



Datamanagementplan voor wetenschappelijk onderzoek

september 2015

Introductie

Onderzoeksdata zijn waardevol. Ze spelen een steeds grotere rol in de wetenschap en dienen goed beheerd te worden om hergebruik en verificatie van onderzoeksresultaten mogelijk te maken.

Steeds vaker dient een onderzoeksproject een datamanagementplan op te stellen waarin wordt aangegeven op welke manier de onderzoeksdata beheerd, gedocumenteerd en gedeeld kunnen worden. Deze brochure bevat aandachtspunten voor het opstellen van een datamanagementplan en dient met name in de beginfase van een dataverzamelingsproject gebruikt te worden. Het laatste hoofdstuk bevat de kernelementen van een datamanagementplan.

De informatie in dit datamanagementplan is o.a. afkomstig uit het datamanagementplan van DANS Data Archiving and Networked Services (januari 2015).

Inhoud

1. Opstellen van een datamanagementplan	3
. 1.1 Waarom een datamanagementplan belangrijk is	3
. 1.2 Wanneer je een datamanagementplan schrijft	3
. 1.3 De betrokkenen	3
2. Voorbereiding van het dataverzamelingsproject	5
. 2.1 Algemene informatie over de dataverzameling	5
. 2.2 Overzicht eerder verzamelde data	5
. 2.3 Keuze voor te gebruiken software en hardware	5
. 2.4 Vaststelling intellectueel eigendom en wettelijke eisen	6
. 2.5 Gebruikersinformatie	6
. 2.6 Interoperabiliteit	7
3. Uitvoering van het dataverzamelingsproject	8
3.1 Datamanagement 3.2 Keuze voor te gebruiken metadata	9
4. Archivering en beschikbaarstelling van data	10
4.1 Data deponeren in een data-archief of repository	10
5. De opzet van een datamanagementplan	11
5.1 Sjabloon voor een datamanagementplan	11

Deel 1 Opstellen van het datamanagementplan

1.1 Waarom een datamanagementplan belangrijk is

Er zijn diverse redenen om een datamanagementplan op te stellen, zoals

- het voldoen aan subsidievereisten van bijvoorbeeld NWO of Horizon2020;
- het stimuleren van hergebruik door de data goed vindbaar te maken en te expliciteren hoe anderen er toegang toe kunnen krijgen;
- het relateren van data aan publicaties en andere onderzoeksinformatie; het mogelijk maken van verificatie en replicatie van het onderzoek;
- en heel letterlijk: het plannen en begroten van alle aspecten van goed datamanagement. Hierdoor wordt bijtijds duidelijk aan welke data-expertise of voorzieningen zoals beveiligde opslag behoefte is.

1.2 Wanneer je een datamanagementplan schrijft

Het is aan te raden om al bij het schrijven van een projectvoorstel rekening te houden met datamanagement. Het opstellen van het datamanagementplan dient uiterlijk in de beginfase van een dataverzamelingsproject te gebeuren.

Datamanagement invoeren als er al veel onderzoeksgegevens verzameld en opgeslagen zijn, is namelijk lastiger en kostbaar, omdat dan veel werk extra gedaan moet worden. Dit neemt niet weg dat er tijdens het project wellicht aanpassingen aan het plan nodig zijn; onderzoek verloopt immers niet altijd geheel volgens plan. In een dergelijke situatie is wel van belang om de wijziging van het datamanagementplan goed te documenteren, zodat de onderzoeker verantwoording kan afleggen aan belanghebbenden, zoals de financier.

1.3 De betrokkenen

Wie hebben met datamanagement en met het bijbehorende plan te maken? In de eerste plaats natuurlijk de onderzoekers die het betreffende onderzoek gaan uitvoeren, maar zij zijn niet de enigen. Ook het management en de ondersteunende diensten in de instelling(en) van de onderzoekers spelen een rol.

Waarom dit belangrijk is

Betrokkenheid van het management en andere sleutelfiguren in de organisatie is noodzakelijk om twee redenen. Ten eerste verdient het aanbeveling om na te gaan voor welke aspecten van datamanagement centraal beleid, richtlijnen en ondersteuning wenselijk en mogelijk zijn. Idealiter hebben onderzoekers bij het opstellen van een datamanagementplan ruimte voor eigen invulling zonder alle aspecten ad hoc te hoeven uitwerken. Ten tweede moeten de opstellers tijd krijgen van het management om te werken aan een datamanagementplan. Dit is momenteel nog geen standaardactiviteit en de ervaring leert dat de doorlooptijd enkele maanden kan bedragen. Dit suggereert eens te meer dat datamanagement op het niveau van een organisatie(onderdeel) geregeld zou moeten worden.

Wanneer de organisatie de functie van datamanager kent, dient die betrokken te worden bij het opstellen van het datamanagementplan; wellicht kan de

datamanager ook het voortouw nemen, aangezien hij/zij waarschijnlijk de meeste ervaring heeft of opdoet met datamanagementplannen. Hetzelfde geldt voor ondersteuners in de instituutbibliotheek. Het is daarnaast van belang om de ICT-afdeling vroeg in het proces te betrekken. Soms is die betrokkenheid erg nodig, soms niet; hoe eerder dit duidelijk is, des te beter het is. Ook het (externe) datarepository of -archief is graag in een vroeg stadium op de hoogte van aard, omvang en toegangsrechten van de toekomstige onderzoeksgegevens.

Deel 2 Voorbereiding van het dataverzamelingsproject

2.1 Algemene informatie over de dataverzameling

Zowel voor de financier van het dataverzamelingsproject als voor de uiteindelijke uitvoerder van het project is het belangrijk een algemene beschrijving te maken van de te verzamelen data, de aard en de omvang ervan; hoe de data verzameld zullen worden; en hoe de data te zijner tijd beschikbaar worden gesteld. In sommige disciplines spreekt men overigens niet van “data”, maar ook daar kent men onderzoeksgegevens en -resultaten die (naast publicaties) voor bewaren en beschikbaar stellen in aanmerking komen. Een datamanagementplan is ook in die disciplines dus relevant.

2.2 Overzicht eerder verzamelde data

Als er eerder verzamelde data zijn die vergelijkbaar zijn met de nieuw te verzamelen data, is het belangrijk een overzicht van de betreffende data te maken en te beschrijven waarom het belangrijk is om nieuwe data te verzamelen. Een dergelijk overzicht kan gemaakt worden met behulp van de catalogi van de belangrijkste data-archieven en repositories.

Waarom dit belangrijk is

De nieuwe dataverzameling kan worden afgestemd op eerder verzamelde data. Daarnaast kan een dergelijke inventarisatie kosten besparen. Mogelijk zijn bepaalde data al eerder verzameld en kunnen de beoogde analyses ook gedaan worden op al bestaande data.

2.3 Keuze voor te gebruiken software en hardware

Vervolgens moet worden vastgesteld welke hardware en software gebruikt zullen worden. Bij de keuze van software is het van belang rekening gehouden met disciplinespecifieke standaard dataformaten. *Waarom dit belangrijk is*

Veroudering van hardware en software betekent dat hergebruik van de data op den duur niet meer mogelijk is. De digitale duurzaamheid van data komt daarmee in gevaar. Het is daarom aan te bevelen software te gebruiken die op termijn door een data-archief of een repository geconverteerd kan worden naar een steeds bruikbaar bestandsformaat.

Aanvullende informatie

Het Meertens Instituut hanteert een lijst waarop de geprefereerde en geaccepteerde bestandsformaten van de volgende soorten bestanden staan:

- Tekstbestanden zoals PDF/A, platte tekst en *mark-up* zoals HTML
- Spreadsheets
- Statistische bestanden
- Databases
- Beeldmateriaal
- Audiovisuele bestanden
- Cartografische gegevens

- Geographic Information System (GIS-data)

Deze lijst met *preferred formats* is overgenomen van DANS (Data Archiving and Networked Services).

2.4 Vaststelling intellectueel eigendom en wettelijke eisen

Voor het vaststellen van het intellectuele eigendom en de wettelijke eisen is het belangrijk een antwoord te vinden op de volgende vragen: welke persoon of instelling bezit het auteursrecht met betrekking tot de verzamelde data, en zullen deze rechten te zijner tijd mogelijk overgedragen worden? Gelden er wettelijke eisen of voorschriften voor het delen en archiveren van de data? Wordt er auteursrechtelijk beschermd materiaal gebruikt, zoals bijvoorbeeld foto's, bepaalde meetinstrumenten of schalen? Indien dit het geval is, hoe verkrijgt het project toestemming om het materiaal te gebruiken en te verspreiden?

Waarom dit belangrijk is

Om de data na afronding van een dataverzamelingsproject via een data-archief of een repository beschikbaar te kunnen stellen, is het noodzakelijk dat bekend is bij wie de intellectuele eigendomsrechten berusten. Indien auteursrechtelijk beschermd materiaal wordt gebruikt bij het verzamelen van data, dient het copyright daarvan gerespecteerd te worden.

Aanvullende informatie

Zowel voor het deponeren als voor het gebruik van data gelden afspraken die gebaseerd zijn op de Nederlandse en Europese wetgeving en op de [Gedragscode van de Vereniging van Universiteiten](#)¹ (VSNU) voor het gebruik van persoonsgegevens in wetenschappelijk onderzoek. Deze code is een uitwerking van de Wet Bescherming Persoonsgegevens.

De afspraken tussen de deponeerder van de data en het Meertens Instituut omtrent het gebruik en het deponeren van data worden vastgelegd in een licentieovereenkomst.

2.5 Gebruikersinformatie

Worden er maatregelen genomen om inzicht te verkrijgen in het gebruik van de data; wie zullen de data bijvoorbeeld gebruiken en voor welk doel?

Waarom dit belangrijk is

De gebruikers voor wie de data verzameld worden, kunnen van invloed zijn op de wijze waarop de data worden beheerd, gedocumenteerd en gedeeld.

¹ Zie <http://www.vsnu.nl/code-pers-gegevens.html>

2.6 Interoperabiliteit

In veel disciplines is het van belang dat de data gekoppeld kunnen worden met data die in beheer zijn bij een externe organisatie en met data die in de toekomst verzameld of gegenereerd zullen worden. Welke standaarden of afspraken inzake de databeschrijving bestaan hiervoor binnen de discipline of de kennisinstelling? Welke voorkeur heeft de financier?

Waarom dit belangrijk is

Deze informatie is van belang om vroegtijdig concrete afspraken te kunnen maken met de beoogde databehurende instellingen.

Deel 3 Uitvoering van het dataverzamelingsproject

3.1 Datamanagement

Bij het uitvoeren van het dataverzamelingsproject is het van belang vast te stellen welke persoon eerstverantwoordelijke is voor het datamanagement, en wie verantwoordelijk is voor de documentatie van alle fases van het dataverzamelingsproject. Is er sprake van een standaard naamgevingconventie? Hoe wordt ervoor gezorgd (indien van toepassing), dat er vertrouwelijk met – persoonsgebonden, commerciële, militaire – gegevens omgegaan wordt? Hoe en waar worden kopieën van de data opgeslagen en hoe wordt de veilige opslag daarvan gegarandeerd? Hoe is het versiebeheer geregeld? Hoe is in elke fase van de dataverzameling de toegang tot de data geregeld voor projectleden binnen en buiten de eigen instelling?

Waarom dit belangrijk is

Beantwoording van bovenstaande vragen betekent dat het beheer van de data gedurende de onderzoeksfase adequaat geregeld is. Digitale gegevens zijn bovendien kwetsbaar en de beste handelswijze is daarom om meerdere kopieën op te slaan op meerdere locaties. De veilige opslag van gevoelige gegevens is, indien van toepassing, ook een belangrijk onderdeel daarvan. Idealiter kan de onderzoeker voor (de informatie over) dergelijke voorzieningen terugvallen op zijn organisatie.

Aanvullende informatie

Een aantal vooraanstaande buitenlandse data-archieven, zoals het Inter-university Consortium for Political and Social Research (ICPSR) in de Verenigde Staten en het UK Data Archive in Engeland, heeft uitgebreide adviezen en sjablonen voor datamanagementplannen gemaakt. Het Britse Digital Curation Center verschaft naast een eigen sjabloon ook verwijzingen naar ingevulde plannen. Ook de werkgroep van Nederlandse universiteitsbibliotheken en de Koninklijke bibliotheek (UKB-werkgroep) heeft een groot aantal sjablonen geïnterpreteerd. Voor gedetailleerde informatie over diverse aspecten van een zorgvuldige omgang met data zie de volgende links:

[UK Data Archive – Managing and sharing data – Best practice for researchers](#)²

[ICPSR – Elements of a Data Management Plan](#)³

[Australian National Data Service – Data Management for Researchers](#)⁴

[Digital Curation Center – Data Management Plans](#)⁵

[UKB-werkgroep – Rapportage Datamanagementplannen](#)⁶

² Zie <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf> of het recentere boek <http://www.uk.sagepub.com/books/9781446267264>

³ Zie <http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/resources.html>

⁴ Zie <http://www.ands.org.au/researchers/manage-data.html>

⁵ Zie <http://www.dcc.ac.uk/resources/data-management-plans>

⁶ Zie <https://wiki.surfnet.nl/display/RD/Datamanagementplannen>

3.2 Keuze voor te gebruiken metadata

Welke metadata zullen gebruikt worden, en behoren die tot een bepaalde metadatastandaard?

Waarom dit belangrijk is

Metadata, dat wil zeggen, data over de onderzoeksdata zelf, zijn nodig om data te kunnen vinden in een (online-)catalogus. Toevoeging van rijke, gedetailleerde metadata betekent bovendien dat de inhoud van een dataverzameling adequaat ontsloten wordt. Geïnteresseerden kunnen hierdoor beter bepalen of een dataverzameling relevant en bruikbaar is voor hun eigen onderzoek. Voor hergebruik zijn metadata dus van groot belang.

Aanvullende informatie

Verschillende wetenschappelijke disciplines hanteren eigen metadatastandaarden; deze kunnen fijnmazig zijn (zoals de Data Documentation Initiative specificatie voor de sociale en gedragswetenschappen) of juist de minimale vereiste informatie vastleggen, bijvoorbeeld de Minimal Information (MI) Standards in de biowetenschappen. Daarnaast zijn er ook generieke metadatastandaarden, zoals Dublin Core. Voor sommige soorten metadata bestaan editors waarin men metadata kan toevoegen aan de data. Het Digital Curation Center (DCC) verwijst naar een groot aantal generieke en disciplinespecifieke metadataformaten.

[Dublin Core Metadata Standaard](#)⁷

[Data Documentation Initiative](#)⁸

[Metadata geodata](#)⁹

[Minimum Information Standards](#)¹⁰

[NESSTAR](#)¹¹

[DCC – Metadata Standards](#)¹²

⁷ Zie <http://dublincore.org/specifications/>

⁸ Zie <http://www.ddalliance.org>

⁹ Zie <http://www.geonovum.nl/onderwerpen/metadata?tab=standaarden>

¹⁰ Zie https://en.wikipedia.org/wiki/Minimum_Information_Standards

¹¹ Zie <http://www.nesstar.com>

¹² Zie <http://www.dcc.ac.uk/resources/metadata-standards>

Deel 4 Archivering en beschikbaarstelling van data

4.1 Data deponeren in een data-archief of repository

Hoe en waar zullen de data voor de lange termijn bewaard worden? Geef aan hoe de data zullen worden geselecteerd voor het archiveren, en hoe lang de data bewaard moeten worden. Zijn de kosten van de archivering bij een data-archief of een repository opgenomen in de begroting van het dataverzamelingsproject?

Waarom dit belangrijk is

Het is van belang om data beschikbaar te kunnen stellen voor nieuw onderzoek; toegang tot onderzoeksdata bevordert tevens de openbaarheid en verificerbaarheid van wetenschappelijk onderzoek. Data-archieven en repositories stellen bepaalde eisen aan de data en de documentatie daarvan ten behoeve van het hergebruik. Dit is vooral ook van belang voor hergebruik op de langere termijn. Het is daarom aan te bevelen hiervoor een deel van het onderzoeksbudget te reserveren.

Aanvullende informatie

Het is aan te bevelen data onder te brengen bij een data-archief dat voldoet aan de richtlijnen van het internationaal vastgestelde Data Seal of Approval.¹³ Het Data Seal of Approval (DSA) benoemt vereisten voor de betrouwbaarheid van data-archieven. De richtlijnen vormen een eerste stap in een reeks internationale certificeringsinitiatieven. Het Meertens Instituut is gecertificeerd met het Data Seal of Approval.

In 2014 heeft de VSNU in de [Nederlandse Gedragscode Wetenschapsbeoefening](#)¹⁴ de minimale bewaartermijn van ruwe data op tien jaar gesteld.

¹³ Zie <http://www.datasealofapproval.org/en/>

¹⁴ Zie

[http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Code_wetenschapsbeoefening_2004_\(2014\).pdf](http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Code_wetenschapsbeoefening_2004_(2014).pdf)

Deel 5 De opzet van een datamanagementplan

5.1 Sjabloon voor een datamanagementplan

Tot slot bevat deze brochure een sjabloon met de belangrijkste elementen van een datamanagementplan. Het sjabloon is discipline-onafhankelijk en beoogt zowel het databewustzijn van onderzoekers te versterken als de basis te leggen voor een bruikbaar plan, dat eventueel wordt aangepast in de loop van het onderzoek.

1	Administratieve informatie
1.a	Projectnaam, hoofdonderzoeker, financier(s), datum van dit plan en van eerdere versies
1.b	Wie is de eerstverantwoordelijke voor het datamanagement?
2	Beschrijving van de data
2.a	Worden bestaande data hergebruikt of nieuwe data gegenereerd?
2.b	Om welke soort(en) data gaat het; omvang van de bestanden; groeitempo?
3	Standaarden en metadata, ofwel alles wat nodig is om de data te vinden en te benutten
3.a	Welke metadatastandaarden worden gebruikt (<i>vindbaarheid</i>)?
3.b	Welke coderingen e.d. worden gebruikt die toekomstige koppeling met andere data mogelijk maken (<i>duiding, interoperabiliteit</i>)?
3.c	Welke software en eventueel hardware wordt er gebruikt (<i>duiding, bruikbaarheid</i>)?
3.d	Wat wordt er gedocumenteerd en bewaard om replicatie mogelijk te maken?
4	Ethisch en juridisch
4.a	Hoe wordt bij het verwerven of genereren van de data de hiervoor eventueel benodigde toestemming verkregen van dataleverancier/ proefpersonen/ ...? Welke beperkingen gelden er eventueel tijdens het onderzoek?
4.b	Hoe worden gevoelige gegevens beschermd tijdens en na het project?
4.c	Wat zijn de afspraken als betrokkenen (voortijdig) vertrekken?
4.d	Zijn de data na het project – eventueel na een embargoperiode – als Open Access beschikbaar? Zo nee, welke voorwaarden gelden er?
5	Opslag en archivering
5.a	Hoe wordt voldoende opslag- en back-upcapaciteit tijdens het project geregeld, inclusief versiebeheer? Zijn de kosten hiervoor gedekt; zo nee...?
5.b	Waar en hoe lang worden de data na afloop van het project beschikbaar gesteld voor vervolgonderzoek en verificatie? Is dit een <i>Trusted Digital Repository</i> , dus met een internationale certificering? Zo niet, hoe worden de data dan vindbaar (<i>denk aan metadata en aan persistent identifiers zoals DOI, Handle en URN</i>) en duurzaam toegankelijk en bruikbaar?
5.c	Zijn de kosten voor (het voorbereiden van de data voor) archivering gedekt; zo nee...?

