

Data Management Policy

Hybrid Intelligence

This policy describes the principles for the careful handling of research data, which is considered very important by HI.

1 Regulations

All researchers adhere to the data management policies of their local universities. In addition, NWO guidelines apply for NWO-funded HI-projects. HI maintains a “code and data” collection, but publication of datasets is always done through the institutional repositories of the participating universities, and must comply with local policies.

2 Data management plan

A Data Management Plan is available at consortium level. Furthermore, all researchers should generate a data management plan per project following the institutional guidelines as each project has distinct data securing and handling needs. For NWO-projects, the [data management plan template](#) can be downloaded from the NWO website.

2.1 Annex1 - consortium DMP



Formulier Datamanagementplan

Dit formulier is bedoeld voor het opstellen van een datamanagementplan, op basis van de datamanagementparagraaf die u eerder bij uw aanvraag hebt ingediend. NWO verwacht dat u eventueel commentaar van referenten en/of de commissie op de datamanagementparagraaf in dit datamanagementplan verwerkt.

NWO vraagt alleen om opslag van voor hergebruik relevante data. Onder de definitie van 'data' vallen zowel verzamelde, onbewerkte alsook geanalyseerde, gegenereerde data. Hierbij zijn vele verschijningsvormen denkbaar; digitaal en niet-digitaal (zoals samples, ingevulde vragenlijsten, geluidsopnames, et cetera).

NWO hanteert voor dit datamanagementplan een format dat zoveel mogelijk aansluit op de richtlijnen voor datamanagement van Horizon 2020. Een toelichting op de vragen vindt u onderaan dit formulier. U wordt dringend geadviseerd om voor het invullen van dit plan ondersteuning te zoeken van het RDM-support office van uw instituut of universiteit.

U wordt verzocht het ingevulde formulier binnen vier maanden na toekenning van de financiering in te dienen. Het datamanagementplan wordt door NWO zo snel mogelijk beoordeeld. Indien nodig roept NWO voor beoordeling de hulp in van (data-)experts uit uw wetenschapsgebied. Zodra het datamanagementplan is goedgekeurd, kan het project van start gaan. NWO beveelt aan het datamanagementplan in de loop van het onderzoek bij te stellen als daar aanleiding toe is.

Het formulier moet worden ingediend via het online aanvraagstelsel [ISAAC](#). Een hoofdaanvrager is verplicht zijn/haar datamanagementplan via zijn/haar eigen ISAAC-account in te dienen. Datamanagementplannen die niet via ISAAC zijn ingediend, worden niet in behandeling genomen.

U kunt dit formulier ook invullen via [DMP-online](#) een web gebaseerde tool ontwikkeld door het Digital Curation Centre (DCC) die u stapsgewijs door het invulproces van dit formulier heen helpt.

1 Algemene informatie		
1.1	Naam aanvrager en projectnummer	Frank van Harmelen - 024.004.022
1.2.	Heeft u bij het opstellen van dit plan de hulp ingeroepen van de afdeling datamanagement support van uw instelling?	Ieder PhD/PD wordt binnen HI verzocht een DMP op te stellen en contact op te nemen met de Data Steward aanwezig bij het eigen instituut.
2 Omschrijving dataset		
2.1	Beschrijf de data die verzameld/gegenereerd worden.	Binnen dit grote en langlopende project (6 universiteiten, 27 wetenschappers, 10 jaar) zullen zeer diverse datasets worden gecreëerd en gebruikt. Het zal gaan om data over de interactie tussen menselijke gebruikers en intelligente machines. Ieder project bepaalt afzonderlijk zijn datasets en legt deze ook vast. Latere updates van het data management plan (jaarlijkse



updates zijn voorzien door het project management) zullen steeds concrete lijsten bevatten met specifieke datasets die zijn gebruikt of gecreëerd.

2.2 Om wat voor type en format data gaat het?

- (semi-)gestructureerde data in CSV, Excel of Linked Data format
- audio data van experimenten
- video data van experimenten

3 Opslag data

Gedurende het onderzoek

3.1 Wat is de omvang (MB/GB/TB) van de data en waar worden de data opgeslagen?

- De (semi)-gestructureerde data zal orde-grootte MB's of GB's zijn
- De audio-data zal orde grootte GB's zijn
- De video-data zal orde grootte GB's of TB's zijn

Gezien de diversiteit van de data per project zal gekozen worden voor opslag van de data bij het eigen instituut en zullen ook de al aanwezige opslag capaciteit hiervoor gebruikt worden indien aanwezig.

3.2 Is er op voldoende opslagcapaciteit tijdens het project?

Ja

Worden tijdens het onderzoek regelmatig back-ups van de data gemaakt? Wie is daarvoor verantwoordelijk?

Ja

De data zal tijdens het onderzoek opgeslagen zijn bij de deelnemende universiteiten (UvA, VU, UU, UG, LU, TUD). De verantwoordelijkheid voor back-ups van data ligt bij de deelnemende universiteiten, veelal de centrale IT dienst.

3.3 Welke kosten denkt u te maken voor de opslag van data tijdens het onderzoek? Specificeer deze en noem een realistisch bedrag. Hoe worden deze kosten gedekt?

Het project voorziet in €150.000 voor de aanschaf van reken- en opslag capaciteit.

In het geval van afwezigheid van voldoende of adequate opslagmogelijkheden, zal dit budget worden ingezet om cloud of vaste opslagmogelijkheden te creëren. Dit budget is een inschatting en op dit moment wordt geïnventariseerd wat er in de eerste vijf jaar nodig zal zijn voor welke projecten en in welke vorm.



Na het onderzoek	
3.4 Geef aan bij welke bestaande trusted repository de data na afloop van het onderzoek worden opgeslagen.	<p>Gezien de diversiteit van de data per project zal gekozen worden voor opslag van de data bij het eigen instituut en zullen ook de al aanwezige opslag capaciteit hiervoor gebruikt worden indien aanwezig.</p> <p>Elk van de deelnemende universiteiten (UvA, VU, UU, UG, LU, TUD) heeft infrastructuur om data langjarig te archiveren na afloop van het onderzoek. In principe wordt de data gearchiveerd bij de universiteit die tijdens het onderzoek de data in eigendom had. Bij geval van twijfel heeft de VU als penvoerder zich bereid verklaard om haar data-opslag infrastructuur (3 lagen voor open, semi-open en closed, en voor MB's, GB's en TB's) hiervoor beschikbaar te stellen.</p>
Wanneer de data niet in een trusted repository worden opgeslagen, geef dan aan hoe de data vindbaar, toegankelijk en herbruikbaar worden gemaakt.	-
3.5 Wordt er een Persistent Identifier gebruikt om de data vindbaar te maken?	<p>Deels wel, deels niet. Dit zal afhangen van de kosten hiervoor. Momenteel zijn de kosten voor data archivering in DataVerse met persistent identifier hoog (€3.50 per Gb per jaar). Tienjarige opslag kost dan €35.000 per terabyte.</p> <p>Opslag zonder persistent identifier is €0.22 per jaar per Gb (€2.200 per terabyte)</p> <p>De coördinator, VU, zal garanderen dat PIs worden aangevraagd voor alle gebruikte dataset en zal ook coördineren dat deze te vinden zijn vanuit een centrale toegang.</p>
3.6 Hoe lang worden de data bewaard?	Data worden bewaard voor minstens 10 jaar, in lijn met de VSNU richtlijn hierover.
3.7 Welke kosten denkt u te maken voor de opslag na het onderzoek? Specificeer deze en noem een realistisch bedrag. Hoe worden deze kosten gedekt?	<p>Kosten worden per project ondergebracht bij het eigen instituut, eventuele extra aanvragen voor extra kosten die gemaakt moeten worden voor opslag etc...worden aangevraagd bij het MT en goedgekeurd door de HI Executive Board.</p> <p>Afhankelijk van het gekozen bewaarmodel variëren de kosten van €35 tot €2.20 per Gb voor 10-jarige opslag. Het project zal hiervoor budget vrijmaken.</p>



4		Standaarden en metadata
4.1	Hoe worden de data gedocumenteerd? Welke metadata standaard wordt gebruikt om de data voor hergebruik geschikt te maken?	De meta-data standaarden variëren sterk per deelgebied en per datatype. Ook dit zal gedocumenteerd worden in de jaarlijkse update van het datamanagement plan.
5		Beschikbaar stellen van data
5.1	Zijn de data na afloop van het project open beschikbaar voor hergebruik? Zo nee, geef aan waarom de data niet geschikt zijn voor hergebruik.	Ja, zowel mogelijk al tijdens en zeker direct na het project voor minstens 10 jaar (VSNU richtlijn). –
5.2	Wanneer data niet of pas op termijn beschikbaar kunnen worden gesteld, geef dan de reden hiervoor aan. Wanneer een gedeelte van de data niet (direct) beschikbaar kunnen worden gesteld, geef aan om welk deel het gaat.	NVT
5.3	Worden er voorwaarden gesteld aan het hergebruik van de data? Indien van toepassing: zijn deze vastgelegd in een <i>consortium agreement</i> ?	De consortium agreement legt vast dat data onder Apache 2.0 licentie wordt vrijgegeven. Deze licentie staat hergebruik toe mits onder vermelding van de bron.



Toelichting

1. Algemene informatie

- 1.1. Vul de naam van de projectleider en het door NWO toegewezen projectnummer in.
- 1.2. NWO beveelt ten stelligste aan dat onderzoekers bij het opstellen van dit data management plan in een vroegtijdig stadium ondersteuning zoeken bij hun onderzoeksinstituut. Alle universiteiten en UMC's beschikken inmiddels over professionele ondersteuning voor research data management soms ondergebracht bij de universiteitsbibliotheek, soms bij de afdeling ICT. Een lijst van contactpersonen is te vinden op de website van NWO.

2. Omschrijving dataset

- 2.1 Beschrijf de data die tijdens het onderzoek gegenereerd zullen worden.
- 2.2 Welk type en format data worden er opgeslagen? NWO verstaat onder 'data' zowel verzamelde, onbewerkte data alsook geanalyseerde, gegenereerde data. Hierbij zijn alle (combinaties van) verschijningsvormen denkbaar; digitaal en niet-digitaal (zoals bijvoorbeeld samples, ingevulde vragenlijsten, geluidsopnames etc.)

3. Opslag data gedurende het onderzoek

- 3.1 Maak een realistische en zo concreet mogelijke schatting van de uiteindelijke omvang van de data en geef aan waar u van plan bent de data op te slaan tijdens het onderzoek. In geval van digitale data geeft NWO de voorkeur aan opslag door het centrale opslagcentrum van uw instelling, bijvoorbeeld de ICT-afdeling en/of de universiteitsbibliotheek.
- 3.2 Het is van belang dat er opslagcapaciteit is en, in het geval van digitale data, ook een back-up van uw data. Een automatisch back-up door de ICT-afdeling is veiliger dan een handmatige back-up. Het opslaan van data op laptops, harde schijven of externe media is in het algemeen riskant en zal daarom in principe door NWO niet worden goedgekeurd. Bij gebruik van externe diensten moet u zorgen dat er geen conflicten optreden met het beleid van onderzoekspartners of medefinancier(s) en met het beleid van uw afdeling of instituut, bijvoorbeeld over de beveiliging van gevoelige data.
- 3.3 Maak een schatting van de kosten die gemaakt gaan worden en geef een realistische schatting. Belangrijke factoren die de kosten bepalen:
 - a. het type data;
 - b. de benodigde capaciteit voor opslag en back-up;
 - c. de mate van handwerk bij het toekennen van metadata en het opstellen van overige documentatie zoals codeboeken en gebruikte queries in het statistische pakket;
 - d. de benodigde mate van beveiliging van de data;
 - e. het inhuren van externe (datamanagement)expertise.

Opslag data na afloop van het project

Bij voorkeur worden de data na afloop van het project langdurig opgeslagen bij een (inter)nationale trusted repository. Wanneer dit niet mogelijk is, dienen de data te worden opgeslagen door de institutionele repository. Neem vroegtijdig contact op met het beoogde data repository of archief over de eisen die gesteld worden aan het deponeren van data.



- 3.4 Geef aan in welke trusted repository de data worden opgeslagen. Er zijn een aantal internationale keurmerken die de mate van betrouwbaarheid van repositories vaststellen. Hiervan is het internationale Data Seal of Approval de eenvoudigste set criteria. Onder Trusted Digital Repository met keurmerk vallen repositories met een Data Seal of Approval, DIN-31644-, ISO-16363- of WDS/ICSU-certificering. Een overzicht van bestaande repositories met Data Seal of Approval kunt u vinden op deze [lijst met repositories](#).
- 3.5 Een persistent identifier helpt om een dataset voor anderen vindbaar en citeerbaar te maken. Bij opslag in een trusted repository zal het repository doorgaans een persistent identifier aanvragen en ter beschikking stellen.
- 3.6 De Nederlandse Gedragscode Wetenschapsbeoefening (2018) bepaalt dat zowel de ruwe data als de bewerkte versies gedurende een voor de discipline en methodologie passende termijn wordt bewaard. NWO beschouwt een termijn van 10 jaar als redelijk en minimaal.

4. Standaarden en metadata

Om data in de toekomst vindbaar, toegankelijk en herbruikbaar te houden moet de dataverzameling voorzien worden van beschrijvende informatie in de vorm van metadata. Zonder metadata – dus zonder een goede documentatie van de data – zijn de meeste datasets niet goed meer te interpreteren of te (her)gebruiken. Verschillende disciplines hanteren hun eigen metadata standaarden. Op die manier wordt ervoor gezorgd dat datasets binnen een discipline aan elkaar gelinkt kunnen worden of met elkaar gecombineerd kunnen worden. Een veel gebruikte standaard is het [Dublin Core Metadata Initiative DCMI](#). Het Digital Curation Centre (DCC) houdt een lijst bij van veel gebruikte metadatastandaarden voor verschillende disciplines: <http://www.dcc.ac.uk/resources/metadata-standards>.

5. Beschikbaar stellen van data

NWO verwacht dat data die voortkomen uit door NWO gefinancierde projecten na afloop van het project in principe voor hergebruik beschikbaar gesteld wordt. NWO hanteert hierbij in navolging van de Europese Commissie het adagium: “zo open als mogelijk, gesloten indien noodzakelijk”. Als er redenen zijn om data na afloop van het project niet (of pas na een embargotermijn) voor hergebruik beschikbaar te stellen, geef daarvoor dan een onderbouwing. Daarnaast is het aan te raden om te bepalen aan welke voorwaarden een onderzoeksgroep die toegang wil verkrijgen tot uw data, moet voldoen. Voorbeelden hiervan zijn afspraken die worden gemaakt rondom methodologie, publicaties, de termijn van toestemming, beschikbaarheid van data, de kosten (handling fee), auteursrechtelijke aspecten en dergelijke.

- 5.3 Geef aan of er embargo's, licenties, commerciële doeleinden of andere voorwaarden zijn die gesteld worden aan het hergebruik van de data. Indien van toepassing: zijn deze vastgelegd in een *consortium agreement*?