

Research Drive

Leidraad mappenstructuur en datadocumentatie

De Haagse Hogeschool en haar onderzoekers hebben de gedeelde verantwoordelijkheid voor het op een correcte en veilige manier opslaan van onderzoeksdata, de data te bewaren voor de vereiste duur en ze beschikbaar te stellen voor review rekening houdend met de geldende (wettelijke) voorwaarden.

Onderzoeksdata zijn van zichzelf van grote waarde en nuttig voor hergebruik. Het oorspronkelijk onderzoek kan worden gereproduceerd wat bijdraagt tot de validatie ervan. Daarbij kunnen subsidieverstrekkingen en uitgeverijen de voorwaarde opleggen dat de data worden gearchiveerd en gedeeld.

Deze leidraad bevat uitgangspunten voor het opslaan, archiveren en delen van onderzoeksdata. Nuttig voor zowel de eigen onderzoekers tijdens het onderzoeksproces als voor tijdens het archiveringsproces na het onderzoek.

Mapstructuur

In Research Drive wordt binnen de hoofdmap van het project de volgende standaard mappen aangeboden:

- *Projectdocumenten*: onderzoeksvoorstel, datamanagementplan, samenwerkings- / consortiumovereenkomst, autorisatiematrix, etc.
 - met sub-map *Toestemmingsformulieren*
[Wees omzichtig met het geven van toegang tot deze map.]
- *Ruwe data*
[Wees omzichtig met het geven van toegang tot deze map. Zorg dat identificeerbare persoonlijke data gescheiden van de bijbehorende onderzoeksdata worden opgeslagen gekoppeld door middel van een uniek referentie.]
- *Geanonimiseerde data*
[Zorg ervoor dat persoonlijke data in de map *Ruwe data* worden vernietigd wanneer ze niet langer nodig zijn om de doelstellingen van het project te bereiken.]
- *Analyses*
- *Publicaties en Producten*

De namen van de mappen worden voorafgegaan door de code van het project.

Binnen deze standaard mappen kan er een eigen mappenstructuur opgezet worden. Zorg voor een solide naamgeving van zowel je mappen als bestanden. Zo maak je het eenvoudiger om de onderzoeksdata te identificeren, te lokaliseren en terug te vinden. Gebruik logische en duidelijke bestands- en mapnamen en wees consequent in je naamgeving. Ook de velden in je spreadsheet verdienen dezelfde aandacht. En maak gebruik van een consistente datumnotatie.

Voorbeeld naamgevingsconventies

Bestandsnamen, geen naamgevingsstrategie toegepast	Bestandsnamen, naamgevingsstrategie toegepast
<ul style="list-style-type: none"> • Test_data_2013.xlsx • Project_Data.docx • Design for project.docx • Lab_work_Eric.docx • Second_test.xlsx • Meeting Notes Oct 23.docx 	<ul style="list-style-type: none"> • 20130503_DOEProject_DesignDocument_Smith_v2-01.docx • 20130709_DOEProject_MasterData_Jones_v1-00.xlsx • 20130825_DOEProject_Ex1Test1_Data_Gonzalez_v3-03.xlsx • 20130825_DOEProject_Ex1Test1_Documentation_Gonzalez_v3-03.xlsx • 20131002_DOEProject_Ex1Test2_Data_Gonzalez_v1-01.xlsx • 20141023_DOEProject_ProjectMeetingNotes_Kramer_v1-00.docx

Voorbeeld mappenstructuur in Research Drive met basismappen en submappen

Ook kan het werken met een gestandaardiseerde lijst van trefwoorden, een thesaurus, die door iedereen binnen het onderzoek gebruikt wordt voordelig zijn. Gestandaardiseerde ontologie en terminologie maakt het tevens mogelijk om verschillende datasets samen te voegen en te analyseren. Op bartoc.org vind je een overzicht van thesauri, ontologieën en classificaties. In Research Drive kan ook gewerkt worden met trefwoorden voor het beter vindbaar maken van de mappen en bestanden. In de handleiding van Research Drive kun je teruglezen hoe dit werkt.

Op [de onderzoeksondersteuningsite](#) (onder het Organiseren van data) vind je handige links met informatie over een goede strategie voor het organiseren van je onderzoeksdata. Volg indien voorhanden de standaardprocedures en werkstromen in het betreffende vakgebied. Communiceer de strategie met alle betrokkenen die data aanleveren of die met de data gaan werken.

Datadocumentatie

Om je onderzoeksdata bruikbaar te houden voor jezelf en anderen, is het verstandig om tijdens je onderzoeksproject regelmatig aandacht te besteden aan het documenteren van je data, en deze informatie bij te houden in een tekstbestand dat je bij je data opslaat. In je documentatie neem je informatie op over het verzamelen van de data, de data-invoer, de dataopslag en de verwerking van de data. Labjournaals of codeboeken fungeren ook als documentatie van een dataset, en worden indien mogelijk bij de data opgeslagen.

README.txt-bestand

Een overzicht van alle bestanden van de dataset met een beschrijving van de inhoud per bestand kun je vastleggen in een Readme.txtbestand. Het 4TU Centre for Research Data biedt [richtlijnen](#) voor het schrijven van zo een README.txt-bestand. Bij deze richtlijnen vind je ook een voorbeeld een goed README-bestand.

Metadata

Metadata is een speciale vorm van gestandaardiseerde datadocumentatie of 'data over data'. Niet alleen mensen maar ook computers kunnen metadata lezen, interpreteren en combineren. Het toekennen van metadata helpt bijvoorbeeld om het tijdstip van het verzamelen van de data, de verzamellocatie, de maker(s) en de gebruiksvoorwaarden van onderzoeksdata (licentie) boven water te krijgen.



De meest voorkomende typen metadata zijn:

Type metadata	Doel	Voorbeeld
Beschrijvende metadata	Dit zijn de minimale metadata, nodig om een digitaal object te kunnen vinden. Als er daarnaast ook nog contextuele metadata aanwezig is, krijgt een gebruiker meer inzicht hoe hij de data zelf kan gebruiken	Auteur, titel, abstract, datum Contextuele metadata zijn bijvoorbeeld locatie, tijd, methoden van dataverzameling (tools)
Structurele metadata	Deze leggen de relatie vast tussen individuele objecten die gezamenlijk een eenheid vormen	Links naar gerelateerde digitale objecten, (bijvoorbeeld het artikel dat geschreven is op basis van de gelinkte onderzoeksdata)
Technische metadata	Informatie over de technische aspecten van de dataset	Dataformaat, gebruikte hardware/software, calibratie, versie, authenticatie, encryptie, metadatastandaard
Administratieve metadata	Metadata die zich richten op gebruik(srechten) en beheer van digitale objecten	Licentie, eventuele redenen voor een embargo, waivers Search logs, user tracking

Daarnaast kunnen metadata worden ingedeeld op bestandsniveau en op dataniveau:

- Metadata op bestandsniveau:
 - De context van de dataverzameling: projectgeschiedenis, doel, doelstellingen en hypothesen
 - Methoden van dataverzameling: bemonstering, dataverzamelingproces, gebruikte instrumenten, gebruikte hardware en software, geografische dekking en gebruikte secundaire datasets
 - De datasetstructuur: relaties tussen bestanden
 - Datavalidatie: de procedures die uitgevoerd zijn m.b.t. controle, opschoning en kwaliteitsborging van de data
 - Wijzigingen die in de loop van de tijd aan data zijn aangebracht sinds hun oorspronkelijke creatie en identificatie van verschillende versies van databestanden
 - De juridische implicaties van het onderzoek inclusief privacy
 - Informatie over toegangs- en (her)gebruiksvoorwaarden of vertrouwelijkheid van de data
 - De archivering (bewaartermijnen en selectiecriteria) van (delen van) de dataset
- Metadata op dataniveau:
 - Benaming, labels en beschrijvingen voor variabelen, records en hun waarden
 - Verklaringen of definities van gebruikte codes en classificatieschema's
 - Definities van gebruikte acroniemen
 - Codes van en redenen voor ontbrekende waarden
 - Afgeleide data gecreëerd na verzameling, met code, algoritme of commandobestand

Metadata op dataniveau worden bij voorkeur ingebed in een databestand zelf. Mocht dit niet mogelijk zijn, dan kan een apart bestand met de benodigde informatie aan het databestand worden gekoppeld. Veel softwarepakketten voor data-analyse hebben faciliteiten voor data-annotatie en -beschrijving, als variabele attributen (labels, codes, datatype, ontbrekende waarden), datatype-definities, tabelrelaties, enz.

Het toekennen van metadata is een activiteit die vooral hoort bij het archiveren van data na het onderzoek. Maar zoals het spreekwoord zegt: een goed begin is het halve werk. Het is dan ook aan te raden om deze activiteit al tijdens het onderzoek op te pakken. Ook kunnen metadata al eerder hun nut bewijzen, bijvoorbeeld wanneer er nieuwe onderzoekers aan het onderzoek worden toegevoegd. Wel is het handig om bij het toekennen van metadata tijdens het onderzoek al rekening te houden met het latere archiveringsproces. Welke metadatavelden verplicht of wenselijk zijn, verschilt namelijk per dataarchief en onderzoekdiscipline. Het is dus wenselijk om met je metadata aan te sluiten bij de opslagplek waar je na het onderzoek je data gaat archiveren. Zo maakt Dans Easy, een Nederlands online archiveringssysteem voor het deponeren en hergebruiken van onderzoeksdata, gebruik van de metadatastandaard Dublin Core. Dit is een internationaal heel gangbaar generieke metadatastandaard die onder andere volgende velden bevat:

Metadataveld	Betekenis
Creator	De belangrijkste onderzoekers die betrokken zijn bij het produceren van de data
Title	Naam of titel van dataset
Date created	Datum waarop de data zijn gecreëerd
Description	Inhoudelijke beschrijving van de data
Audience	Publiek voor wie de dataset interessant is, beschreven in termen van onderzoeksdomeinen
Rights holder	De persoon of organisatie die houder is van het auteursrecht of intellectuele eigendomsrechten
Access Rights	Een basale keuze tussen Open Access of Restricted Access en een verplichte keuze voor het soort licentie indien Open Access wordt gekozen (CC0-1.0; CC-BY-4.0 etc)

Naast een generieke metadatastandaard kun je ook kijken naar de gangbare metadatastandaard in je eigen vakgebied. Op [de onderzoeksondersteuningsite](#) (onder het kopje Data archiveren) vind je handige links die een overzicht van metadatastandaarden per vakgebied geven.

Bron en meer info: [Essentials 4 Data Support](#) door [Research Data Netherlands](#)

- Mappenstructuur: <https://datasupport.researchdata.nl/start-de-cursus/iiionderzoeksfase/organiseren>
- Datadocumentatie: <https://datasupport.researchdata.nl/start-de-cursus/iiionderzoeksfase/datadocumentatie/>
- Gestandaardiseerde metadata: <https://datasupport.researchdata.nl/start-de-cursus/ivoogstfase/data-archieven/gestandaardiseerde-metadata>