

Data labs

Inleiding

Tijdens je onderzoek komt er onherroepelijk een moment waarop je je onderzoeksdata veilig en toegankelijk voor anderen wilt opslaan.

- Je promotor wil zich snel een beeld vormen van je data ;
- Je wilt weten wat een collega vindt van je data ;
- Je wilt je data door studenten laten gebruiken in een cursus ;
- Je wilt je data mogelijk uitwisselen met wie je samenwerkt in je project ;
- Je data zijn uniek en waardevol of kunnen niet makkelijk opnieuw verzameld worden ;
- Je data worden verder gebruikt in een vervolgproject van je onderzoeksgroep.

Datamanagement tijdens je onderzoek heeft een tweeledig doel: het behoud van je data - waarbij het in dezen niet alleen gaat om het loutere bestaan van je data maar ook om het zeker stellen van de kwaliteit en betrouwbaarheid ervan - en de beschikbaarstelling van je data aan anderen.

Deze 2 doelen vereisen in ieder geval:

- Opslagruimte en -apparatuur voor je data ;
- Metadata, dit is een voor anderen heldere en duidelijke beschrijving van je data ;
- Gecontroleerde toegang tot je data: als niet iedereen je data mag inzien of bewerken wie dan wel?

Data labs

In een data lab worden deze onderdelen van datamanagement ondersteund of gefaciliteerd. Het is een elektronische omgeving of platform waar je je onderzoeksdata veilig kunt opslaan, duidelijk kunt beschrijven en de beschikbaarstelling ervan kunt regelen. Vaak zie je ook dat een data lab de mogelijkheid biedt om met anderen gezamenlijk te werken aan je onderzoeksdata.

Hieronder wordt aan de hand van het data lab *Dataverse Network* toegelicht hoe daar de hierboven genoemde onderdelen van datamanagement bijeenkomen en gerealiseerd worden.

Dataverse Network

Dataverse Network (DVN) is een oorspronkelijk door Harvard University ontwikkelde data-labomgeving (<http://thedata.org>). In Nederland is in 2010 door de Universiteitsbibliotheek Utrecht een lokale applicatie van DVN geïnstalleerd en in beheer. Sinds 2013 bestaat er het landelijke samenwerkingsverband *Dutch Dataverse Network* (DDN, <http://www.dataverse.nl>) waar een aantal instellingen zich bij hebben aangesloten en die gebruik maken van de bij Utrecht geïnstalleerde applicatie.

Opslag van data en DVN

Technisch gezien zijn er geen beperkingen aan de hoeveelheid data die door een onderzoeker in DVN kan worden opgeslagen. Tot 2 GB kan gratis opgeslagen worden. Backups van de gegevens worden verzorgd door het technisch beheer van DVN.

Een dataverse is gewoon een virtuele container om data-files in te plaatsen. Binnen een dataverse zijn Studies de manier om de eigenlijke data-files te organiseren en beheren. Een Study is een beschrijving (zie hieronder) van een onderzoek(sonderdeel) waarbij de bijbehorende data-files geplaatst kunnen worden.

Metadata en DVN

Het toekennen van (beschrijvende) metadata aan je onderzoeksdata is er in de eerste plaats op gericht dat deze begrepen kunnen worden door anderen. Daarnaast zijn metadata ook functioneel voor het kunnen terugvinden van je data door anderen maar dat is voor data in een fase dat je onderzoek nog niet is afgerond minder relevant.

Metadateren van data kan op verschillende niveaus: het niveau van het onderzoek of studie en het niveau van de data zelf. In het laatste geval moet gedacht worden aan bijvoorbeeld correcte, heldere en volledige omschrijvingen van de onderzochte variabelen (de kolommen in een tabel), de records (de rijen) en de gebruikte codes (1=man, 2=vrouw). Is er sprake van een survey dan is het makkelijk als in de data file een verwijzing naar de vragen wordt opgenomen. Etc.

Meestal zijn deze metadata ingebed in de data-file zelf. Vaak wordt hierin ook voorzien door de gebruikte data-analysesoftware. Dit niveau van metadatering – belangrijk voor een mogelijk later hergebruik van je data door anderen - komt in DVN niet aan bod.

Op het internet zijn diverse tools te vinden die je helpen bij het metadateren van je onderzoeksdata zelf. *DataUp* (<http://dataup.cdlib.org/>) bijvoorbeeld helpt je bij het beschrijven van (tabellarische) data in een Excel-file en bij het opsporen van zaken in een Excel-file die het archiveren en hergebruik ervan in de weg staan.

DVN ondersteunt het metadateren op het niveau van het onderzoek of studie. Voordat in DVN een data-file opgeslagen kan worden moet deze eerst beschreven worden ('cataloging information' geheten in DVN). Metadateren gaat in DVN dus vooraf aan opslaan. Het toevoegen van beschrijvende informatie gebeurt in DVN met behulp van een formulier waarop een aantal velden ingevuld kan worden. Het aantal velden dat ingevuld kan worden is groot maar slechts een veld – de titel van de data-file - is verplicht. Van een aantal velden wordt aangeraden deze in te vullen, een aantal hoeft niet per se.

Het gaat om onder meer de volgende gegevens:

- formele gegevens: gegevens als titel, auteur/maker, producent en datum van opslag van de data-file. Deze worden ook gebruikt bij het eventueel citeren van de data-file ;

- inhoudelijke gegevens: omschrijving, trefwoorden, eventuele geografische en dekking in tijd van de data-file ;
- gegevens over de wijze van data-verzameling en onderzoeksmethodologie. Meestal gegevens die standaard ook vermeld worden in de methodologie-sectie van een wetenschappelijk paper ;
- gegevens over vertrouwelijkheid en gebruiksvoorwaarden van de data. Relevant voor een mogelijk later hergebruik van de data door anderen.

Samengevat is het in DVN mogelijk om uitgebreid metadata toe te voegen aan een data-file. In de praktijk zal een onderzoeker dat niet altijd doen en bijvoorbeeld volstaan met het aangeven van de titel en auteur van de data-file. Zolang de data ‘voor eigen gebruik’ zijn hoeft dat geen probleem te zijn. Maar het zal duidelijk zijn dat dat met het oog op mogelijke beschikbaarstelling van je data aan anderen tijdens je onderzoek – maar ook aan jezelf als je bijvoorbeeld in een latere fase van je onderzoek moet terugkomen op je data – onvoldoende is.

Toegangscontrole en DVN

Nadat een data-file is beschreven kan de data-file zelf toegevoegd worden aan de metadata. Hiermee is de data-file nog niet beschikbaar voor anderen. Dat is pas het geval als de data-file plus metadata (‘Study’ geheten in DVN) vrijgegeven (‘released’) wordt. Er kunnen meerdere data-files aan een Study toegevoegd worden. Het uploaden van meerdere data-files tegelijk naar een Study is nog niet mogelijk in DVN.

Een eenmaal vrijgegeven Study krijgt in DVN automatisch een persistente URL en kan hierna niet meer verwijderd worden. Wel kunnen vrijgegeven Studies onzichtbaar (‘deaccessioned’) gemaakt worden.

Een vrijgegeven studie is in DVN default voor iedereen toegankelijk (‘public’). Wil je dat niet, dan moet de toegang tot de Study beperkt worden (‘restricted access’) waarbij aangegeven kan worden wie er toegang toe hebben. Alleen personen die bekend zijn bij DVN - een account hebben – kunnen toegangsrechten tot een Study krijgen. Dit kan ook iemand zijn die niet via een universitaire instelling is aangesloten bij DVN.

Toegangsrechten kunnen alleen door de eigenaar van de Study (‘Curator’) aan anderen toegekend worden. Toegang tot een data-file houdt in dat je de data-file mag downloaden. In DVN kan ook aangegeven worden dat downloaden pas mag nadat je toestemming hebt gevraagd en gekregen van de eigenaar van de data (of een door hem aangestelde contactpersoon).

Een eigenaar van een Study kan behalve downloadrechten (‘toegangs- of leesrechten’) ook andere rechten toekennen aan anderen, bijvoorbeeld het recht om data-files toe te voegen aan of zelfs te verwijderen van zijn Study (‘schrijf- of editrechten’). In dit opzicht heeft DVN dus kenmerken van een samenwerkingsomgeving. Het recht om een door anderen gewijzigde Study vrij te geven (en daarmee beschikbaar te stellen) blijft echter voorbehouden aan de eigenaar van de Study. DVN kent versiebeheer en bewaart eventuele verschillende versies van Studies (niet van de data-files sec). Onhandig in DVN is dat van een nieuwe versie opnieuw de rechten aangegeven moeten worden.

Samengevat is het in DVN mogelijk om gedetailleerd en op verschillende niveaus (Dataverse, Study en data-file) toegangs- en editrechten toe te kennen aan anderen.

Samenwerken en DVN

Het is hierboven al even genoemd: DVN als samenwerkingsomgeving, als elektronische omgeving dus die je in staat stelt om gezamenlijk te werken aan onderzoeksdata. In dit verband is het belangrijk je te realiseren dat DVN samenwerking mogelijk maakt op het niveau van de Study en niet op het niveau van de data-file zelf. Met andere woorden: data-samenwerking bestaat hieruit dat anderen data-files kunnen toevoegen aan je Study. Dit kan natuurlijk een - na downloaden - bewerkte versie zijn van de oorspronkelijke data-file in de Study maar deze kan in de Study niet onder dezelfde file-naam bewaard worden. Het gezamenlijk bewerken van een data-file waarbij ook de verschillende versies van de data-file automatisch bewaard worden is niet mogelijk in DVN.

Data-labachtigen

Als gezegd: een data lab stelt je in staat je onderzoeksdata (veilig) op te slaan, te beschrijven, aan anderen beschikbaar te stellen en er gezamenlijk aan te werken. Behalve DVN zijn er meer applicaties die, hoewel niet ontworpen met het oog op samenwerking en hergebruik van onderzoeksdata, deze functionaliteit aanbieden, althans gedeeltelijk. Dropbox en Google Drive bijvoorbeeld worden veel gebruikt door onderzoekers om ook data (tijdelijk) op te slaan en beschikbaar te stellen aan anderen. Dit kan tot op bepaalde hoogte voldoende zijn om samenwerking mogelijk te maken.

Leon Osinski
(07-02-2014)

Beschikbaar onder een [CC-BY licentie](#)