

Op weg naar hergebruik

De wetenschap worstelt met een enorme uitdaging: onderzoeksgegevens moeten even toegankelijk en transparant worden als publicaties dat nu al zijn. Data die door research verkregen is mag niet langer in een la verstoven, maar moet beschikbaar komen voor hergebruik door andere onderzoekers. Het is tijd voor herbruikbaarheid van data.

Dit vereist een professioneel datamanagement dat gebaseerd is op heldere regels en een goede infrastructuur. Bij elke kennisinstelling wordt nu gewerkt om dit doel binnen bereik te brengen. Met de Landelijke Bijeenkomst Datamanagement op 10 september 2014 wilde NWO zicht krijgen én geven op de voortgang van deze inspanningen. Bovendien zal zij de uitkomsten van de dag gebruiken bij het opstellen van de eisen die ze straks stelt aan het datamanagement van onderzoeksprojecten.

Naast een vijftal plenaire presentaties bestond het programma uit vier sessies over deelonderwerpen, plus een collectieve inventarisatie van de uitkomsten.

De koers van NWO

“Deze dag had een lange aanloop”, zo verklaarde NWO-bestuurslid Franciska de Jong. Voor de nationale onderzoeksfinancier vloeit goed datamanagement namelijk voort uit de keuze voor Open Access: we moeten nu daadwerkelijk de vruchten kunnen plukken van eerder onderzoek.

NWO heeft zich grondig georiënteerd op de beste aanpak, onder meer door te kijken bij internationale zusterorganisaties. Uiteindelijk besloot ze inhoudelijk aan te sluiten bij Horizon 2020, het EU-programma voor onderzoek en innovatie. Door die aansluiting krijgen Nederlandse onderzoekers automatisch een betere uitgangspositie voor financiering vanuit Brussel.

Maar er liggen nog veel vragen op de weg naar het ideaal, zo benadrukte De Jong. “Hoe staat het met de benodigde infrastructuur voor datamanagement? En met de kennis en de ondersteuning voor onderzoekers? Er lopen in het land allerlei initiatieven en die vertonen veel voortschrijdend inzicht.” Daarom was De Jong erg blij met de grote opkomst van deze dag en de zeer diverse bijdragen. “De bijeenkomst zal op allerlei manieren doorwerken, ook in ons beleid.”

In januari 2015 start NWO met breed opgezette pilots. Onder meer wordt VICI-aanvragers verzocht ook een datamanagementparagraaf in te vullen. “Ons doel is de mindset van onderzoekers meteen bij de start al te beïnvloeden. Na honorering van het project moeten de aanvragers een datamanagementplan opstellen, dat getoetst wordt door externe experts.”

De opzet voor de genoemde datamanagementparagraaf en -plan was nog in ontwikkeling. Daarbij zouden de uitkomsten van deze dag zeker meespelen.



Europees beleid

Data is de nieuwe olie, volgens EU-commissaris Neelie Kroes. Maar wat zijn nu eigenlijk de plannen van de Europese Commissie op dit gebied? En hoe kun je er invloed op uitoefenen? Marjolein van Griethuysen vertegenwoordigt de Erasmus Universiteit Rotterdam in Brussel en lichtte enkele tippen van de sluier op.

“De Commissie”, legde ze uit, “wil belemmeringen voor integratie uit de weg ruimen, ook op wetenschapsgebied, maar u en ik weten dat de werkelijkheid weerbarstig is.” Het EU-programma Horizon 2020 moet als breekijzer fungeren voor wetenschappelijke excellentie, economische innovatie en duurzame, leefbare samenlevingen. Openheid en transparantie, met name ‘open digital science’, zijn daarbij steeds terugkerende begrippen. Zorgwekkend is echter dat de betrokkenheid van het bedrijfsleven bij Horizon 2020 gestaag afneemt. Wellicht, suggereerde Van Griethuysen, liggen hier kansen voor universiteiten en hogescholen die andere communicatiemiddelen weten te ontwikkelen dan publicaties met datasets.

Een andere relevante Brusselse discussie draait om Science 2.0, specifiek modeling en data. Een van de plannen is om datamanagement te koppelen aan de loopbaan van onderzoekers. De Commissie nodigt iedereen uit om een vragenlijst over dit thema in te vullen.

De discussie rond datamanagement wordt verder beïnvloed door enkele actuele issues. Ten eerste databescherming: enerzijds is privacy belangrijk in de EU, maar anderzijds kunnen persoonsgegevens veel waarde hebben voor de wetenschap. Die spanning vestigt de aandacht op de kwestie van data-ownership. Ook de vraag, wat nu eigenlijk mag gelden als onderzoeksdata is niet beslecht.

Verder is er de implementatie van de EU-richtlijn over ‘public sector information’ (PSI): gegevens die gebruikt worden door publieke organen en nu openbaar gemaakt moeten worden. Van Griethuysen stelde droogjes vast dat Nederland een fantastische infrastructuur heeft, maar dat New York al tien keer zo veel datasets openbaar heeft gemaakt als Rotterdam.

Afsluitend concludeerde ze, dat er in Brussel veel in beweging is en dat de Nederlandse onderzoekswereld, met zijn uitstekende reputatie, daar wel degelijk invloed op kan uitoefenen.

Datamanagement wereldwijd

Het streven naar professioneel datamanagement is een wereldwijde trend. Peter Doorn (directeur van DANS) liet zien hoe veelvormig dit fenomeen inmiddels is. Kort gezegd: er is heel veel overeenstemming op hoofdlijnen en heel verschil in details, afhankelijk van de instantie die de eisen stelt en de grondigheid waarmee dat gebeurt.

In de VS kwam het Witte Huis al in 2011 met een verordening inzake researchdata-management (RDM). Sindsdien is er hard gewerkt aan uitvoering.

Ontvangers van NSF-grants moeten nu bijvoorbeeld in een datamanagementplan van niet meer dan twee pagina’s vertellen wat voor gegevens ze zullen produceren, in welke formats, hoe de toegang, de beveiliging en de (her)gebruiksrechten geregeld zijn (inclusief de productie van derivaten) plus de archivering en de toegang tot gearchiveerde bestanden. De Europese Unie heeft, zoals bekend, haar programma Horizon 2020. Haar richtlijnen voor researchdata-management zijn nog in ontwikkeling, maar de trend is duidelijk: net zoals regels voor citeren normaal zijn geworden, zullen er straks regels zijn voor RDM.

Australië volgt een iets andere benadering, gebaseerd op haar Code for Responsible Conduct of Research. Die bevat een heel hoofdstuk over RDM, waarin de verantwoordelijkheden van onderzoekers en van instellingen worden opgesomd. Deze zijn weliswaar nauwelijks afdwingbaar, maar het land is wel op weg, zo concludeerde Doorn.



Zo heeft elk land, en bovendien elke onderzoeksfinancier, zijn eigen normen en regels. De London School of Hygiene & Tropical Medicine heeft een staatje gemaakt, waaruit blijkt dat bijvoorbeeld de Bill & Melinda Gates Foundation heel andere eisen stelt dan Cancer Research UK. Het zal niet verrassen dat voor dit probleem inmiddels ICT-oplossingen zijn ontwikkeld, in de vorm van applicaties (zoals DMPTool) met templates en toelichtingen voor specifieke funders. Een betere oplossing zou echter zijn dat funders elkaars regels overnemen. Nederland heeft dan het grote voordeel dat het een relatief klein land is, met een cultuur van openheid en samenwerking in het hoger onderwijs en onderzoek.

Datamanagement bij de universiteiten

Die open cultuur kwam goed naar voren in de presentatie van Floor Frederiks (Universiteit Leiden). Ze vertelde dat eind augustus een eerste bijeenkomst was gehouden waarin vertegenwoordigers van haast alle universiteiten plus de NFU (de medische centra) keken naar mogelijkheden om het RDM-beleid onderling af te stemmen. Tussen UB- en ICT-medewerkers was al eerder overleg op dit gebied, maar nu deden beleidsmensen ook mee.

Het bleek dat bij vijf universiteiten RDM-beleid is vastgesteld door het College, inclusief bewaartermijnen, ondersteuning et cetera. De invulling verschilt echter, niet alleen in de mate van detail maar ook in het niveau waar de verantwoordelijkheid belegd wordt. Alle andere universiteiten werken nog aan hun RDM-beleid, waarbij veel ervaring wordt opgedaan in pilots.

De belangrijkste drijfveer is overal wetenschappelijke integriteit, maar ook eisen van financiers en tijdschriften spelen mee, evenals de groeiende omvang van databestanden. Hergebruik krijgt nog weinig aandacht.

Knelpunten zijn de overgang van beleid naar implementatie, de verschillen tussen de vele disciplines, de afstemming met de universitaire medische centra (die allang eigen protocollen hebben) en het feit dat de buitenwereld op dit gebied ook niet stilstaat.

De deelnemers aan het overleg zagen perspectief voor samenwerking in het uitwisselen van beleidsdocumenten. Verder moest afstemming binnen de disciplines worden gestimuleerd. Om de juridische aspecten gezamenlijk onder de knie te krijgen wilde de EUR een bijeenkomst van universitaire specialisten organiseren, terwijl de regels van NWO en EU hanteerbaar gemaakt moesten worden via een webtool en stroomschema's voor onderzoekers.

Data delen bij ZonMw

In de laatste plenaire presentatie vertelde Margreet Bloemers hoe het RDM-beleid tot stand komt bij ZonMw, de nationale financier van gezondheidsonderzoek. Snelle en brede toepassing van resultaten in de praktijk staat bij ZonMw voorop, dus het delen van kennis is een groot goed. "Data delen = goud in handen", zo vatte Bloemers het samen.

De realiteit blijkt echter anders: gegevens over de Nederlandse volksgezondheid zijn onvoldoende beschikbaar en worden ook onvoldoende benut. ZonMw wil hier verandering in brengen en heeft daarom twee doelstellingen geformuleerd. Ten eerste: onderzoekers vullen alleen nog databronnen die vindbaar, toegankelijk en uitwisselbaar zijn, en duurzaam opgeslagen. Ten tweede: het gebruik van bestaande databronnen is een vanzelfsprekend onderdeel in wetenschappelijk onderzoek en het informeren van beleid en praktijk.

Momenteel loopt al een pilot met RDM, niet alleen om ervaring op te doen maar ook om zowel eigen medewerkers als onderzoekers te laten wennen aan de principes van datamanagement. Elke onderzoeker die subsidie ontvangt van ZonMw, moet nu een datamanagement-plan ter goedkeuring voorleggen. Daarbij krijgt hij ondersteuning, onder meer via een checklist.

Zo'n datamanagement-plan is bedoeld als groeidocument: sommige punten hoeven pas naderhand te worden gerealiseerd. Maar verplicht zijn ze wel, dus ze vormen een vast punt in de voortgangsbesprekingen. Bloemers: "We willen geen bureaucratische moloch creëren, maar een aantal zaken moet echt goed geregeld zijn om het uiteindelijke doel – het hergebruik van data – mogelijk te maken."

Daarbij wil ZonMw graag samenwerken met andere organisaties in de onderzoeksweld, om gezamenlijk te komen tot een sterke data-infrastructuur.

Sessie 1: de financiering van datamanagement

Over het belang van datamanagement is weinig discussie. Maar wie betaalt de rekening? Het gaat hierbij niet alleen om het veilig en goed opslaan van data tijdens onderzoek, maar ook om het beschikbaar en toegankelijk blijven van data op de lange termijn. “Het is te kort door de bocht om voor de financiering van datamanagement alleen naar de onderzoeksfinancier te kijken”, zo stelde Axel Berg van SURFSara. “Het vinden van een oplossing vraagt om creativiteit en nieuwe manieren van denken.” Vervolgens vroeg hij Peter Doorn (DANS), Cas Maessen (NWO), Barend Mons (DTL) en Esther van Tienhoven (UMC Utrecht) om hun visie op de financiering van datamanagement kort te presenteren.

NWO maakt kosten datamanagement subsidiabel

Cas Maessen van NWO vertelde dat NWO de kosten voor datamanagement als subsidiabel gaat beschouwen. NWO neemt een dataparagraaf op in alle aanvragen. Na honorering van de aanvraag wordt een uitgebreid datamanagementplan gevraagd. Hierin staat in detail hoe data tijdens en na het onderzoek wordt opgeslagen, beschikbaar en toegankelijk blijft en wat daarvoor nodig is. Maessen hamerde op samenwerking. “Laten we één lijn trekken in Nederland of beter nog Europa en het wiel niet vijftientig keer uitvinden.” Ook ging hij in op de vraag waarin we nog moeten investeren: “Op sommige vakgebieden is datamanagement al goed intern georganiseerd en wordt gebruikgemaakt van internationale repositories. Ook op nationaal niveau zijn goede repositories beschikbaar. Als onderzoekers deze bestaande infrastructuur gebruiken, keurt NWO dat goed. Laten we gebruikmaken van bestaande (infra)structuren en vooral investeren in de vakgebieden waar datamanagement nog niet goed georganiseerd is. De instellingen zouden op dit punt een rol kunnen spelen.” De investering die hiervoor nodig is, ligt niet alleen op het vlak van infrastructuur. Het vergroten van het bewustzijn dat goed datamanagement onderdeel is van goed onderzoek, het stimuleren van samenwerking en goede ondersteuning van de onderzoekers, noemt Maessen minstens even belangrijk.

Naast hardware en software ook aandacht voor humanware

Peter Doorn van DANS stelde dat naast hardware en software ook humanware (organisatie, advies, ondersteuning en training) een belangrijke component is van datamanagement. Tijdens de onderzoeksfase acht hij de onderzoeksorganisaties verantwoordelijk voor goed datamanagement. Voor het beschikbaar en toegankelijk maken van onderzoeksdata op de lange termijn is een federatieve oplossing nodig. Aandachtspunt hierbij is software-sustainability. “We moeten ervoor zorgen dat data toegankelijk blijft, ook als bepaalde software niet langer beschikbaar is of ondersteund wordt.” Financiering is vooral een probleem voor de very big data en minder voor longtail-data: daarbij speelt meer een probleem van organisatie, verantwoordelijkheden en naleving. Oplossingen zijn in principe beschikbaar, denk aan het Dutch Dataverse Network. Een oplossing voor de financiering ziet Doorn in het subsidiabel maken van de kosten voor datamanagement. Hij ziet weinig mogelijkheden voor bijdragen van private partijen. Tijdschriften zouden wel een rol kunnen spelen door ‘data availability policies’ in te voeren.

De vervuiler betaalt

Barend Mons had direct de aandacht van de aanwezigen door te beginnen met de stelling dat datamanagement verboden moet worden. Deze term wordt namelijk te veel geassocieerd met het ‘netjes omgaan’ met data tijdens een onderzoek. Mons pleitte voor de term ‘data-stewardship’. Dit omvat ook ‘FAIR data-publishing’ (FAIR staat voor Findable, Accessible, Interoperable and Re-usable) en langetermijnopslag. Hij hangt het principe aan van ‘de vervuiler betaalt’ en stelde dat data genererende projecten ten minste vijf procent van hun budget zouden moeten reserveren voor data-stewardship.

Zorg dat projecten op elk niveau in lijn met visie zijn

Esther van Tienhoven van het UMC Utrecht liet zien hoe haar instelling werkt aan goed data-stewardship. Het UMC Utrecht investeert in processen en architectuur rond onderzoeksdata en in infrastructuur en expertise. Wie dat betaalt? Deels loopt financiering via projectgebonden subsidiegelden en afdelings- of onderzoeksbudgetten. Steeds vaker wordt er samengewerkt met het bedrijfsleven (bijvoorbeeld zorgverzekeraars). Als het om de lange termijn of infrastructuur draait, wordt financiering deels op instellingsniveau geregeld via innovatieprogramma’s of continue budgetten zoals voor bulkstorage. Ook is er ondersteuning van organisaties als SURF en DTL en van grote infrastructurele programma’s zoals ESFRI en de nationale roadmap.

Er lopen veel projecten op verschillende niveaus. Het is belangrijk dat elk project, klein of groot, in lijn is met een overkoepelende visie op data-stewardship en dat opgedane kennis en ervaring gebundeld en gedeeld worden.

Discussie

In het gesprek dat hierop volgt, lijkt iedereen het erover eens dat het alleen verantwoord is om te investeren in research-datageneratie met een solide data-stewardshipplan. Dit vraagt om investeringen in bewustzijn, kennis, ondersteuning en infrastructuur. Voor opslag en publieke toegang op de lange termijn zien de aanwezigen geen grote rol voor private partijen. De financiering zal moeten plaatsvinden op diverse niveaus: van internationaal, nationaal tot instellings- en projectniveau. Belangrijk daarbij is dat er één gezamenlijke, algemene visie en standaard nagestreefd wordt. Onderwerpen waarvoor in de discussie aandacht werd gevraagd waren data-eigenaarschap, licentiëring, software-sustainability en juridische aspecten. Tot besluit werd gezegd dat de financiering wel een uitdaging, maar zeker niet de bottleneck zal zijn. “Technisch weten we hoe het moet. Financieel weten we die vijf procent ook wel bij elkaar te toveren. Maar zie iedereen maar eens mee te krijgen. Daar ligt de werkelijke uitdaging!”

Sessie 2: naar meer afstemming in datamanagementbeleid

Deze sessie begon met presentaties over de RDM-koers van een viertal organisaties.

NWO: luisteren naar de onderzoekers

Christa Hooijer presenteerde de visie van NWO op datamanagement. Het hoofddoel van NWO is hergebruik van onderzoeksgegevens, liefst over de grenzen van disciplines heen. Daarom wil NWO zo veel mogelijk één lijn trekken. Een deugdelijk datamanagement-plan zal ook voorwaarde zijn voor het vrijgeven van onderzoeksgelden.

Maar bij de vormgeving van haar RDM-beleid luistert NWO graag naar de onderzoekers, want ze wil de administratieve last zo laag mogelijk houden. Daarom wil ze ook waar mogelijk aansluiten bij de bestaande praktijken.

Erasmus Universiteit: duidelijkheid scheppen

Bij de Erasmus Universiteit staat niet hergebruik maar wetenschappelijke integriteit voorop, zo vertelde Paul Plaatsman. Researchdata-management is daarbij slechts één – zij het essentieel – aandachtspunt. Opvallend is dat de verantwoordelijkheden bij de EUR zeer duidelijk belegd zijn. Zo is er één centrale coördinator aangesteld voor RDM. Faculteiten hebben de taak jonge onderzoekers integere werkwijzen bij te brengen, terwijl UB en ICT verantwoordelijk zijn voor ondersteuning (frontoffice) respectievelijk opslag (backoffice).

In de praktijk blijkt dit een grootschalig en lastig proces, maar de duidelijkheid helpt wel. Samenwerking tussen universiteiten kan ook helpen, en daarbij ziet de EUR vooral een taak voor NWO.

Radboud Universiteit: samenwerking

Ook in Nijmegen was integriteit de aanleiding, maar hier werd researchdata-management een zelfstandig beleidspunt voor de universiteit, aldus Mijke Jetten. Daarbij werken beleidsmakers en ondersteuners eendrachtig samen.

Opslag van data en metadata wordt verplicht, uiterlijk vanaf het moment van publicatie. Ook de data van goedgekeurde scripties wordt opgeslagen, voor minstens tien jaar. De universiteit zorgt voor de randvoorwaarden, maar toch moet er nog veel beleid worden ingevuld door de afzonderlijke onderzoeksinstituten. Daarbij gaat het bijvoorbeeld om training van onderzoekers.

Er lopen drie pilots met RDM. Die betreffen een repository voor big data, koppeling van bestaande infrastructuren voor opslag, en de opzet van een expertisecentrum bij de UB, dat de individuele onderzoekers zal ondersteunen.

Wageningen: initiatief vanuit de UB

Bij Wageningen University biedt de UB eveneens ondersteuning aan de onderzoekers. Sinds april '14 moeten alle AIO's een datamanagement-plan in hun projectvoorstel opnemen. Ook elke leerstoelgroep moet een RDM-plan hebben, zo verklaarde Jacqueline Ringersma.

In 2010 begonnen de UB te praten met wetenschappers over datamanagement. Ringersma: “We dachten dat data verstoften en ontoegankelijk waren. We hebben daar roadshows over gehouden en er kwamen wat schandaaltjes die ook hielpen. Tot onze verbazing bleken we al een goede ‘code of

conduct' te hebben waarin ook stond dat data vijf jaar bewaard moest worden en op verzoek aan anderen beschikbaar gesteld. Maar feitelijk deed niemand dat.” Samen met de graduate-schools werd een echt RDM-beleid ontwikkeld, dat vervolgens centraal werd overgenomen en aangescherpt.

Discussie

Na deze presentaties konden de toehoorders zich uitspreken door plakkerijtjes (in drie kleuren voor 'eens', 'oneens' en 'neutraal') aan te brengen op flipovervellen met een zestal stellingen. Op de plakkerijtjes was ook ruimte voor vragen of commentaar.

- Stelling 1: “NWO dient duidelijke randvoorwaarden te stellen waaraan RDM-beleid van de instellingen moet voldoen: een sturende rol is geboden.” Over deze stelling waren de meningen verdeeld. In de discussie viel te horen dat NWO wel heldere randvoorwaarden mag stellen, maar goed moet praten met het veld en ook speelruimte moet geven.
- Stelling 2: “Er is behoefte aan duidelijke taal over juridische aspecten van onderzoeksdata. SURF is de uitgelezen partner om dit nationaal te faciliteren.” Hier was bijna niemand op tegen, al werden in de discussie wel vraagtekens gesteld bij de rol van SURF.
- Stelling 3: “Het is voor alle partijen gunstig dat er een open cultuur ontstaat over RDM-beleid, waarin beleidsstukken, concept of definitief, zonder bezwaar gedeeld worden.” Dit werd vrijwel unaniem ondersteund – maar nog niet in praktijk gebracht.
- Stelling 4: “Om de administratieve last voor onderzoekers te verminderen zou NWO ook DMP-sjablonen van andere instellingen of 'learned societies' moeten accepteren bij aanvragen.” Dit leverde weer zeer verdeelde stemmen op. Men zag liever minder dan meer DMP-sjablonen in omloop komen. NWO is de enige partij die dat kan opleggen. En het is ook eerlijker als ze op dit gebied een vlak speelveld biedt.
- Stelling 5: “Een goed beleidsstuk staat niet op zichzelf, maar gaat gepaard met het bieden van infrastructuur en ondersteuning.” Dit werd massaal bevestigd. Een reactie op een plakkerijtje: “Beleid zonder implementatie = hallucinatie!”
- Stelling 6: “NWO moet onderzoekers met raad en daad ondersteunen bij het opzetten van een dataparagraaf en DMP, als onderdeel van hun aanvraag.” Driekwart was hier tegen: de onderzoekers vallen onder hun eigen universiteiten. NWO gaat ze toch ook niet adviseren bij het invullen van onderzoeksvorstellen? Belangrijker is dat datamanagement ook handig is voor de onderzoeker zelf. NWO kan dan ook leren van de ervaringen die de instellingen hiermee opdoen. Wel viel de vraag of er geen neutraal data-expertisebureau moest komen in plaats van dat elke instelling zelf het wiel probeert uit te vinden.



Sessie 3: professionalisering en kwaliteitsverbetering

Om professionalisering en kwaliteitsverbetering op het vlak van datamanagement te realiseren, moeten we de competenties in kaart brengen die dit vraagt van onderzoekers en hun ondersteuners. Daarnaast willen we weten of zij al beschikken over deze competenties. Zodat duidelijk wordt of, en welke, actie nodig is en hoe verschillende partijen hierbij kunnen samenwerken.

Enquête en stemmen

Voorafgaand aan de parallelsessie 'Professionalisering en kwaliteitsverbetering' vulden 23 onderzoekers een enquête in. Deze was onderverdeeld in stellingen over datamanagement vóór, tijdens en na onderzoek. Dezelfde stellingen werden aan het begin van de sessie voorgelegd aan de aanwezigen in de zaal (bijna allemaal onderzoeksondersteuners). Zij konden ter plekke stemmen met behulp van stemkastjes. Na het stemmen werden de uitkomsten vergeleken.

Bij elke stelling konden de respondenten een keuze maken uit de volgende opties.

- A. De competentie is minder belangrijk.
- B. De competentie is belangrijk en er is voldoende aandacht voor aan mijn instelling.
- C. De competentie is belangrijk, maar er is nog nauwelijks aandacht voor aan mijn instelling.

Belangrijke competenties

Opvallend was dat bij geen enkele stelling antwoord A de meerderheid van stemmen kreeg. Hieruit blijkt dat de meerderheid van zowel de onderzoekers als de aanwezigen in de zaal alle genoemde competenties van belang achten. Bij de stelling 'De onderzoeker heeft kennis van algemene of disciplinespecifieke datacentra die geschikt zijn voor archivering van zijn onderzoeksdata' kozen wel relatief veel van de aanwezigen (38 procent) voor antwoord A: de competentie is minder belangrijk. Terwijl 59 procent van de onderzoekers antwoord C koos: de competentie is belangrijk, maar er is nauwelijks aandacht voor.

Bij nadere bespreking van dit punt kwam naar voren dat vooral promovendi minder bekend zijn met datacentra. Zij zijn gericht op het promoveren en 'het stukje opslag' komt pas helemaal aan het eind in beeld. De onderzoeksondersteuners die voor antwoord A kozen verklaarden hun keuze nader: dataopslag vinden zij een minder belangrijke competentie voor de onderzoeker. Dit is iets wat de universiteit zou moeten regelen. "Zo lang onderzoekers de weg naar het juiste loket binnen de instelling kunnen vinden, is er niets aan de hand."

Privacy en veiligheid

Bij stellingen over privacy en veiligheid kozen veel onderzoekers voor antwoord B (belangrijk en voldoende aandacht voor) terwijl de aanwezigen in meerderheid voor antwoord C (belangrijk, maar onvoldoende aandacht voor) kozen. Dit verschil kan worden veroorzaakt door verschillende opvattingen over veiligheid en privacy. Terwijl de één data op een USB-stick in zijn broekzak veilig vindt, denkt een ander bij veilige opslag aan het onderbrengen van data bij een gespecialiseerd datacentrum. In de zaal werd de opmerking geplaatst dat "onderzoekers op sommige punten niet weten dat ze iets niet weten." Een andere reden voor het verschil in perceptie kan het verschil in rollen en tussen disciplines zijn: "Wij benaderen data vanuit de invalshoek van integriteit en duurzaamheid, terwijl onderzoekers data vanuit hun wetenschappelijke activiteit bekijken."

Beleid in ontwikkeling

Bij zes van de twaalf stellingen gaf meer dan 50% van de onderzoekers aan dat er te weinig aandacht voor de genoemde competentie is. Bij het stemmen in de zaal werd dit antwoord zelfs acht van de twaalf keer gekozen. Dit resultaat werd niet opvallend of verontrustend gevonden. Datamanagement is volop in ontwikkeling binnen de instellingen. Veel is nog niet geregeld en vastgelegd. De verwachting is dat onderzoekers beter op de hoogte zullen zijn wanneer het beleid eenmaal duidelijk is. Het zou interessant zijn om de enquête over enkele jaren opnieuw te laten invullen. De ambitie is dat dan op meer punten antwoord B wordt gekozen, door zowel onderzoekers als onderzoeksondersteuners.

Loketfunctie

Zouden subsidiegevers zoals NWO onderzoekers moeten ondersteunen bij datamanagement? Nee, zeiden de aanwezigen. Zij zien dit als een rol van de instelling. Op dit moment wordt bij de instellingen beleid rond datamanagement ontwikkeld. Het is belangrijk dat duidelijk wordt wie op welk punt verantwoordelijk is voor datamanagement. Daarnaast moeten onderzoekers weten bij wie zij terecht kunnen met vragen en voor ondersteuning. Bij veel instellingen vervult de universitaire bibliotheek een loketfunctie. De aanwezigen vragen zich af of onderzoekers de weg naar dit loket weten te vinden.

Subsidiegevers moeten onderzoekers die om informatie over datamanagement vragen bij voorkeur terugverwijzen naar de instelling. Het zou mooi zijn als subsidiegevers op hun website of in hun informatie over dataparagrafen en het datamanagementplan een overzicht plaatsen van contactpersonen binnen de instellingen. In dit overzicht staat wie zich binnen welke instelling met datamanagement bezighoudt. Bepaalde instellingen zullen specialismen ontwikkelen. Als dit duidelijk in het overzicht vermeld wordt, kunnen universiteiten ook naar elkaar verwijzen en hoeft niet elke instelling op alle vlakken van datamanagement het wiel uit te vinden.

Opleidingen en cursussen

Datamanagement moet een integraal onderdeel in de opleidingen voor onderzoekers worden. Er bestaan per vakgebied verschillende standaarden en aandachtspunten en het is een veelomvattend onderwerp. Daarom zal een algemene cursus voor onderzoekers niet volstaan. Wel is een cursusaanbod rond datamanagement gericht op onderzoeksondersteuners en intermediairs iets dat subsidiegevers, databanken en instellingen gezamenlijk zouden kunnen ontwikkelen.

Sessie 4: de onderzoeksdata-infrastructuur in Nederland

Datamanagement móét, maar hoe dan? Er komt zo veel bij kijken, aan diensten en verantwoordelijkheden van allerlei partijen. Om dit soort complexe zaken inzichtelijk en hanteerbaar te maken, worden doorgaans modellen bedacht. Zo ook hier: inmiddels zijn in Nederland twee verschillende modellen voor de gewenste data-infrastructuur in beeld gebracht. Doel van deze sessie was een 'battle of the models'. Welk model zou het beste werken? En hoe komen we uiteindelijk tot de beste inrichting van een data-infrastructuur?

NFU-model: architectuur

Het eerste model werd gepresenteerd door Jan-Willem Boiten namens de NFU, de Nederlandse Federatie van Universitaire Medische Centra. Deze medische centra hebben bij hun onderzoekstaak te maken met een enorm breed speelveld waar ze weinig of geen grip op hebben: van twee ministeries en allerhande fondsen tot patiëntenorganisaties en farmaceutische bedrijven.

In dat warrige speelveld wil de NFU komen tot een goede data-infrastructuur. Dat is, in haar eigen woorden: 'een geïntegreerde set met precompetitieve services en tools die onderzoekers nodig hebben om data te verzamelen, te verwerken, te analyseren en te archiveren'.

Het beeld dat de NFU daarvan heeft ontwikkeld bestaat uit lagen, passend genoeg weergegeven als verdiepingen van een ziekenhuis. De onderste laag is de techniek, zoals opslag en netwerken.

Daarboven komen achtereenvolgens applicaties (waaronder ook authenticatie), data (inclusief metadata en anonimisering van patiëntgegevens) en processen (zoals ondersteuning en samenwerking). De bovenste laag is de 'business', waar bijvoorbeeld onderzoeksaanvragen plaatsvinden.

Dit model moet UMC-overstijgende projecten vergemakkelijken. Voor alle niveaus zijn al groepjes van medische centra aan de slag.

RDNL-model: stakeholders

Marjan Grootveld presenteerde het model van Research Data Netherlands (RDNL), het samenwerkingsverband van DANS, 3TU Datacentrum en SURFsara. Anders dan het eerste heeft dit model betrekking op de hele onderzoeksweld: het geeft een Nederlandse uitwerking van het model uit het EU-rapport 'Riding the wave: how Europe can gain from the rising tide of scientific data'.

Grootveld beschreef de behoeften van de diverse betrokken partijen. Financiers eisen toegankelijkheid van data. Kennisinstellingen willen integer onderzoek. Onderzoekers ook, maar liefst zonder veel rompslomp. UB's tenslotte zoeken nieuwe rollen.

In de praktijk gaat dit niet zonder problemen. Grootveld: "Researchers moeten geprikkeld worden om aan datamanagement te doen. Voor UB's is het beheren van data bijzonder leuk, maar het vraagt wel veel van ze. Op hun beurt kennen de ICT'ers die de opslag verzorgen de disciplines niet en willen ze al die mensen niet individueel begeleiden."

RDNL zoekt de oplossing in een scheiding tussen front- en backoffice. Het frontoffice is een informatieportal voor de onderzoekers. Het staat hun bij met advies, training, handige onderzoekstools en deugdelijke datamanagement-plannen. Het frontoffice wordt daarbij zelf ondersteund door het backoffice: hier zitten de experts op RDM-gebied en de voorzieningen voor langetermijntoegang tot data. RDNL wil het frontoffice beleggen bij de UB's en bij de disciplinespecifieke onderzoeksvoorzieningen. Het backoffice ligt dan bij instellingsoverstijgende organisaties zoals haar leden DANS, 3TU Datacentrum en SURFsara.

Onder dit alles ligt de lokale en landelijke ICT-infrastructuur, met partijen zoals de universitaire computercentra, SURFsara en Target. Bovenin het model zitten tenslotte de onderzoekers als producenten en consumenten van de data. Essentieel voor het welslagen van deze opzet, zo benadrukte Grootveld, is vertrouwen tussen de vele stakeholders. Dat vertrouwen moet gezamenlijk worden opgebouwd.

Discussie

Nadat de beide modellen waren gepresenteerd kon de zaal losbranden. “We gaan in een uur niet de hele wereld redden”, waarschuwde Marjan Grootveld vooraf.

In de zaal werd opgemerkt dat governance in de modellen ontbrak. Boiten gaf toe: “Tja, dan gaat er van alles mis; dat is natuurlijk duidelijk.” Ook het wettelijk kader en de duurzame financiering van de data-infrastructuur zouden toegevoegd kunnen worden.

Een UB-medewerker vond het RDNL-model nuttig om te weten op wie je een beroep kan doen. Dat gold voor de UB zelf evenzeer als voor de gebruikers. Het NFU-model had andere bedoelingen, verklaarde Boiten: dat was vooral bedoeld om in kaart te brengen wie met wie zou moeten praten over RDM-ontwikkeling. De front/backofficekwestie was voor Boiten geen hoofdzaak: “Ik droom ervan dat we processen zodanig kunnen inrichten dat je vanzelf bij de backoffice uitkomt: dan heb je onderzoekers echt ontzorgd van archivering.”

Zo bleken de beide modellen weliswaar inzicht te geven in de problematiek waar de beide omgevingen mee worstelen, maar moeilijk verenigbaar te zijn. Het was het wat tegen het wie. Beter konden we spreken van twee perspectieven op de kwestie van hoe je een coherente data-infrastructuur bouwt en bruikbaar maakt.

Er werd dan ook voorgesteld om het idee van een model los te laten: “Laten we gewoon elementen benoemen die belangrijk zijn, en waar je voor die zaken moet zijn. Dan kan de infrastructuur per universiteit of discipline worden vormgegeven.” Een andere spreker gaf de hoop op een model nog niet op, maar daarin zou dan de levenscyclus van de data zelf centraal moeten staan.

Uiteindelijk werden enkele deelnemers bereid gevonden om als een ‘bende van vijf’ te verkennen of een gedeelde terminologie of zelfs perspectief zinvol en haalbaar is.

Conclusies

Aan het eind van de dag werden de uitkomsten van de sessies plenair gepresenteerd. Daarna was het tijd voor het moeilijkste onderdeel: de conclusies. Die werden geformuleerd in een open debat van de aanwezigen, onder leiding van Peter Doorn.

1. Samenwerking is niet voldoende: er moet een regiefunctie worden belegd.
Er zijn veel samenwerkingsverbanden, maar per saldo werkt men langs elkaar heen. “Je ontmoet steeds weer nieuwe mensen die iets blijken te doen.”
2. Er moet een duurzaam financieringsmodel komen voor datamanagement.
Financiering van datamanagement door NWO betekent per saldo minder geld voor onderzoek.
3. Er moet één expertisepunt komen voor researchdata-management.
Of anders meerdere punten voor juridische en andere deelexpertises.
4. De geïdentificeerde competenties van onderzoekers moeten verder ontwikkeld worden.
5. De diverse datamanagement-‘talen’ moeten worden afgestemd.
Dit brengt de bestaande perspectieven en modellen dichterbij en maakt het gesprek over datamanagement eenvoudiger.
6. Er moet een gezamenlijke catalogus komen van nationale en internationale, publieke en private datamanagement-diensten en -voorzieningen.
Per instelling moet dan wel worden aangegeven welke voorwaarden voor haar medewerkers gelden bij het gebruik van de diensten.