

Tilburg University

Witboek dataprofessionals Nederland

Grim, R.; van der Heijden, M.; de Smaele, M.; Verbakel, E.

Publication date:
2011

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Grim, R., van der Heijden, M., de Smaele, M., & Verbakel, E. (2011). *Witboek dataprofessionals Nederland*. SURF.
http://www.surf.nl/binaries/content/assets/surf/nl/kennisbank/2011/Witboek_Dataprofessionals_in_NL_def.pdf

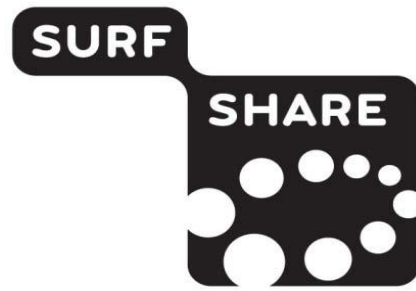
General rights

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 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 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.



**Witboek
Dataprofessionals
in Nederland**

Colofon

*Witboek Dataprofessionals in Nederland
Verslag van de Werkgroep Deskundigheidsbevordering*

SURFfoundation
PO Box 2290
NL-3500 GG Utrecht
T + 31 30 234 66 00
F + 31 30 233 29 60

info@surf.nl
www.surf.nl

Auteurs

Rob Grim – Universiteit van Tilburg
Marianne van der Heijden – NIOO-KNAW
Madeleine de Smaele – Technische Universiteit Delft
Ellen Verbakel – Technische Universiteit Delft

Partner instellingen

Universiteit van Tilburg, 3TU.Datacentrum, NIOO-KNAW.

Eindredactie

Keith Russell - SURFfoundation

SURF is de ICT-samenwerkingsorganisatie van het hoger onderwijs en onderzoek (www.surf.nl).
Deze publicatie is digitaal beschikbaar via de website van SURFfoundation:
www.surffoundation.nl/publicaties

© Stichting SURF
Augustus 2011

Deze publicatie verschijnt onder de Creative Commons licentie Naamsvermelding 3.0 Nederland.



Inhoudsopgave

Managementsamenvatting	5
Management summary	7
Inleiding: Opmars van de dataprofessionals	9
1 De omgeving van de Dataprofessional	11
1.1 Data en databeheer van onderzoeksdata	11
1.2 Data-archivering, metadata en digitale duurzaamheid	11
1.3 Training en opleiding van dataprofessionals.....	12
1.4 Data Generic en Data Specific Intelligence	13
2 Dataprofessionals aan zet	15
2.1 Werkwijze	15
2.2 Databeheer bij de voorbereiding van onderzoek.....	15
2.2.1 Situatie bij het NIOO-KNAW	15
2.2.2 Situatie bij TU Delft.....	16
2.2.3 Situatie bij Tilburg University	17
2.3 Databeheer tijdens onderzoek	18
2.3.1 Situatie bij het NIOO-KNAW	18
2.3.2 Situatie bij TU Delft.....	19
2.3.3 Situatie bij Tilburg University	19
2.4 Databeheer na afronding van onderzoek.....	20
2.4.1 Situatie bij het NIOO-KNAW	20
2.4.2 Situatie bij TU Delft.....	21
2.4.3 Situatie bij Tilburg University	22
2.5 Samenvatting databeheer in de praktijk	22
3 Aanbevelingen	25
3.1 Aanbevelingen voor training en opleiding in wetenschappelijke curricula	25
3.2 Praktische handvatten	26
3.3 De invulling van trainingen en cursussen	27
Lijst van afkortingen	29
Referenties.....	31
Bijlage 1 - Functieomschrijvingen van dataprofessionals	33
Bijlage 2 - Voorstel toevoeging variant onderzoeksdata aan UFO-profiel Informatie- /Collectiespecialist	35
Bijlage 3 - Functieomschrijving Research Data Specialist.....	37
Bijlage 4 - Voorbeelden van workshops en trainingen.....	39
Bijlage 5 - 3TU.Datacentrum interviewverslagen	41
Bijlage 6 - NIOO-KNAW interviewverslagen	42
Bijlage 7 - Tilburg University interviewverslagen	43

Managementsamenvatting

Ook in Nederland komt steeds meer aandacht voor onderzoeksdata. Om wetenschappers goed te kunnen ondersteunen bij het omgaan met onderzoeksdata is een nieuw type professional nodig, in dit witboek gemakshalve aangeduid met de algemene term 'dataprofessional'. Aan de hand van interviews met onderzoekers en dataspecialisten is de praktijk van het databeheer in enkele wetenschapsdisciplines gescand op de kennis, ervaring en competenties die nodig zijn om in een ondersteunende functie met onderzoeksdata te kunnen werken.

De kern van dit witboek bestaat uit een beschrijving van het databeheer in de praktijk bij wetenschapsdisciplines van drie verschillende instellingen: NIOO, TU Delft en Tilburg University. Bij deze instellingen is onderzocht hoe wetenschappers en ondersteuners in de praktijk omgaan met databeheer en wat nodig is voor een goed beheer van onderzoeksdata. Het databeheer binnen een discipline is daarbij steeds vanuit drie invalshoeken beschreven: een fase waarin het onderzoek wordt voorbereid, de uitvoeringsfase van het onderzoek en het beheer van data nadat onderzoek is afgerond.

Het eerste deel van het witboek gaat echter in op de opkomst van de dataprofessional en het training- en scholingsaanbod dat beschikbaar is voor ondersteuning van onderzoekers en *good practices* voor databeheer. We hebben daarbij ook gekeken naar de omgeving waarin de dataprofessional werkt en ons verdiept in opvattingen over datamanagement. Beheer van onderzoeksdata kent drie aandachtsgebieden: metadata, data-archivering en digitale duurzaamheid. Elk van deze elementen is van belang voor het werk van een dataprofessional. In verschillende landen om ons heen is inmiddels een aanbod van trainingen, cursussen en scholing ontwikkeld gericht op de kennis en vaardigheden die een dataprofessional nodig heeft. Opvallend is dat de inhoud van deze trainingen en scholingprogramma's voor dataprofessionals vaak slechts in algemene termen zijn omschreven. Wanneer de inhoud van een training wel in detail is aangegeven ontstaat al snel het beeld dat een dataprofessional een IT-specialist met *soft skills* is.

Het onderwerp beheer en management van onderzoeksdata is duidelijk nog in ontwikkeling. Gerichte onderwerpspecifieke (vaardigheids)trainingen, workshops en cursussen over eigendomsrecht van data, databankrecht, publiceren van data, opstellen van datamanagement plannen voor onderzoeksaanvragen, citaties van datasets, kwaliteitseisen, privacy en ethische aspecten van data, dataproducenten en samenwerkingsomgevingen voor multidisciplinaire datasets (cloud computing) kunnen naar verwachting rekenen op belangstelling uit het veld. Speciale aandacht verdienen ook de beginnende onderzoekers voor wie het aanleren van nieuw gedrag nog geen vanzelfsprekendheid is. Juist bij deze groep is een cultuur van open data makkelijk aan te leren en kan worden gestuurd op groepsprestaties waarin ook de dataprofessie zichtbaar is.

Bij het NIOO-KNAW wordt overwogen om in de nabije toekomst afspraken te maken over eigendomsrechten van onderzoeksdata en periodiek voorlichting te geven over dit onderwerp. In aanvulling hierop bestaan plannen om in het arbeidscontract vast te leggen dat de onderzoeker gehouden is onderzoeksdata goed gedocumenteerd over te dragen aan een leidinggevende. Afspraken over onderzoeksdata in een arbeidscontract bieden een houvast voor leidinggevendenden om bij functioneringsgesprekken de voortgang en de kwaliteit van het werk te toetsen.

Bij Tilburg University wordt het belang van een goede data-infrastructuur genoemd in het strategisch meerjarenplan en blijkt uit de laatste organisatiemonitor dat meer aandacht nodig is voor de ondersteuning van data-intensief onderzoek. Het verrijken van publicaties met onderzoeksdata is een vaste activiteit binnen de bibliotheek en er is aandacht voor scholing van informatie- en collectiespecialisten.

Het 3TU.Datacentrum richt zich expliciet op de archivering en permanente toegankelijkheid van technisch-wetenschappelijke onderzoeksdata. 3TU.Datacentrum wil op korte termijn meer aandacht schenken aan de interne scholing van dataprofessionals. Zowel Library en Information Services van de Universiteit van Tilburg als 3TU.Datacentrum stellen zich ten doel om in 2011 te voldoen aan de voorwaarden die DANS stelt aan een Trusted Digital Repository.

Het witboek wordt afgesloten met aanbevelingen voor het opzetten van een training- en scholingsaanbod voor dataprofessionals in Nederland.

Management summary

Increasing attention is being paid to research data in the Netherlands. Helping researchers deal with their data requires a new kind of professional, referred to in this White Paper for the sake of convenience by the general term "data professional". A number of researchers and data specialists were surveyed on data management practice in a various scientific disciplines, focusing on the know-how, experience, and competencies needed to provide support regarding research data.

This White Paper provides a description of data management practices in scientific disciplines at three different institutes: the Netherlands Institute of Ecology (NIOO-KNAW), Delft University of Technology, and Tilburg University. A survey was carried out of how researchers and support staff actually manage data at these institutions and what good research data management involves. Data management within each discipline was described from three different perspectives: the preparatory phase of the research, the implementation phase, and the post-completion phase.

The first part of the White Paper deals with the advent of the data professional and the available education and training to support researchers and good data management practices. We also considered the context in which the data professional works and investigated the various views on data management. Managing research data involves three different focuses: metadata, data archiving, and digital sustainability. Each of these elements is relevant to the work of a data professional. A number of neighbouring countries have developed a range of educational and training programmes focusing on the knowledge and skills that such professionals need to have. It is striking that the actual content of these programmes is often specified only in general terms. When the content of training programme is in fact described in detail, it quickly gives the impression that a data professional is an IT specialist with "soft" skills.

Research data management is clearly still under development. We can expect those in the field to be interested in subject-specific skills training, workshops, and courses on data ownership rights, database rights, publication of data, data management plans for research applications, citation of datasets, quality requirements, privacy and ethical aspects, data producers, and collaboration environments for multidisciplinary datasets (cloud computing). Special attention needs to be paid to researchers who are just starting out and who are not yet adept at learning a new approach. In this group especially a culture of open data can easily be developed, with a focus on group performance in which the value of the data profession is apparent.

NIOO-KNAW is considering making arrangements in the near future regarding data property rights and providing information periodically on this topic. There are also plans to specify in researchers' employment contracts that their research data must be transferred to their superiors together with full documentation. Arrangements regarding research data in employment contracts provide a basis for senior staff to monitor progress and quality in the context of performance appraisal interviews.

Tilburg University refers to the importance of a good data infrastructure in its long-term strategic plan, and the latest organisation survey indicates that greater attention needs to be paid to providing support for data intensive research. Enhancing publications by adding research data is now a standard activity at the library, and attention is being paid to training for information and collection specialists.

3TU.Datacentrum focuses explicitly on the storage and permanent accessibility of scientific and technological research data. It intend to devote greater attention to internal training for data professionals in the near future. Both the Library and Information Services department at Tilburg University and 3TU.Datacentrum have set themselves the target of complying with DANS's conditions for a Trusted Digital Repository by the end of 2011.

The White Paper concludes with recommendations regarding the setting up of education and training programmes for Dutch data professionals.

Inleiding: Opmars van de dataprofessionals

De laatste jaren bestaat nadrukkelijk aandacht voor het (permanent) beschikbaar houden en toegankelijk maken van onderzoeksdata. De data-infrastructuur voor onderzoeksdata is verre van ontwikkeld en in de samenleving en bij financiers van onderzoek groeit het besef dat publiek gefinancierde onderzoeksopbrengsten zoveel mogelijk beschikbaar moeten komen en worden gedeeld en hergebruikt [1, 2]. De recente belangstelling voor Open Data kan in dit licht worden gezien [2]. Maar ook uitgeverij en reviewers eisen steeds vaker toegang tot de onderzoeksdata. Daarnaast is wetenschap in de 21^e eeuw data-intensief met ongekeerd grote datavolumes, grenzeloze mogelijkheden om data te delen en indrukwekkende nieuwe data-analyse- en visualisatietechnieken. Dit alles heeft gevolgen voor de kennis en vaardigheden die nodig zijn om data-intensieve wetenschap te beoefenen, maar ook om het wetenschappelijk beheer van onderzoeksdata efficiënt en effectief te kunnen ondersteunen.

Dit witboek is bedoeld om meer inzicht te bieden in de kennis en vaardigheden die nodig zijn voor het ondersteunen van het datamanagement zoals dat plaatsvindt binnen een aantal onderzoeksdisciplines in Nederland. Door literatuurstudie en interviews met onderzoekers schetsen we een beeld van de huidige situatie. Dit leidt tot enkele gerichte adviezen voor het ontwikkelen van een training- en scholingsaanbod voor dataprofessionals. Het onderzoek is uitgevoerd in opdracht van SURFfoundation in het kader van het SURFshare programma. Met dit programma wil SURFfoundation een gemeenschappelijke infrastructuur realiseren die de toegankelijkheid en de uitwisseling van onderzoeksinformatie bevordert. Dit programma loopt van 2008 tot en met 2011 en heeft tot doel om samen met alle Nederlandse universiteiten en hogescholen, NWO en KNAW te werken aan het optimaal delen van onderzoeksresultaten, gebruik makend van de nieuwste ICT mogelijkheden.

We beginnen in hoofdstuk een met een korte toelichting op enkele kernbegrippen uit dit witboek. Vervolgens bespreken we de opmars van de dataprofessional en opkomst van de databeroepen om ons heen. Daarna doen we in hoofdstuk twee verslag van de gesprekken die de werkgroep leden voerden met wetenschappers en geven we op deze manier inzicht in de procedures, richtlijnen en 'datacultuur' bij het NIOO-KNAW, Tilburg University en de TU Delft. Het witboek wordt in hoofdstuk drie afgerond met enkele concrete handreikingen voor de opzet van een training- en scholingsaanbod voor dataprofessionals in Nederland.

Verantwoording

Bij de opzet van dit witboek zijn we ervan uitgegaan dat een wetenschappelijk onderzoek grofweg drie fasen doorloopt: een fase waarin het onderzoek wordt voorbereid en bijvoorbeeld de projectaanvraag wordt gedaan, de uitvoeringsfase waarin gegevens worden verzameld of gegenereerd, bewerkt en geanalyseerd, en een fase waarin de resultaten voor publicatie worden aangeboden en de projectadministratie afgerond [3]. Elke fase kent specifieke aandachtspunten die van belang zijn voor het beheer van onderzoeksdata en het profiel van de dataprofessional. In de voorfase van een onderzoek kan het bijvoorbeeld nodig zijn een datamanagement plan op te stellen op verzoek van de financier van het onderzoek. Maar ook het geheel van procedures en richtlijnen dat binnen een instelling of organisatie is ontwikkeld voor databeheer, archivering en lange-termijn-toegankelijkheid van onderzoeksdata rekenen we in dit witboek tot de voorfase van een onderzoek.

Tijdens de voorfase van een onderzoek volstaat doorgaans de inzet van een beleidsmedewerker die beschikt over een academische opleiding, enige onderzoekservaring en een aantal algemene vaardigheden, zoals een goede schrijfvvaardigheid, kennis van subsidiemogelijkheden en affiniteit met een onderzoeksveld (*generic data intelligence*). De beleidsmedewerker hoeft dan niet specifiek te zijn opgeleid binnen de wetenschapsdiscipline.

Anders ligt het voor werkzaamheden die tijdens een wetenschappelijk onderzoek op data worden uitgevoerd. Deze werkzaamheden vereisen doorgaans een hoog niveau van domeinspecifieke kennis en vaardigheden (*data specific intelligence*). Onze interesse ligt hier in het opsporen van kennis, vaardigheden en competenties die van belang zijn voor een efficiënte, effectieve en

innovatieve multidisciplinaire onderzoeksondersteuning. Duidelijk is dat een dataprofessional hier minimaal dient te beschikken over een gedegen informatica ondergrond.

Het belang om ook een afrondende fase te onderscheiden is onder meer gelegen in het feit dat juist in deze fase vaak het lot van de onderzoeksdata en de bijbehorende documentatie wordt bezegeld. Daarbij speelt niet alleen dat de infrastructuur voor onderzoeksdata onderontwikkeld is, maar ook dat doorgaans de middelen en de tijd ontbreken om data te documenteren en te archiveren. Soms wordt ook het belang van het beschikbaar houden en toegankelijk maken van data betwijfeld. Verder geldt dat slechts een beperkt aantal tijdschriften de eis heeft om onderzoeksdata mee te leveren of beschikbaar te houden. Tegelijkertijd stellen financiers steeds vaker als voorwaarde dat onderzoeksdata na afloop beschikbaar worden gehouden, of gedeponereerd in een data-archief, zoals DANS of 3TU.Datacentrum. Maar ook copyright en databankrecht kunnen verhinderen dat data niet worden gearhiveerd, gepubliceerd of gedeeld.

1 De omgeving van de Dataprofessional

1.1 Data en databeheer van onderzoeksdata

Onderzoeksdata

Met onderzoeksdata bedoelen we in dit witboek: digitale gegevens die gebruikt (kunnen) worden voor wetenschappelijk onderzoek, alsook de gegevens die uit wetenschappelijk onderzoek worden gegenereerd. Het woord 'kunnen' staat tussen haakjes om het belang aan te geven van gegevens die misschien niet voor wetenschappelijk onderzoek zijn verzameld, maar daar (op termijn) wel voor gebruikt kunnen worden.

De interesse en aandacht voor een beter beheer van onderzoeksdata hangt onder meer samen met de overgang naar een digitale wetenschap, het overweldigende aanbod van data voor de wetenschap, de ongekende mogelijkheden om onderzoeksdata te delen met anderen en het feit dat de data-infrastructuur voor onderzoeksdata nog in ontwikkeling is [1, 4]. De ongekende mogelijkheden om data te delen, maken hergebruik van data eenvoudig en stimuleren interdisciplinair onderzoek. De mogelijkheden om datakwaliteit te garanderen zijn eveneens ongekend, evenals de mogelijkheden om de herkomst en de integriteit van wetenschappelijke brondata te manipuleren [5, 6]. Duidelijk is dat dit alles grote gevolgen heeft voor het wetenschappelijk databeheer en de deskundigheid die hiervoor vereist is.

Databeheer, datamanagement en digitaal cureren

In de Engelstalige literatuur worden databeheer (*data management*) en digitaal cureren (*digital curation*) door elkaar gebruikt. Ook in dit witboek maken we geen strikt onderscheid tussen databeheer en digitaal cureren en gebruiken we de begrippen voor alle werkzaamheden die te maken hebben met het archiveren, collectioneren, beschrijven, toegankelijk maken en publiceren van onderzoeksdata in digitale vorm.

Een tweetal opmerkingen is echter nog van belang. Databeheer (*data management*) heeft in dit witboek een andere en bredere betekenis dan binnen de ICT gebruikelijk is, bijvoorbeeld waar de term wordt gebruikt om de functie van een databasebeheerder te beschrijven. Tegelijkertijd is het belangrijk vast te stellen dat de kennis en vaardigheden die nodig zijn voor wetenschappelijk databeheer en datamanagement doorgaans een sterk beroep doen op kennis van ICT en inzicht in de data-infrastructuur.

1.2 Data-archivering, metadata en digitale duurzaamheid

Bij datamanagement van onderzoeksdata zijn drie hoofdthema's van belang: data-archivering, metadata en digitale duurzaamheid. We lichten elk thema hieronder kort toe.

Data-archivering

Onderzoeksdata bestaan uit gegevens die digitaal zijn opgeslagen in een of meer bestanden. Belangrijke onderzoeksdata dienen te worden opgeslagen in een datacentrum of data-archief. Daarbij kan gebruik gemaakt worden van een kwaliteitskeurmerk zoals het Data Seal of Approval dat door DANS wordt verleend aan centra en archieven die voldoen aan criteria op het gebied van kwaliteit, duurzaamheid en toegankelijkheid van onderzoeksdata. DANS en het 3TU.Datacentrum zijn voorbeelden van centra die in Nederland zijn ingericht om onderzoeksdata te archiveren.

Metadata

Metadata geven informatie over de data. Metadata van onderzoeksdata leggen kenmerken van de onderzoeksdata vast, zoals de titel van een bestand, de naam van de producent, trefwoorden en informatie over gebruikte methoden en instrumenten, etc. Maar ook alle overige documentatie van de onderzoeksdata rekenen we tot de metadata. Metadatering van onderzoeksdata is noodzakelijk om deze hervindbaar te maken. Metadata zijn een belangrijk onderdeel van het wetenschappelijk databeheer en datamanagement. Onderzoeksdata zijn zonder een goede metadatering niet of

slechts beperkt bruikbaar. Er zijn verschillende standaarden voor metadata in omloop. De bekendste is wel het Dublin Core Metadata Initiative (DCMI, [7]). DCMI wordt in Nederland toegepast voor metadatering van overheidsinformatie [8]. Een andere veelgebruikte metadata standaard voor onderzoeksdata is ontwikkeld door het Data Documentation Initiative (DDI [9]).

Digitale duurzaamheid (digital preservation)

Voor de digitale wetenschap is het van belang dat gegevens ook op lange termijn nog benaderbaar en bruikbaar zijn. Bestandsformaten, opslagmedia in combinatie met de juiste hard- en software en de aanwezigheid van metadata zijn van essentieel belang voor behoud van toegang tot digitale databronnen. Duurzame opslag en toegankelijkheid van onderzoeksdata vereist specialistische kennis. DANS speelt in Nederland een belangrijke rol bij het archiveren en waarborgen van de toegankelijkheid van onderzoeksdata in de sociale wetenschappen en de humaniora. De Nationale Coalitie Digitale Duurzaamheid (NCDD) vervult eveneens een belangrijke verbindende en discipline overstijgende rol bij het bevorderen van digitale duurzaamheid [10].

1.3 Training en opleiding van dataprofessionals

De laatste jaren zien we een groeiend aantal functies waarbij de term data expliciet in de functiebenaming voorkomt [11-13]. Op het Internet worden nu vacatures en functies aangeboden waarin gezocht wordt naar datawetenschappers (*data scientists*), data-architecten, datamanagers, data consultants en data stewards, onderzoeksdata specialisten (*research data specialists*) en databibliothecarissen (*data librarians, science data services librarians*, [13-24]). Opmerkelijk daarbij is dat datawetenschappers veel gevraagd zijn bij jonge hightech bedrijven [11]. Maar ook de uitspraak van Jacob de Vlieg, voorzitter van het Nationaal eScience Center, is in dit verband belangwekkend. De Vlieg stelt dat het zoeken naar een nieuw soort wetenschapper wellicht de grootste uitdaging is voor het centrum in oprichting [25].

Tegelijkertijd met de groeiende vraag naar dataprofessionals is in de Angelsaksische landen een indrukwekkend cursus- en scholingsaanbod ontwikkeld voor het beheer van onderzoeksdata [26, 27]. Belangrijke spelers in het Verenigd Koninkrijk (UK) zijn het Digital Curation Centre (DCC), JISC en UK Data Archive. Deze organisaties zijn de drijvende kracht achter de data curation en data management trainingen voor onderzoekers en dataprofessionals en het aanbod van 'train the trainer' cursussen [28] in de UK. In de VS is bij MIT Libraries een workshop over databeheer ontwikkeld die door informatiespecialisten aan onderzoekers wordt gegeven. Bij Cornell University Library is bijvoorbeeld een online tutorial beschikbaar voor *digital preservation management* [29].¹

Gesteld kan worden dat de bovengenoemde cursussen voor databeheer en digitaal cureren doorgaans een algemeen karakter bezitten waardoor ze voor onderzoekers op zijn best een *nice to know* karakter hebben in plaats van *need to know karakter*. Daarnaast is het van belang op te merken dat de bovengenoemde cursussen ook niet bedoeld zijn om dataprofessionals op te leiden die onderzoekers kunnen ondersteunen. Dit laatste geldt niet voor het aanbod van Liddy, Madnick et al. en Corral waarop we hieronder kort ingaan [15, 30-32].

Liddy [15] heeft een 2 jarige mastersopleiding ontworpen aan de Library and Information Science School van Syracuse University die informatiespecialisten opleidt tot bijvoorbeeld *Cyber Infrastructure Facilitator*. Studenten volgen gedurende twee jaar vakken waarin ze een combinatie van technische -, informatica - en communicatievaardigheden opdoen. Na afstuderen zijn deze specialisten direct inzetbaar in de natuur- en technologie wetenschappen en wiskundelaboratoria. Tijdens de opleiding wordt veel aandacht besteed aan het ontwikkelen van een gids- en ondersteuningsfunctie die deze dataspecialisten moeten vervullen in het contact met onderzoekers.

Aanknopingspunten voor het uitwerken van curricula voor specialistische dataprofessionals vinden we ook bij Madnick et al. [30, 31], Corral [32] en enkele andere initiatieven [15, 33, 34]. Corral geeft een aantal voorbeelden van trainingen en opleidingen waarin aandacht is voor data skills. De

¹ In Bijlage 4 is een overzicht opgenomen met verwijzingen naar bijeenkomsten, workshops en trainingen binnen en buiten de UK voor het beheeren en cureren van onderzoeksdata.

voorbeelden hebben betrekking op informaticacursussen die onderdeel zijn van opleidingen binnen gezondheids- en sociale wetenschappen, chemie en technologie en digital library management. Daarnaast geeft Corral ook voorbeelden van informatievaardigheidstrainingen waarin databeheer centraal staat.

Situatie in Nederland

Ook in Nederland zien we initiatieven ontstaan om trainingen voor wetenschappers en datamanagers op te zetten. Er is bijvoorbeeld een korte algemene datamanagement cursus ontwikkeld door DANS en NIOO-KNAW en ook de Universiteit van Amsterdam heeft enkele workshops voor onderzoekers over dit thema georganiseerd. Zowel DANS en 3TU.Datacentrum werken daarnaast aan een aanvullend trainingsaanbod voor ondersteuning van dataprofessionals in een academische setting [35].

Van belang is dat de noodzaak voor meer aandacht voor databeheer door onderzoekers niet of onvoldoende wordt herkend en daardoor geen prioriteit heeft. Natuurlijk speelt hier ook dat onderzoekers worden afgerekend op publicaties en niet op goed databeheer of de publicatie van datasets. De inhoudelijke uitdagingen liggen deels in het verbinden van onderzoek en onderzoeksondersteuning en deels in het verbinden van beleidsmakers, dataprofessionals en de IT onderzoeksondersteuning binnen een vakgroep of discipline.

Verder geldt dat het trainings- en cursusaanbod voor databeheer en digitaal cureren doorgaans een algemeen karakter heeft waardoor deze voor onderzoekers op zijn best een *nice to know* karakter hebben in plaats van *need to know karakter*. Ook is het huidige aanbod onvoldoende uitgewerkt om dataprofessionals op te leiden die onderzoekers kunnen ondersteunen. Bij de ontwikkeling van een cursus- en opleidingsaanbod in Nederland dienen we hiermee rekening te houden.

1.4 Data Generic en Data Specific Intelligence

Door het groeiend aanbod van functiebenamingen voor dataprofessionals is het voor niet-ingewijden moeilijk helder te krijgen wat nu de essentiële kwalificaties, competenties en vaardigheden zijn van een dataprofessional. De kortste weg naar een antwoord is dat het aankomt op de juiste mix van IT- en 'soft skills'. In de praktijk is de inhoud van deze mix vaak vertroebeld doordat een onwaarschijnlijke combinatie van ervaring, competenties en vaardigheden op de dataprofessional voor onderzoeksdata wordt geprojecteerd [14, 15, 30, 32, 36-40]. Waar vind je immers iemand die beschikt over programmeerkennis, kennis van data-analyse, datastandaarden, onderzoekservaring en de juiste sociale vaardigheden om als 'buitenstaander' dataprojecten bij vakgroepen te gaan uitvoeren?

In een poging om een deel van de vertroebeling weg te nemen en om tot een gericht training- en scholingsaanbod van dataprofessionals te komen, maken we in het witboek onderscheid tussen 'generic data intelligence' en 'data specific data intelligence'. Deze begrippen zijn moeilijk één op één te vertalen in het Nederlands. Voor het gemak vertalen we 'data specific data intelligence' in data-specifieke vaardigheden en 'generic data intelligence' in niet-specifieke of algemene data vaardigheden.

Dataspecifieke vaardigheden zijn vereist om de technische aspecten van opslag van data te kunnen begrijpen, de werking van zoekmachines te kunnen doorgronden, technieken voor 'data retrieval' te kunnen gebruiken en informatie-visualisatie-technieken te kunnen toe passen. Niet-specifieke en algemene datavaardigheden zijn van belang voor het digitaal cureren van onderzoeksdata (beschrijven, ontsluiten, publiceren), het beheren van specifieke bestanden of collecties van bestanden en eenvoudige veredeling van onderzoeksdata. Een data librarian, data steward of datamanager zal doorgaans kunnen volstaan met een goede bagage van algemene vaardigheden, terwijl een data-architect, data-analist of datawetenschapper vaker over dataspecifieke vaardigheden dient te beschikken.

Het gebruik van dataspecifieke versus algemene vaardigheden is ook te illustreren door de werkzaamheden voor databeheer in samenhang te zien met aspecten van datakwaliteit. Heeft de

organisatie bijvoorbeeld een beleid voor databeheer? Hoe is het databeheer in de praktijk geregeld? Hoe wordt databeheer en datakwaliteit vanuit de techniek ondersteund en hoe is de beveiliging van de toegang tot data en de identificatie van de gebruiker geregeld? Duidelijk is dat een organisatie bij de eerste twee vragen baat heeft bij een dataprofessional die beschikt over voldoende *generic data intelligence* en dat vraagstukken met overwegend technische componenten om *data specific intelligence* vragen [41-43].²

Om een indruk te krijgen van de activiteiten voor databeheer die we tot *generic data intelligence* kunnen rekenen staan we even stil bij de inventarisatie die Liddy [7] maakte van de werkzaamheden die feitelijk worden uitgevoerd bij wetenschappelijk databeheer. Volgens Liddy bestaat wetenschappelijk databeheer uit de volgende activiteiten: verzamelen en genereren van onderzoeksdata, toevoegen van metadata aan onderzoeksdata, schonen van onderzoeksdata, uitvoeren van berekeningen, converteren, samenvoegen, visualiseren, archiveren en uitwisselen van onderzoeksdata. Elke dataprofessional moet in staat zijn eenvoudige taken en bewerkingen voor elk van deze activiteiten uit te voeren.

Al met al kan gesteld worden dat een dataprofessional die breed inzetbaar is, dient te beschikken over kennis, ervaring en vaardigheden op de volgende drie gebieden:

1. Kennis van data-archivering, beschikbaarheid van data, metadata en methoden van dataverwerking.
2. Kennis van informatie- en communicatie technologie.
3. Kennis van en ervaring met wetenschappelijk onderzoek in minimaal één discipline.
4. Goede sociale vaardigheden ('soft skills'). Soft skills zijn vereist om ondersteuning te kunnen bieden aan studenten en onderzoekers en overtuigingskracht en overwicht te tonen in situaties waarin de dataprofessional een meerwaarde kan bieden.

Tot slot geven we een voorbeeld van *data specific intelligence* aan de hand van een profiel van een datawetenschapper binnen wetenschappelijke bibliotheek [45].³ Kenmerkend voor de datawetenschapper is dat deze persoon kwalificaties en ervaring heeft op verschillende terreinen, zoals kennis van programmeren, databases, multivariate data-analyse, webservices, semigestructureerde databases en data retrieval technieken. Kort gezegd zijn datawetenschappers moeilijk te vinden omdat voor elk van de genoemde deelgebieden (nu nog) afzonderlijke opleidingen bestaan. Dit staat haaks op de kerncompetentie van de datawetenschapper. We verwachten dat deze in staat moet zijn traditioneel gescheiden aandachtsgebieden te verbinden en discipline overstijgend te werken. Volgens Mason en Wiggins [46] dient de datawetenschapper daarom specifieke ervaring te bezitten op de volgende vijf gebieden:

1. Methoden van dataverzameling via het web.
2. Kennis van data-exploratie technieken en data mining.
3. Kennis van het efficiënt en slim schoonmaken van data ('data scrubbing').
4. Gebruikswensen kunnen vertalen in datastructuren, datamodellen en visualisaties.
5. Kennis van data-analyse technieken en verslaglegging.

Soft skills worden in het bovenstaande overzicht niet genoemd en lijken daarom misschien voor de datawetenschapper minder van belang. Niets is echter minder waar aangezien een datawetenschapper vooral zal worden aangesproken op zijn soft skills wanneer hij of zij in plaats van een strikt uitvoerende rol de verbindende schakel is in een multi-disciplinair team of technisch projectleider van een discipline overstijgend onderzoek. De soft skills die nodig zijn om als dataprofessional te kunnen functioneren in een dataspecifieke - of een meer algemene setting zijn op hoofdlijnen vergelijkbaar. Maar het is tegelijkertijd niet onrealistisch om een proactieve houding en overtuigingskracht te vragen van een dataprofessional die zich moet 'waar maken' bij een vakgroep, afdeling of faculteit die zich buiten de eigen kring bevindt.

² Een andere manier om inzicht te krijgen in het vereiste niveau van specialisatie is de (fictieve) dataprofessional de 20 belangrijkste vragen voor te leggen die de afdeling beantwoord wil zien 44. Hey, A.J.G., S. Tansley, and K. Tolle, *The Fourth Paradigm: Data-intensive Scientific Discovery.*, A.J.G. Hey, S. Tansley, and K. Tolle, Editors. 2009, Microsoft Research..

³ In bijlage 1 zijn definities opgenomen voor verschillende typen van dataprofessionals; bijlage 2 en 3 toont een concreet voorbeeld van een profielschets van een data librarian.

2 Dataprofessionals aan zet

2.1 Werkwijze

In dit hoofdstuk wordt verslag gedaan van de gesprekken die zijn gevoerd met wetenschappers van NIOO-KNAW, TU-Delft en Tilburg University. Elk werkgroeplid heeft minimaal drie gesprekken gevoerd met onderzoekers en ondersteuners die intensief gebruik maken van onderzoeksdata. In elk van de gesprekken is ingegaan op zoveel mogelijk relevante aspecten van het gebruik en het beheer van onderzoeksdata bij een wetenschappelijk onderzoek volgens bovenstaande opzet.⁴

Bij het NIOO-KNAW zijn gesprekken gevoerd met een onderzoeker, een onderzoeksassistent en een datamanager. Het onderzoek dat bij het NIOO wordt uitgevoerd op het gebied van de ecologie betreft vooral veldstudies en onderzoek waarbij tijdreeksen en herhaalde opnames van veldmonsters worden verzameld. Daarnaast worden experimenten gedaan in het veld en in het laboratorium. Er zijn ook studies waarin veld- en laboratoriumexperimenten worden gecombineerd en er worden literatuurstudies uitgevoerd. De voorbereidende fase omvat activiteiten die in het kader van een specifieke projectaanvraag worden gedaan en aspecten van de workflow van een onderzoek. Het maakt verschil of een onderzoek is ingebed in een vooraf opgestelde structuur, zoals bij het koolmezenonderzoek. Het koolmezenonderzoek is een langlopend project waarbij al vanaf 1955 vier populaties koolmezen in het wild worden bestudeerd en waarvan de nest- en broedgegevens worden gerapporteerd en bestudeerd.

In Tilburg is een wetenschapper geïnterviewd die biofeedback-onderzoek verricht en de hoofdonderzoeker geïnterviewd achter het EDACwowe datacentrum [47]. Ook is het hoofd van de ICT-infrastructuur van Tilburg University geïnterviewd. Deze persoon heeft vanuit zijn functie te maken met opslag-, beschikbaarheid- en een veilige toegang tot bedrijfskritische informatie en de centrale data-infrastructuur voor Tilburg University. Verder zijn ook de inzichten meegenomen uit de gesprekken die gevoerd zijn in het kader van de dataverkenning economische wetenschappen in Nederland.

3TU.Datacentrum heeft enkele TU Delft wetenschappers geïnterviewd van de faculteit Civiele Techniek en Geowetenschappen. Het betreft zowel universitair docenten als aio's. Tevens zijn de bevindingen meegenomen uit de interviews die voor het 3TU.Datacentrum/DANS-project 'Waardevolle diensten en data' zijn gehouden.

2.2 Databeheer bij de voorbereiding van onderzoek

2.2.1 Situatie bij het NIOO-KNAW

De voorbereidende fase omvat activiteiten die in het kader van een specifieke projectaanvraag worden gedaan en aspecten van de workflow van het onderzoek. Het maakt daarbij verschil of een onderzoek is ingebed in een vooraf opgestelde structuur. Bij het koolmezenonderzoek is bijvoorbeeld tamelijk strak omschreven hoe, wanneer en in welke vorm de onderzoeksdata worden opgeslagen. Bij een volledig nieuw onderzoeksthema vindt in de voorbereidende fase overleg plaats met de databasemanager om de gegevens op een voor onderzoekers bruikbare manier op te slaan. In andere gevallen kan het zijn dat de workflow van een onderzoek de inzet van verschillende medewerkers vraagt, bijvoorbeeld om de overdracht van gegevens te coördineren. In alle gevallen wordt met de onderzoeksassistenten en werkbegeleiders vooraf tot overeenstemming gekomen over waar, wanneer en in welke vorm de onderzoeksgegevens worden vastgelegd en overgedragen.

Het NIOO-KNAW zoekt naar een vorm waarbij databeheer als een natuurlijk proces is opgenomen in de workflow van de onderzoeker: "We willen een vorm vinden waarbij de ijkpunten voor het

⁴ De gespreksverslagen zijn opgenomen in bijlage x, y en z.

datamanagement worden vastgelegd, bijvoorbeeld bij de start van een onderzoek en als vast onderdeel bij voortgangsgesprekken met een aio. Ook is het belangrijk om bij de financieringsaanvraag voor nieuw onderzoek standaard ruimte voor datamanagement aan te vragen.”

Tabel 1: Overzicht van voorbereidende werkzaamheden bij het NIOO-KNAW

Richtlijn in gebruik tot 2007: elke vertrekkende onderzoeker (vaak een PhD-student, die zijn promotie heeft afgerond) dient zijn databestanden op een Cd-rom in te leveren bij de bibliotheek (<i>richtlijnen en procedures</i>) ⁵ .
In 2008 is tweemaal samen met VLIZ een workshop 'algemeen wetenschappelijk databeheer' georganiseerd (<i>training en scholing</i>).
De rol van de bibliotheek is vooral gericht op de bewustwording van het belang van goede metadata en het gestructureerd opslaan van onderzoeksgegevens. Ook proberen wij een infrastructuur aan te bieden en bij de communicatie met de ICT-afdeling te bemiddelen (<i>beleid, voorlichting/informatievoorziening, intermediair en opslag</i>).
De rol van de 'data librarian' kan alleen maar op een algemeen en organisatorisch niveau uitgevoerd worden, het benadrukken van het belang van datamanagement, en het omgaan met digitale bestanden en formulieren (<i>data librarian</i>).
Vanuit de directie van het NIOO-KNAW is altijd het standpunt gehuldigd dat alleen een onderzoeker zijn/haar eigen data kan beheren. Ook is een onderzoeker zelf verantwoordelijk voor de documentatie bij zijn onderzoeksdata (<i>richtlijnen, procedures, cultuur</i>).
Sinds 2007 worden de mogelijkheden om onderzoeksdata efficiënt en duurzaam op te slaan onderzocht (<i>procedures, data-infrastructuur</i>).
Vanuit de KNAW wordt de mogelijkheid onderzocht om bij elk nieuw arbeidscontract expliciet te stellen dat de pas aangestelde onderzoeker tekent voor de erkenning dat het instituut eigenaar van de onderzoeksdata is en dat een onderzoeker gehouden is om de onderzoeksgegevens bij afronding of bij voortijdig verlaten van het onderzoeksproject over te dragen (<i>richtlijnen en procedures</i>).
De geïnterviewde onderzoeksassistent ziet wel wat in de mogelijkheid om via opslag van ruwe data ook het proces van metingen te stroomlijnen en te waarborgen. [...] Dat zou een mooi punt zijn voor waarborg bij de overdracht van gegevens. Bovendien als het data-archiveren wordt opgenomen in de workflow dan wordt het geïntegreerd in het natuurlijk gedrag van de onderzoeker (<i>opslag, procedures</i>).
Het maakt verschil of een onderzoek ingebed wordt in een vooraf opgestelde structuur, zoals bij onze koolmezen-onderzoekers (<i>workflow</i>).
Bij experimenteel onderzoek is het moeilijker om vooraf te bepalen wanneer en hoe de gegevens worden vastgelegd en overgedragen. Veelal zal in een projectbeschrijving of onderzoeksplan de toekomstige onderzoeker moeten aangeven wat, hoe en wanneer hij de onderzoeksgegevens archiveert (<i>workflow</i>).

2.2.2 Situatie bij TU Delft

3TU.Datacentrum stelt zich ten doel om dé faciliteit te worden voor permanente toegankelijkheid van technisch-wetenschappelijke onderzoeksdata. Accountmanagers en informatiespecialisten spelen een belangrijke rol in het attenderen van wetenschappers op het belang van een goede en duurzame opslag van onderzoeksdata. Tevens wordt veel aandacht gegeven aan advies en ondersteuning met betrekking tot databeheer.

⁵ In alle tabellen zijn de uitspraken van de geïnterviewden van trefwoorden voorzien met het oog op verdere analyse.

Tabel 2: Overzicht van voorbereidende werkzaamheden bij TU Delft

In de fase waarin het onderzoek wordt voorbereid is er geen sprake van een data managementplan. Onderzoekers hanteren eigen richtlijnen of aanpak voor het bewaren van data (beleid: afwezigheid, richtlijnen).
Voor het opslaan van metadata worden geen duidelijke richtlijnen gehanteerd. Het belang van metadata voor de terugvindbaarheid en hergebruik van data wordt wel erkend. Onderzoekers zouden daar meer in willen worden gefaciliteerd, bijv. in de vorm van een metadata formulier (datadiensten, richtlijnen).
Beginnende onderzoekers hebben behoefte aan een training op het gebied van data opslag en beheer (datadiensten).
Een deel van het onderzoeksbudget reserveren voor datamanagement is niet gebruikelijk (voorwaarden, beleid: afwezigheid).

2.2.3 Situatie bij Tilburg University

Tilburg University kent vijf faculteiten waarbinnen sociaal-wetenschappelijk onderzoek wordt uitgevoerd. Binnen de faculteiten is een groot aantal specialisaties aanwezig die elk een geheel eigen workflow kennen. Bij de keuze van de interviews voor dit witboek is een inschatting gemaakt van de mate waarin een discipline als data-intensief is aan te merken. Daarnaast is gekeken naar die onderzoeksgroepen waar het beheer van onderzoeksdata een vast onderdeel is van de bedrijfsvoering.

Typisch voor het onderzoek naar biofeedback zijn langdurige, longitudinale experimenten. Personen worden getraind om een bepaalde fysiologische activiteit te volgen en te controleren. Een ontwikkeling in het onderzoek naar biofeedback is het onderzoek naar multimodale responsystemen waarbij databestanden ontstaan met bijvoorbeeld gegevens over hersenactiviteit, autonome- en spieractiviteit. De trend naar Open Access wordt onderkend maar vooral voor publicaties en niet zozeer voor het delen en publiceren van data.

EDACwowe is onderdeel van een 'network of excellence' waarin 27 instituten en 280 onderzoekers vertegenwoordigd zijn. Het netwerk bestaat uit politicologen, sociologen, economen, juristen, filosofen en incidenteel onderzoekers van een andere discipline[47]. Meestal wordt gebruik gemaakt van secundaire onderzoeksdata maar binnen het netwerk worden soms ook zelf data verzameld met enquêtes.

Tabel 3: Overzicht van voorbereidende werkzaamheden bij Tilburg University⁶

Er wordt gebruik gemaakt van gevalideerde stimulussets die in eigen beheer worden ontwikkeld of op de markt gekocht worden (I1) (<i>procedures</i>).
Er worden specifieke en aspecifieke hypothesen getoetst. Beide vormen van onderzoek hebben een eigen aanpak. Specifiek doet een beroep op standaardisatie; a-specifiek onderzoek vereist het zoeken naar patronen in onderzoeksdata (I1) (<i>workflow</i>).
Nagegaan wordt of een experiment al eens eerder is gedaan. Een ervaren onderzoeker weet dat, voor een phd is dat minder makkelijk (I1) (<i>workflow</i>).
Onderzoekers zijn minder afhankelijk geworden van wat lokaal wordt ontwikkeld (I1) (<i>data-infrastructuur, datadiensten</i>).
Bij Library and IT services is geen infrastructuur voor onderzoeksdata aanwezig. De infrastructuur is afgestemd op de kritische onderdelen van de bedrijfsvoering. De infrastructuur voor onderzoeksdata wordt decentraal geregeld (I1) (<i>data-infrastructuur</i>).
Toppublishaties resulteren vooral uit de keuze voor een goed onderwerp en zijn niet zozeer afhankelijk van stimulusmateriaal of een statistische analyse (I1) (<i>voorwaarden</i>).
Er is veel interesse in onderzoek naar lichaamstaal (zie bijv. De Gelder). Ook social neuro science is in opkomst: sociale psychologie combineren met kennis van hersenen (I1) (<i>workflow</i>).
Elke onderzoeker, of onderzoeksgroep is zelf verantwoordelijk voor databeheer. Met SpITs, de ICT-ondersteuningsunit worden afspraken gemaakt over opslag van gegevens (I1) (<i>procedures, opslag</i>).

⁶ Voor Tilburg University zijn de uitspraken van de geïnterviewden gemarkeerd met I1 t/m I3.

De eerste stappen in medisch- en neuro-psychologisch onderzoek zijn zeer domeinspecifiek. Tegen de student hebben we gezegd dit zijn 'microvolts per seconden', maar doe maar net alsof het 'aandelenkoersen per maand zijn' (I1) (*ondersteuning, data specific intelligence*).

Ik vind dat bij de opleiding tot onderzoeker ook een stukje datamanagement hoort. Dat verschilt nu per onderzoeker of daar aandacht voor is bij het onderwijs. Het gaat hierbij om organisatie en opslag van data en documentatie van stappen die voor analyse van belang zijn (I1) (*beleid*).

In toenemende mate hebben we te maken met primaire data-instituten; die instituten verzamelen data en maken dit beschikbaar via het internet. [...] In toenemende mate zag men door de bomen het bos niet meer (I2) (*data-infrastructuur, data kwaliteit*).

We maken voor EDACwowe een onderscheid tussen surveys, micro data, statistische data en kwalitatief inhoudelijke data (I2) (*workflow*).

De opzet van de portal is voorgelegd aan een extern panel van gebruikers en collega's, zo van schiet er maar op. Daarna is bepaald zo gaan we het doen. Gedacht wordt aan sociaal risico als een nieuwe categorie waaronder data over pensioenen, arbeidsmarkt, werkloosheid, gezondheidsvraagstukken en regelgeving over ziektewet en re-integratie. (I2) (*workflow*).

De data die je nodig hebt zijn heel erg afhankelijk van de specifieke onderzoeksvragen die je hebt (I2) (*workflow*).

Vanaf het begin is gesteld dat alleen kwalitatieve mensen voor de dataportal worden ingezet om vergelijkende informatie over beleid en beleidsinstrumenten, sociale regelingen etc. in te voeren (I2) (*voorwaarden, richtlijnen*).

2.3 Databeheer tijdens onderzoek

2.3.1 Situatie bij het NIOO-KNAW

In de uitvoerende fase van het onderzoek is het van belang dat alle onderzoeksgegevens direct op de juiste manier worden vastgelegd, zodat er geen vervuiling van gegevens kan optreden. NIOO-KNAW pleit voor een tussentijds, bijvoorbeeld een jaarlijks vastleggen van de onderzoeksgegevens, - niet alleen als waarborg en back-up voor de onderzoeksactiviteiten -, maar ook als signalering voor stagnatie en voortgang.

Jaarlijks organiseert het NIOO-KNAW vanuit de bibliotheek een datadag waarbij in samenwerking met VLIZ de onderzoekers per afdeling in de gelegenheid worden gesteld om hun data te archiveren. Vanaf 2011 zal ook per afdeling het aantal gearchiveerde datasets als indicator aan de KNAW worden gerapporteerd.

Tabel 4: Databeheer in de praktijk bij het NIOO-KNAW (direct van belang voor primair proces)

De onderzoeker krijgt van de werkgroep leider en/of van de assistent/databasebeheerder begeleiding bij de manier waarop onderzoek hoort te worden uitgevoerd, inclusief de vastlegging daarvan (*workflow: cultuur*).

Zelden werd en wordt er naar de Cd-roms omgekeken, omdat de wetenschappelijke onderzoeksgegevens onvoldoende of niet toegankelijk zijn (*opslag en toegankelijkheid*).

Het ontbreken van de nodige documentatie bij de bestanden is veeleer de regel dan uitzondering (*toegankelijkheid*).

Misschien wel de belangrijkste kwestie: er zou een duidelijke lijn moeten zijn binnen de werkgroepen hoe zij met onderzoeksdata omgaan. Hoe past het specifieke onderzoek binnen de werkgroep en het grote geheel daarbuiten. Welke data zijn relevant en kunnen eventueel van elders meegenomen worden? Het zou goed zijn als werkgroepen een eigen database zouden aanleggen waarin alle data die verzameld worden ingebracht worden (*procedures en richtlijnen*).

Library and Information Services kan assisteren bijv. bij het opzetten van een database in MS-Access en bij het uploaden van bestanden naar het Data Archive (*opslag, toegang*).

Door de variatie aan gegevens is het onmogelijk om één relationeel databankstelsel naar voor te schuiven dat aan alle functionele eisen voldoet. Wenselijk is de verschillende (bestaande en nieuwe) systemen naast elkaar te gebruiken waarbij ook aandacht is voor lange termijn opslag, centrale beschikbaarheid en herbruikbaarheid van de gegevens. Om dit te realiseren zijn er niet alleen systemen nodig, maar dienen er ook bepaalde procedures en afspraken te worden vastgelegd en zal er meer aandacht worden besteed aan bewustwording en deskundigheidsbevordering (*voorwaarden, procedures, richtlijnen*).

De ondersteuning moet primair van binnen de werkgroep komen en vooral in het technische – fysische vlak liggen. Bijv. voor het inwerken van Matlab (of nu R) scripts is computationele vaardigheid nodig. Eigen expertise en eigen ontwerpen plus voldoende tijd zijn een belangrijke factor daarin (*ondersteuning, data specific intelligence*).

Databestanden kunnen bestaan uit gegevens die door laboratoriuminstrumenten gegenereerd worden, maar omvatten ook zowel relationele databanken als lijsten of inventarissen en ruwe datafiles. Belangrijke voorbeelden hiervan zijn de primaire-productie-database, de DNA-sequentiedatabase, de biomarkers-database, een overzichtsbestand cultuurcollectie algen en een inventaris van beschikbare primers. Voorts zijn er de volgende datatypen: reeksen van wetenschappelijke foto's (blimp, oblique, image tracking,) remote sensing data (lucht- en satelliet fotos, rasterbeelden), hydrodynamische flume metingen, hydrodynamische veldmetingen, overige veld experimenten en metingen, historische vegetatieopnames, modelberekeningen. Vele wetenschappers werken met een eigen databestand voor hun eigenonderzoek. Deze databestanden hebben elk hun eigen design en formaat en zijn opgesteld hetzij in spreadsheets, hetzij in relationele databanken (*databeheer, data formaten, workflow*).

Tijdens de eerste inventarisatie werden 21.393 bestanden bekeken, daarvan bleken uiteindelijk 3421 bestanden bruikbaar, die in 91 datasets beschreven konden worden (*datakwaliteit*).

De metingen worden gedaan door de laboratoriumassistenten en die dragen de metingen over aan de onderzoekers. Dat zou een mooi punt zijn om de gegevens te borgen. Bovendien als het data-archiveren wordt opgenomen in de workflow dan wordt het geïntegreerd in het natuurlijk gedrag van de onderzoeker (*workflow, databeheer*).

Een afdeling binnen het NIOO-KNAW onderhoudt al jaren een grote databank met gegevens. De gegevens worden in veldboekjes verzameld en later in het systeem ingevoerd. Bijna alle onderzoeksgegevens van die afdeling kunnen in de database worden opgenomen (*opslag en toegang*).

Waar VLIZ ons adviseerde tegen een grote relationele databank, is de databankmanager daar juist heel positief over. Zijns inziens zou dat ook voor andere afdelingen een oplossing zijn, omdat je op die manier meer structuur krijgt en overzicht in het onderzoek en de resultaten van een afdeling. " Breng alles in een geheel, want dan kun je ook andere elementen en randgebieden zien en ev. nieuw onderzoek zal uitlokken (*opslag, vindbaarheid, toegang, data kwaliteit*) "

Onderzoekers hebben ondersteuning nodig bij computationele activiteiten: je moet onderzoekers faciliteren (*voorwaarden, opleiding*).

2.3.2 Situatie bij TU Delft

Tabel 5: Databeheer in de praktijk bij TU Delft

Metingen, databewerking en data-opslag worden door de onderzoeker zelf gedaan. In veel gevallen worden de data opgeslagen op de harde schijf van de eigen computer en op dvd's voor de back-up (*opslag en toegang*).

Een goede documentatie van de data wordt als belangrijk gezien maar in de praktijk is hier niet altijd voldoende aandacht voor (*datakwaliteit: toegankelijkheid*)

De geïnterviewden is gevraagd welke taken m.b.t. databeheer overgenomen zouden kunnen worden door medewerkers zonder kennis van het vakgebied maar met kennis van data-archivering. Mogelijkheden hiervoor ziet men vooral bij het stroomlijnen van het proces, controle op metadata, versiebeheer en conversie (*procedure, opslag*).

Beelddata kosten veel opslagcapaciteit. In het verleden werden alleen bewerkte gegevens opgeslagen vanwege de opslagcapaciteit (PhD onderzoek: 400 GB). Nu is die capaciteit geen probleem meer maar omdat de gegevens slecht toegankelijk zijn, kan men er weinig mee (*opslag en toegankelijkheid*).

Wanneer data al tijdens het onderzoek in een datacentrum worden opgeslagen zou dit veel tijdswinst kunnen opleveren, evenals de zekerheid dat de data direct bewaard blijven (*opslag, toegankelijkheid, datadiensten*).

2.3.3 Situatie bij Tilburg University

Tabel 6: Databeheer in de praktijk bij Tilburg University

Stimulussets worden alleen gedeeld met bekende onderzoekers en instituten en bij een Europees project. Er gaat veel tijd op aan het ontwikkelen van stimulussets (I1) (*workflow, delen van data, cultuur*).

Er wordt gebruik gemaakt van stimulussets en registraties bij personen (fysiologische activiteit, hersenactiviteit, spierspanning, bloeddorstroming) (I1) (*workflow*).

Er is behoefte aan voldoende flexibiliteit bij de opslag van onderzoeksdata. Een centrale voorziening als SARA schiet in de praktijk te kort op snelheid en kosten (I1) (*opslag, data-infrastructuur, datadiensten*).

Onderzoekers zitten in multidisciplinaire teams verspreid over verschillende universiteiten die allemaal toegang willen tot dezelfde data. Voorzieningen moeten niet per universiteit worden geregeld maar bijv. door SURFnet (I1) (<i>data-infrastructuur, datadiensten</i>).
Lichtpaden worden ingezet om dezelfde werkomgeving (desktop) te hebben op verschillende locaties. Lichtpaden zijn ingericht voor TIASNimbas en FKT ⁷ in Utrecht (I3) (<i>workflow, toegang, data-infrastructuur</i>).
We hebben een kluis met dvd's waarop de data staan. De databestanden worden steeds groter en worden nu opgeslagen op externe hard disks (I1) (<i>opslag</i>).
Rond 1990 is een nieuwe hersengolf ontdekt door een onderzoeker die geen geld had om nieuwe data te verzamelen. Normaal gesproken kijk je in ons vak kijk je naar de goede trials; deze onderzoeker heeft analyses uitgevoerd op de foute trials en een ontdekking gedaan dat een heel nieuw onderzoeksgebied heeft gestimuleerd (I1) (<i>workflow</i>).
Tussen de ruwe data die bij een experiment op disk komen en de data die statistisch geanalyseerd kunnen worden liggen veel stappen; je moet data filteren, je moet er algoritmes op los laten, je moet componenten extraheren. Op het moment dat je van je ruwe data tot een score bent gekomen ben je los van je eigen domein. Vanaf dat moment maakt het niet uit of je een hersengolfje hebt, of een vragenlijst (I1) (<i>workflow</i>).
Ook voor domein-specifiek onderzoek zijn componenten te onderscheiden waar vakken elkaar kunnen helpen (I1) (<i>datadiensten</i>).
Medische psychologie en neuro psychologie liggen niet zo heel ver uit elkaar maar die vakken hebben toch wel andere eisen wat betreft data opslag en dat soort dingen (I1) (<i>opslag, databeheer</i>).
Er zijn minder beperkingen aan de data-infrastructuur dan voorheen; nu begin je een onderzoek, en later kijk je eens van waar moet ik dan die data laten; dan koop je een keer een diskje (I1) (<i>opslag</i>).
Beperkingen voor het delen van data hangen ook samen met het feit dat de resultaten sterk afhankelijk zijn van precieze manipulaties en keuzes die de onderzoeker heeft gemaakt (I1) (<i>voorwaarden, delen van data</i>).
De gebruikers zijn van mening dat wanneer ze EDACwowe gebruikt hebben, dat ze alles hebben gezien wat er beschikbaar is, en als we het niet hebben kunnen vinden dan hebben we het idee dat het er dus niet is. Dat is een groot voordeel ten opzichte van Google, want dan weet je nooit of er niet meer is, of je niet iets gemist hebt (I2) (<i>data- diensten, data-infrastructuur, volledigheid</i>).
We proberen data zo toegankelijk mogelijk te maken, door per databron een standaard informatieformat te maken zodat de gebruiker per bron precies kan zien wat de inhoud is, dus welke landen deelnemen, welke jaren het betreft, waar ik toegang heb tot publicaties, webcontent etc. We hebben er uiteindelijk anderhalf jaar over gedaan voordat we volledig de lucht in gingen. We wilden eerst overeenstemming over wat per databron vermeld moest worden (I2) (<i>datakwaliteit, data-infrastructuur, datadiensten</i>).
Het zijn vooral secundaire data die verzameld worden. Een typisch voorbeeld is dat onderzoekers zeggen, we hebben data nodig voor die landen op deze kenmerken, etc. Wij proberen dan voor zoveel mogelijk Europese landen de data in een datamatrix te krijgen (I2) (<i>workflow, hergebruik, vindbaarheid</i>).
Een gevaar van standaard datasets is dat sommige collega's met die dataset gaan werken, omdat het zo gemakkelijk is. En zich daardoor sterk beperken in de vragen die ze kunnen stellen (I2) (<i>workflow, beschikbaarheid</i>).

2.4 Databeheer na afronding van onderzoek

2.4.1 Situatie bij het NIOO-KNAW

Het NIOO hecht eraan dat bij afronding van een wetenschappelijk onderzoek alle relevante gegevens integraal worden overgedragen aan de afdelingshoofden. Dit standpunt geldt nadrukkelijk ook voor promovendi. Verder vindt het NIOO het belangrijk dat onderzoeksdata gekoppeld worden aan de publicatie; de publicatie is daarbij leidend.

⁷ FKT: Faculteit Katholieke Theologie.

Tabel 7: Databeheer in de praktijk bij het NIOO-KNAW (na afronding onderzoek)

Een zeer groot deel van de datasets die worden gegenereerd zijn dan ook niet of nauwelijks bekend buiten de kring van mensen die onmiddellijk bij de opzet en het gebruik van de databestanden betrokken zijn (<i>vindbaarheid, bekendheid</i>).
Vraag is of metadata moeten worden opgeslagen in een eigen systeem voor toelevering aan nationale en internationale instanties (<i>procedures, richtlijnen, metadata management, opslag</i>).
De informatie over de data moest worden afgeleid uit de overeenkomstige publicatie, maar dit gaf problemen bij data die niet rechtstreeks in de literatuur gebruikt (<i>data kwaliteit: toegankelijkheid, documentatie</i>).
Een wetenschappelijke bibliotheek heeft ook een rol in het registreren van de 'output', te weten de eigen onderzoekspublicaties. Het lijkt logisch om die registratie uit te breiden naar de onderzoeksdata (<i>datamanagement: opslag, archivering</i>).
De rol van de 'data librarian' kan alleen maar op een algemeen en organisatorisch niveau uitgevoerd worden, d.w.z. het benadrukken van het belang van datamanagement, en het omgaan met digitale bestanden en formulieren (<i>functieprofiel data librarian</i>).
De bibliotheek is vanouds bekend met ontsluitingssystemen en metadatasystemen. Toegankelijk maken van informatie is een standaard functiebeschrijving in de bibliotheek. Evenals het assisteren bij het terugzoeken van opgeslagen informatie. Als een verlengde van die traditionele rol wordt de bibliotheek bij datamanagement betrokken (<i>toegang en data-kwaliteit</i>).
Bij het uitvoeren van goed onderzoek en het afleveren van een goede publicatie hoort dat de data goed gedocumenteerd is. De overige geïnterviewden (inclusief de interviewer) onderschrijven die uitspraak. Wel is het belangrijk dat de onderzoeker weet wat er van hem verwacht wordt en daarvoor ook de geëigende middelen krijgt (in geld en faciliteiten). (<i>procedures, richtlijnen en voorwaarden</i>).
Tijdens de workshops is gediscussieerd over het gebruik van standaard templates voor dataverzameling. Een senior onderzoeker stelt: "Nu pas na 11 jaar weet ik wat ik van een AIO moet vragen bij dit soort onderzoek". Hij stelt dat het niet mogelijk is om vooraf al te gaan omlijnen wat je vastlegt (<i>procedures, richtlijnen</i>).
Voor de andere medewerkers is het soms erg lastig als gegevens niet goed beschreven zijn. Ik ben er zelfs voorstander van om dit als onderwerp op te nemen in functioneringsgesprekken. Dan wordt goed duidelijk dat documentatie van onderzoeksdata een verantwoordelijkheid is van de onderzoeker en een integraal onderdeel vormt van zijn onderzoeksproject (<i>procedures, voorwaarden</i>).
Een persoon schetst een nationale situatie waarbij een instelling als bijv. SURF zorg draagt voor de ICT-eisen en een centrale opslag die in principe interdisciplinair moet zijn met een helpdesk en een contactpersoon op het onderzoeksinstituut. Die contactpersoon moet een mandaat hebben en zich bezig houden met nacontrole en dienen als vraagbaak voor onderzoekers. (<i>data-infrastructuur</i>).
Het is van belang dat een financier duidelijk te kennen geeft wat hij vindt dat er met de onderzoeksgegevens moet gebeuren. Bij alle onderzoeken die gefinancierd zijn uit de subsidies voor het International Polar Year is uitdrukkelijk geëist dat de gegevens worden opgeslagen en toegankelijk gemaakt (<i>beleid, voorwaarden, procedures</i>).
Promovendi leverden traditioneel bij het afronden van hun dissertatie de gedrukte dissertatie in vijfvoud aan de bibliotheek inclusief een Cd-rom met hun onderzoeksgegevens. Nu willen we, behalve hun dissertatie (bij voorkeur ook digitaal), ook dat de onderzoeksgegevens, goed gedocumenteerd worden overgedragen (<i>procedures</i>).

2.4.2 Situatie bij TU Delft

Eén van de kerntaken van TU Delft Library is het registreren van de wetenschappelijke output van de TU Delft. Behalve wetenschappelijke publicaties vallen hier ook datasets onder. Volgens afspraak leveren promovendi hun dissertatie digitaal aan de TU Delft Library. De Library stimuleert dat ook de bijbehorende datasets worden aangeleverd.

Tabel 8: Databeheer in de praktijk bij TU Delft (na afronding onderzoek)

Zolang het onderzoek niet is afgerond, of er nog niet over is gepubliceerd, is men van mening dat de data niet of beperkt toegankelijk moeten zijn voor anderen. Eventueel onder embargo. Wanneer in het project wordt samengewerkt met commerciële partners gelden er strikte regels m.b.t. het openbaar maken van de data (<i>delen van data, voorwaarden</i>).
Als belangrijk voordeel van het openbaar maken van onderzoeksdata wordt de toename van het aantal citaties gezien (<i>citatie van data, datadiensten</i>).

Onderzoekers zouden meer gestimuleerd kunnen worden hun data beschikbaar te stellen door daar een vorm van beloning tegenover te stellen, bijv. financieel of puntensysteem (*voorwaarden, prikkels*).

Hergebruik van data kan ook een financieel voordeel opleveren. Immers, men hoeft het onderzoek niet zelf uit te voeren, en kan door met de resultaten van de ander (*delen van data, efficiency*).

Hergebruik van data wordt gestimuleerd door in de publicatie te refereren naar de data (*citatie van data, datadiensten*).

Ook op internationaal niveau zijn data van belang. "Binnen Nederland wordt er relatief weinig hydrologisch onderzoek gedaan. Data van andere internationale instituten zijn ook slecht toegankelijk" (*datakwaliteit: toegankelijkheid, bruikbaarheid*).

2.4.3 Situatie bij Tilburg University

Tabel 9: Databeheer in de praktijk bij Tilburg University (na afronding onderzoek)

Het delen van stimulussets is een groot goed met een sterke wetenschapspolitieke lading (I1) (*cultuur*).

Databestanden worden zelden of niet hergebruikt en komen dus in een soort donker archief terecht. Er is steeds minder tijd om eens te kijken naar een oude dataset (I1) (*cultuur, workflow*).

Studenten worden nog wel eens ingezet om voor hun bachelor onderzoek nog eens naar een stuk data kijken (I1) (*workflow*).

Masterstudenten moeten nieuw en origineel onderzoek doen; oude data zijn dan niet bruikbaar (I1) (*workflow*).

[Over het aggregeren van gegevens]. Je ziet dat vaker in de biologische hoek; hoe 'biologischer' het wordt hoe meer data (vgl. genetica waar centrale databases gebruikt worden) (I1) (*data-infrastructuur, datadiensten*).

[Over het delen van eigen onderzoeksgegevens] Ik heb er op zich niets op tegen, laat daar geen misverstand over ontstaan. Maar ik zie er relatief weinig in omdat de data nogal specifiek is. Ik zou er geen problemen mee hebben om bruikbare data openbaar te maken, voor iedereen om er mee te werken, wanneer ik zelf uitgepubliceerd met die data (I1) (*voorwaarden, delen van data*).

Het zijn nu vooral Europese projecten waarin vastligt dat data bewaard toegankelijk en evt. gedeeld of openbaar moeten worden (vgl. Tango-project) (I1) (*beleid*).

Het onderhoud van de portal gebeurt door een coördinator van het datacentrum (EDACwowe) die twee studentassistenten aanstuurt. In totaal is met de portal de inzet van 1 FTE gemoeid (I2) (*data disseminatie*).

Ik zoek naar bijvoorbeeld naar werkeloosheidscijfers, ik vind dan twee of drie bronnen, vertel mij dan eens welke cijfers het meest betrouwbaar, of het meest valide zijn? (*data kwaliteit, provenance, data-diensten*).

[Over een kwaliteitsoordeel over onderzoeksdata]. Maar dat kan ik niet, want dat is helemaal afhankelijk van jouw onderzoeksvraag (I2) (*procedures, datakwaliteit*).

2.5 Samenvatting databeheer in de praktijk

NIOO

Vanuit de directie van het NIOO is altijd het standpunt gehuldigd dat alleen een onderzoeker zijn/haar eigen data kan beheren. Ook is een onderzoeker zelf verantwoordelijk voor de documentatie van de onderzoeksdata. Een dataprofessional heeft dus een faciliterende rol. De competenties voor een goed ondersteuner zijn als volgt verwoord: goed kunnen omgaan met iedereen, met assistenten, onderzoekers en PhD's, niet bedreigend maar wel voor elkaar krijgen wat noodzakelijk is, kunnen overtuigen. De nadruk moet liggen op de verbindende schakel vormen tussen ICT en onderzoek. De dataprofessional moet een mandaat hebben, zich bezig houden met nacontrole en dienen als vraagbaak. Bij het NIOO-KNAW gaan gedachten op om in de nabije toekomst afspraken te maken over eigendomsrechten van onderzoeksdata en periodiek voorlichting te geven over dit onderwerp. In aanvulling hierop bestaan plannen om in het arbeidscontract vast te leggen dat de onderzoeker gehouden is onderzoeksdata data goed gedocumenteerd over te dragen aan een leidinggevende. Afspraken over onderzoeksdata in een arbeidscontract bieden een houvast voor leidinggevendenden bij functioneringsgesprekken om de voortgang en de kwaliteit van het werk te toetsen.

Hoofdconclusie is dat datamanagement nog niet als structureel onderdeel is opgenomen in het werk van de onderzoeker. Onduidelijkheid over de juridische status (wie is eigenaar van de data?) en de manier waarop de onderzoeker, het management, de KNAW en de hele wetenschappelijke wereld aankijken tegen data management zijn hierbij van belang. Als er geen ruimte in geld en tijd wordt opgenomen bij de financiering van een onderzoeksproject wordt het erg moeilijk voor de altijd in tijdnood verkerende onderzoeker om extra tijd in te ruimen voor het goed documenteren en archiveren van zijn onderzoeksdata.

TU Delft

Bij de TU Delft zijn geen richtlijnen beschikbaar hoe om te gaan met de uit het onderzoek verkregen data. Iedere groep volgt zijn eigen weg. De onderzoeker wil graag scholing voor datamanagement ontvangen en wil ook ondersteund worden. Ook tijdens het onderzoek wordt door de onderzoeker eigen inbreng gevraagd. Er is wel behoefte aan ondersteuning, met name bij controle op metadata en datamanagement. Data worden pas na het onderzoek gepubliceerd en veelal bij de eigen onderzoeksgroep opgeslagen. Publiceren van data kan gestimuleerd worden door een vorm van beloning. De TU Delft stimuleert het Open Access beschikbaar stellen van onderzoeksresultaten.

Tilburg University

Databestanden die door onderzoekers worden gegeneerd, worden lokaal bij de faculteit opgeslagen. Er zijn grote verschillen tussen vakgroepen hoe met data wordt omgegaan en het belang dat wordt toegekend aan het beschikbaar houden van data voor hergebruik. Soms stelt een tijdschrift als eis dat data en programmacode worden opgestuurd. Voor onderzoek dat gepubliceerd wordt in APA tijdschriften gelden richtlijnen voor het beschikbaar houden van data. Er zijn geen cijfers voorhanden waaraan de naleving van de richtlijnen kan worden getoetst. Er is zeker interesse om meer aandacht te vragen voor dataopslag met het oog op hergebruik. Financiering van de activiteiten en infrastructuur is een voorwaarde om databeheer structureel in te bedden. Binnen Library and IT services (LIS) is de afgelopen jaren ervaring opgedaan met het verrijken van publicaties met onderzoeksdata. Het verzamelen en beschikbaar maken van andere bronnen van wetenschappelijke output zoals datasets is een vaste activiteit geworden en informatiespecialisten worden getraind in het collectioneren en beschrijven van datasets en aanvullende materialen.

Tilburg University heeft een sterke empirische traditie en huisvest ook enkele datacentra. CentERData geldt internationaal als een best practice voor documentatie en disseminatie van onderzoeksdata. Ze hebben een niveau bereikt dat alleen mogelijk is door substantieel te investeren in de documentatie van data. Datacentrum EDACwowe is opgezet voor en door onderzoekers en sluit daarom naadloos op de behoefte van onderzoekers. De financiering van het datacentrum is tijdelijk waardoor onzekerheid bestaat over de continuïteit van het centrum. European Values Study (EVS) is eveneens een bekende studie waarbij de data opgeslagen liggen in een archief in Keulen. EVS is een goed voorbeeld hoe gedetailleerde informatie over de data en de operationalisaties bij een publicatie kunnen worden aangeboden.

3 Aanbevelingen

Een van de doelstellingen van dit witboek is aanbevelingen te formuleren voor de ondersteuning van wetenschappelijk databeheer op basis van gesprekken met onderzoekers, ervaringen in de landen om ons heen en literatuurstudie. De aanbevelingen zijn daarom als eerste gericht op de kennis en ervaring die de dataprofessional in wording nodig heeft om in een ondersteunende functie te kunnen werken. Maar het witboek is ook bedoeld om aanbevelingen te geven voor een cursus en scholingsaanbod dat studenten en jonge onderzoekers van nu stimuleert om onderzoeksdata toegankelijk te maken en te delen met anderen. Tot slot proberen we met dit witboek ook door de ogen van ervaren onderzoekers de kwetsbaarheden, knelpunten en verbeterpunten voor wetenschappelijk databeheer bloot te leggen.

In de volgende paragraaf geven we aanbevelingen voor het inrichten van een cursus- en scholingsaanbod voor wetenschappelijke opleidingen. In paragraaf 3.2 bespreken we daarna enkele praktische handvatten die specifiek van belang zijn bij de opzet van een cursus- en trainingsaanbod voor databeheer. In paragraaf 3.3 noemen we tot slot enkele inhoudelijke hoofdlijnen voor het invullen van een cursus- en scholingsaanbod voor wetenschappelijk databeheer.

3.1 Aanbevelingen voor training en opleiding in wetenschappelijke curricula

De digitale en data-intensieve wetenschap is zowel een bron van zorg als een bron van vernieuwing en innovatie. Hetzelfde kan gezegd worden over het wetenschappelijk beheer van onderzoeksdata. Bewustwording is nodig bij onderzoekers om onderzoeksdata toegankelijk te maken, te delen en te waarderen als een zelfstandige wetenschappelijke opbrengst. Tegelijkertijd zal meer systematische aandacht voor databeheer ook bijdragen aan innovatie van wetenschappelijk onderzoek.

Uit de interviews die voor dit witboek zijn gehouden en de literatuur wordt duidelijk dat onderzoekers ondersteuning van onderzoek dicht bij huis willen organiseren [48]. Daarbij wordt als eerste gedacht aan een directe lijn naar ICT- en technische ondersteuning. De vraag naar inhoudelijke ondersteuning, voor bijvoorbeeld databeheer, wordt niet genoemd. Dit is ook niet verwonderlijk omdat een inhoudelijk aanbod voor onderzoeksondersteuning ontbreekt.

Voor de leden van de werkgroep is duidelijk dat een training- en scholingsaanbod voor ondersteuning en innovatie van wetenschappelijk databeheer grenst aan een braakliggend terrein. Door aandacht te vragen voor opleidingen en cursussen kan dit braakliggend terrein bouwrijp worden gemaakt. Volgens de werkgroep zijn de volgende aanbevelingen daarbij van belang:

1. In nagenoeg alle bachelor opleidingen zou een introductiecursus data-science gegeven moeten worden in combinatie met een skills werkgroep. Bij de cursus en de werkgroep dienen de belangrijkste recente ontwikkelingen rondom data-intensieve wetenschap aan de orde te komen, evenals uitgangspunten voor goed data beheer, een typering van de vaardigheden en competenties en inzicht in mogelijke specialisaties.
2. Het is van belang om minimaal één masteropleiding in Nederland te starten naar het voorbeeld van Syracuse University [15] waarbij diepgaand aandacht wordt besteed aan de kennis en vaardigheden die nodig zijn voor de ondersteuning van data-intensieve wetenschap en geavanceerde toepassingen van databeheer. Aanknopingspunten voor het uitwerken van curricula voor specialistische dataprofessionals die inzetbaar zijn voor onderzoeksondersteuning zijn te vinden bij Liddy [15] en Madnick et al. [30, 31].
3. Databeheer en digitaal cureren verdient ook nadrukkelijk aandacht bij de GO-opleiding en IDM-opleiding in samenhang met de competenties en vaardigheden waarover een professional dient te beschikken. Verder is het van belang dat een aanbod wordt ontwikkeld waar praktijkstages gevolgd kunnen worden. Het inrichten van een werkgroep die de mogelijkheden voor een opleidingsaanbod voor dataprofessionals verkent, wordt eveneens aanbevolen.

4. Het is daarnaast wenselijk een module datamanagement te ontwikkelen die kan worden ingepast in een algemeen programma voor academische informatievaardigheden en die toegespitst kan worden op masterstudenten en promovendi. Een module dient minimaal enkele discipline specifieke thema's te bevatten.
5. Verder is het wellicht van belang een aanbod voor nascholing en trainingen te ontwerpen voor informatieprofessionals die met onderzoeksdata en databeheer te maken krijgen. Ook dit is een nagenoeg onontgonnen terrein in Nederland. Het is belangrijk dat de kennis die in Nederland op dit terrein al is opgedaan, wordt gedeeld en uitgewisseld.
6. Een haalbaarheidsstudie valt te overwegen om binnen de opleiding tot informatieprofessional een specialisatie in te richten voor *research information support*, met als belangrijkste hoofdlijnen: methoden en technieken van onderzoek, *open access* en *open data*, *data curation* en *repositories*.
7. Een cursusaanbod is wenselijk voor data- en informatieprofessionals die trainingen kunnen verzorgen voor onderzoekers.

3.2 Praktische handvatten

Datamanagement vs. ICT-vaardigheden

Als we het aanbod aan cursussen en workshops bekijken valt op dat ze vaak bestaan uit een (onduidelijke) mengeling van algemene vaardigheden voor databeheer en ICT-vaardigheden. De werkgroep pleit ervoor om algemene vaardigheden die bijdragen aan blijvende toegankelijkheid van onderzoeksdata duidelijk te scheiden van (specifieke) informatie- en ICT-vaardigheden die van belang zijn om data te kunnen verwerken en analyseren, zoals een cursus Excel of Access.⁸ Dit laatste hoeft overigens niet te betekenen dat beide onderwerpen per definitie niet in dezelfde cursus aan de orde kunnen komen. Noordegraaf laat bijvoorbeeld zien hoe met een modulaire opzet een training voor databeheer kan worden opgezet die bestaat uit een evenwichtige combinatie van algemene - en specifieke vaardigheden [35].

Planning van een training of cursus

1. Het belang van de juiste *doelgroep*. Omdat bachelor- en masterstudenten, promovendi en onderzoekers een verschillende benadering vereisen is het van belang dat een training- en scholingsaanbod op de juiste doelgroep wordt afgestemd.
2. Het belang van het juiste *moment* voor het aanbieden van een training of cursus. Het is belangrijk om - bijvoorbeeld in overleg met onderwijsdirecteuren - vast te stellen wanneer een cursus of training het best kan worden aangeboden, zodat het aanbod optimaal past binnen een curriculum.
3. Het belang van het juiste *formaat* waarin de training, cursus of scholing wordt aangeboden. Een training, cursus of workshop kan op tal van manieren worden ingevuld. DataONE, het Data Observation Network for Earth, biedt bijvoorbeeld een Summer Internship Opportunity en de Universiteit van Edinburgh kent ochtend- en middagseminars, eendaagse cursussen en 'smaaksessies informatievaardigheden'.⁹ Er is geen eenduidige richtlijn te geven voor het best passende formaat voor een cursus of training. Een zorgvuldige afstemming op de praktijk en de doelgroep is belangrijk om een betekenisvolle en duurzame relatie aan te gaan.

Academische vaardigheden

Universiteiten bieden doorgaans voor elke discipline een cursusaanbod van informatie- en academische vaardigheden. Een cursus, training of module voor databeheer kan naar onze mening aansluiten op dit aanbod. Dataprofessionals kunnen natuurlijk worden ingezet om dit aanbod in een instelling vorm en inhoud te geven.

⁸ Edinburgh University plaatst bijvoorbeeld cursussen MS Access en - Excel onder data management en informatievaardigheden. Maar ook de workshop datamanagement van het VLIZ voor het NIOO uit 2008 was vooral gericht op het werken met Access en Excel.

⁹ Bijlage 4 geeft een overzicht van cursussen en trainingen die beschikbaar zijn.

3.3 De invulling van trainingen en cursussen

Hieronder noemen we enkele onderwerpen die als uitgangspunt kunnen dienen voor de inhoudelijke opzet van een datamanagement training of cursus. Aanknopingspunten voor de opleiding en training van *data librarians* zijn te vinden bij Grim en Noordegraaf [35, 49, 50]. De volgende onderwerpen zijn van belang ongeacht de wetenschapsdiscipline.

1. Organisatorische aspecten van *data governance* [51].¹⁰ Hiertoe rekenen we het opzetten van een databeleid of het stimuleren van *good practices* voor databeheer. Maar ook het bijhouden van data-incidenten, bij gebrek aan een beleid voor databeheer, is een onderwerp dat aandacht verdient.
2. Cureren van databestanden. Een vast onderdeel van een cursus databeheer is het beschrijven en documenteren van een databestand of een collectie van bestanden. Een training kan bijvoorbeeld ingaan op het gedetailleerd beschrijven van een bestand volgens discipline specifieke richtlijnen [52].
3. Schrijven van een datamanagement plan en een subsidieaanvraag waarvoor een dataparagraaf vereist is.
4. Een theoretische module waarin aandacht besteedt wordt aan metadata voor onderzoeksdata en de kennis van de (lokale) data-infrastructuur voor onderzoeksdata. Na een training zou een cursist bijvoorbeeld in staat moeten zijn om een beschrijving van de lokale data-infrastructuur te geven.
5. Aandacht voor *soft skills*. Overtuigingskracht, onderhandelingsvaardigheden, en 'acquisitie vaardigheden' zijn bijvoorbeeld van belang voor dataprofessionals die in gesprek gaan met onderzoekers met eigen opvattingen over databeheer.

¹⁰ Met data governance wordt hier bedoeld het aansturen van processen, techniek en mensen.

Lijst van afkortingen

CARDS	Controlled Access Remote Data Storage (http://www.surffoundation.nl/nl/projecten/Pages/CARDS.aspx)
DANS	Data Archiving Networked Services (http://www.dans.knaw.nl/)
DCC	Digital Curation Centre (http://www.dcc.ac.uk/)
KNAW	Koninklijke Nederlandse Academie voor Wetenschappen (http://www.knaw.nl/)
ODAP	Open Data and Publications (http://www.surffoundation.nl/nl/projecten/Pages/OpenDataandPublications.aspx)
LIS	Library and IT Services - Universiteit Tilburg (http://www.tilburguniversity.edu/nl/over-tilburg-university/bibliotheek/)
NEEO	Network of Economists Online (http://www.nereus4economics.info/projectneeo.html)
NIOO-KNAW	Nederlands Instituut Onderzoek der Zee (http://www.nioo.knaw.nl/)
SARA	Stichting Academisch Rekencentrum Amsterdam (http://www.sara.nl/)
VLIZ	Vlaams Instituut voor Onderzoek der Zee (http://www.vliz.be/)

Referenties

1. Borgman, C., L., *Data, Disciplines and Scholarly Publishing*. Learned Publishing, 2008. 21(1): p. 29-38.
2. EC, *i2010- Digital Libraries*, D.I.S.a. Media, Editor. 2008, EC.
3. Owen, J.M., *The scientific article in the age of digitization*. 2005, UvA: Amsterdam.
4. Shuichi Iwata, et al., eds. *Communications and Discoveries from Multidisciplinary Data*. Studies in Computational Intelligence, ed. J. Kacprzyk. Vol. 123. 2008, Springer: Berlin.
5. SDMX. *Statistical Data and Metadata eXchange*. 2010; Available from: <http://sdmx.org/>.
6. ESFRI. *European Roadmap on Research Infrastructures*. 2006; Available from: <http://cordis.europa.eu/esfri/roadmap.htm>.
7. DCMi, *DCMi*.
8. *e-Overheid voor burgers*. Available from: <http://www.e-overheidvoorburgers.nl/standaarden,metadata/Over-metadata.html>.
9. DDI. *Data Documentation Initiative*. Available from: <http://www.ddialliance.org/>.
10. NCDD. *Nationale Coalitie Digitale Duurzaamheid*. Available from: <http://www.ncdd.nl/en/index.php>.
11. Loriga, B. *How to nurture Data Scientists*. 2010; Available from: <http://practicalquant.blogspot.com/2010/07/how-to-nurture-data-scientists.html>.
12. *The Biological Data Scientist*. Available from: <http://mndoci.com/2010/06/22/the-biological-data-scientist/>.
13. Granville, V. *Data Week: Becoming a data scientist*. 2010; Available from: <http://www.analyticbridge.com/profiles/blogs/data-week-becoming-a-data>.
14. Swan, A. and S. Brown. *The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs*. 2008.
15. Liddy, E.D. *Designing Curriculum to Educate Digital Library Professionals in iSchools*. [ppt] 2010; Available from: http://webcache.googleusercontent.com/search?q=cache:PvrNNoEx3GcJ:www.ffos.hr/lida/dato+teke/lida_2009_liddy.ppt+E+Liddy+designing+curriculum&cd=3&hl=en&ct=clnk.
16. RDSWG. *Stewardship of Research Data in Canada: A Gap Analysis*. [Report] 2008.
17. RIN, *Stewardship of digital research data: a framework of principles and guidelines*. 2008.
18. Macdonald, S. and L. Martinez. *e-Research and the Data Librarian*. In Support for e-Research: Filling the Library Skills Gap [ppt] 2007; Available from: http://www.nesc.ac.uk/talks/770/Stuart_MacDonald_NeSC_Jun_07_whole.ppt.
19. OCLC, *A Slice of Research Life: Information Support for Research in the United States*. 2010.
20. Dumbill, E. *Becoming a data scientist*. Data week 2010; Available from: <http://radar.oreilly.com/2010/09/data-week-becoming-a-data-scie.html>.
21. FlowingData. *Rise of the Data Scientist*. 2009; Available from: <http://flowingdata.com/2009/06/04/rise-of-the-data-scientist/> en <http://flowingdata.com/>
22. Foursquare. *Data Scientist at Foursquare in New York, NY*. 2010; Available from: <http://foursquare.jobscore.com/jobs/foursquare/data-scientist/bZUOhKbTGr4k1UeJe4bk1X?sid=312&name=Twitter&rid=cZizBObVur4kgweJe4bk1X>.
23. Loukides., M. *What is data science? Analysis: The future belongs to the companies and people that turn data into products*. 2010; Available from: <http://radar.oreilly.com/2010/06/what-is-data-science.html>.
24. *Data, data, everywhere: A special report on managing information*. The Economist, 2010(February).
25. Vlieg, J.D., *Wetenschappelijke doorbraken versnellen met ICT*. SURF, 2011(Maart): p. 6-8.
26. DCC. *Digital Curation Centre (DCC)*. Available from: <http://www.dcc.ac.uk/>.
27. DigiCurve. Available from: <http://www.dpconline.org/newsroom/latest-news/676-dpc-and-digicurve-a-new-framework-for-vocational-education>.
28. *DCC training material*. Available from: <http://www.dcc.ac.uk/training/data-management-courses-and-training>.
29. MIT Libraries. Available from: <http://libraries.mit.edu/guides/subjects/data-management/Managing%20Research%20Data%20101.pdf>.
30. Qin, J. and J. D'Ignazioy, *Lessons learned from a two-year experience in science data literacy education*. 2010.

31. D'Ignazio, J., et al. *Access to scientific data: The Social and Technical Challenges and strategies*. in *Proceedings of the American Society for Information Science and Technology*. 2006.
32. Corral, S. *Research Data Management: Professional Education and Training Perspectives*. Paper presented at the RDMF2: Roles and Responsibilities for Effective Data Management 2008 March 18 2011]; Available from: <http://www.dcc.ac.uk/sites/default/files/documents/RDMF/RDMF2/07%20Corral.pdf>
33. Loshin, D., *Master Data Management*. 2009, Elsevier Morgan Kaufmann OMG Press.
34. SDL. *The Science Data Literacy Project*. Available from: http://sdl.syr.edu/?page_id=19.
35. Marina Noordegraaf, *Data intelligence for librarians* 3TU.Datacentrum, Editor. 2011.
36. Altman, M. *Data Management Plans: Approaches to Dissemination and Preservation*. IASSIST [ppt] 2007.
37. Austin, R.D. and R.L. Nolan, *Bridging the gap between stewards en creators*. MIT Sloan Management Review., 2007. 48(2): p. 29-36.
38. DCC. *RDMF2: Roles and Responsibilities for Effective Data Management*. 2008 [cited 2011; Available from: <http://www.dcc.ac.uk/events/research-data-management-forum/roles-and-responsibilities>.
39. Swan, A. *Research Data Skills*. [pdf] 2008 [cited 2011; Available from: <http://www.dcc.ac.uk/sites/default/files/documents/RDMF/RDMF2/02Swan.pdf>.
40. FIRST. *Research Data Management Online Workshop*. [cited 2011 March 18]; This online workshop is part of the University of Minnesota's Fostering Integrity in Research, Scholarship, and Teaching (FIRST) program. It is one of four online workshops that comprise the Responsible Conduct of Research (RCR), part 2.]. Available from: http://cflegacy.research.umn.edu/datamgtq1/IN_020.html.
41. Madnick, S. and H. Zhu, *Improving Data Quality Through Effective Use of Data Semantics*. *Data Knowl. Eng.*, 2006. 59(2): p. 460-475.
42. Madnick, S.E. and Y.W. Lee, *Editorial Letter for the Special Issue on Data Quality in Databases and Information Systems*. *J. Data and Information Quality*, 2009: p. 1-2.
43. IMF, *Kingdom of the Netherlands, in Report on the observance of standards and codes (ROSC)-data module*, IMF, Editor. 2008.
44. Hey, A.J.G., S. Tansley, and K. Tolle, *The Fourth Paradigm: Data-intensive Scientific Discovery.*, A.J.G. Hey, S. Tansley, and K. Tolle, Editors. 2009, Microsoft Research.
45. Hammerbacher, J. (2010) *Information Platforms and the Rise of the Data Scientist*. Scribd.
46. Hilary, M. and W. Chris. *A taxonomy of data science*. 2010; Available from: <http://www.dataists.com/2010/09/a-taxonomy-of-data-science/>.
47. EDACwowe. Available from: <http://www.edacwowe.eu/>.
48. Maccoll, J., *Supporting research: environments, administration and libraries*. 2011, OCLC.
49. Grim, R., *Data Librarian: een nieuw informatiespecialisme?* *Informatie Professional*, 2010. 14(11): p. 36-38.
50. Grim, R. *Data-Bibliothecaris: tijd voor een nieuw informatiespecialisme? Ervaringen uit de praktijk*. Presentatie voor de landelijke werkgroep Sociaal-Wetenschappelijke informatie [ppt] 2010 20 Oktober; Available from: <http://slidesha.re/klHc6J>.
51. Tarantino, A. and D. Cernauskas, *Risk management in finance*, in *Six Sigma and Other Next-Generation Techniques*. 2009, Wiley Finance.
52. Lacroix, Z. and T. Critchlow, *Bioinformatics: Managing Scientific Data*. The Morgan Kaufmann series in multimedia information and systems. 2003, San Francisco, CA [etc.]: Morgan Kaufmann. XXI, 441 p.
53. DataOne. *2011 Summer Internship Program*. Available from: <http://www.dataone.org/content/2011-summer-internship-program>.
54. Grim, R. and M.v.d. Berg, *Voorstel toevoeging variant onderzoeksdata aan UFO-profiel Informatie-/Collectiespecialist*. 2011.

Bijlage 1 - Functieomschrijvingen van dataprofessionals

In deze bijlage tonen we een aantal voorbeelden van functies en profielen van dataspecialisten die zijn toegespitst op het beheer van onderzoeksdata en het ondersteunen van onderzoek en onderwijs. Swan [14] onderscheidt bijvoorbeeld vier functies voor onderzoeksdata specialisten: de *data creator*, de *data scientist*, de *data manager* en de *data librarian*, die als volgt worden omschreven:

- Data Creator: researchers with domain expertise who produce data. These people may have a high level of expertise in handling, manipulating and using data.
- Data Scientist: people who work where the research is carried out – or, in the case of data centre personnel, in close collaboration with the creators of the data – and may be involved in creative enquiry and analysis, enabling others to work with digital data, and developments in data base technology.
- Data Manager: computer scientists, information technologists or information scientists and who take responsibility for computing facilities, storage, continuing access and preservation of data.
- Data Librarian: people originating from the library community, trained and specialising in the curation, preservation and archiving of data.

In aanvulling op de typering van Swan tonen we ook de functies die door Liddy [15] genoemd worden:

Research Data Specialist:

- Key role in e-science, digital data curation & supporting cyber-scholarship for all STEM¹¹ disciplines.
- Staying abreast of scientific data trends, data documentation tools & standards for data exchange, re-use & interoperability
- Utilizing information technology tools for metadata manipulation & script execution
- Consultation, assessment & support services to facilitate all aspects of digital data curation.
- MLS or Masters degree in science
- 1 year research lab experience
- Strong communication, inter-personal & communication skills
- Aptitude for & consistency in detailed work
- Ability to analyze & solve problems creatively & flexibly in a complex & rapidly changing environment
- Strong service orientation & interest in users' values & needs

Science Data Services Librarian:

Responsibilities:

- Work with primary research community & other librarians to develop & sustain services for accessing & analyzing research data with focus on bio-informatics & chemo-informatics.
- Act as liaison between campus & library initiatives and the research community to promote, facilitate, and support the development of services for collecting & archiving research data.
- Maintain awareness of tools & methodologies for computationally centered, data-driven science
- MLIS from ALA-accredited program
- Excellent communication skills
- Superior organizational & analytical skills
- Life sciences subject background
- Experience with computer & telecommunications technologies for information management

¹¹ STEM: Science Technology Engineering, Medicine.

Bijlage 2 - Voorstel toevoeging variant onderzoeksdata aan UFO-profiel Informatie-/Collectiespecialist

Hieronder is ter illustratie het voorstel opgenomen dat is ingebracht in de werkgroep van bibliothecarissen in Nederland die zich onder meer bezighoudt met aanvullingen en wijzigingen van bestaande UFO-profielen van informatie en collectiespecialisten[54]. Met nadruk zij erop gewezen dat het hier voornamelijk gaat om een voorstel [49, 50].

Onderzoeksdata (variant)

Kernactiviteit

- Informeren, adviseren en assisteren van onderzoekers met betrekking tot het creëren, beschrijven, opslaan, vinden, gebruiken en hergebruiken, behandelen, verzorgen en archiveren van resultaten van wetenschappelijk onderzoek in de vorm van digitale gegevens (data en data sets).

Kader

- Doelstellingen wetenschappelijke informatievoorziening en -infrastructuur
- Algemene en disciplinespecifieke onderzoeks- en onderwijsondersteuning
- Nationale afspraken, internationale standaarden

Resultaat

- Wetenschappelijke data van de instelling wordt structureel en toekomstvast beschreven, hervindbaar en herbruikbaar gemaakt.
- Optimale afstemming op - en ondersteuning van het primaire proces.

Activiteiten

- Advisering
- Ondersteuning bij het gebruik van databanken in het onderwijs, zoals bijvoorbeeld het verzorgen van databank-demonstraties tijdens colleges.
- Verzorgen van data lectures waarin een overzicht wordt geboden van de databanken die onder licentie of vrij toegankelijk zijn. De data lecture gaat in op de inhoud, gebruiksmogelijkheden en -voorwaarden voor gebruik van de databanken.
- Inrichten en ondersteunen van een informatiepunt voor onderzoeksdata (afhankelijk van de zwaarte en het aandachtsgebied is dit bijvoorbeeld een datacentrum, een data helpdesk, of een financial data support center).
- Ontwikkelen van trainingen voor het gebruik van databanken.
- Deelname aan projecten
- Deelname aan nationale en internationale samenwerkingsverbanden
- Bijdragen leveren aan de ontwikkeling van de onderzoeksinfrastructuur.
- Verder levert de IC-specialist een bijdrage aan de inhoud en invulling van een nieuw vakgebied binnen de bibliotheek.

Kwalificaties en competenties van een data librarian

- Afgeronde academische opleiding.
- Brede interesse in wetenschappelijk onderzoek, overtuigingskracht, pioniersmentaliteit, generalist en specialist binnen ten minste één discipline.
- Gedegen kennis van onderzoeksmethoden, statistiek en data-analyse technieken.
- Gedegen kennis van software voor dataverwerking (SPSS, Stata, SAS, R, Matlab, Excel), database management software (Access) en dataformaten (NetCDF).
- Kennis van een of meer programmeertalen (JAVA, Perl, Python).
- Kennis van XML-technologieën (XPath, XSLT, XML-Schema), RDF-toepassingen (Linked Data) en UML voor het maken van data modellen.

- Kennis van bibliotheekstandaarden, standaarden voor data-archivering (OAIS) en voor beschrijven van wetenschappelijke data (DDI).
- Kennis van repository software (Dataverse Network, Fedora, EPrints).
- Kennis van Geografische Data Systemen (GIS).
- Elementaire kennis van information retrieval en data-mining technieken.

Bijlage 3 - Functieomschrijving Research Data Specialist

Profile¹²

We are looking for an enthusiastic, pro-active colleague to facilitate and assist in the (re)use of qualitative and quantitative research data, in close cooperation with researchers and research groups.

Important tasks

You encourage easy access to and publication of research data generated at Tilburg University.

You assist and coordinate placing these data sets in scientific context and in archiving them.

You are actively involved in acquisition of and providing access to external data sets.

You play an active role in supporting students and staff using data sets.

You actively contribute to development of data archiving and publication at international level.

Your work is closely related to the scientific research at Tilburg University. A close cooperation between you and the researchers and research groups is therefore mandatory.

Requirements

You have completed an academic degree course. You have several years of research experience, preferably in the social or economic sciences. You have been deeply involved in data acquisition and data analysis. You have proven experience with software packages for statistical analysis, such as SPSS. You are customer-driven, and have excellent oral and written communicative skills, both in Dutch and in English. You have a communicative, cooperative, and service-oriented attitude.

¹² Deze vacature was beschikbaar eind 2006 bij de Universiteit van Tilburg

Bijlage 4 - Voorbeelden van workshops en trainingen

De onderstaande tabel toont de resultaten van een zoekopdracht in Google met de zoektermen: 'research data management', 'curriculum data' en 'data skills'. De tabel is niet opgezet vanuit het oogpunt van volledigheid.

Tabel 1: Overzicht van datamanagement trainingen, opleidingen en cursussen

Bron	Title	Url
Northumbria University	Data management – the skill every researcher needs	http://www.northumbria.ac.uk/browse/ne/uninews/datamanagement
The Skills Library (Jennifer Leonard)	Data management skills	http://www.skillslibrary.com/techskills/datamgmt.htm
RIN	Data management and information literacy	http://www.rin.ac.uk/our-work/researcher-development-and-skills/data-management-and-information-literacy
Ghent University	University Curriculum in Marine Data Management (2011)	http://classroom.oceanteacher.org/course/view.php?id=102
JISC	Research data management training materials (RDMTrain)	http://www.jisc.ac.uk/whatwedo/programmes/~link.aspx?id=677B9C0D0E8F4B12A7E2ACC86FD9D736&z=z
Incremental Project	Passing on data management skills	http://incrementalproject.wordpress.com/2011/03/30/passing-on-data-management-skills/
Google custom search	Research Data Management Training	http://www.google.com/cse/home?cx=000734657432312665985:qb4iquvqvia
International Oceanographic and Data and Information Exchange (IODE), the POGO Centre of Excellence Student Training in Observational Oceanography sponsored by the Japan Foundation	University Curriculum in Marine Data Management (2011).	http://ioc-unesco.org/index.php?option=com_oe&task=viewEventRecord&eventID=891
Statistics for Education: Essential data skills	Essential Data Skills for Business and Management	http://www.ac777.dial.pipex.com/essentialdata/essential_data.htm
National curriculum mathematics (UK)	Handling data	http://curriculum.qcda.gov.uk/key-stages-1-and-2/subjects/mathematics/keystage2/index.aspx
WRC	Spatial / GIS & Business Data Management concerning Water and Environmental industries	http://www.wrcplc.co.uk/default.aspx?item=853
DCC	Data Management Courses and Training (in-career training, university courses with data-related components)	http://www.dcc.ac.uk/training/data-management-courses-and-training
DCC	Skills framework	http://www.dcc.ac.uk/training/data-management-courses-and-training/skills-frameworks

Knowledge Exchange	Vier landen overzicht (NK,UK, GE, DEN) van trainingsactiviteiten	http://www.knowledge-exchange.info
University of New Mexico	Spatial Data Management in Environmental Sciences	http://elibrary.unm.edu/courses/
DCC	RDMF2: Roles and Responsibilities for Effective Data Management	http://www.dcc.ac.uk/events/research-data-management-forum/roles-and-responsibilities
Alma Swan	DCC RIN workshop (Research Data Skills)	http://www.dcc.ac.uk/sites/default/files/documents/RDMF/RDMF2/02%20Swan.pdf

Bijlage 5 - 3TU.Datacentrum interviewverslagen

Opgenomen in de zipfile die beschikbaar is via:

| <http://www.surffoundation.nl/nl/publicaties/Pages/WitboekDataprofessionals.aspx>

Bijlage 6 - NIOO-KNAW interviewverslagen

Opgenomen in de zipfile die beschikbaar is via:

| <http://www.surffoundation.nl/nl/publicaties/Pages/WitboekDataprofessionals.aspx>

Bijlage 7 - Tilburg University interviewverslagen

Opgenomen in de zipfile die beschikbaar is via:

<http://www.surffoundation.nl/nl/publicaties/Pages/WitboekDataprofessionals.aspx>