

Research Data Management

Richtlijnen voor de duurzame opslag van onderzoeksdata

Vastgesteld in CvB d.d. 4 november 2014
U2014/06793/JRI



1. Inleiding

De noodzaak voor beleid met betrekking tot datamanagement van onderzoeksgegevens wordt ingegeven door ontwikkelingen op twee verschillende gebieden, te weten het terrein van *open acces* (i.e. de optimale verspreiding en toepassing van de resultaten van publiek gefinancierd onderzoek)¹ en wetenschappelijke integriteit². Centraal hierbij staan de transparantie met betrekking tot de verwerving en analyse van onderzoeksgegevens en de permanente toegankelijkheid van diezelfde gegevens en analysetools voor doeleinden van hergebruik en/of verificatie.

Onder datamanagement vallen alle handelingen die ertoe leiden dat onderzoeksgegevens (duurzaam) vindbaar, toegankelijk en bruikbaar zijn. Dit omvat de organisatie, documentatie, opslag en archivering van onderzoeksgegevens. Deze notitie voorziet in richtlijnen met betrekking tot de inrichting van een centrale faciliteit voor duurzame opslag van digitale onderzoeksdata en de bijbehorende definiërende gegevens, na afronding van het onderzoek, met als doel het optimaal mogelijk maken van verificatie en hergebruik van de data. Elementen die buiten het bestek van deze notitie vallen zijn de opslag van data *tijdens* onderzoek en de opslag van onderzoeksdata verkregen door studenten. Voor deze onderwerpen wordt apart beleid voorbereid.

2. Richtlijnen voor de duurzame opslag van onderzoeksdata

De richtlijnen in deze notitie sluiten aan bij het door een interne werkgroep opgestelde advies '*Onderzoeksgegevens en Auteursrechten*' (2013)³. Als centraal uitgangspunt dienen hierbij de nationale en internationale richtlijnen en aanbevelingen op het gebied van dataopslag⁴.

2.1 Definitie onderzoeksdata

Onder onderzoeksdata worden alle gegevens verstaan die tijdens wetenschappelijk onderzoek worden verzameld. Het gaat hierbij bijvoorbeeld om tekst, beeld, geluid, spreadsheets, databases, statistische gegevens, geografische data, etc. Hierbij hoort ook de beschrijving van de verzamelde data (de zgn. metadata), zoals wat de data representeert, hoe en wanneer deze is verzameld, de gehanteerde methodologie, etc.

2.2 Eigenaarschap

Uitgangspunt van het auteursrechtenbeleid⁵ van de Open Universiteit (OU) is dat, op basis van artikel 7 van de Auteurswet (het werkgeversauteursrecht), het auteursrecht van werken van onderwijs en onderzoek, vervaardigd door wetenschappelijk en ondersteunend personeel in het kader van hun dienstverband bij de OU, berust bij de OU. Hieruit vloeit voort dat de OU, indachtig subsidievoorwaarden van onderzoeksfinanciers of wettelijke vereisten, de zeggenschap heeft over de wijze waarop met onderzoeksdata in het kader van het datamanagementbeleid wordt omgegaan.

Wanneer van dit uitgangspunt afgeweken dient te worden, bijvoorbeeld in het kader van samenwerkingsverbanden (onder andere 2^{de} en 3^{de} geldstroom), worden tussen partners in

¹ Zie onder andere de *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* (2003), inmiddels getekend door 480 instellingen waaronder de Nederlandse universiteiten, KNAW en NWO, en het huidige kabinetsbeleid: <http://www.government.nl/documents-and-publications/parliamentary-documents/2014/01/21/open-access-to-publications.html>

² Zie onder andere het rapport van de KNAW '*Zorgvuldig en integer omgaan met wetenschappelijke onderzoeksgegevens*' (2012); advies van de KNAW commissie Onderzoeksgegevens

³ Advies '*Onderzoeksgegevens en auteursrechten*' van de werkgroep o.l.v. prof. dr. P. Kirschner, vastgesteld door het CvB op 10 september 2013.

⁴ Het KNAW rapport '*Zorgvuldig en integer omgaan met wetenschappelijke onderzoeksgegevens*' (2012); DANS: datakeurmerk.nl – kwaliteitsrichtlijnen voor digitale onderzoeksdata in Nederland; European research Council: *The European Code of Conduct for Research Integrity*

⁵ Het beleid rondom intellectueel eigendom dient nog nader uitgewerkt te worden.

een schriftelijke overeenkomst nadere afspraken gemaakt over de geldende zeggenschap met betrekking tot de data. Daarnaast wordt vastgelegd hoe onder andere wordt omgegaan met de opslag, registratie van metadata, bewaartermijnen, hergebruik en de mate waarin de data vrij toegankelijk zijn. Deze schriftelijke overeenkomst maakt integraal deel uit van het Data Management Plan dat per project opgesteld wordt (zie paragraaf 2.9).

2.3 Bewaartermijn

Standaard wordt een opslagtermijn van minimaal 10 jaar aangehouden. Uitzonderingen worden gevormd door bijvoorbeeld wettelijk bepaalde maximale bewaartermijnen. Materiaal dat opgeslagen dient te worden omvat:

- Ruwe data
- Secundaire data
- Analyse protocollen
- Gedocumenteerde transformaties op ruwe data in de vorm van syntaxfiles, codeboeken, etc
- Specifieke software die noodzakelijk is voor analyse van resultaten
- Benodigde metadata

Wanneer een werknemer de universiteit verlaat dienen in het exit-gesprek afspraken te worden gemaakt over de opslag van de data.

2.4 Infrastructuur

De OU richt een centrale registratie en repository voorziening in die ter beschikking wordt gesteld aan de faculteiten, te weten DANS Easy⁶, het online archiveringssysteem van het nationale instituut Data Archiving and Networked Services (DANS). Deze opslagvoorziening is alleen geschikt voor duurzame opslag van onderzoeksdata. Per faculteit is een zogenaamde datafunctionaris verantwoordelijk voor de kwaliteit en integriteit van de in de repository opgeslagen data en het onderhouden van de contacten met DANS.

2.5 Metadata

Voor hergebruik en/of verificatie van de opgeslagen data is metadata noodzakelijk. Met betrekking tot de metadata geldt dat van elk onderzoek naast dataset(s) ook de metadata worden opgeslagen. De toegang tot datasets is in principe open, tenzij er zwaarwegende redenen zijn om dat niet te doen (zie ook paragraaf 2.7). De metadata zijn na opslag altijd toegankelijk, ook als (delen) van de dataset niet zondermeer als toegankelijk zijn bestempeld. De documentatie die onder metadata valt wordt deels bepaald door de vereisten van DANS Easy en deels door de betrokken onderzoeker. Naast de standaard metadata kan het bijvoorbeeld nodig zijn om een uitgebreidere beschrijving te geven van de data, om de onderlinge samenhang tussen datasets te beschrijven, de methode en omstandigheden waaronder de data zijn verzameld. Daarnaast hanteert een aantal wetenschappelijke disciplines eigen standaarden met betrekking tot de metadata⁷.

2.6 Moment van opslag

De voor permanente opslag aangemerkte data worden bij voorkeur opgeslagen per datum van indiening van de betreffende publicatie bij de tijdschriftredactie, dan wel promotiecommissie.

⁶ <http://www.dans.knaw.nl/content/online-archiveringssysteem-easy>. Zie ook de notitie 'opslagonderzoeksgegevens en registratie onderzoeksoutput' (2014). De keuze voor DANS is ingegeven door het feit dat DANS een instituut is van KNAW en NWO met landelijke erkenning; het feit dat DANS enkele jaren geleden aan de wieg heeft gestaan van het internationaal erkende datakeurmerk Data Seal of Approval en dat het repository systeem van DANS (EASY) naar verwachting in 2015 ISO gecertificeerd wordt als betrouwbaar archief.

⁷ Zie voor een overzicht van standaarden met betrekking tot metadata: <http://www.dcc.ac.uk/resources/metadata-standards>

2.7 Toegankelijkheid

Alle data worden in principe openbaar toegankelijk opgeslagen. De uitzondering wordt bepaald door:

- gedeelde zeggenschap of door zeggenschap van derden.
- embargo tot publicatie datum, met een maximum van 18 maanden

De afspraken met betrekking tot toegankelijkheid staan beschreven in het bij het project behorende Data Management Plan (zie paragraaf 2.9).

2.8 Vertrouwelijkheid persoonsgegevens

Bij de opslag wordt de vertrouwelijkheid van de persoonsgegevens in acht genomen conform de Wet Bescherming Persoonsgegevens (WBP). Bij interviewdata geldt aanvullende regelgeving: deze data mogen uitsluitend na toestemming van de betrokkene (de geïnterviewde) aan DANS ter beschikking worden gesteld.

Bij het hergebruik van data met persoonsgegevens in de zin van de WBP geldt dat deze conform de richtlijnen van DANS uitsluitend ten behoeve van wetenschappelijk, statistisch of historisch onderzoek gebruikt mogen worden. Een werkwijze conform de Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek van de VSNU is hierbij verplicht gesteld.

2.9 Facultair handboek datamanagement en het Data Management Plan

Facultair handboek datamanagement

Met het oog op de verschillen tussen de diverse wetenschapsgebieden (type onderzoeksdata en eigen gebruiken) dienen de faculteiten in een zogenaamd 'Handboek Datamanagement' aan te geven hoe zij, op basis van deze OU-brede kaderstelling, het specifieke facultair beleid voor datamanagement invullen⁸. In dit handboek wordt aangegeven aan welke centraal gestelde richtlijnen wordt voldaan en op welke wijze. Bij een keuze voor het niet implementeren van de centrale richtlijnen moet beargumenteerd worden aangegeven waarom niet.

Data Management Plan

Een Data Management Plan (DMP) of dataparagraaf wordt in toenemende mate een verplicht onderdeel van een subsidieaanvraag⁹. Dit betekent dat elk onderzoeksvorstel vanuit de OU dat wordt ingediend ter financiering intern of extern moet worden voorzien van een datamanagement paragraaf. Na toewijzing van de financiering wordt deze paragraaf uitgewerkt tot een Data Management Plan (DMP). Een DMP is een levend document, d.w.z. dat het gedurende het onderzoek kan worden aangepast. In het DMP worden de afspraken rond de data nader vastgelegd. De indiener van het projectvoorstel beschrijft hierbij onder meer wat voor soort data verzameld wordt (omvang en format), hoe en waar de data tijdens en na afloop van het onderzoek worden opgeslagen en voor wie de data toegankelijk is¹⁰.

Zoals eerder aangegeven beperkt datamanagement zich niet tot het archiveren van digitale data. Alle gegevens die nodig zijn om de opgeslagen data te kunnen interpreteren (o.a. codeboeken, analyse protocollen, specifieke software) vallen ook onder het datamanagement beleid en dienen daarom in het DMP aandacht te krijgen.

⁸ Bij het opstellen van dit handboek kan mede gebruik worden gemaakt van de richtlijnen opgesteld door DANS en het 3TU Datacentrum ten aanzien van de selectiecriteria voor het archiveren van data. Zie: *Selection of Research Data, Guidelines for appraising and selecting research data*; Heiko Tjalsma en Jeroen Rombouts (3TU.Datacentrum. DANS Studies in Digital Archiving 6. 2011).

⁹ Bijvoorbeeld: ZonMW, NWO en Europese subsidies.

¹⁰ Voorbeeld van sjablonen, zie: <https://wiki.surfnet.nl/display/RD/Datamanagementplannen>