

ZonMw (Nederlands): Datamanagement ZonMw-template 2016-2018 - Datamanagement ZonMw

1. Kenmerken van het project en de dataverzameling

1.1 Contactgegevens projectleider

1.2 Ik heb mijn DMP opgesteld in samenwerking met een expert op het gebied van datamanagement. Noem naam, functie, organisatie/afdeling, telefoonnummer, e-mailadres.

- De expert is van buiten mijn vakgroep/instituut.
- De expert is verbonden aan mijn vakgroep/instituut.

1.3 Bij het verzamelen van de data voor mijn project ga ik als volgt te werk:

- Bestaande data gebruiken (noem)
- Nieuwe data genereren
- Databestanden koppelen (noem)
- Nieuwe data toevoegen aan een bestaande dataset (noem)
- Een MDS (Minimale Dataset) gebruiken/ afnemen

Guidance:

Dataverzamelingen zijn te vinden in online (metadata)catalogi en data-archieven. Enkele voorbeelden zijn:

- De [DANS website](#) levert algemene informatie over het vinden van data op het gebied van gezondheidsonderzoek en life sciences
- [Zorggegevens.nl](#) van het RIVM is een website met informatie over databestanden in de publieke gezondheid en zorg
- Het Centraal Bureau voor Statistiek ([CBS](#)) heeft veel [databestanden](#) op het gebied van gezondheid en welzijn
- [Datacentrum](#), o.a. voor biomedisch onderzoek.
- Biobanking and Biomolecular Research Infrastructure (BBMRI-NL) bevat 200 Nederlandse **biobanken** ([nlgenome.nl](#) & [www.biobanken.nl/](#)). BBMRI-NL verschaft ook een infrastructuur voor datamanagement voor grootschalige data-analyse genomicsdata en biobanken.
- [Nederlands Instituut voor Beeld en Geluid](#) voor beeldmateriaal.
- Ook [Bio-Medical Imaging Archive or BMIA](#) is een archief voor beeldmateriaal.
- [neuroinformatics.nl](#) is een platform waar neuroinformatici samenwerken en data delen.

NB 1 Deze opsomming is niet uitputtend.

NB 2 Bij het maken van plannen voor hergebruik van bestaande data is het raadzaam om advies te vragen over de voorwaarden waaronder dit mag. Denk bijvoorbeeld aan o.a. kosten, koppelingsvariabelen en informed consent.

1.4 Ik maak binnen mijn onderzoek gebruik van:

- Uitsluitend kwantitatieve data
- Uitsluitend kwalitatieve data
- Anders (leg uit)
- Een combinatie van kwantitatieve en kwalitatieve data

Guidance:

Het verifiëren en hergebruiken van data is van belang voor zowel kwantitatieve als kwalitatieve data. Het datamanagement van beide typen data verschilt op enkele punten. U

wordt in deze vragenlijst daarop gewezen.

1.5 Ik ga bestaande data hergebruiken en/of koppelen en ik heb toestemming van de data-eigenaar/eigenaren voor het gebruik/ koppelen van zijn/hun data.

- Toestemming is niet nodig, omdat de data open beschikbaar zijn
- Nee, ik ga geen bestaande data hergebruiken
- Ja, ik heb toestemming voor het gebruik van zijn/hun data
- Ja, ik heb toestemming voor het gebruik van zijn/hun data, maar moet deze na afloop van het project vernietigen

Guidance:

Bedenk bij gebruik van **bestaande data**:

- of er condities zijn voor het gebruik ervan en of er conflict kan komen met bestaand onderzoek.
- dat u voor de reproduceerbaarheid van resultaten vastlegt welke versie van de bestaande databestanden u gebruikt.

Converteer of vertaal verschillende standaarden in reeds beschikbare databestanden om te kunnen koppelen of databestanden te verrijken.

1.6 Bij het verzamelen van nieuwe data werk ik samen met andere partijen.

- Nee
- Ja, we hebben afspraken gemaakt over gebruiksrechten van de data uit het project.
- Ja, de te verzamelen nieuwe data worden (deels) door een partner van het project of toeleverancier aangeleverd.
- Ja, het verzamelen van nieuwe data doe ik samen met andere onderzoekers, onderzoeksgroepen.

Guidance:

Indien u samenwerkt met andere (onderzoeks)organisaties bij het verzamelen van onderzoeksgegevens, is het van belang dat u:

- duidelijk afspraken maakt over de toegankelijkheid, herbruikbaarheid, uitwisselbaarheid en controleerbaarheid van de (nieuwe) dataset;
- **afspraken** maakt over **eigenaarschap of co-producentschap** van data;
- **voorwaarden** opstelt voor gebruik van data door derden (bijv. co-auteurschappen, toestemming voor onderzoek, informeren over nieuwe onderzoeksrapporten en artikelen);
- de **verantwoordelijkheden** verdeelt in het onderzoeksproces;
- de afspraken schriftelijk vastlegt in een **samenwerkings- of consortiumovereenkomst**.

Meer informatie over ZonMw publiek-private samenwerking en co-financiering.

<https://www.zonmw.nl/nl/onderzoek-resultaten/life-sciences-health/cofinanciering/>

1.7 Het project voer ik uit in een consortium met twee of meer partners. Binnen het consortium zijn duidelijke afspraken gemaakt over datamanagement en intellectueel eigendom. (Denk hierbij ook aan wat het effect is van veranderingen binnen het consortium voor datamanagement en intellectueel eigendom)

- Ja, er zijn (nog) geen duidelijke afspraken gemaakt m.b.t. datamanagement en intellectueel eigendom (leg uit)

- Nee, ik werk niet met 2 of meer partners samen.
- Ja, er zijn duidelijke afspraken gemaakt m.b.t. datamanagement en intellectueel eigendom d.m.v. een consortiumovereenkomst.

1.8 Ik kan een inschatting maken van de omvang van de dataverzameling, namelijk het aantal deelnemers of subjects ("n=") van de dataverzameling en de grootte in giga-/terabytes.

- Ja (noem)
- Nog niet (leg uit)

Guidance:

Met de "n=" (voor het aantal deelnemers of subjects) en het aantal giga-/terabytes per deelnemer (of voor de totale verzameling) heeft u een schatting van **de omvang van de dataverzameling**. Deze heeft u nodig voor de planning van onder meer de opslagcapaciteit en het budget (*zie onderdeel 6 Data herbruikbaar maken en duurzaam opslaan*).

1.9 De volgende eindproducten van het project stel ik beschikbaar voor vervolgonderzoek en verificatie. (licht kort toe)

- Syntaxen
- Documentatie over de data
- Ruwe data
- Software
- Biobank
- Beeldmateriaal/ audiovisueel materiaal
- Documentatie van het onderzoeksproces, incl dat van alle betrokkenen
- Andere producten (noem)
- (verschillende versies van) bewerkte data

1.10 Gedurende het project heb ik voldoende opslaglocaties en -capaciteit en heb ik een back-up van de data beschikbaar. (Geef een korte toelichting.)

- Nee, ik heb nog geen opslaglocatie geregeld. (Leg uit)
- Ja, voor de opslag en back-up van mijn data maak ik gebruik van een externe dienstverlener.
- Ja, de opslag en back-up van mijn data doe ik in eigen beheer
- Ja, voor de opslag en back-up van mijn data maak ik gebruik van de standaardvoorzieningen van mijn instituut.

Guidance:

Let bij het opslaan van data op de volgende punten:

- Een goed georganiseerde **opslag** en **back-up** van data is nodig om te voorkomen dat data verloren gaan door technische storingen of menselijke fouten.
- Laat u adviseren over de **opslagfaciliteiten van uw onderzoeksinstituut. Dat kan bij** uw centrale ICT-afdeling, bij SURF, of de externe dienstverlener die u eventueel inhuurt.
- Opslag op laptops, harde schijven of externe media is in het algemeen riskant. Het gebruik van robuuste, beheerde opslagmogelijkheden die door de **ICT-afdeling van de instelling** worden aangeboden, verdient de voorkeur. Ook is automatische back-up door de ICT-afdeling veiliger dan handmatige back-up.
- Bij gebruik van **externe diensten** moet u ervoor zorgen dat er geen conflicten optreden met het beleid van de onderzoeksfinancier(s) (zoals een goede toegankelijkheid van data) of het beleid van uw afdeling of instituut (bijvoorbeeld de beveiliging van gevoelige data).

Indien u de opslag en back-up van uw onderzoeksgegevens **in eigen beheer** doet, denkt u

dan aan: de omvang en groei van de data, opslaglocaties en -capaciteit, versiebeheer, back-up en technische en organisatorische maatregelen ter beveiliging van de onderzoeksgegevens.

2. Wet- en regelgeving (incl privacy)

2.1 Ik ga mensgebonden onderzoek doen en ik ben op de hoogte van en mij houd aan de wet- en regelgeving betreffende privacygevoelige data

- Ja, ik ga mensgebonden onderzoek doen en houd mij aan de Algemene Verordening Gegevensbescherming (AVG)
- Mijn onderzoek is niet WMO-plichtig.
- Wet op de Geneeskundige Behandelingsovereenkomst
- Nee, ik ga geen mensgebonden onderzoek doen. Ga door naar onderdeel 3 (Data vindbaar maken)
- Ik verzamel en/of hergebruik lichaamsmateriaal en houd mij aan de Gedragscode Goed Gebruik van lichaamsmateriaal
- Mijn onderzoek is WMO-plichtig en ik laat mijn project toetsen door de Medisch Ethische Toetsingscommissie. Ik houd mij aan de Kwaliteitsborging Mensgebonden Onderzoek.

Guidance:

Achtergrondinformatie bij deze vraag vindt u op de [webpagina van Toegang tot data over wet- en regelgeving](#), in het bijzonder onder de tabs:

Bescherming van persoonsgegevens: *Achtergrond*

Bescherming van persoonsgegevens: *Stappen in de bescherming van persoonsgegevens*

Bescherming van persoonsgegevens: *Toetsing door METC of CCMO*

2.2 Ik ga mensgebonden onderzoek doen en ik heb geregeld dat ik de data verkrijg met (een vorm van) toestemming van de deelnemers.

- Ja. (noem de vorm van toestemming)
- Ja, de toestemingsvorm maakt hergebruik van data mogelijk (In de gedragscode Gezondheidsonderzoek wordt voor 'hergebruik' ook wel de term 'nader gebruik' gebruikt)

Guidance:

Achtergrondinformatie bij deze vraag vindt u op de [webpagina van Toegang tot data over wet- en regelgeving](#), in het bijzonder onder de tab:

Bescherming van persoonsgegevens: *Toestemming*

2.3 Ik ga mensgebonden onderzoek doen en ik ga privacygevoelige data beschermen tegen datalekken en/of misbruik.

- Anders (leg uit)
- Ja, ik laat de data pseudonimiseren. (leg uit hoe en door welke instantie dat gebeurt)
- Ja, ik laat de data anonimiseren. Ik realiseer me dat ik daarmee de mogelijkheden voor hergebruik van data beperk. (leg uit)

Guidance:

Achtergrondinformatie bij deze vraag vindt u op de [webpagina van Toegang tot data over wet- en regelgeving](#), in het bijzonder onder de tab:

Bescherming van persoonsgegevens: Beschermingsmaatregelen

2.4 Ik houd me aan het privacyreglement van de organisatie waaraan ik verbonden ben.

- Ja
- Mijn organisatie heeft geen regeling zoals hier bedoeld. Daarom controleer ik zelf de naleving van de maatregelen die zijn genomen om de privacy en de niet-herleidbaarheid van de data te waarborgen

3. Data vindbaar maken

3.1 De dataverzameling uit mijn project is vindbaar voor vervolgonderzoek. (Let op: Dit is een kernegegeven dat u aan het einde van uw project aan ZonMw door moet geven.)

- Nee, ik heb nog geen keuze gemaakt voor een archief of catalogus/webportal
- Ja, via een online (metadata-)catalogus of webportaal (noem)
- Ja, via de zoekmachine van het archief (repository) waarin hij is opgeslagen (noem)

Guidance:

Databestanden zijn te vinden via **data-archieven (repositories)**, **(metadata-)catalogi** en **webportals**. Zij geven een overzicht van de databestanden in combinatie met beschrijvende informatie (data over data, ofwel **metadata**).

Een data-archief of repository is bedoeld voor de **lange termijn-opslag** van uw data na afloop van uw project. Meer informatie en vragen hierover vindt u in *onderdeel 6 Herbruikbare data en duurzame opslag* van dit DMP.

U wordt aangeraden om al in een vroeg stadium van uw project een data-archief te kiezen. U kunt dan bij het opzetten van de dataverzameling **rekening houden met de vereisten** van het archief, zoals een passend *file format* en eventueel voorgeschreven metadata.

Indien u gewend bent om uw data in **internationale repositories** op te slaan, kunt u dat blijven doen. Zeker wanneer dit wordt vereist door co-financiers of wetenschappelijke tijdschriften. Meer informatie staat op o.a. [EMBL EBI](#). In dat geval hoeft u ook niet nog uw data op te slaan in een Nederlands data-archief.

(Zie ook vraag 1.3)

3.2 Voor de beschrijving van de dataverzameling gebruik ik een metadataschema.

- Ja, ik kies een metadataschema uit het overzicht van Datacite (noem)
- Ja, ik kies een metadataschema dat specifiek is voor mijn vakgebied (noem)
- Nee, ik heb nog geen keuze gemaakt voor een metadataschema

Guidance:

Een metadataschema gebruik je voor de beschrijving van:

- het databestand als geheel (daar gaat deze vraag 3.2 over);
- de beschrijving van het databestand (op recordniveau) (daar gaat vraag 5.2 over).

Bij deze vraag (3.2) gaat het over metadata die informeren over: Wie heeft de data verzameld? Onder welke omstandigheden? Wat is er gemeten/vastgelegd?

Er bestaan verschillende metadataschema's:

- generiek of vakspecifiek
- veel of weinig gedetailleerd.

Idealiter is de beschrijving van de data zo gedetailleerd, dat een potentiële gebruiker kan beslissen of de data passend zijn voor zijn onderzoek.

Om databestanden nu en in de toekomst goed vindbaar en herbruikbaar te laten zijn, moeten

de metadata volgens de principes van [FAIR-data](#) ook computer leesbaar ('machine readable') zijn.

Een **archief** (of repository) kan een metadataschema voorschrijven. Als dat niet zo is, kunt u er zelf een kiezen.

Er zijn verschillende **overwegingen voor een metadataschema** :

- Kies een voor uw vakgebied **specifiek** (passend) **metadataschema**. Dit maakt een gedetailleerde beschrijving mogelijk en bevordert de vindbaarheid en het hergebruik van de data binnen een vakgebied. Ook verbetert het de mogelijkheid om te koppelen met andere databestanden.
- Voorbeelden van **internationale schema's** zijn te vinden op [DataCite](#). DataCite biedt metadataschema's voor verschillende soorten databestanden. Ieder schema geeft een aantal data-items die voor de vindbaarheid nodig zijn. De data-items kunnen worden gevuld met keywords of als free text.
- Het [Data Documentation Initiative Alliance](#) is onder meer voor het documenteren van sociaalwetenschappelijk en gezondheidsonderzoek.
- Kies ook een **generiek metadataschema**, zodat de dataverzameling "geogost" kan worden door een dataportal. Het NARCIS-portaal van de KNAW is hiervan een voorbeeld. Bekend zijn de [Dublin Core standaard](#).

3.3 Ik zal gebruik maken van een Persistent Identifier (PI) om duurzaam naar het databestand te verwijzen. (Let op: Dit is een kerngegeven dat u aan het einde van uw project aan ZonMw door moet geven.)

- Ja, de DOI-code
- Nee, ik maak geen gebruik van een PI (leg uit)
- Ja, naast de DOI-code gebruik ik nog een PI (noem)

Guidance:

Gebruik een **Persistent Identifier** (PI) om duurzaam naar het databestand te verwijzen. Een PI wordt aangemaakt door duurzame archieven (of repositories). Deze moet u aan het einde van uw project aan ZonMw doorgeven.

Voorbeelden van PI's zijn bijvoorbeeld [DOI](#), Handle, URN of ARK.

Het voordeel van een Persistent identifier boven gewone links (URL's) is, dat uw databestand **vindbaar** blijft, ook wanneer de locatie verandert.

Daarnaast kan uw databestand met een PI worden **gecteerd** (datacitatie).

Meer over persistent identifier en DOI's is te vinden bij de [International DOI Foundation \(IDF\)](#), [Datacite](#), bij [TUDelft](#) of in de [RDNL cursus](#).

Tot slot

Verwar de persistent identifier niet met de identificerende code (*zie onderdeel 5 Interoperabel*). Dat is namelijk een code voor de records binnen een databestand.

4. Data toegankelijk maken

4.1 Na afloop van het project zullen de data toegankelijk zijn voor verificatie en vervolgonderzoek.

- Nee (leg uit)
- Ja, direct na afloop van het project.
- Ja, na een embargoperiode (leg uit)

Guidance:

Het kan zijn dat het nodig is om een embargoperiode in acht te nemen, waarin uw data voor een zekere periode niet volledig openbaar zijn. Bijvoorbeeld vanwege publicaties, openbare veiligheid, privacy, intellectueel eigendom dat berust bij betrokken bedrijven, of commerciële

belangen die een rol spelen bij de exploitatie van onderzoeksresultaten. De subsidiebepalingen van ZonMw noemen een embargoperiode van 3 maanden en bij patenten maximaal 9 maanden.

Leg uit wat de reden is voor de embargoperiode en de duur ervan.

ZonMw zal indien nodig in overleg met u de embargoperiode vaststellen.

Voor meer informatie zie:

- Informatie over [cofinanciering](#) op de ZonMw website.
- [Toelichting verschillende eigendomsrechten](#)

4.2 Na afloop van het project wordt het databestand openbaar toegankelijk, zonder aanvullende voorwaarden (open access).

- Nee, ik verbind voorwaarden aan de toegang tot het databestand (restricted access) (leg uit)
- Ja. U kunt naar het volgende onderdeel van dit DMP (5. Interoperabel)

4.3 Ik heb gebruiksvoorwaarden beschikbaar waarmee ik de voorwaarden voor toegang tot mijn databestand uitleg na afloop van het project. Geef een link of Persistent Identifier (Let op: Dit is een kernegegeven dat u aan het einde van uw project aan ZonMw door moet geven.)

- Ja, mijn instituut heeft de gebruiksvoorwaarden in samenwerking met een jurist opgesteld
- Nee, mijn instituut gaat de gebruiksvoorwaarden in samenwerking met een jurist opstellen.
- Ja, mijn instituut maakt gebruik van internationaal beschikbare standaard gebruiksvoorwaarden.

Guidance:

Bij **restricted access** moet u nauwkeurig omschrijven aan welke **voorwaarden** een onderzoeksgroep, die toegang wil verkrijgen tot uw onderzoeksgegevens, moet voldoen.

Volgens de [FAIR guiding principles](#) moet de **juridische** positie van de licenties en de voorwaarden voor gebruik helder zijn om hergebruik van data ook werkelijk mogelijk maken.

U kunt internationaal beschikbare **standaarden** nemen, of u stelt de voorwaarden zelf op in samenwerking met een **jurist**.

De gebruiksvoorwaarden moeten door uw **instituut of vakgroep** beschikbaar gesteld worden en mogen niet persoonsgebonden zijn.

ZonMw verzoekt u in de gebruiksvoorwaarden zoveel mogelijk van de criteria te regelen die in de volgende vraag staan genoemd.

4.4 In de voorwaarden die ik stel aan het gebruik van mijn data (restricted access), heb ik in ieder geval de hieronder aangekruiste punten opgenomen.

- Of de dataset mag worden gekoppeld aan een andere dataset (privacy)
- Samenwerking bij het gebruik van de dataset, inclusief afspraken over publicaties, auteurschappen.
- Afspraken over methodologie.
- Anders (leg uit)
- Voorwaarden met betrekking tot dataveiligheid.
- De manier waarop de dataset beschikbaar wordt gemaakt.
- De periode van toestemming voor gebruik van de dataset.
- Over de goedkeuring van data-aanvragen zal een stuurgroep, programmacommissie of projectleider beslissen.
- Toestemming van de deelnemers laat vervolgonderzoek met de dataset toe

- Het delen van data voor commerciële doeleinden. Daarbij houd ik rekening met de bepalingen van het staatssteunrecht
- De vergoeding van kosten, bijvoorbeeld voor het verkrijgen van data.

Example Answer:

Zie ook uitleg bij vraag 4.3

5. Data interoperabel maken (uitwisselbaar, koppelbaar)

5.1 Ik kies een dataformat zodat mijn data leesbaar zijn voor andere onderzoekers en hun computers ('machine actionable').

- Ja (noem)
- Nee (leg uit)

Guidance:

Om uw onderzoeksdata **interoperabel** te maken is het raadzaam dat u een **ICT-standaard (of data-format)** kiest voor het vastleggen en coderen van uw data die in **uw vakgebied** gangbaar is. Anders gezegd, let er bij de keuze van de standaard op dat uw data zowel leesbaar zijn voor andere onderzoekers (eenheid van taal), als hun computers ('machine readable' of 'actionable').

De keuze die u maakt, is verder afhankelijk van de mogelijkheden van uw instituut. Daarom moet u zich zeker laten adviseren door uw ICT- en/of datamanager.

Daarnaast kunt u informatie inwinnen bij experts en/of op de website van hun organisaties, bijvoorbeeld:

- [Nictiz](#); zie 'vierde laag van interoperabiliteit' (applicatie)
- Voorbeelden van **internationale schema's** zijn te vinden op [DataCite](#). Zie ook de guidance bij 3.2.
- Research Data Netherlands [RDNL](#) (of de daarin samenwerkende instituten DANS, 3TU.Datacentrum, SURF),
- [Dutch Techcentre for Lifesciences DTL](#)
- Indien u gebruik maakt van het datamanagementsysteem Castor of Open Clinica, zijn uw data leesbaar en uitwisselbaar met andere computers

Deze organisaties bieden (online) informatie, dienstverlening en faciliteiten voor ICT(-standaarden) en data-opslag en overleggen regelmatig met de ICT-afdelingen van alle Nederlandse onderzoeksinstellingen.

ZonMw heeft met hen afspraken over de mogelijkheid om advies te vragen over datamanagement in ZonMw-projecten.

5.2 Ik kies een metadatastandaard zodat mijn data gekoppeld kunnen worden aan andere data. (Let op: Dit is een kerngegeven dat u aan het einde van uw project aan ZonMw door moet geven).

- Nee (leg uit)
- Ja, ik kies een metadatastandaard uit het overzicht van Biosharing (noem)

Guidance:

Om onderzoeksdata interoperabel te maken is het raadzaam om gebruik te gebruiken van een metadatastandaard die gangbaar is in uw vakgebied. (zie 'derde laag van interoperabiliteit' van [Nictiz](#)) Denk hierbij aan terminologiestandaarden, classificaties en informatiestandaarden.

Let op dat u dit vanaf het begin van uw project goed regelt. Achteraf is het veel meer werk!

Veel gebruikte metadatastandaarden zijn:

- [SNOMED CT](#)
- [DataCite](#). DataCite biedt metadataschema's voor verschillende soorten databestanden.
- De website van [Biosharing](#) biedt een overzicht van metadatastandaarden in de medische wetenschappen.
- Het datamanagementsysteem [Castor EDC](#) maakt gebruik van de aanbevolen metadatastandaarden.

Let op

In vraag 3.2 werd u ook gevraagd naar een metadataschema, welke betrekking had op de beschrijving van het hele databestand. Vraag 5.2 daarentegen betreft de uitwisselbaarheid van gegevens op record niveau. Om twee soortgelijke databestanden aan elkaar te kunnen koppelen, moeten de variabelen op dezelfde wijze gedocumenteerd en gecodeerd zijn.

5.3 Ik ga mensgebonden onderzoek doen en heb bij de privacybescherming rekening gehouden met hergebruik van de data en eventuele koppeling met andere databestanden.

- Nee (leg uit)
- Ja, de deelnemers hebben toestemming gegeven voor hergebruik van de data en de data zijn gepseudonimiseerd

Guidance:

Hier wil ZonMw nagaan of u de bescherming van de privacy zo heeft geregeld dat koppeling van uw data met andere databestanden in principe mogelijk is.

Bij **mensgebonden onderzoek** is koppeling gebonden aan ethische en juridische voorwaarden. Als de persoon uitdrukkelijk **toestemming** heeft gegeven voor nader gebruik van zijn data (inclusief koppeling aan andere datasets), mag de **identificerende** code een BSN-nummer, geboortedatum of postcode zijn. Als er geen toestemming is, moet de code een **pseudoniem** zijn, die niet op de persoon herleidbaar is.

Wanneer de data **geanonimiseerd** zijn, is er geen toestemming nodig, maar is koppeling ook niet mogelijk.

Een trusted third party zoals [ZorgTTP](#) kan ondersteunen bij het uitwisselen en ontsluiten van databestanden met privacygevoelige informatie.

6. Data herbruikbaar maken en duurzaam opslaan

6.1 Ik zorg voor een goede kwaliteit en documentatie van de data opdat andere onderzoekers ze kunnen interpreteren en gebruiken ('replication package').

- Ik neem daarnaast nog andere maatregelen t.b.v. de kwaliteit (noem)
- Ik documenteer het onderzoeksproces (leg uit)
- Ik leg informatie vast over de gebruikte software (noem)
- Ik voer kwaliteitscontroles uit op de data opdat ze compleet, correct en consistent zijn (leg uit)

Guidance:

Met het oog op hergebruik van onderzoeksgegevens en de mogelijke replicatie van het onderzoek moet u minimaal de volgende **data** en **documentatie** archiveren:

- de ruwe data (indien u bestaande data hergebruikt, is er in het onderzoek wellicht geen sprake van "ruwe" data, maar bewerkte data; in dit geval moeten de bestaande data duurzaam gearchiveerd zijn of alsnog gearchiveerd worden);
- de data die ten grondslag liggen aan publicaties;
- documentatie over de gehanteerde onderzoeksmethodologie (zoals codeboeken, data manuals, metadatering, machine instellingen, gebruik van SOP's, versiebeheer, e.d.), projectvoorstel, goedkeuringen van de METC e.d., alle betrokkenen (onderzoekers,

laboranten, assistenten, proefpersonen e.d.), kortom, alles wat nodig is om “het spoor terug” te volgen.

- Beschrijf voor **kwalitatieve data** de procedure die is gebruikt voor het transcriberen van data (inclusief conventies en symbolen). Kijk ook eens naar [Atlas Ti](#).

Dit geheel wordt wel aangeduid als “**replication package**”.

Een goede uitwerking van dit minimum pakket aan informatie, dat is toegesneden op onderzoek dat valt binnen de reikwijdte van de Wet Medisch-wetenschappelijk onderzoek met mensen (WMO), is te vinden in de publicatie “[Kwaliteitsborging mensgebonden onderzoek 2.0](#)” (NFU, 2012). Bijlage 3 hiervan bevat de lijst te archiveren documenten bij de voorbereiding, de uitvoering en de voltooiing van dit type onderzoek.

6.2 Ik heb selectiecriteria om aan het eind van het project te bepalen welk deel van de data moet worden bewaard. (zie uw antwoord op vraag 1.9)

- Nog niet
- Ja
- De data (of een deel ervan) moeten na afloop van het project vernietigd worden, vanwege een contract of de wet.

Guidance:

Onderzoeksdata dient u te bewaren voor hergebruik of om onderzoeksresultaten te verifiëren, valideren. Maar dit geldt niet voor alle data. In sommige situaties is het opnieuw genereren van exact dezelfde data eenvoudiger of minder kostbaar dan het bewaren van die data. Het opnieuw genereren van data zal onder omstandigheden ook minder privacygevoelig kunnen zijn dan het bewaren ervan.

Om te bepalen welke onderzoeksdata waardevol zijn als (bron)materiaal voor verder onderzoek heeft DANS een **checklist** opgesteld met ‘[Algemene richtlijnen voor het selecteren van onderzoeksdata om te bewaren](#)’. Deze bevat de belangrijkste redenen voor het opslaan van onderzoeksdata voor de lange termijn. De checklist kan gebruikt worden door individuele onderzoekers, onderzoeksgroepen, instellingen, beheerders van archieven en door financiers.

Ook heel inzichtelijk is het **beslisschema** voor de **selectie van data** in de [cursus van Research Data Netherlands](#) (RDNL).

6.3 Aan het einde het project kan ik, na de selectie van de data, een inschatting maken van de omvang van het databestand (in Gb/Tb) dat ik voor lange termijn ga opslaan/ archiveren.

- Nog niet (leg uit)
- Ja (noem)

6.4 Ik zorg dat ik aan het eind van het project een keuze heb gemaakt voor een archief of repository voor duurzame lange termijnarchivering (gecertificeerd) van mijn databestand. (Let op: Dit is een kernegeven dat u aan het einde van uw project aan ZonMw door moet geven.)

- Ja, het archief heeft een data seal of approval (noem het archief)
- Nog niet
- Ja, het archief heeft een andere certificering (noem certificering en archief)
- Ja, het archief volgt de criteria (en heeft de intentie zich te laten certificeren) (leg uit hoe uw data op lange termijn toegankelijke en herbruikbaar blijven)

Guidance:

In *onderdeel 3 (Data vindbaar maken)* heeft u al het archief (of repository) genoemd waar uw data **vindbaar** zijn. Is het archief van uw keuze geschikt voor een duurzame opslag / archivering op lange termijn, zodat data **duurzaam opgeslagen** en **herbruikbaar** zijn? Voor **duurzame lange termijnarchivering** zijn internationale richtlijnen beschikbaar. De

eenvoudigste set criteria is die van het internationale [Data Seal of Approval](#) (DSA). Deze **criteria** en het bijbehorende **keurmerk** zijn onafhankelijk van het vakgebied. Een archief dat data duurzaam opslaat is een **trusted digital repository**. In Nederland hebben **CentERdata**, **DANS**, **4TU.Datacentrum** en **SURF** dit certificaat.

Specifieke **zorgdatarepositories** met het DSA zijn er nog nauwelijks. Een onderzoeker kan echter bij bestaande brede repositories terecht (zie de lijst met [Seals](#) op genoemde website). Opslag en archivering bij **uw eigen instituut** is ook mogelijk, maar let daarbij op dat zoveel mogelijk rekening wordt gehouden met de richtlijnen.

Als u kiest voor een repository die niet beschikt over het Data Seal of Approval, DIN-31644-, ISO-16363- of WDS/ICSU-certificering, maak dan duidelijk hoe ervoor gezorgd wordt dat de data na afloop van het project effectief toegankelijk en herbruikbaar zijn.

Wij raden u aan om bij uw instituut na te gaan wat er al mogelijk is aan duurzame opslag en archivering: in hoeverre werkt deze opslagfaciliteit volgens de criteria? En in hoeverre heeft hij de intentie om certificering voor duurzame opslag aan te vragen? Wijs de beheerder van de interne opslag eventueel op deze mogelijkheid.

6.5 Na afloop van het project, zal ik voor mijn data de aanbevolen bewaartermijn van minimaal 10 jaar hanteren.

- ja volgens de VNSU richtlijnen (noem aantal jaren)
- Ja, het aantal jaren is [noem], volgens de richtlijn [noem] (leg uit)
- Nee, noem het aantal jaren en richtlijn (leg uit)

Guidance:

Verschillende codes en richtlijnen hanteren verschillende bewaartermijnen. ZonMw vraagt u om tenminste die van de **Nederlandse Gedragscode Wetenschapsbeoefening (VSNU)** te volgen: “De bewaartermijn van ruwe onderzoeksgegevens is minimaal 10 jaar” (*zie ook onderdeel 2. Wet- en regelgeving*).

Ook in de publicatie “**Kwaliteitsborging mensgebonden onderzoek 2.0**” (NFU, 2012) staan diverse bewaartermijnen voor uiteenlopend materiaal. (zie vraag 6.1)

6.6 De kosten voor datamanagement gedurende het project en de voorbereiding van archivering mogen op de projectbegroting. De kosten zijn:

- Bedrag is [noem] (licht toe)
- Onbekend (leg uit)

6.7 De kosten van lange termijn archivering van de dataset, na afloop van het project, zijn gedekt.

- Ja (leg uit)
- Nog niet (leg uit)

Guidance:

Bij het **beramen van de kosten** moet u onderscheid maken tussen:

- (1) de kosten van datamanagement **tijdens het project** en het voorbereiden van de data voor archivering en
- (2) de kosten voor duurzame opslag en archivering **na afloop** van het project.

Ad (1) Deze kosten kunt u op de begroting van uw subsidieaanvraag zetten. De kosten van datamanagement hebben voor een deel betrekking op:

- de tijd die besteed moet worden aan metadatering en documentatie tijdens het onderzoek;
- **het type data** en de **omvang** van de verzameling (u heeft daarvan een inschatting moeten maken in *onderdeel 1. Kenmerken van het project*);

- de opslag en beveiliging van de data tijdens het project.

Ad (2) De kosten voor lange termijn opslag worden nu door de instituten gedragen. Er bestaat echter geen goed beeld van de kosten en waaruit ze bestaan. De kosten worden in ieder geval bepaald door het type data en de omvang (aantal deelnemers en giga-/terabytes) van de **geselecteerde dataset** (*zie het selectieproces in vraag 2 en 3*).

Diverse partijen werken aan **kostenmodellen** voor duurzame dataopslag, maar kennis om te komen tot passende financieringsmodellen voor opslag van data ontbreekt meestal nog. Via de [cursus van RDNL](#) kunt u informatie inwinnen over de kosten van archivering.

ZonMw (Nederlands): Datamanagement ZonMw-template 2016-2018 - Data management ZonMw (English version)

1. General features of the project and data collection

1.1 Project leader contact details

1.2 I have composed my DMP with the assistance of a data management expert. List his or her name, function, organisation/department, phone number and email address.

- The expert is not connected to my department or institution
- The expert is connected to my department or institution

1.3 In collecting data for my project, I will do the following:

- Use an MDS (Minimal Data Set)
- Generate new data
- Add new data to an existing data set (please specify)
- Use existing data (please specify)
- Merge different data files (please specify)

Guidance:

Data collections can be found in online metadata catalogues and data archives. Examples:

- The [DANS website](#) provides general information on how to find data on health research and life sciences
- RIVM's zorggegevens.nl is a website with information on data sets on public health and healthcare
- Statistics Netherlands ([CBS](#)) has lots of [data sets](#) on health and welfare
- [Datacentrum](#), incl. for biomedical research.
- Biobanking and Biomolecular Research Infrastructure (BBMRI-NL) contains 200 Dutch **biobanks** (nlgenome.nl & www.biobanken.nl/). BBMRI-NL also provides an infrastructure for data management for large-scale data analysis of genomics data and biobanks.
- The [Netherlands Institute for Sound and Vision](#) for information regarding image libraries.
- [Bio-Medical Imaging Archive or BMIA](#) is an image library.
- [Neuroinformatics.nl](#) is a platform where neuroinformatics experts collaborate and share data.

NB 1 This list is not comprehensive.

NB 2 When drawing up plans to reuse existing data it is wise to seek advice concerning the terms and conditions that apply. This will include things like costs, linkage variables and informed consent.

1.4 In my research, I will use:

- Exclusively quantitative data
- A combination of quantitative and qualitative data
- Other (please specify)
- Exclusively qualitative data

Guidance:

The verification and reuse of data is important for both quantitative and qualitative data.

Management of these two types of data differs in some respects. The questionnaire draws attention to this fact.

1.5 I will be reusing or combining existing data, and I have the owner's permission for using or combining their data.

- Yes, I have permission to use the data
- No, I will not be reusing or combining existing data
- Yes, I have permission to use the data, but I am required to destroy them at the end of the project
- No permission is required, since the data are openly accessible

Guidance:

When using existing data consider:
whether conditions apply to the use of the data and whether conflicts could arise with existing research;
the fact that for the reproducibility of results you need to note which version of the existing data sets you use.

Convert or translate different standards in pre-existing data sets in order to link or enhance them.

1.6 In collecting new data, I will be collaborating with other parties.

- No
- Yes, we have reached agreements on the user rights of the data used in the project
- Yes, the new data will be (partly) provided by a project partner or supplier
- Yes, I will collect the new data in conjunction with other researchers or research groups

Guidance:

If you are collaborating with other organisations on the gathering of research data, it is important to:

- make clear arrangements concerning the accessibility, reusability, exchangeability and verifiability of the new data set;
- **agree on ownership or co-producership** of data;
- draw up **terms and conditions** for the use of data by third parties (e.g. co-authorship, permission for research, information on new research reports and papers);
- allocate **responsibilities** in the research process;
- record all these arrangements in writing in the form of a **collaboration or consortium agreement**.

For more information on ZonMw public-private partnership and co-funding:

<https://www.zonmw.nl/nl/onderzoek-resultaten/life-sciences-health/cofinanciering>

1.7 I am a member of a consortium of 2 or more partners. Clear arrangements have been made regarding data management and intellectual property. (also consider the possible effect of changes within the consortium on issues of data management and intellectual property)

- Yes, clear arrangements have been made regarding data management and intellectual property through a consortium agreement
- No, I am not working with 2 or more partners
- Yes, I am a member of a consortium of 2 or more partners, but clear arrangements have not (yet) been made regarding data management and intellectual property (please explain)

1.8 I can give an estimate of the size of the data collection; specifically, the number of participants

or subjects ("n=") in the collection and its size in GB/TB

- Not yet (please explain)
- Yes (please specify)

Guidance:

Use "n=" (number of participants or subjects) and the number of giga-/terabytes per participants (or for the entire collection) to estimate **the size of the data collection**. You will need this to plan for things like storage capacity and budget (*See part 6 Making data reusable and permanent storage*).

1.9 The following end products I will make available for further research and verification (please elaborate briefly)

- Several versions of processed data
- Raw data
- Other products (please specify)
- Data documentation
- Audiovisual material/ Images
- A biobank
- Syntaxes
- Software
- Documentation of the research process, including documentation of all participants

1.10 During the project, I will have access to sufficient storage capacity and sites, and a backup of my data will be available. (please elaborate briefly)

- Yes, I will personally facilitate the storage and backup of my data
- No, I have not yet secured a storage location (please explain)
- Yes, I will make use of an external provider's services for storage and backup of my data
- Yes, I will make use of my institution's standard facilities for storage and backup of my data

Guidance:

Please consider the following when it comes to data storage:

- The properly organised **storage** and **backup** of data is necessary to prevent data from becoming lost due to technical problems or human error.
- Seek advice on **your research institute's storage facilities**. Contact your IT department, SURF, or any external service provider your organisation uses.
- Storage on laptops, hard disks or external media is generally risky. It is preferable to use robust, properly managed storage facilities provided by **the institution's IT department**. Automatic backups by the IT department are also safer than manual backups.
- When using **external services** you must ensure that there is no conflict with the policy of the body/ies funding the research (e.g. accessibility of data) or with the policy of your department or institute (e.g. security of sensitive data).

If you **manage** the storage and backup of your research data **yourself**, always consider: the scale of and growth in your data set, storage locations and capacity, version management, backups and technical and organisational measures to secure the data.

2. Legislation (including privacy)

2.1 I will be doing research involving human subjects, and I am aware of and compliant with laws

and regulations concerning privacy sensitive data.

- The Wet Medisch-Wetenschappelijk Onderzoek met Mensen (WMO, or Medical Research (Human Subjects) Act) does not apply to my project.
- No, I will not be doing research involving human subjects; proceed to section 3 (Making data findable)
- Yes, I will involve human subjects in my research. I will comply with the Algemene Verordening Gegevensbescherming (AVG)
- Wet op de Geneeskundige Behandelingsovereenkomst (Medical Treatments Contracts Act)
- Gedragscode Goed gebruik van lichaamsmateriaal (Code of Conduct for Responsible Use of Human Tissue)
- The Wet Medisch-Wetenschappelijk Onderzoek met Mensen (WMO, or Medical Research (Human Subjects) Act) applies to my project; I will have it reviewed by a Medical Research Ethics Committee. In addition I will comply with the Kwaliteitsborging Mensgebonden Onderzoek (Quality Assurance for Research Involving Human Subjects)

Guidance:

Achtergrondinformatie bij deze vraag vindt u op de [webpagina van Toegang tot data over wet- en regelgeving](#), in het bijzonder onder de tabs:

Bescherming van persoonsgegevens: *Achtergrond*

Bescherming van persoonsgegevens: *Stappen in de bescherming van persoonsgegevens*

Bescherming van persoonsgegevens: *Toetsing door METC of CCMO*

2.2 I will be doing research involving human subjects, and I have (a form of) informed consent from the participants for collecting their data.

- Yes, and this informed consent allows for the reuse of data (note that in the Code of Conduct for Medical Research, 'reuse' is also referred to as 'further use')
- Yes (please describe the form this consent takes)

Guidance:

Achtergrondinformatie bij deze vraag vindt u op de [webpagina van Toegang tot data over wet- en regelgeving](#), in het bijzonder onder de tabs:

Bescherming van persoonsgegevens: *Toestemming*

2.3 I will be doing research involving human subjects, and I will protect my data against misuse.

- Yes, the data will be anonymised. I realise that this will limit the options for re-use of my data. (explain)
- Other (please explain)
- Yes, the data will be pseudonymised. (please explain how this will be done, and by which organisation) and

Guidance:

Achtergrondinformatie bij deze vraag vindt u op de [webpagina van Toegang tot data over wet- en regelgeving](#), in het bijzonder onder de tabs:

Bescherming van persoonsgegevens: *Beschermingsmaatregelen*

2.4 I will stick to the privacy regulations of my organisation

- Yes
- My organisation does not have any regulations in this field, so I will personally make sure that the measures taken to guarantee the privacy and non-traceability of data are upheld

3. Making data findable

3.1 The data collection of my project will be findable for subsequent research (note: this is a key item, which you should report to ZonMw at the end of your project).

- Yes, it can be found through an online (metadata) catalogue or web portal (please specify)
- No, I have not yet chosen an archive or catalogue/web portal
- Yes, it can be found through the search engine of the archive or repository in which it is stored (please specify)

Guidance:

Data sets can be found in **data repositories, catalogues (including metadata catalogues)** and **web portals**. They provide a list of the data sets in combination with descriptive information (data on data, or **metadata**).

A data repository is intended for the **long-term storage** of your data once your project is complete. For more information and answers to questions on this see *part 6, Reusable data and permanent storage*.

You are advised to choose a data repository at an early stage of your project. This will allow you to **consider the requirements set by the repository** when setting up your data collection, including a suitable file format and any metadata required.

If you are used to storing data in **international repositories**, you may continue to do so, particularly if this is a requirement stipulated by co-financers or scientific journals. More information can for example be found on [EMBL EBI](#). In this case, you need not also store your data in a Dutch repository.

(See also question 1.3)

3.2 I will use a metadata scheme for the description of my data collection.

- No, I have not yet chosen a metadata scheme
- Yes, I will select a metadata scheme from the list published by Datacite (please specify)
- Yes, I will use a metadata scheme specific for my field of research (please specify)

Guidance:

Use a metadata scheme for describing:

- the data set as a whole (this is what question 3.2 refers to);
- the individual records in the data set (this is what question 5.2 refers to).

This question (3.2) concerns metadata that provide information on: who gathered the data, under what circumstances, and what was measured/recorded.

Various metadata schemes exist:

- generic or specialist
- detailed or general.

Ideally, the data description should give sufficient detail for potential users to decide whether the data are appropriate for their research.

To ensure data sets are findable and reusable, both now and in the future, metadata also have to be machine-readable, in accordance with the [FAIR-data](#) principles.

A **repository** may prescribe a certain metadata scheme. If not, you are free to choose your own.

There are several things to **consider when choosing a metadata scheme** :

- Choose a **metadata scheme specific to** (appropriate for) your field. This allows you to produce a detailed description and enhanced the findability and reusability of data within the field. It also improves the potential for linking with other data sets.
- Examples of **international schemes** can be found on [DataCite](#). DataCite offers metadata schemes for various types of data set. Every scheme gives a number of data items that are needed to make the data findable. The data items can be completed with keywords or as free text.
- The [Data Documentation Initiative Alliance](#) is for the documentation of, among other things, social science and health research.

Also choose a **generic metadata scheme** so that the data collection can be 'harvested' by a data portal. The KNAW's NARCIS portal is one example. The [Dublin Core standards](#) are another well-known example.

3.3 I will be using a persistent identifier as a permanent link to my data collection (note: this is a key item, which you should report to ZonMw at the conclusion of your project).

- No, I will not be using a persistent identifier (please explain)
- Yes, in addition to the DOI code I will be using another persistent identifier (please specify)
- Yes, I will be using the DOI code

Guidance:

Use a **persistent identifier** (PI) to provide a permanent link to the data set. A PI is generated by a repository, and must be passed on to ZonMw at the end of the project.

Examples of PIs include [DOI](#), Handle, URN and ARK.

The advantage of a persistent identifier over a normal link (URL) is that your data set will remain findable even if its location changes.

A data set can also be **cited using a PI** (data citation).

More information about persistent identifiers and DOIs can be obtained from the [International DOI Foundation \(IDF\)](#), [DataCite](#), [TUDelft](#) or the [RDNL course](#).

Finally

Do not store the persistent identifier with the identifying code (*see part 5, Interoperability*). This is a code for the records in the data set.

4. Making data accessible

4.1 Once the project has ended, my data will be accessible for further research and verification.

- Yes, immediately
- No (please explain)
- Yes, after an embargo period (please explain)

Guidance:

It could be that an embargo period needs to be observed, during which your data will not be

fully public, in connection with publication, public safety, privacy, intellectual property belonging to certain companies or commercial interests associated with exploitation of the research results, for example. ZonMw's grant terms and conditions stipulate a 3-month embargo, and a maximum of 9 months for patents.

Explain the reasons for the embargo and how long it will last.

If necessary, ZonMw will help you determine the duration of the embargo.

For more information see:

- Information on [cofinancing](#) on the ZonMw website.
- [Guidance on property rights](#)

4.2 Once the project has ended, my data collection will be publicly accessible, without any restrictions (open access).

- Yes, proceed to section 5 (Making data interoperable)
- No, there will be access restrictions to my data collection (please explain)

4.3 I have a set of terms of use available to me, which I will use to define the requirements of access to my data collection once the project has ended (please provide a link or persistent identifier; also note that this is a key item, which you should report to ZonMw at the conclusion of your project).

- Yes, my institution employs internationally available terms of use
- Yes, my institution has drafted a set of terms of use with the help of a legal advisor
- Not yet, my institution will draft a set of terms of use with the help of a legal advisor

Guidance:

In the event of **restricted access** you must give a precise description of the **conditions** that a research group seeking access to your research data must comply with.

In accordance with the [FAIR guiding principles](#) the **legal** status of the licences and the terms and conditions of use must be clear to ensure that data can actually be reused.

You can use internationally available **standards**, or draw up your own terms and conditions with the assistance of a **lawyer**.

The terms and conditions of use must be made available by your **institute or department** and may not be issued on an individual basis.

ZonMw requests that you ensure the terms and conditions cover as many of the criteria listed in the following question as possible.

4.4 In the terms of use restricting access to my data, I have included at least the following:

- Other (please explain)
- The reimbursement of costs, for example in obtaining the data
- The permitted period of use of the data set
- The manner in which the data set can be accessed
- Collaboration in using the data set, including agreements on publication and authorship
- Whether or not the data set may be linked with another data set (for reasons of privacy)
- Agreements on methodology
- Conditions related to data security
- The sharing of data for commercial purposes, taking into account the provisions of state aid law
- A steering committee, programme committee or project leader will be charged with approving data requests

- The approval of the participants allows for further research using this data set

Guidance:

See guidance on question 4.3

5. Making data interoperable

5.1 I will select a machine actionable data format, which will allow other researchers and their computers to read my data collection.

- No (please explain)
- Yes (please specify)

Guidance:

To make your research data **interoperable** it is advisable to use an **IT standard (or data format)** commonly used in your **field** for recording and coding your data. In other words, when choosing the standard, ensure that data are readable both by other researchers (unity of language) and by their computers ('machine readable' or 'actionable').

Your choice will also depend on the technical capabilities of your institute. You must therefore always seek the advice of your IT and/or data manager.

You can also obtain information from experts and/or their organisation's website, such as:

- [Nictiz](#); see 'fourth layer of interoperability (application)'
- Examples of **international schemes** can be found on [DataCite](#). See also the guidance to question 3.2.
- Research Data Netherlands [RDNL](#) (or the participating institutes DANS, 3TU.Datacentrum, SURF),
- [Dutch Techcentre for Life Sciences DTL](#)
- If you use the Castor or Open Clinica data management system, your data are readable by and exchangeable with other computers.

These organisations provide information, IT services and facilities, IT standards and data storage, and they hold regular consultations with the IT departments of all Dutch research institutions.

ZonMw has made arrangements with these organisations concerning advice on data management in ZonMw projects.

5.2 I will select a metadata standard to allow my data collection to be linked to other collections (note: this is a key item, which you should report to ZonMw at the conclusion of your project).

- No (please explain)
- Yes, I will select a metadata standard from the list published by Biosharing (please specify)

Guidance:

To make your data interoperable, you are advised to use a metadata standard commonly used in your field (see [Nictiz](#) 'third layer of interoperability'). Consider standard terminology, classifications and information standards.

Make good arrangements for this aspect of your study from the start of the project. It is much more work to do it in retrospect!

Commonly used metadata standards including:

- [SNOMED CT](#)
- [DataCite](#). DataCite provides metadata schemes for various types of data set.
- The [Biosharing](#) website lists metadata standards used in the medical sciences.
- The [Castor EDC](#) data management system uses recommended metadata standards.

NB

Question 3.2 also asks for a metadata scheme, which refers to the description of the entire data set. This question, by contrast, concerns the exchangeability of data at the level of individual records. In order to link two similar data sets, the variables must be documented and coded in the same way.

5.3 I will be doing research involving human subjects, and I have taken into account the reuse of data and the potential combination with other data sets when taking privacy protection measurements.

- No (please explain)
- Yes, the participants have given their permission for reuse of the data, and the data have been pseudonymised

Guidance:

Here ZonMw wishes to establish whether you have arranged privacy protection in such a way that your data cannot in principle be linked with other data sets.

In **research involving human subjects** certain ethical and legal requirements pertain to the linking of data. If the person has given their express **permission** for the secondary use of their data (including linking it with other data sets), the **identifying code** may be their BSN (Dutch tax and social security number), date of birth or postcode. If no permission has been given, the code must be a **pseudonym** that cannot be traced to the individual.

If the data are **anonymised** no permission is required, but it is not possible to link the data.

A trusted third party such as [ZorgTTP](#) can provide support with the exchange and accessibility of data sets containing privacy-sensitive information

6. Making data reusable

6.1 I will ensure that the data and their documentation will be of sufficient quality to allow other researchers to interpret and reuse them (in a replication package).

- I will document the software used in the course of the project (please specify)
- I will document the research process (please explain)
- In addition, I will take further quality assurance measures (please specify)
- I will perform quality checks on the data to ensure that they are complete, correct and consistent (please explain)

Guidance:

With a view to the reuse of data and possible replication of the research, you must archive at least the following **data and documentation**:

- the raw data (if you have reused existing data, you may have processed data rather than 'raw' data; in that case, the existing data must already be permanently archived,

- or must now be archived);
- the data on which publications are based;
- documentation on the research methodology used (such as code books, data manuals, metadata compilation, machine settings, use of SOPs, version management etc.), project proposal, approval from ethics committees such as METC, all stakeholders (researchers, laboratory assistants, test subjects etc.)) in short, everything needed to ‘retrace the steps’.
- When storing **qualitative data**, describe the procedure used to transcribe the data (including conventions and symbols). Also look at [Atlas Ti](#).

Together, this is known as a ‘**replication package**’.

Details on this minimum set of information, focused specifically on research covered by the Medical Research (Human Subjects) Act (WMO), can be found in the publication “[Quality Assurance for Research Involving Human Subjects 2.0](#)” (NFU, 2012). Appendix 3 to this publication lists the documents that must be archived for the preparation, implementation and completion of this type of research.

6.2 I have a number of selection criteria, which will allow me to determine which part of the data should be preserved once the project has ended. (see also question 1.9)

- No
- Yes
- Some or all of the data must be destroyed once the project has ended, because of a contract or law

Guidance:

You must store research data for reuse or to verify and validate research results. This does not however apply to all data. In some situations it is easier or less costly to generate exactly the same data again than to store that same data. In some circumstances, it may also be in the interests of privacy to generate data again rather than storing it.

To help you determine which research data constitute valuable source material for further research, DANS has drawn up a **checklist** setting out ‘[General guidelines for the selection of research data for storage](#)’. The guidelines set out the main reasons for storing research data in the long term. The checklist can be used by individual researchers, research groups, institutions, archive managers and funding bodies.

The [Research Data Netherlands](#) course (RDNL) also includes a user-friendly **decision diagram for data selection**.

6.3 Once the project has ended and the data has been selected, I can make an estimate of the size of the data collection (in GB/TB) to be preserved for long-term storage or archival.

- Not yet (please explain)
- Yes (please specify)

6.4 I will select an archive or repository for (certified) long-term archiving of my data collection once the project has ended. (note: this is a key item, which you should report to ZonMw at the conclusion of your project)

- Yes, and this archive has a data seal of approval (please specify the archive)
- Not yet

- Yes, and this archive has a different form of certification (please specify the archive and certification)
- Yes, and this archive meets certification criteria and intends to get certified (please explain how your data will remain accessible and reusable in the long term)

Guidance:

In part 3 (Making data findable) you will have specified a repository where your data are **findable**. Is the repository of your choice suitable for permanent storage / long-term archiving, ensuring your data are **permanently stored** and **reusable**?

International guidelines are available for permanent long-term archiving. The simplest set of criteria is that of the international [Data Seal of Approval](#) (DSA). These **criteria** and the associated **quality label** are independent of subject area. A repository that stores data permanently is known as a **trusted digital repository**. In the Netherlands, **CentERdata**, **DANS**, **4TU.Datacentrum** and **SURF** have this status.

There are barely any specific **healthcare data repositories** with the DSA. Researchers can however use an existing general repository (see the list of [Seals](#) on the website).

You may also store and archive your data at **your own institute**, but please ensure maximum compliance with the guidelines.

If you choose a repository that does not have the Data Seal of Approval, DIN-31644-, ISO-16363- or WDS/ICSU certification, please make clear how you will ensure that the data remain effectively accessible and reusable after the project.

You are advised to ask your institute what possibilities it offers for permanent storage and archiving. To what extent does the storage facility operate in accordance with the criteria? And does it intend to apply for permanent storage certification? You might like to make the internal storage administrator aware of this option.

6.5 Once the project has ended, I will uphold the recommended data preservation period of at least 10 years.

- No, specify the guidelines and the number of years (please explain)
- Yes, in accordance with VNSU guidelines (please specify the number of years)
- Yes, in accordance with other guidelines (please explain, and specify the guidelines and the number of years)

Guidance:

Different codes and guidelines specify different storage periods. ZonMw requests that you at least comply with the **Dutch Code of Conduct for Academic Practice (VSNU)**: 'Raw data must be stored for a period of at least 10 years' (*see also part 2, Legislation*).

The publication '**Quality Assurance for Research Involving Human Subjects 2.0**' (NFU, 2012) also specifies minimum storage periods for a range of material. (see question 6.1)

6.6 Data management costs during the project and preparations for archival can be included in the project budget. These costs are:

- Amount (please elaborate)
- Unknown (please explain)

6.7 The costs of archiving the data set once the project has ended are covered.

- Yes (please elaborate)
- Not yet (please explain)

Guidance:

When **estimating the costs** you must distinguish between:

- (1) the costs of data management **during the project** and the preparation of data for archiving, and
- (2) the costs of permanent storage and archiving **after** the project.

Re. (1) These costs can be included in the budget submitted as part of your grant application. The costs of data management refer in part to:

- the time spent on metadata compilation and documentation during the study;
- **the type of data** and the **size** of the collection (you will have made an estimate of this in *part 1. General features of the project*);
- storage and securing of data during the project.

Re. (2) The costs of long-term storage will be borne by the institutes. A good picture exists of the costs of this and how they break down. They are at any rate determined by the type of data and the size (number of subjects and giga-/terabytes of the **selected dataset** (*see selection process in questions 2 and 3*)).

Several parties are working on a **costing model** for permanent data storage, though the knowledge to devise a suitable funding model for the storage of data is currently lacking. You can obtain information on the cost of archiving from the [RDNL course](#).