, waarin in 2005 een proefschrift verscheen, geschreven door DANS-medewerker René van Horik.

3.8 Internationaal

DANS is de nationale vertegenwoordiger op het gebied van data-archivering en beschikbaarstelling in een aantal internationale dataorganisaties zoals CESSDA, ICPSR, IFDO-IASSIST). In Europees verband streeft CESSDA naar versterking van de onderzoeksinfrastructuur voor de sociale wetenschappen (ESFRI, Zevende Kaderprogramma). DANS heeft het initiatief genomen om met de Britse *Arts and Humanities Data Service*, de Duitse *Max Planck Gesellschaft* (MPG) en het *Centre National de Recherche Scientifique* (Frankrijk) de kern te vormen van een Europese data-infrastructuur voor de kunsten en humaniora: DARIAH.

Bijlagen

Tijdsinzet onderzoek, onderwijs en collecties	2005
<i>Personele inzet voor onderzoek</i>	
Onderzoeksinzet – totaal (fte's)	1,20
waarvan vast	1,2
waarvan tijdelijk, excl. promovendi	0
waarvan promovendi	0,00
<i>Personele inzet voor onderwijs</i>	
Onderwijsinzet – totaal (fte's)	0,2
<i>Personele inzet voor collecties</i>	
Inzet collectietaken – totaal (fte's)	
waarvan collectievorming	
waarvan beheer en ontsluiting	
waarvan dienstverlening	
<i>Personele inzet Consultancy</i>	
Consultancy totaal	
waarvan vast	
waarvan tijdelijk	
<i>Gastonderzoekers</i>	
Inkomende gastonderzoekers/fellows (aantal)	
Uitgaande gastonderzoekers/fellows (aantal)	
<i>Leerstoelen</i>	
Gewone leerstoelen – bezoldigd (aantal)	
Gewone leerstoelen – onbezoldigd (aantal)	
Bijzondere leerstoelen (aantal)	
Tijdsinzet onbezoldigde en bijzondere leerstoelen (fte's)	

Wetenschappelijke en overige output	2005
Wetenschappelijke publicaties	
<i>Artikelen in 'refereed journals' – totaal</i>	0
waarvan impactfactor ≥ 20	
waarvan $10 \leq$ impactfactor < 20	
waarvan impactfactor < 10	
<i>Artikelen in 'non-refereed journals'</i>	0
<i>Boeken</i>	4
waarvan proefschriften	1
waarvan monografieën	
waarvan geredigeerde bundels, congres proceedings	3
<i>Hoofdstukken in boeken, bundels en proceedings</i>	0
Begeleiding promovendi	
<i>Begeleide interne (afgeronde) promoties – totaal</i>	0
waarvan als 1e begeleider	
waarvan als medebegeleider	
<i>Begeleide externe (afgeronde) promoties – totaal</i>	0
waarvan als 1e begeleider	
waarvan als medebegeleider	
Vakpublicaties	
<i>Recensies</i>	
<i>Overige vakpublicaties</i>	2
Wetenschappelijke voordrachten	
<i>Wetenschappelijke voordrachten en lezingen</i>	17
waarvan als invited en/of keynote speaker	11
waarvan als selected speaker	
overig	6
Wetenschappelijke bijeenkomsten	
<i>Wetensch. bijeenkomsten (mede)georganiseerd door instituut (congressen, symposia, workshops, etc.)</i>	2
waarvan met overwegend nationale deelnemers	2
waarvan met overwegend internationale deelnemers	0

Instituutspecifieke output – DANS	2005
Bezoekers website unieke bezoekers	61.010
<i>DANS</i>	8.500
<i>eDNA</i>	1.729
<i>Nederlandse Volkstellingen</i>	37.848
<i>Amsterdams Kooplieden</i>	1.758
<i>Nationale rekeningen</i>	11.175
<i>EASY</i>	-
Webpagina's geraadpleegd	621.721
<i>DANS</i>	115.148
<i>eDNA</i>	82.197
<i>Nederlandse Volkstellingen</i>	369.581
<i>Nationale rekeningen</i>	54.795
<i>EASY (dataset hits)</i>	-
Datasets gedownload	742
<i>Nesstar</i>	143
<i>ICPSR</i>	-
<i>Steinmetz (inclusief CBS bestanden)</i>	599
<i>EASY – files gedownload</i>	-
<i>waarvan – Humanities</i>	-
<i>waarvan – Sociale sciences</i>	-
<i>waarvan – Behavioral sciences</i>	-
<i>waarvan – Socio-cultural sciences</i>	-
<i>waarvan – Life sciences and Medicine</i>	-
<i>overig</i>	-
Datasets geraadpleegd	2.222
<i>Nesstar</i>	222
<i>Volkstellingen (of gedownload)</i>	-
<i>Nationale rekeningen (of gedownload)</i>	-
<i>eNHDA (of gedownload)</i>	2.000
Datasets gedocumenteerd	274
<i>NHDA</i>	
<i>Steinmetz</i>	72
<i>eDNA</i>	202
<i>EASY (incl. conversie)</i>	-

Colofon

Tekstbijdragen: Peter Doorn, Martijn de Groot, Henk Harmsen, René van Horik,
Peter Schroeder en Laurents Sesink

Productie: Martijn de Groot

Eindredactie: Martijn de Groot

Foto's: John Thuring (pagina 5), NWO (pagina 14)

Vormgeving: Ellen Bouma