

DATA: DIGITALE DIAMANTEN

Datamanagement voor onderzoeksprojecten gesubsidieerd door ZonMw

RIVM, NIVEL, DANS, CentERdata en BBMRI-NL



Datum: December 2013

Auteurs: Drs. R. Nugteren (RIVM)
Dr. C. Veenhof (NIVEL)
Drs. S. el Markhous (NIVEL)
Drs. E.A. Balster (CentERdata)
Dr. M.J. Grootveld (DANS)
Dr. M.A. Swertz (BBMRI-NL)

Algemene gegevens

1. Projecttitel: Datamanagement voor onderzoeksprojecten gesubsidieerd door ZonMw
2. Subsidieverstrekker: ZonMw
3. Deelnemende instellingen: RIVM, NIVEL, DANS, CentERdata en BBMRI-NL
4. Samenwerking: ZonMw (coördinator)
5. Projectleider/contactpersoon: R. Nugteren (RIVM)
6. Coördinatoren intern: E.A. Balster (CentERdata), C. Veenhof (NIVEL)
7. Uitvoerders: M.A. Swertz (BBMRI-NL), E.A. Balster, M. De Bruijne, L.H.J.M. Wijnant (CentERdata), M. Grootveld (DANS), C. Veenhof, S. el Markhous (NIVEL), R. Nugteren (RIVM)
8. Aanvang: 1 juli 2013

Inhoudsopgave

Algemene gegevens.....	2
Inhoudsopgave.....	3
Beleidssamenvatting	4
1 Inleiding.....	6
1.1 Achtergrond	6
1.2 Doelstelling project.....	7
1.3 Leeswijzer.....	8
2 Datamanagement in ontwikkeling	9
2.1 Ontwikkelingen, initiatieven en samenwerkingsverbanden.....	9
2.2 Zorgvuldig en integer onderzoek	9
2.3 Open data en data-infrastructuur in binnen- en buitenland	10
2.4 Samenvattend	13
3 Hoe zijn we te werk gegaan?	14
3.1 Checklists voor dataparagraaf en datamanagementplan	14
3.2 Literatuuronderzoek	14
3.3 Consultatie van experts.....	14
3.4 Expert meeting	15
3.5 Workshop datamanagement.....	15
3.6 Datamanagementvragen.....	15
4 Dataparagraaf en datamanagementplan.....	16
4.1 Checklist Dataparagraaf bij ZonMw-subsidieaanvragen	17
4.2 Checklist Datamanagementplan voor ZonMw-projecten	20
4.3 Workshop checklist datamanagementplan: projectleiders aan het woord	28
5 Management van onderzoeksdata.....	31
5.1 Hoe kunnen bestaande databronnen beter vindbaar gemaakt worden?.....	31
5.2 Hoe kan duurzame opslag van data gefaciliteerd worden met bijbehorende metadata volgens bestaande (internationale) standaarden?.....	32
5.3 Wat is het beste financieringsmodel voor (duurzame) opslag van data?	33
5.4 Aan welke minimale eisen rondom privacy en gebruiksvoorwaarden dienen aanvragen te voldoen?	34
5.5 Wat moet specifiek voor biobanken en beeldmateriaal toegevoegd worden?	34
5.6 Overige informatie: de expert meeting	35
6 Conclusies en adviezen.....	38
Literatuur.....	42
Bijlage 1: Aanzet kennisnetwerk datamanagement	43

Beleidsamenvatting

Effectief en succesvol volksgezondheidsbeleid en wetenschappelijk gezondheidsonderzoek zijn beide sterk afhankelijk van de beschikbaarheid van data. In het kader hiervan ontstaat er steeds meer maatschappelijke en politieke druk om onderzoeksbestanden die gefinancierd worden met publiek geld beschikbaar te stellen voor onderzoek door derden. Een recent en aansprekend voorbeeld hiervan is de brief van de minister van VWS uit oktober 2013 aan de Tweede Kamer. In deze brief geeft ze aan dat standaardisatie en het beschikbaar stellen van zorggegevens essentieel is voor een zogenaamd 'duurzaam informatiestelsel' met actuele en betrouwbare gegevens over de volksgezondheid en zorg. Momenteel zijn er zowel nationaal als internationaal initiatieven om data beter beschikbaar te stellen.

Zo werkt ZonMw met het programma "Toegang tot data" aan de verbetering van de beschikbaarheid en benutting van onderzoeksdata. Om een subsidieontvanger verantwoordelijk te maken voor het vindbaar, toegankelijk en uitwisselbaar maken van databestanden verplicht zij onderzoekers een dataparagraaf in de subsidieaanvraag in te vullen. Het doel van het project was de ontwikkeling van checklists met toelichtingen voor het invullen van een dataparagraaf en het schrijven van een datamanagementplan als onderdeel van subsidieaanvraag bij ZonMw. Een werkgroep bestaande uit RIVM, NIVEL, DANS, CentERdata en BBMRI-NL hebben deze checklists inclusief toelichting ontwikkeld voor de onderzoekers. Bij subsidietoekenning verwacht ZonMw dat de onderzoeker de dataparagraaf uitwerkt tot een datamanagementplan.

De checklists zijn tot stand gekomen door inventarisatie van belangrijke items op basis van literatuuronderzoek, schriftelijke/telefonische informatie van experts en een expertmeeting. Belangrijke onderwerpen in de checklists zijn: privacy, gebruiksvoorwaarden, vindbare data en metadata, kwaliteit en duurzame opslag. Na toetsing van de checklists onder een groep projectleiders bleken ze goed bruikbaar en een goed instrument voor bewustwording te zijn.

Ook bleek uit de bijeenkomsten en interviews in het ontwikkeltraject dat onderzoekers positief staan tegenover het uitwisselen van data en tegenover de ontwikkelde checklists. Dat neemt niet weg dat de invoering van dit instrument een ambitieuze stap is. Het moge duidelijk zijn dat er aan datahergebruik en datamanagement vragen kleven over bijvoorbeeld bevoegdheden, structurele kosten en bewaartermijnen. De invoering van dataparagraaf en datamanagementplan is dan ook een eerste stap richting verantwoord databeheer bij wetenschappelijk onderzoek. Evaluatie van de ervaringen in de praktijk en het gaandeweg beantwoorden van bovengenoemde vragen, samen met de andere partijen in het datalandschap, zijn belangrijke vervolgstappen.

De werkgroep doet in dit rapport aanbevelingen aan ZonMw, waarvan dit de belangrijkste zijn:

1. Beschouw de ontwikkelde checklists voor de dataparagraaf en voor het datamanagementplan als eerste versies, test het gebruik in pilots en evalueer het gebruik. Bepaal hier ook bij voor welke projecten de checklists werkzaam zijn.
2. Geef voorbeelden van best practices en zet een netwerk van experts op het gebied van datamanagement op.
3. Neem langetermijnarchivering mee als vaste kostenpost in de projectbegroting.
4. Denk na over de status van een datamanagementplan en de mogelijkheden van toetsing ervan. Vraag bijvoorbeeld in de eindevaluatie van een subsidieproject naar de vindbaarheid van de data.
5. Adviseer universiteiten en onderzoeksinstituten zoveel mogelijk criteria op instituutsniveau in te vullen en een centraal intern loket in te richten voor vragen over datamanagement.

1 Inleiding

1.1 Achtergrond

Effectief volksgezondheidsbeleid en succesvol wetenschappelijk gezondheidsonderzoek zijn beide afhankelijk van gegevens over de volksgezondheid, epidemiologische, genetische en moleculaire profielen, socio-economische en milieufactoren, proefdieren, in-vitro experimenten, etc. Beschikbaarheid van kwalitatief hoogwaardige onderzoeksbestanden is daarbij van groot belang. In Nederland en internationaal is een aantal gegevensbronnen beschikbaar op basis waarvan onderzoek, vaak onder specifieke voorwaarden, kan worden verricht (zie bijvoorbeeld [zorggegevens.nl](#)). Door onderzoeksbestanden beschikbaar te stellen voor wetenschappelijk onderzoek door derden kunnen andere onderzoeksvragen (efficiënt) worden beantwoord en kunnen verschillende gegevensbronnen worden gecombineerd om onderzoeksvragen te beantwoorden die uit het primaire onderzoeksbestand alleen niet zouden kunnen worden beantwoord. Voorbeelden hiervan zijn onderzoek bij mensen met zeldzame aandoeningen of onderzoek naar de relatie tussen woon- en werksituatie en gezondheidsproblemen. De beschikbaarheid van dergelijke gegevensbronnen laat nog te wensen over. De gegevens die er zijn, worden bovendien niet altijd goed benut (Bron=[ZonMw.nl](#)). Momenteel is het beter beschikbaar en uitwisselbaar maken van onderzoeksgegevens een actueel thema in binnen- en buitenland.

In het kader hiervan wordt steeds vaker gesteld dat onderzoeksbestanden die gefinancierd worden met maatschappelijk geld beschikbaar moeten worden gemaakt voor derden (Raad voor Gezondheidsonderzoek/[RGO, 2008](#)). Voorbeeld is dat de Koninklijke Nederlandse Academie van Wetenschappen (KNAW) en Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) al een dataparagraaf of datacontract hebben opgenomen in hun subsidiebepalingen/-voorwaarden ([KNAW, 2011](#); [NWO & DANS, 2010](#)). Ook ZonMw stimuleert de bredere beschikbaarheid van onderzoeksgegevens en de infrastructuur die hiervoor nodig is. Hiertoe dient het [Programma "Toegang tot Data"](#) oftewel #TTD. ZonMw stimuleert hiermee de brede beschikbaarheid van onderzoeksgegevens (zie bijvoorbeeld het [Goed Gebruik Geneesmiddelen rapport, 2012](#)) en werkt aan het zo goed mogelijk beschikbaar krijgen van door hen gesubsidieerde gegevensverzamelingen. De einddoelen van het #TTD-programma zijn als volgt geformuleerd (Bron=[ZonMw.nl](#)):

1. "Onderzoekers vullen alleen nog databronnen die vindbaar, toegankelijk en uitwisselbaar zijn.
2. Het gebruik van bestaande databronnen is een vanzelfsprekend onderdeel in wetenschappelijk onderzoek en het informeren van beleid en praktijk."

Het programma 'Toegang Tot Data' streeft naar de volgende winstpunten (Bron=[ZonMw.nl](#)):

1. Besparing van onderzoeksbudget, proefdieren, administratiedruk en tijd;
2. Onderzoek kunnen verifiëren, wetenschapsfraude voorkomen;
3. Genereren van nieuwe onderzoeksvragen;
4. Informatie voor beleid (impact van interventies, trends in volksgezondheid);
5. Informatie voor de zorgpraktijk (kwaliteit);
6. Versterking data-infrastructuur.

Om het voorafgaande zo goed mogelijk te kunnen realiseren heeft ZonMw Artikel 20 opgenomen in haar subsidievoorwaarden, [ZonMw.nl TTD](#).

Artikel 20 Databestanden

1. Bij databestanden worden ZonMw en de subsidieontvanger, samen beschouwd als de ‘producent van de databank’ zoals bedoeld in de Databankenwet.
2. De subsidieontvanger stelt een datamanagementplan op. In dit plan geeft de subsidieontvanger aan hoe data gedeeld worden, wanneer de data aan derden beschikbaar worden gesteld en de wijze waarop de data toegankelijk worden gemaakt. Dit plan wordt ter goedkeuring aan ZonMw voorgelegd.
3. Databanken met bijbehorende toelichting worden in elk geval op door ZonMw aan te geven wijze beschikbaar gesteld voor gebruik ten behoeve van verder wetenschappelijk onderzoek. De binnen een project verzamelde gegevens dienen daarom volgens de geldende standaarden toegankelijk te worden gedocumenteerd en opgeslagen.

Om een subsidieontvanger verantwoordelijk te maken voor het vindbaar, toegankelijk en uitwisselbaar maken van databestanden dient de subsidieaanvrager een dataparagraaf in de subsidieaanvraag in te vullen en deze – na toekenning van de subsidie – tot een datamanagementplan uit te werken. Aangezien de meeste onderzoekers hiermee weinig ervaring hebben, wil ZonMw hun structuren en toelichtingen ter beschikking stellen, gebruik makend van expertise en werkwijzen die al beschikbaar zijn in het veld. Hiertoe dient het onderhavige rapport.

1.2 Doelstelling project

Het belangrijkste doel van dit adviesproject is om checklists met toelichtingen te ontwikkelen voor het schrijven van

1. een dataparagraaf: het onderdeel van de subsidieaanvraag dat gericht is op het vindbaar, toegankelijk en uitwisselbaar maken van data die in het aangevraagde project wordt gegenereerd;
2. een datamanagementplan: de uitwerking van de dataparagraaf die wordt geschreven na het verkrijgen van subsidie. Het datamanagementplan (DMP) beslaat onderwerpen als de locatie waar de gegevens zullen worden opgeslagen, hoe dit zal gebeuren, hoe wordt omgegaan met privacy, de datastandaarden die gebruikt gaan worden en wie verantwoordelijk zal zijn voor het goed vindbaar maken en houden van de data, evenals het duurzaam archiveren van de data.

De vraagstellingen van dit rapport zijn in tweeën gesplitst. Allereerst zijn er vragen over de ontwikkelde dataparagraaf en datamanagementplan en ten tweede zijn er vragen over goed datamanagement in het algemeen.

Vragen over de dataparagraaf en het datamanagementplan

De dataparagraaf en het datamanagementplan – die zelf dus geen doel zijn van het onderhavige adviesproject – hebben tot doel om na te gaan of en in welke mate de subsidieaanvrager in staat en bereid is om de binnen het onderzoek verzamelde of gebruikte gegevens, beschikbaar te stellen. Het onderhavige adviesproject toetst de uitgebreide checklist voor het datamanagementplan aan de hand van de volgende vragen:

1. Wat is de praktische bruikbaarheid van het ontwikkelde concept-datamanagementplan?
2. Op welke onderdelen voldoen huidige (ZonMw) projecten niet aan de elementen uit het datamanagementplan?
3. Is hier een logische verklaring voor en in hoeverre is het haalbaar dat projecten in de toekomst hier wel aan voldoen?

Vragen over goed datamanagement

1. Hoe kunnen bestaande databronnen beter vindbaar gemaakt worden? Welke metadata is hiervoor nodig en hoe ontsluiten we die het beste?
2. Hoe kan de (duurzame) opslag van data gefaciliteerd worden met bijbehorende metadata volgens bestaande (internationale) standaarden?
3. Wat is het beste financieringsmodel voor de duurzame opslag van data; hoeveel ruimte (tijd en geld) moet worden gereserveerd voor adequaat datamanagement?
4. Aan welke minimale eisen rondom privacy en gebruiksvoorwaarden dienen aanvragen te voldoen? Hoe worden de privacy en gebruiksvoorwaarden op dit moment geregeld door organisaties die zich bezighouden met het delen van onderzoeksbestanden?
5. Wat moet er specifiek voor biobanken en beeldmateriaal toegevoegd worden?

1.3 Leeswijzer

Hoofdstuk 2 geeft een bloemlezing van diverse nationale en internationale initiatieven om onderzoeksgegevens vindbaar, toegankelijk en uitwisselbaar te maken. Hoofdstuk 3 beschrijft hoe er in dit project te werk is gegaan. In hoofdstuk 4 worden de checklists gepresenteerd, samen met de bevindingen uit een workshop met projectleiders van lopend onderzoek. Hoofdstuk 5 beantwoordt de praktische vragen over datamanagement die in paragraaf 2 van dit hoofdstuk zijn gesteld en doet verslag van een expert meeting. In het afsluitende hoofdstuk 6 worden conclusies getrokken en aanbevelingen gedaan voor het gebruik van de checklists en de benodigde vervolgstappen.

Het beheer van onderzoeksgegevens is een veld in beweging, met beleid, organisaties en technologie. Bovendien is het een veld met veel – internationale – spelers. Daardoor is jargon niet te vermijden en ontbreekt het geregeld aan eenduidige termen die voor iedereen bruikbaar zijn. Alleen al het centrale begrip “data” wekt niet bij iedereen dezelfde associaties; wie wil, vervang het door “onderzoeksgegevens”. In dit rapport worden “datamanagement”, “data stewardship” en “beheer van onderzoeksgegevens” als equivalenten gebruikt. Ook “catalogus”, “(web)portal” en “portaal” komen door elkaar voor als online-locaties waar informatie over data (metadata dus) te vinden is; het is niet noodzakelijk dat de data op diezelfde locatie zijn opgeslagen.

Tot slot dient u zich ervan bewust te zijn dat het rapport tweeledig is, enerzijds geeft het een beschrijving van de ontwikkeling van de checklist en anderzijds worden de ervaringen met datamanagement in de praktijk beschreven.

2 Datamanagement in ontwikkeling

2.1 Ontwikkelingen, initiatieven en samenwerkingsverbanden

Momenteel is het beter vindbaar, beschikbaar en uitwisselbaar maken van onderzoeksgegevens een actueel thema zowel in binnen- als buitenland. Hieronder beschrijven we een aantal aansprekende ontwikkelingen, initiatieven en samenwerkingsverbanden die tot doel hebben om gegevens optimaal beschikbaar te stellen (gratis of tegen zo laag mogelijke kosten), te delen (met afdoende beschrijvende gegevens, ofwel goede metadata) en vindbaar te maken (via internet door middel van bestaande portalen en websites). Dit overzicht is een momentopname en streeft niet naar compleetheid; bij het opstellen ervan is gezocht naar ontwikkelingen die relevant zijn voor de ambitie van “Toegang tot data”. Het is duidelijk dat bij het pleidooi voor open data en het zoveel mogelijk beschikbaar krijgen van gegevens afkomstig uit bijvoorbeeld de zorg, de Wet bescherming persoonsgegevens (WBP) altijd zeer hoog in het vaandel staat. Onder de WBP is gebruik voor secundaire analyses van onderzoeksgegevens bijvoorbeeld alleen toegestaan op voorwaarde dat dit niet onverenigbaar is met het doel waarvoor de onderzoeksgegevens verzameld zijn. Op basis van de WBP en daaruit voortvloeiende Gedragscode Gezondheidsonderzoek worden uitsluitend geanonimiseerde/gepseudonimiseerde gegevens gebruikt voor onderzoeksanalyses.

Zorggegevens essentieel voor onderzoek en beleid

Binnen de context van volksgezondheid- en zorgonderzoek is duidelijk een ontwikkeling gaande van ‘gesloten tenzij’ naar ‘open tenzij’. In een recent verschenen [brief \(oktober 2013\)](#) van de minister van VWS geeft de minister aan dat standaardisatie en het beschikbaar stellen van zorggegevens essentieel is voor een zogenaamd ‘duurzaam informatiestelsel’. Daarvoor is het noodzakelijk om samen met alle partijen in de zorg te werken aan de ontwikkeling, inrichting, het beheer, en het onderhoud van een samenhangend geheel van afspraken, standaarden en voorzieningen. En daarmee voor het monitoren van de zorg en beheersen van de kosten in de zorg (VWS, 2013).

Omvangrijke en belangrijke gegevensverzamelaars zoals de academische ziekenhuizen (UMC's) hebben recentelijk afspraken gemaakt over het delen van ICT-infrastructuur en het gestandaardiseerd uitwisselen van patiëntgegevens. Om het valide meervoudig gebruik van medische patiëntgegevens te bevorderen werken de UMC's aan een zuivere en uniforme zorgdocumentatie aan de ‘voorkant’ met secundair hergebruik van de gegevens aan de ‘achterkant’ van het informatiesysteem, zie het strategisch document [“Registratie aan de bron”](#) van de Nederlandse Federatie van Universitaire Ziekenhuizen (NFU, 2013).

2.2 Zorgvuldig en integer onderzoek

Recente schandalen in de academische wereld waarbij onderzoekers de hand lichtten met onderzoeksdata, zoals de zogenaamde “Stapel-affaire”, hebben het proces van het beschikbaar stellen van de onderzoeksdata versneld. Het is in het kader van wetenschappelijke integriteit belangrijk dat gegevens uit wetenschappelijk onderzoek beschikbaar komen voor derden en onder kwaliteitscontrole staan. Dit wordt o.a. geadviseerd in het KNAW-[rapport van de commissie Schuyt](#) (KNAW, 2012).

Gedragscode wetenschapsbeoefening

Iedere individuele wetenschapsbeoefenaar is gehouden aan [De Nederlandse Gedragscode Wetenschapsbeoefening](#) van de Vereniging van Universiteiten (VSNU) uit 2004 en herzien in 2012. De gedragscode beschrijft de principes van goed wetenschappelijk onderwijs en onderzoek. Een belangrijk element in de VSNU gedragscode is de ‘controleerbaarheid’:

Als onderzoeksresultaten openbaar worden gemaakt, blijkt duidelijk waar de gegevens en de conclusies op zijn gebaseerd, waaraan ze zijn ontleend en waar ze te controleren zijn. De kwaliteit

van dataverzameling, data-invoer, dataopslag en dataverwerking wordt goed bewaakt. Goede verslaglegging van alle stappen en controle op de uitvoering is noodzakelijk (labjournaals, voortgangsrapportages, documentatie van afspraken en beslissingen enz.). De bewaartermijn van ruwe onderzoeksgegevens is minimaal 5 jaar. Deze gegevens worden op aanvraag ter beschikking gesteld aan andere wetenschapsbeoefenaren. Ruwe onderzoeksgegevens worden zodanig gearhiveerd dat deze te allen tijde met een minimum aan tijd en handelen kunnen worden geraadpleegd (VSNU, 2012).

Onderzoeksbestanden beschikbaar voor derden

Onderzoeksbestanden gefinancierd met maatschappelijk geld moeten zo veel mogelijk beschikbaar komen voor derden, al dan niet beschermd door een password ([RGO Van gegevens verzekerd \(2008\)](#) en het kabinetsstandpunt daarop in 2009). De KNAW en NWO stellen voorwaarden, bijvoorbeeld in het [KNAW-beleid voor open access en digitale duurzaamheid](#) (“Alle KNAW-onderzoeksdata worden duurzaam opgeslagen en voor iedereen toegankelijk gemaakt, tenzij zwaarwegende redenen daaraan in de weg staan”) en in de [NWO-regeling subsidies 2011](#) (“databanken met bijbehorende toelichting worden op door NWO aan te geven wijze beschikbaar gesteld voor gebruik ten behoeve van verder wetenschappelijk onderzoek. De binnen een onderzoeksproject verzamelde gegevens dienen daarom goed en toegankelijk te worden gedocumenteerd en opgeslagen”, artikel 30).

NWO heeft tevens een Nationale Roadmap ontwikkeld voor grootschalige onderzoeksfaciliteiten, zie <http://www.nwo.nl/>. Deze Nationale Roadmap is een gedetailleerd strategisch plan om richting en invulling te geven aan het opzetten of verbeteren van grote onderzoeksfaciliteiten. Hiermee kan Nederland internationaal een belangrijke positie innemen op het gebied van onderzoeksinfrastructuur. Daarnaast neemt Nederland deel aan de [Europese roadmap](#) van het European Strategy Forum on Research Infrastructures (ESFRI).

Ook ZonMw heeft, als onderdeel van het project TTD, haar [subsidiebepalingen](#) aangescherpt door de volgende elementen op te nemen:

1. “Voor de verbetering van de beschikbaarheid van gegevens moet een onderzoeker die subsidie ontvangt voor een nieuwe dataverzameling, dit ‘state of the art’ doen, opdat de data vindbaar, toegankelijk en uitwisselbaar zijn.
2. Voor het beter benutten van bestaande gegevens wordt een onderzoeker die subsidie aanvraagt voor een project waarvoor data nodig zijn, gevraagd om te zoeken naar bestaande dataverzamelingen die voor het project gebruikt kunnen worden.”

2.3 Open data en data-infrastructuur in binnen- en buitenland

Onderzoekers stellen zelf in toenemende mate hun data (vrijwillig) beschikbaar voor andere onderzoekers. Dit gebeurt bijvoorbeeld via zogeheten verrijkte publicaties (rich publications) of als [open data](#). Daarnaast zijn er vele lokale, nationale en internationale initiatieven die een goede basis bieden voor onderzoekers, instituten en overheidsorganisaties om data stewardship verder in te richten:

Overheid

Een belangrijk initiatief vanuit de overheid is het “Kennisnetwerk Open Data”. Dit netwerk wordt door ‘Open Data NEXT’ opgezet vanuit de behoefte om een bredere samenwerking tussen overheid en hergebruikers te organiseren. Voorbeelden hiervan zijn [Data.overheid.nl](#) en het [Open Data Innovatie Netwerk \(ODIN\)](#) van het ministerie van Economische Zaken, wat een zogenaamd driesporenaanpak heeft: 1. overheidsdata worden beschikbaar gemaakt; 2. overheidsdata worden eenvoudig toegankelijk; 3. gebruik van overheidsdata wordt gestimuleerd.

Federatieve data-infrastructuur

Voor de uitvoering van databeleid zijn een goede ondersteuning van de onderzoekers en een adequate technische infrastructuur nodig. In het invloedrijke rapport "[Riding the wave \(2010\)](#)", dat door Eurocommissaris voor de Digitale Agenda Neelie Kroes is omarmd, wordt een dergelijke "collaborative data infrastructure" aangeprezen.

Het fundament bestaat uit een technische basisinfrastructuur, waarin onder meer zorg wordt gedragen voor dataopslag en back-upvoorzieningen. In Nederland spelen computercentra zoals [SURFsara](#) en [Target](#) al deze rol.

Daarboven bevindt zich een laag van backoffice-datadiensten voor langetermijnarchivering en toegankelijkheid van data. Deze functies worden in Nederland vervuld door [trusted digital repositories](#) zoals [DANS](#) en [3TU.Datacentrum](#), die samenwerken in Research Data Netherlands.

Nog weer een niveau hoger bevinden zich de frontoffice-diensten, die de eerstelijnscontacten met onderzoekers verzorgen: ondersteuning, advies en training in verantwoord databeheer. Deze frontoffices bevinden zich bij universiteiten (UB's, lokale datacentra), onderzoeks- en kennisinstituten, nationale en internationale onderzoeksinfrastructuren (ESFRI/Nationale Roadmap) en deels ook bij de financiers (NWO, ZonMw, ministeries).

De bovenste laag ten slotte bestaat uit onderzoekers: de mensen die data voortbrengen en benutten, en hun werk kunnen doen dankzij deze infrastructuur¹.

Bij het Data Integration & Stewardship Center [DISC/DTL](#) komen enkele lagen samen. Dit nationale samenwerkingsverband voor universiteiten en UMC's beoogt data stewardship (ook datamanagement genoemd) en dataintegratievoorzieningen (zogenaamde 'hotels') beschikbaar te stellen voor mensgebonden, plantgebonden en molecuulgebonden onderzoek in samenwerking met infrastructuraanbieders zoals SURFsara en Netherlands BioInformatics Center [NBIC](#) (respectievelijk 'rode', 'groene', 'witte' en 'paarse' DISC genoemd). De voor ZonMw belangrijke rode DISC is gecombineerd met het research-ICT coördinatie-initiatief van de NFU en slaat bruggen tussen infrastructuren van [BBMRI-NL](#), [CTMM TraIT](#) en de UMC's.

Ontwikkelingen in (zorg)dataland op Europees niveau

Er zijn verschillende buitenlandse initiatieven op het gebied van het bevorderen van het beschikbaar stellen van data zoals de Royal Society, [Royalsociety.org](#) of bijvoorbeeld het Nordic eScience Globalisation Initiative, een samenwerkingsverband van de Scandinavische landen gericht op delen van zowel infrastructuur als data over volksgezondheid en zorg, zie [Nordforsk.org](#). De universiteit van Oxford heeft een model voor "good practice" ontwikkeld in [Research Data Management](#) en ook het data-archief van het Verenigd Koninkrijk (UKDA) heeft veel documentatie en best practices online beschikbaar.

Vanuit Europa (de Europese Unie en de WHO Europe) worden steeds meer maatregelen getroffen om gegevens te delen en (gestandaardiseerd) uit te wisselen, binnen maar ook tussen de verschillende EU lidstaten. Deze maatregelen zijn soms bindend (zoals bij een EU verordening) maar veelal eerder richtinggevend en stimulerend. We noemen eerst twee belangrijke algemene voorbeelden om vervolgens in te gaan op initiatieven specifiek gericht op het structureren, harmoniseren en/of beschikbaar krijgen van volksgezondheid- en zorgdata.

In verband met het beschikbaar krijgen en houden van wetenschappelijke publicaties en onderzoeksdata is de internationale declaratie [Berlin Declaration on Open Access to Knowledge in](#)

¹ Een uitwerking van het [frontoffice-backoffice-model](#) is te vinden bij DANS.

[the Sciences and the Humanities](#) belangrijk. De Berlin Declaration: „has contributed to a world-wide coordination of national and international academic organizations in the field of Open Access: By now, more than 440 institutions worldwide have signed the “Berlin Declaration”.“ Daarnaast zijn er EU infrastructuur in ontwikkeling zoals BBMRI-ERIC (<http://bbmri-eric.eu/>) en ELIXIR (<http://www.elixir-europe.org/>) met sterke data stewardship componenten voor biobanking en biologische data en mede gedragen door NWO namens de Nederlandse overheid.

Het [EU Research & Innovation funding programme for 2014-2020](#) stimuleert in het kader van EU Horizon 2020 het beschikbaar stellen van zowel data als wetenschappelijke artikelen. Zie in dat kader ook [Cordis.europa.eu](#). Een andere ontwikkeling is het bevorderen van harmonisatie van wetgeving binnen de lidstaten op het gebied van (het uitwisselen van) persoonsgebonden data middels de EU General Data Protection Regulation (GDPR). Bedoeld om een gelijk ‘speelveld’ te creëren voor alle EU lidstaten, bijvoorbeeld als het gaat om gezondheidsonderzoek op basis van persoonsgebonden gegevens. Ondanks het evidente nut van een dergelijk gelijk speelveld binnen de EU zijn er ook belangrijke bezwaren. In het bijzonder de ‘koplopers’ zoals de Scandinavische landen, worden misschien vooral achterop gezet, zie e.g. [Oxfordjournals.org art1](#) en [Oxfordjournals.org art2](#). De GDPR verordening is voor dit schrijven belangrijk omdat het niet voldoet aan deze verordening als lidstaat (e.g. bestaande gerelateerde privacy wetgeving dient volledig herzien te worden), maar ook als universiteit, (kennis)instituut, of andere organisatie (rechtspersoon) kan leiden tot zeer hoge boetes en andere dwingende maatregelen. Over het hoe en waarom van het delen van persoonsgebonden gegevens moeten dus, liefst landelijk, heldere en bindende afspraken gemaakt worden.

Internationale voorbeelden vanuit volksgezondheid en zorgperspectief

Tot slot noemen we in deze paragraaf een aantal zorgspecifieke voorbeelden:

1. het op initiatief van de EU ontwikkelen van nationale strategieën, waarbij internationale samenwerking en uitwisseling van gegevens, richtlijnen en standaarden hoog in het vaandel staan. Medische zorgkennis en -informatie worden immers niet begrensd door nationale grenzen, zie bijvoorbeeld het Nationale Plan Zeldzame Ziekten, www.npzz.nl/.
2. het ontwikkelen van bindende EU richtlijnen zoals de [Inspire richtlijn](#). Deze richtlijn bevordert de harmonisatie en standaardisatie van gegevens zoals gezondheid- en zorggegevens die gerelateerd kunnen worden aan de leefomgeving of (milieu)incidenten. Deze gegevens dienen als ‘open data’ beschikbaar gesteld te worden binnen een Geo-portaal², zodat milieu- en omgevingseffecten beter in kaart gebracht kunnen worden en landen beter met elkaar vergeleken kunnen worden;
3. Europese indicatoren projecten zoals het European Community Health Indicator ([ECHI](#)) project van de EU en het Environment and Health Information System ([ENHIS](#)) van de WHO stimuleren harmonisatie, standaardisatie en uitwisseling van gegevens op Europese schaal zodat er geen ‘appels met peren’ vergeleken worden;
4. de toegenomen EU inmenging op het gebied van de kwaliteit van de zorg en cross-border healthcare en diens gevolge uitwisseling van zorggegevens (“EU interference has increased from sharing information to standardisation and even to signs of enforcement”, Vollaard et al., 2013), zie de [Cross-border Health Care Directive](#);
5. het stimuleren door de EU van gestandaardiseerde en geharmoniseerde zorgregistratie van zorg die een blik over de landgrens vereisen, zoals de zorg voor mensen met een zeldzame aandoening (waar soms maar enkele patiënten binnen een land bekend zijn) en medische technieken, zie bijvoorbeeld het Cross-Border Patient Registries Initiative ([PARENT](#)).

² <http://nl.wikipedia.org/wiki/Geo-portaal>

2.4 Samenvattend

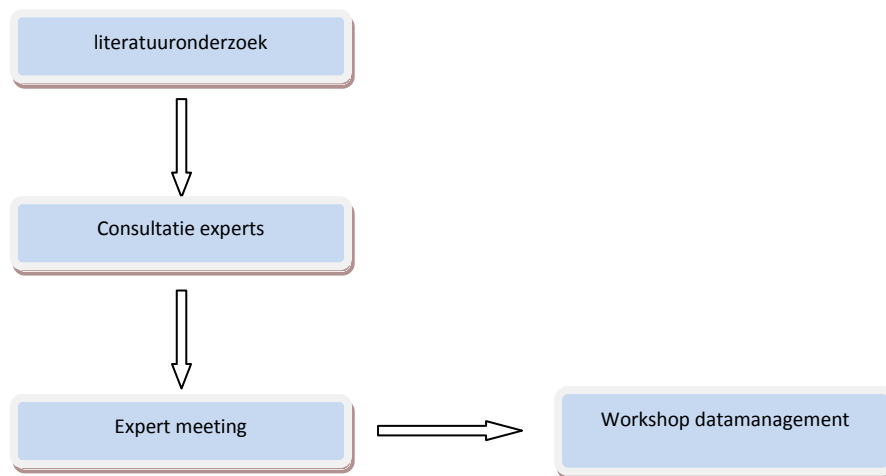
De bovenstaande ontwikkelingen en initiatieven laten zien dat er op grote schaal en binnen allerlei voor dit project zeer relevante werkvelden en gremia en op verschillende manieren gewerkt wordt aan betere vindbaarheid, toegankelijkheid en uitwisselbaarheid van gegevens. De ontwikkelingen gaan in de richting van standaardisatie en het in meerdere of mindere mate afdwingen of stimuleren van het beschikbaar stellen en uitwisselen van gegevens. De voorwaarden waaronder dit moet gebeuren, worden steeds meer centraal vastgelegd (bijvoorbeeld door de nationale overheid of de EU) en voor een deel zelfs opgelegd.

Dit vergt veel van zowel onderzoekers als organisaties en instituten. Positief is dat er steeds meer gegevens beschikbaar komen voor onderzoek, dat er nieuwe en soms onverwachte inzichten ontstaan, bijvoorbeeld doordat datasets op een nieuwe manier gecombineerd kunnen worden, en dat op deze manier minder nieuwe data verzameld hoeven te worden. Verder kan dit er met de juiste stimulans toe leiden dat financiers, universiteiten en kennisinstituten niet meer hoeven te vragen om data, maar dat het beschikbaar stellen van data een vanzelfsprekend onderdeel is van goede en correcte wetenschapsuitoefening. Het breed beschikbaar zijn van data heeft uiteraard ook economische voordelen doordat men niet telkens onnodig hoeft te investeren in het verkrijgen van data (door individuele onderzoekers of instituten) en doordat innovatie gestimuleerd wordt.

3 Hoe zijn we te werk gegaan?

3.1 Checklists voor dataparagraaf en datamanagementplan

In dit project stond het ontwikkelen van de checklist ten behoeve van het datamanagementplan (in het vervolg: "Checklist DMP") centraal. Het ontwikkelen van de checklist DMP was een iteratief proces, waarbij een aantal stappen gezet is en het product vervolgens werd bijgesteld. Er zijn hierbij verschillende fasen doorlopen, namelijk literatuuronderzoek, consulten met experts en een expert meeting. Een workshop datamanagement met een aantal projectleiders had vervolgens tot doel de checklist DMP te toetsen op begrijpelijkheid en de bruikbaarheid ervan te valideren. Het gehele proces is schematisch weergegeven in figuur 1. Hieronder wordt dit proces nader uitgelegd.



Figuur1. Stroomschema ontwikkelproces checklist DMP

De checklist dataparagraaf bevat uitsluitend items uit de uitgebreide checklist DMP, maar is veel korter en bedoeld om er als subsidieverstrekker zeker van te zijn dat de subsidieaanvrager kan voldoen aan bepaalde minimale eisen op het gebied van datamanagement.

3.2 Literatuuronderzoek

Om inzicht te krijgen in de huidige situatie van data delen en van mogelijke vergelijkbare initiatieven om handvatten te bieden voor datamanagement in Nederland, is er een quick scan van de literatuur uitgevoerd door middel van deskresearch. Verschillende bronnen zijn geraadpleegd voor relevante literatuur. Ook is er literatuur geraadpleegd die al bekend was bij de betrokken partijen in dit project. Enkele documenten die gebruikt zijn voor het ontwikkelen van de checklist DMP zijn de checklist van het [Goed Gebruik Geneesmiddelen programma](#) van ZonMw en bestaande modellen voor datamanagementplannen zoals het [sjabloon van DANS](#) en dat van het [Digital Curation Center](#).

3.3 Consultatie van experts

Na de literatuurstudie is een vragenlijst ontwikkeld waarin de volgende aspecten werden onderscheiden: vindbare data en metadata, duurzame opslag, privacy en gebruiksvoorwaarden. Het doel van de vragenlijst was enerzijds om te inventariseren hoe deskundigen op het gebied van datamanagement en zorggegevens in Nederland omgaan met databeheer en anderzijds om een basis te leggen voor de checklist DMP. Ook werd hiermee beoogd het kennisniveau op het gebied van

datamanagement en privacy(wetgeving) te meten, inzicht te verschaffen in ervaren knelpunten bij databeheer en in eventuele beschikbaarheid van ondersteunende diensten binnen organisaties.

De vragenlijst is per e-mail verstuurd naar 45 experts op het gebied van datamanagement en zorggegevens in Nederland, afkomstig uit verschillende organisaties. In totaal zijn negen vragenlijsten geretourneerd. Aanvullend zijn 32 deskundigen telefonisch benaderd om verdere verdieping op de vragenlijst te krijgen en inzicht te verschaffen in wat er al aan datamanagement wordt gedaan. De antwoorden zijn verwerkt, geanalyseerd en in een verslag gebundeld. Deze analyse was belangrijke input voor het ontwikkelen van de checklist DMP.

3.4 Expert meeting

Nadat de checklist DMP is doorontwikkeld is deze per e-mail verstuurd naar 76 experts, onder wie beleidsmakers, onderzoekers, juristen en andere experts op het gebied van bijvoorbeeld datamanagement. De experts zijn vervolgens uitgenodigd voor een expert meeting, van wie in totaal 28 aanwezig waren. De expert meeting had tot doel een algemene indruk te geven van de conceptchecklist DMP, waarbij nader is ingegaan op drie vragen 1) *Wat zijn de specifieke uitdagingen van zorggegevens?* 2) *Wat kan/mag ZonMw eisen van onderzoekers?* 3) *Wat moet er komen aan ondersteunende diensten en instrumenten?* Deze drie vragen zijn belicht vanuit zowel de kant van de onderzoeker als de kant van de beleidsmaker-financier. De reacties van deze experts hebben geresulteerd in een nieuwe versie van de checklist DMP. Zie paragraaf 5.6 voor meer informatie over deze bijeenkomst.

3.5 Workshop datamanagement

De checklist DMP is uiteindelijk verstuurd naar 32 projectleiders. Hiertoe zijn door ZonMw projecten aangeleverd met verschillende datatypes (zoals: randomized controlled trials, beeldmateriaal, biobanken en vragenlijsten): evenredig verdeeld over net toegekende projecten en projecten die al in een ver gevorderd stadium zijn.

De projectleiders hebben de checklist voor een datamanagementplan per e-mail ontvangen met de vraag of zij de checklist voor hun project wilden invullen ter voorbereiding op de workshop. Uiteindelijk waren 12 projectleiders aanwezig op de workshop. Deze projectleiders zijn wel voorlopers op het gebied van datamanagement, wat ertoe kan leiden dat deze aantallen niet geheel representatief zijn voor de gehele groep experts in Nederland. Tijdens de workshop is de checklist DMP getoetst op bruikbaarheid en uitvoerbaarheid. De checklist is inhoudelijk beoordeeld en er is gekeken naar de mogelijkheden voor ondersteuning en hoe de checklist in de praktijk gebruikt kan worden. Na de workshop zijn aanpassingen gedaan in de checklist DMP. Een jurist heeft vervolgens de juridische items in de checklist DMP van commentaar voorzien. Zie paragraaf 4.3 voor meer informatie over deze workshop.

3.6 Datamanagementvragen

De praktische vragen over goed datamanagement (zie paragraaf 1.2 en hoofdstuk 5) zijn beantwoord door de experts die zitting hadden in de projectgroep, aangevuld met informatie uit de bovengenoemde interviews en literatuurstudie.

4 Dataparagraaf en datamanagementplan

In dit hoofdstuk komen eerst de beide checklists aan bod, namelijk die voor de dataparagraaf (de korte checklist) en die voor het datamanagementplan (de uitgebreide). In hoofdstuk 2 is beschreven hoe ze tot stand zijn gekomen. Hier benadrukken we dat beide checklists – inclusief de toelichtingen die daarin staan – een eerste versie zijn en dat ZonMw in de loop der tijd ongetwijfeld aanleiding zal zien tot wijzigingen. De projectwerkgroep heeft gezocht naar een haalbare versie. Daarom is ook feedback gevraagd van projectleiders van lopend onderzoek. Paragraaf 4.3 doet verslag van de workshop met deze projectleiders.

Een ander aandachtspunt is dat de toelichting op de checklist momenteel in de vorm van voetnoten wordt gepresenteerd. Dat is niet ideaal. In een online-versie valt dan ook zeker te denken aan een presentatie waar de toelichtende teksten – per aspect – op te roepen zijn via een icon, bijvoorbeeld een vraagteken of een i-teken. Het implementeren en onderhouden van de beide checklists en de toelichting daarop is in handen van ZonMw.

Hieronder worden achtereenvolgens de checklists voor de dataparagraaf en het datamanagementplan letterlijk weergegeven zoals wij aan ZonMw hebben geadviseerd om deze te implementeren:

4.1 Checklist Dataparagraaf bij ZonMw-subsidieaanvragen

ZonMw vindt het essentieel dat onderzoeksgegevens vindbaar, toegankelijk en uitwisselbaar zijn, daarom dient u binnen deze ZonMw-subsidieaanvraag de onderstaande paragraaf in te vullen³. In de volgorde van het onderzoeksproces – data verzamelen, analyseren, opleveren, bewaren en beschikbaar stellen – verschaft u informatie over de volgende aspecten. De VSNU heeft gedragscodes voor wetenschapsbeoefening en persoonsgegevens die gelden voor iedereen die in Nederland onderzoek uitvoert en die ook datamanagementaspecten omvatten. Wanneer nieuwe data verzameld of bestaande data hergebruikt worden, moeten de privacy en gevoelige gegevens van betrokken personen of zorginstellingen worden beschermd en het proces zodanig gedocumenteerd dat replicatie mogelijk is. Elders in de aanvraag beschrijft u de beoogde projectresultaten. Ook data vormen een projectresultaat en ze moeten zo worden beschreven en aangeboden dat ze voor hergebruik in de toekomst te lezen en te interpreteren zijn. Voor de beschikbaarheid tijdens en na afloop van het project is het belangrijk om te weten hoe de data gedurende het onderzoek opgeslagen en geback-up worden en waar de data voor langere termijn beschikbaar zullen worden gemaakt. Na afloop van het project en een eventueel door de financier toegestane embargoperiode dienen uw data beschikbaar te zijn voor andere onderzoeksgroepen. Hierbij is het van belang dat de onderzoeksgegevens goed te vinden zijn en onder welke voorwaarden deze hergebruikt kunnen worden.

In het onderstaande verwijst “ik” naar de aanvrager.

1. Uitgangspunten		
De VSNU heeft gedragscodes voor wetenschapsbeoefening en persoonsgegevens die gelden voor iedereen die in Nederland onderzoek uitvoert en die ook datamanagementaspecten omvatten:		
	JA	N.V.T.
Ik houd me aan de actuele Nederlandse Gedragscode Wetenschapsbeoefening (VSNU-code) en zie er in redelijkheid op toe dat ook de andere projectleden dit doen ⁴ .	<input type="checkbox"/>	<input type="checkbox"/>
Ik houd me aan de actuele Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek (VSNU-code) en zie er in redelijkheid op toe dat ook de andere projectleden dit doen.	<input type="checkbox"/>	<input type="checkbox"/>

³ Beschikbaarheid van kwalitatief hoogwaardige onderzoeksbestanden is van groot belang voor gezondheidsonderzoek. In Nederland en daarbuiten is een groot aantal gegevensbronnen beschikbaar op basis waarvan onderzoek kan worden verricht; zie bijvoorbeeld zorggegevens.nl. Door onderzoeksbestanden beschikbaar te stellen voor wetenschappelijk onderzoek door derden kunnen andere onderzoeksvragen (efficiënt) worden beantwoord. Ook kunnen gegevensbronnen worden gecombineerd om vragen te beantwoorden die uit het primaire onderzoeksbestand alleen niet te beantwoorden zijn.

Steeds vaker stellen onderzoeksfinanciers eisen aan de beschikbaarheid van data. Ook ZonMw wil dat (op termijn) onderzoekers die subsidie ontvangen van ZonMw alleen nog databestanden opleveren die vindbaar, toegankelijk en uitwisselbaar zijn en dat het gebruik van bestaande dataverzamelingen een vanzelfsprekend onderdeel is in wetenschappelijk onderzoek en het informeren van beleid en praktijk. In de notitie *Toegang tot data: de ontwikkeling van datamanagement bij ZonMw* is meer informatie te vinden.

⁴ In de uitwerking van het principe dat onderzoeksresultaten controleerbaar dienen te zijn, noemt de gedragscode onder meer de noodzaak van repliceerbaarheid van onderzoek, de kwaliteit van dataverzameling en dataopslag, goede verslaglegging van alle stappen, vijf jaar als minimale bewaartermijn van alle ruwe data en het archiveren van ruwe data.

2. Data verzamelen en analyseren

Wanneer nieuwe data verzameld of bestaande data hergebruikt worden, moeten de privacy en gevoelige gegevens van betrokken personen of zorginstellingen worden beschermd en het proces zodanig gedocumenteerd dat replicatie mogelijk is:

	JA	NEE	N.V.T.
Bestaande data hergebruiken			
Ik ga bestaande onderzoeksgegevens of datasets hergebruiken ⁵ . Zo ja, specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	
Nieuwe data verzamelen			
Ik verkrijg de onderzoeksgegevens van deelnemers met (een vorm van) toestemming ⁶ . Zo ja, specificeer:	<input type="checkbox"/>		<input type="checkbox"/>
Gevoelige gegevens beschermen			
Ik neem maatregelen om gevoelige gegevens te beschermen ⁷ . Zo ja, specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De kwaliteit van de onderzoeksgegevens waarborgen⁸			
Ik documenteer het onderzoeksproces ⁹ om replicatie mogelijk te maken. Zo nee, verklaar waarom niet:	<input type="checkbox"/>	<input type="checkbox"/>	

3. Data opleveren: projectproducten

Elders beschrijft u de beoogde projectresultaten. Ook data vormen een projectresultaat. De data moeten zo worden beschreven en aangeboden dat deze voor hergebruik in de toekomst te lezen en te interpreteren zijn.

	JA	NEE	N.V.T.
Ik bied beschrijvende informatie aan bij de verschillende projectproducten zodat de data in de toekomst te lezen en te interpreteren zijn ¹⁰ . Zo ja, specificeer de informatie en de te gebruiken metadatastandaarden, zo nee, leg uit waarom niet:	<input type="checkbox"/>	<input type="checkbox"/>	

⁵ Behandel hier in ieder geval of u al toestemming heeft voor het hergebruik en of u deze data voor uw onderzoek moet koppelen aan andere data. Indien u moet koppelen dan graag ook de bron vermelden, aangeven of u al toestemming heeft om te koppelen, welke techniek u gaat gebruiken en hoe u de privacy gaat borgen.

⁶ Denkt u hierbij aan informed consent, opt-out en dergelijke. In beginsel worden alleen onderzoeksgegevens met kennis en toestemming van de betrokken personen en/of zorginstellingen verkregen. Op basis van de Gedragscode Gezondheidsonderzoek zijn enkele uitzonderingen mogelijk op de noodzaak van toestemming, zie art.5 en art.6.

⁷ Om onderzoeksgegevens te kunnen delen met derden moet de privacy van betrokken personen of zorginstellingen gewaarborgd worden. U moet aan kunnen tonen dat u zich bewust bent van het belang van de privacy van betrokken personen of zorginstellingen en eveneens dat u daar naar handelt. Hierbij moet u denken aan het vragen van toestemming voor het gebruik van gegevens (informed consent, opt-out), anonimiseren (het verwijderen van tot persoon of zorginstelling herleidbare gegevens zodanig dat anonieme gegevens overblijven; koppelen is dan meestal niet meer mogelijk) en pseudonimiseren (een vorm van versleuteling van identificerende gegevens waarmee identificatie van de persoon redelijkerwijs wordt voorkomen dan wel onmogelijk wordt, maar waarbij de mogelijkheid tot koppelen open wordt gehouden) van onderzoeksgegevens. In de Gedragscode Gezondheidsonderzoek kunt u al deze aspecten terug vinden.

⁸ Het is essentieel dat de door u aangeboden data goed te interpreteren zijn en voldoen aan kwaliteitsstandaarden. Verder moeten de data aan allerlei wettelijke en vakinhoudelijke criteria voldoen.

⁹ Zulke documentatie volgt uit de Nederlandse Gedragscode Wetenschapsbeoefening (VSNU). Hierbij kunt u bijvoorbeeld denken aan labjournaals, automatische scripts, query syntax en Standard Operating Protocols.

¹⁰ Denk bijvoorbeeld aan gestandaardiseerde metadata zoals Dublin Core en het Data Documentation Initiative DDI, aan informatie over bewerkingssoftware en aan contextinformatie.

4. Data bewaren tijdens en na het project

Voor de beschikbaarheid tijdens en na het afloop van het project is het belangrijk om te weten hoe de data gedurende het onderzoek opgeslagen en geback-up worden en waar de data voor langere termijn beschikbaar zullen worden gemaakt.

	JA	NEE	N.V.T.
Ik heb voldoende opslaglocaties en -capaciteit en ik heb back-up van de data geregeld. Zo ja, specificeer, zo nee, specificeer hoe u dit gaat oplossen en bekostigen:	<input type="checkbox"/>	<input type="checkbox"/>	
Ik archiveer/publiceer de data in een repository of archief voor minimaal 5 jaar (zie Uitgangspunten en de toelichting daarop). Zo ja, specificeer het archief en of het een internationale certificering heeft: Zo nee, specificeer waarom niet en hoe de data niettemin duurzaam toegankelijk en bruikbaar zijn:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De kosten van (het voorbereiden van de data voor) archivering zijn gedekt binnen deze aanvraag. Zo nee, specificeer hoe u dit oplost:	<input type="checkbox"/>	<input type="checkbox"/>	

5. Data beschikbaar stellen

Na afloop van het project en een eventueel door de financier toegestane embargoperiode dienen uw data beschikbaar te zijn voor andere onderzoeksgroepen. Hierbij is het van belang dat de onderzoeksgegevens goed te vinden zijn en onder welke voorwaarden deze hergebruikt kunnen worden¹¹.

	JA	NEE	N.V.T.
Ik stel mijn data na afloop van het onderzoeksproject (of eerder) beschikbaar voor verificatie en vervolgonderzoek. Zo ja, beschrijf onder welke voorwaarden. Zo nee, beschrijf waarom niet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De onderzoeksgegevens en beschrijvende informatie die ik beschikbaar stel, zijn goed te vinden op het internet (wat niet impliceert dat de data openbaar toegankelijk zijn of moeten zijn). Zo ja beschrijf waar de onderzoeksgegevens te vinden zijn. Zo nee, leg uit waarom niet:	<input type="checkbox"/>	<input type="checkbox"/>	

¹¹ Om onderzoeksgegevens te kunnen delen met derden moet u nauwkeurig omschrijven aan welke voorwaarden een onderzoeksgroep, die toegang wil verkrijgen tot uw onderzoeksgegevens, moet voldoen. Daarbij kunt u denken aan afspraken die worden gemaakt rondom methodologie, publicaties, de termijn van toestemming, beschikbaarheid van gegevens, de kosten, auteursrechtelijke aspecten en dergelijke.

4.2 Checklist Datamanagementplan voor ZonMw-projecten

Beschikbaarheid van kwalitatief hoogwaardige onderzoeksbestanden is van groot belang voor gezondheidsonderzoek. In Nederland en daarbuiten is een groot aantal gegevensbronnen beschikbaar op basis waarvan onderzoek kan worden verricht; zie bijvoorbeeld zorggegevens.nl. Door onderzoeksbestanden beschikbaar te stellen voor wetenschappelijk onderzoek door derden kunnen andere onderzoeksvragen (efficiënt) worden beantwoord. Ook kunnen gegevensbronnen worden gecombineerd om vragen te beantwoorden die uit het primaire onderzoeksbestand alleen niet te beantwoorden zijn. Voorbeelden hiervan zijn onderzoek bij mensen met zeldzame aandoeningen of onderzoek naar de relatie tussen woon- en werksituatie en gezondheidsproblemen.

Steeds vaker stellen onderzoeksfinanciers eisen aan de beschikbaarheid van data. Ook ZonMw wil dat (op termijn) onderzoekers die subsidie ontvangen van ZonMw alleen nog databestanden opleveren die vindbaar, toegankelijk en uitwisselbaar zijn en dat het gebruik van bestaande dataverzamelingen een vanzelfsprekend onderdeel is in wetenschappelijk onderzoek en het informeren van beleid en praktijk. In de notitie “Toegang tot data: de ontwikkeling van datamanagement bij ZonMw” is meer informatie te vinden.

Met het oog op vindbaarheid, toegankelijkheid en uitwisselbaarheid dient een subsidieontvanger de zogeheten dataparagraaf uit de subsidieaanvraag – na toekenning van de subsidie – uit te werken tot een datamanagementplan. De aspecten die hierin aan bod moeten komen, staan in deze checklist en omvatten in het kort het volgende. Er zijn diverse gedragscodes en er is wet- en regelgeving voor wetenschapsbeoefening (VSNU, WMO, WBGO, CBP) en voor het opslaan en gebruik van persoonsgegevens die gelden voor iedereen die in Nederland onderzoek uitvoert. Wanneer nieuwe data verzameld of bestaande data hergebruikt worden, moet de privacy van betrokken personen of zorginstellingen worden beschermd en het proces zodanig gedocumenteerd dat replicatie mogelijk is. Bij hergebruik en/of koppeling van datasets moet er toestemming zijn. Daarnaast moet verzameling/analyse voldoen aan geldende regelgeving.

In de goedgekeurde aanvraag hebt u de beoogde projectresultaten beschreven. Ook data vormen een projectresultaat en ze moeten zo worden beschreven en aangeboden dat ze voor hergebruik in de toekomst te lezen, te interpreteren en zo nodig te koppelen zijn. Voor de beschikbaarheid tijdens en na afloop van het project is het belangrijk om te weten hoe de data gedurende het onderzoek opgeslagen en geback-up worden en waar de data voor langere termijn beschikbaar zullen worden gemaakt. Aspecten hierbij zijn omvang, groei, versiebeheer, beveiliging, bewaartijd en kosten. Na afloop van het project en een eventueel door de financier toegestane embargoperiode dienen uw data beschikbaar te zijn voor andere onderzoeksgroepen. Hierbij is het van belang dat de onderzoeksgegevens goed te vinden zijn en onder welke voorwaarden deze hergebruikt kunnen worden.

In het onderstaande verwijst “ik” naar de hoofdonderzoeker of andere contactpersoon die aan de hand van deze checklist het datamanagementplan schrijft.

Uitgangspunten

Er zijn diverse gedragscodes en regelgeving voor wetenschapsbeoefening (VSNU, WMO, WGBO, CBP) en persoonsgegevens die gelden voor iedereen die in Nederland onderzoek uitvoert en die ook datamanagementaspecten omvatten.

	JA	N.V.T.
Ik houd me aan de actuele Nederlandse Gedragscode Wetenschapsbeoefening (VSNU-code) en zie er in redelijkheid op toe dat ook de andere projectleden dit doen ¹² .	<input type="checkbox"/>	<input type="checkbox"/>
Ik houd me aan de actuele Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek (VSNU-code) en zie er in redelijkheid op toe dat ook de andere projectleden dit doen.	<input type="checkbox"/>	<input type="checkbox"/>
Ik houd me aan de actuele Gedragscode Gezondheidsonderzoek en zie er in redelijkheid op toe dat ook de andere projectleden dit doen.	<input type="checkbox"/>	<input type="checkbox"/>
Ik houd me aan de Wet Medisch-Wetenschappelijk Onderzoek met mensen (WMO) en zie er in redelijkheid op toe dat ook de andere projectleden dit doen.	<input type="checkbox"/>	<input type="checkbox"/>
Ik houd me aan de Wet op de Geneeskundige Behandelingsovereenkomst (WGBO) en zie er in redelijkheid op toe dat ook de andere projectleden dit doen.	<input type="checkbox"/>	<input type="checkbox"/>
Ik houd me aan de Richtsnoeren voor de informatiebeveiliging van het College Bescherming Persoonsgegevens (CBP) en zie er in redelijkheid op toe dat ook de andere projectleden dit doen.	<input type="checkbox"/>	<input type="checkbox"/>
Ik houd me aan richtlijnen en voorbeelden die in mijn organisatie bestaan en zie er in redelijkheid op toe dat ook de andere projectleden dit doen. Zo ja, specificeer de richtlijnen en voorbeelden:	<input type="checkbox"/>	<input type="checkbox"/>

1. Administratieve informatie

1.1	Referentienummer	Het referentienummer of het door ZonMw toegewezen projectnummer.
1.2	Onderzoeksfincancier	Door wie wordt het onderzoek gefinancierd? Noem hierbij ook eventuele cofinanciers.
1.3	Projectnaam	De naam van het project.
1.4	Projectbeschrijving	Vat kort het type onderzoek (of onderzoeken) samen om anderen het doel van de dataverzameling uit te leggen. <i>Max. 50 woorden</i>
1.5	Hoofdonderzoeker	De naam van de hoofdonderzoeker van het project
1.6	Gegevens contactpersoon	Naam telefoon en e-mailadres.
1.7	Datum eerste versie	De datum waarop de eerste versie van het datamanagementplan gereed was
1.8	Datum huidige versie	De datum wanneer het datamanagementplan voor de laatste keer is aangepast.

¹² In de uitwerking van het principe dat onderzoeksresultaten controleerbaar dienen te zijn, noemt de gedragscode onder meer de noodzaak van replicerbaarheid van onderzoek, de kwaliteit van dataverzameling en dataopslag, goede verslaglegging van alle stappen, vijf jaar als minimale bewaartermijn van alle ruwe data en het archiveren van ruwe data.

2. Data verzamelen en analyseren

Wanneer nieuwe data worden verzameld of bestaande data hergebruikt moeten de privacy en gevoelige gegevens van betrokken personen of zorginstellingen worden beschermd en het proces zodanig gedocumenteerd dat replicatie mogelijk is. Bij hergebruik en/of koppeling van datasets moet er toestemming zijn. Daarnaast moet verzameling/analyse voldoen aan geldende regelgeving.¹³

Indien u gebruik maakt van gegevens die (direct of indirect) tot een individu herleidbaar zijn, gaat ZonMw ervan uit dat u voldoet aan onderstaande punten. Indien u hier niet aan voldoet, zal uw onderzoeksvoorstel nader bekeken worden.

2.1 Bestaande data hergebruiken		JA	NEE	N.V.T
2.1.1	Ik ga bestaande onderzoeksgegevens of datasets hergebruiken. Zo nee, ga verder bij 2.2. Zo ja, specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	
2.1.2	Ik heb toestemming gekregen om deze onderzoeksgegevens of datasets te gebruiken. Zo nee, specificeer hoe dat opgelost gaat worden:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.1.3	Voor het hergebruik moet ik data koppelen ¹⁴ . Zo ja, geef aan met welke data gekoppeld moet worden, zo nee, ga naar 2.1.6.	<input type="checkbox"/>	<input type="checkbox"/>	
2.1.4	Ik heb toestemming gekregen om de data te koppelen. Zo ja, specificeer onder welke voorwaarden, zo nee, leg uit:	<input type="checkbox"/>	<input type="checkbox"/>	
2.1.5	Het is technisch mogelijk om de gegevens verzameld in dit project te koppelen met andere onderzoeksgegevens ¹⁵ . Zo ja, zijn er nog aanvullende voorwaarden waaraan moet worden voldaan:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.1.6	Ik waarborg de privacy van toeleveranciers van onderzoeksgegevens. ¹⁶ Specificeer:	<input type="checkbox"/>		<input type="checkbox"/>
2.1.7	Voor het gebruik van deze data gelden beperkingen ¹⁷ . Zo ja, specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.2 Nieuwe data verzamelen				

¹³ Om onderzoeksgegevens te kunnen delen met derden moet de privacy van betrokken personen of zorginstellingen gewaarborgd worden. U moet aan kunnen tonen dat u zich bewust bent van het belang van de privacy van betrokken personen of zorginstellingen en eveneens dat u daar naar handelt. Het gaat er bij dit onderdeel om dat u aangeeft welke maatregelen er worden getroffen om de privacy van betrokken personen of zorginstellingen te borgen. Hierbij moet u denken aan het vragen van toestemming voor het gebruik van gegevens (informed consent, opt-out), anonimiseren (het verwijderen van tot persoon of zorginstelling herleidbare gegevens zodanig dat anonieme gegevens overblijven; koppelen is dan meestal niet meer mogelijk) en pseudonimiseren (een vorm van versleuteling van identificerende gegevens waarmee identificatie van de persoon redelijkerwijs wordt voorkomen dan wel onmogelijk wordt, maar waarbij de mogelijkheid tot koppelen open wordt gehouden) van onderzoeksgegevens. In de Gedragscode Gezondheidsonderzoek kunt u al deze aspecten terug vinden.

¹⁴ Het gaat er hier om dat u, zowel op technisch, juridisch als op ethisch gebied, beschrijft wat uw koppelmogelijkheden zijn. Koppeling van gegevens over dezelfde persoon die in verschillende onderzoeken of registers zijn verzameld, is een manier om verzameling van gegevens beter te benutten.

¹⁵ Met "technisch" wordt bedoeld dat het überhaupt in technische zin mogelijk is om de gegevens te koppelen. Dat betekent dat er op de een of andere manier een identifier in een dataset is opgenomen waarmee je de data van een persoon uit de ene set kunt koppelen aan data van diezelfde persoon in een andere set. Die identifier kan een pseudoniem zijn, een BSN-nummer, of een combinatie van variabelen waarmee je met redelijke zekerheid een persoon kunt vaststellen (geboortedatum, geslacht, postcode).

¹⁶ Voeg de passages uit het privacyreglement toe waarin dit wordt uitgewerkt.

¹⁷ Denk bijvoorbeeld aan bescherming van de privacy, openbare veiligheid, intellectueel eigendom, commerciële belangen, of andere vereisten zoals een bepaalde expertise bij nieuwe gebruikers van de data.

		JA	NEE	N.V.T
2.2.1	Ik verkrijg de onderzoeksgegevens van deelnemers met (een vorm van) toestemming ¹⁸ . Zo ja, specificeer:	<input type="checkbox"/>		<input type="checkbox"/>
2.2.2	Ik heb het consent zo opgesteld/ ik informeer deelnemers zodanig dat het consent, naast het oorspronkelijke doel van de gegevensverzameling, de mogelijkheid biedt dat de onderzoeksgegevens ook voor ander (ZonMw-) onderzoek kunnen worden gebruikt. ¹⁹ Zo nee, specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2.3 Gevoelige gegevens beschermen				
		JA	NEE	N.V.T
2.3.1	Een ieder die betrokken is bij de verwerking van persoonsgegevens tekent een geheimhoudingsverklaring. ²⁰	<input type="checkbox"/>		<input type="checkbox"/>
2.3.2	Ik leg het project ter toetsing voor aan een Medisch Ethische Toetsingscommissie. ²¹	<input type="checkbox"/>		<input type="checkbox"/>
2.3.3	Ik meld het project aan bij het CBP.	<input type="checkbox"/>		<input type="checkbox"/>
2.3.4	Ik anonimiseer/ pseudonimiseer (privacygevoelige) onderzoeksgegevens. ²² Zo ja, specificeer:	<input type="checkbox"/>		<input type="checkbox"/>
2.3.5	Ik houd me aan het privacyreglement van de organisatie waaraan ik verbonden ben. Zo nee, leg uit:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.3.6	De organisatie waaraan ik verbonden ben, heeft beleid voor <i>information security</i> . Zo ja, wie is hiervoor verantwoordelijk?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.3.7	Ik controleer zelf de naleving van de maatregelen die genomen zijn om de privacy en de niet-herleidbaarheid van de onderzoeksgegevens te waarborgen. Dit item is m.n. van belang indien 2.3.5 en 2.3.6 negatief worden beantwoord. Zo ja, specificeer ²³ :	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

¹⁸ Denkt u hierbij aan informed consent, opt-out en dergelijke. In beginsel worden alleen onderzoeksgegevens met kennis en toestemming van de betrokken personen en/of zorginstellingen verkregen. Op basis van de gedragscode gezondheidsonderzoek zijn er enkele uitzonderingen mogelijk. Met dit document wordt de Gedragscode Gezondheidsonderzoek versie 2004 bedoeld. Een nieuwe versie van de Gedragscode is op komst. Op basis van de Gedragscode Gezondheidsonderzoek zijn enkele uitzonderingen mogelijk op de noodzaak van toestemming, zie art.5 en art.6.

¹⁹ Onder de WBP is gebruik voor secundaire analyses van onderzoeksgegevens toegestaan op voorwaarde dat dit niet onverenigbaar is met het doel waarvoor de onderzoeksgegevens verzameld zijn. Zie ook verklaring van Helsinki over ethische principes.

²⁰ Een geheimhoudingsverklaring is iets anders dan een beroepsgeheim. Een beroepsgeheim is wettelijk geregeld in de relatie tussen arts en patiënt. Een geheimhoudingsverklaring is een belofte dat de persoonsgegevens die iemand onder ogen krijgt, niet doorverteld zullen worden. U hoort een vastgesteld beveiligingsbeleid te hebben. Zie ook Gedragscode Gezondheidsonderzoek.

²¹ Onderzoek dat WMO-plichtig is, hoort getoetst te worden door een Medisch Ethische Toetsingscommissie. Zie ook Gedragscode Gezondheidsonderzoek.

²² Op basis van de WBP en daaruit voortvloeiende Gedragscode Gezondheidsonderzoek gebruikt u geanonimiseerde/gepseudonimiseerde gegevens voor onderzoeksanalyses.

²³ Indien u een privacyreglement heeft, voeg dit toe als bijlage. Als u geen privacyreglement heeft, beschrijf dan hoe wordt gecontroleerd op de naleving van de maatregelen die genomen zijn om de privacy te waarborgen.

2.4 De kwaliteit van de onderzoeksgegevens waarborgen ²⁴				
		JA	NEE	N.V.T.
2.4.1	Ik documenteer het onderzoeksproces ²⁵ om replicatie mogelijk te maken. Zo nee, verklaar waarom niet:	<input type="checkbox"/>	<input type="checkbox"/>	
2.4.2	Ik voer kwaliteitscontroles uit ten aanzien van volledigheid, correctheid en consistentie van de gegevens. Specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	
2.4.3	Er bestaat een wettelijke basis, richtlijn of verplichting om de gegevens te registreren/onderzoeken Zo ja, specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	

2.5 Samenwerken met andere (onderzoeks)organisaties. ²⁶				
		JA	NEE	N.V.T.
2.5.1	Ik werk bij de verzameling van onderzoeksgegevens samen met partners of toeleveranciers van gegevens. Zo ja, specificeer afspraken over gebruiksrechten van de onderzoeksgegevens uit het project:	<input type="checkbox"/>	<input type="checkbox"/>	

3. Data opleveren: projectproducten

In de goedgekeurde aanvraag hebt u de beoogde projectresultaten beschreven. Ook data vormen een projectresultaat. De data moeten zo worden beschreven en aangeboden dat deze voor hergebruik in de toekomst te lezen, te interpreteren en zo nodig te koppelen zijn:

Welke producten levert het project op. ²⁷				
		JA	NEE	N.V.T.
3.1	Ik voorzie de volgende eindproducten van het project: Specificeer:			
3.2	Ik bied beschrijvende informatie aan bij de verschillende projectproducten zodat de data in de toekomst te lezen en te interpreteren zijn. ²⁸ Zo ja, specificeer de informatie en de metadatastandaarden: Zo nee, leg uit waarom niet:	<input type="checkbox"/>	<input type="checkbox"/>	

²⁴ Het is essentieel dat de door u aangeboden data goed te interpreteren zijn en voldoen aan kwaliteitsstandaarden. Verder moeten de data aan allerlei wettelijke en vakinhoudelijke criteria voldoen.

²⁵ Zulke documentatie volgt uit de Nederlandse Gedragscode Wetenschapsbeoefening (VSNU). Hierbij kunt u bijvoorbeeld denken aan labjournaals, automatische scripts, query syntax en Standard Operating Protocols.

²⁶ Wanneer bij de verzameling van onderzoeksgegevens wordt samengewerkt met andere (onderzoeks)organisaties is het van belang dat een aantal zaken wordt afgesproken met betrekking tot eigenaarschap of co-producentschap van data, voorwaarden voor gebruik van gegevens door derden (bijv. auteurschappen, toestemming voor onderzoek, informeren over nieuwe onderzoeksrapporten en artikelen) en verantwoordelijkheden in dit proces. De afspraken die u maakt met samenwerkingspartners dient u schriftelijk vast te leggen. Indien er sprake is van een samenwerkingsovereenkomst, voeg deze als bijlage toe.

²⁷ Het is belangrijk om bijtijds te bedenken welke producten het project oplevert. Denk hierbij bijvoorbeeld aan ruwe data, (verschillende versies van) bewerkte data, documentatie over de data, documentatie van het onderzoeksproces inclusief alle betrokkenen, syntaxen, software.

²⁸ Denk bijvoorbeeld aan gestandaardiseerde metadata zoals Dublin Core en het Data Documentation Initiative DDI, aan informatie over bewerkingssoftware en aan contextinformatie.

3.3	Het is mogelijk om de gegevens te koppelen met andere onderzoeksgegevens. Zo ja, specificeer standaard(en) en/of de betreffende koppelingsvariabele(n): Zo nee, leg uit hoe u dit oplost:	<input type="checkbox"/>	<input type="checkbox"/>	
3.4	Ik voorzie hergebruik van de uitkomsten in ander onderzoek. Zo ja, specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	
3.5	Ik heb criteria om te bepalen welk deel van de uitkomsten het bewaren waard is. ²⁹ Zo ja, specificeer. Zo nee, leg uit hoe u dit oplost:	<input type="checkbox"/>	<input type="checkbox"/>	
3.6	Een deel van de uitkomsten moet na afloop van het project vernietigd worden, vanwege een contract of de wet. Zo ja, specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	

4. Data bewaren tijdens en na het project

Voor de beschikbaarheid tijdens en na het afloop van het project is het belangrijk om te weten hoe de data gedurende het onderzoek opgeslagen en geback-up worden en waar de data voor langere termijn beschikbaar zullen worden gemaakt. Aspecten hierbij zijn omvang, groei, versiebeheer, beveiliging, bewaartijd, en kosten.

4.1 Data bewaren gedurende het onderzoek.³⁰

		JA	NEE	N.V.T.
4.1.1	Voor opslag en back-up van mijn gegevens maak ik gebruik van de standaardvoorzieningen van mijn instituut. Zo ja, ga naar 4.2.	<input type="checkbox"/>	<input type="checkbox"/>	
4.1.2	Voor opslag en back-up van mijn gegevens maak ik gebruik van een externe dienstverlener. Zo ja, noem de dienstverlener en ga dan naar 4.2.	<input type="checkbox"/>	<input type="checkbox"/>	
4.1.3	Ik weet bij benadering wat de omvang en eventuele groei van de data zijn. Zo ja, specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	
4.1.4	Ik heb voldoende opslaglocaties en -capaciteit. Zo ja, specificeer (incl. bekostiging): Zo nee, specificeer hoe u dit gaat oplossen en bekostigen:	<input type="checkbox"/>	<input type="checkbox"/>	
4.1.5	Ik heb versiebeheer van de data geregeld. Zo ja, specificeer: Zo nee, specificeer hoe u dit gaat oplossen:	<input type="checkbox"/>	<input type="checkbox"/>	
4.1.6	Ik heb back-up van de data geregeld. Zo ja, specificeer: Zo nee, specificeer hoe u dit gaat oplossen:	<input type="checkbox"/>	<input type="checkbox"/>	

²⁹ 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 gegevens zal onder omstandigheden ook minder privacygevoelig kunnen zijn dan het bewaren. Zie de [checklist van DANS](#).

³⁰ Beschrijf hoe vaak en waar de data geback-up worden en wie waarvoor verantwoordelijk is. 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 in het algemeen 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) of het beleid van uw afdeling of instituut, bijvoorbeeld over de beveiliging van gevoelige data.

4.1.7	Ik neem de volgende technische en organisatorische maatregelen ter beveiliging van onderzoeksgegevens ³¹ :			
4.1.8	Ik neem fysieke maatregelen voor toegangsbeveiliging ³² Zo ja, specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.1.9	Ik maak gebruik van logische toegangscontroles met behulp van iets wat iemand weet (wachtwoorden, pincodes)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.1.10	Ik maak gebruik van logische toegangscontroles met behulp van iets wat iemand bij zich draagt (pasjes)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.1.11	Ik maak gebruik van logische toegangscontroles met behulp van iets wat bij iemand hoort (biometrisch kenmerk) en overige logische toegangscontroles	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.1.12	Ik maak gebruik van overige beveiligingsmaatregelen Zo ja, specificeer ³³ :	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4.2 Het is belangrijk te weten hoe en waar de data voor de lange termijn bewaard zullen worden. ³⁴				
		JA	NEE	N.V.T.
4.2.1	Ik archiveer/publiceer de data in een repository of archief. Zo ja, specificeer het archief en of het een internationale certificering heeft: Zo nee, specificeer waarom niet en hoe de data niettemin duurzaam toegankelijk en bruikbaar zijn.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.2.2	De bewaartermijn van de data is: Specificeer:			
4.2.3	De kosten van (het voorbereiden van de data voor) archivering zijn gedekt. Zo nee, specificeer hoe u dit oplost:	<input type="checkbox"/>	<input type="checkbox"/>	

5. Data beschikbaar stellen

Na afloop van het project en een eventueel door de financier toegestane embargoperiode dienen uw data beschikbaar te zijn voor andere onderzoeksgroepen. Hierbij is het van belang dat de onderzoeksgegevens goed te vinden zijn en onder welke voorwaarden deze hergebruikt kunnen worden.

5.1 Data beschikbaar stellen en gebruiksvoorwaarden ³⁵				
		JA	NEE	N.V.T.
5.1.1	Er zijn onderzoeksgegevens die ik niet beschikbaar ga stellen. Zo ja, leg uit welke en waarom. Kijk eerst naar mogelijke gebruiksvoorwaarden in dit blok.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

³¹ Indien u dit heeft vastgelegd in een document, voeg dit als bijlage toe. Indien u dit niet in een document heeft vastgelegd, beschrijf dan alle fysieke en technische maatregelen die zijn genomen om de onderzoeksgegevens te beveiligen.

³² U kunt hierbij denken aan organisatorische controle, inbraakalarm en kluis voor opslag van gegevensbestanden.

³³ Indien de beveiligingsmaatregelen in een document zijn opgenomen, voeg dan het document (ook) toe.

³⁴ Vroegtijdig overleg met het beoogde repository of archief is aan te bevelen, bijvoorbeeld over geschikte bestandsformaten en benodigde metadata. Als u kiest voor een repository dat niet beschikt over het Data Seal of Approval, DIN-31644-, ISO-16363- of WDS/ICSU-certificering, maak dan duidelijk hoe ervoor gezorgd is dat de data na afloop van het project effectief toegankelijk en herbruikbaar zijn.

³⁵ Om onderzoeksgegevens te kunnen delen met derden moet u nauwkeurig omschrijven aan welke voorwaarden een onderzoeksgroep, die toegang wil verkrijgen tot uw onderzoeksgegevens, moet voldoen. Daarbij kunt u denken aan afspraken die worden gemaakt rondom methodologie, publicaties, de termijn van toestemming, beschikbaarheid van gegevens, de kosten, auteursrechtelijke aspecten en dergelijke.

5.1.2	Ik stel voorwaarden aan het gebruik van de onderzoeksgegevens. Zo ja:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.1.3	Ik deel onderzoeksgegevens voor commerciële doeleinden Zo ja, specificeer voorwaarden:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.1.4	Ik stel voorwaarden met betrekking tot dataveiligheid Zo ja, specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.1.5	Ik maak afspraken rondom publicaties, auteurschappen, en de periode van toestemming. Zo ja, specificeer welke afspraken:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.1.6	Er geldt een handling fee voor het verkrijgen van onderzoeksgegevens. Zo ja, geef een indicatie van de hoogte:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.1.7	Ik breng andere kosten dan een handling fee in rekening. Zo ja, specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.1.8	Ik ga een samenwerkingsovereenkomst aan met de onderzoekspartner(s).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.1.9	Ik stel beperkingen ten aanzien van de te beantwoorden onderzoeksvragen. Zo ja, specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.1.10	Ik stel overige gebruiksvoorwaarden aan derden Zo ja, specificeer:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.1.11	Over de goedkeuring van gegevensaanvragen beslist een stuurgroep, programmacommissie of projectleider. Zo ja, specificeer ³⁶ :	<input type="checkbox"/>	<input type="checkbox"/>	

5.2 De data en beschrijvende informatie die u beschikbaar stelt, zijn goed te vinden op het internet (wat niet impliceert dat de data openbaar toegankelijk zijn of moeten zijn).

		JA	NEE	N.V.T.
5.2.1	De data zijn te vinden via een catalogus, zoekmachine of webportaal dat de metadata en/of andere beschrijvende informatie benut. Zo nee, leg uit:	<input type="checkbox"/>	<input type="checkbox"/>	
5.2.2	Ik attendeer actief op de beschikbaarheid van de data. Zo nee, leg uit ³⁷ :	<input type="checkbox"/>	<input type="checkbox"/>	
5.2.3	Ik maak gebruik van een Persistent Identifier om duurzaam naar de data te verwijzen ³⁸ . Zo nee, leg uit:	<input type="checkbox"/>	<input type="checkbox"/>	

³⁶ Denk ook aan het voorleggen van een datahergebruikaanvraag aan de METC.

³⁷ Denk bijvoorbeeld aan een data hub, een helpdesk, gebruikersgroep of forum.

³⁸ Denk bijvoorbeeld aan een DOI, Handle, URN of ARK. Zulke identifiers zijn duurzamer dan gewone links (URL's) en worden aanbevolen voor datacitatie. Duurzame repositories genereren zulke identifiers.

4.3 Workshop checklist datamanagementplan: projectleiders aan het woord

Op 31 oktober 2013 is een workshop georganiseerd met twaalf projectleiders van lopende ZonMw-projecten. Tijdens de workshop met de projectleiders is de checklist DMP tegen het licht gehouden. Op basis van de bovengenoemde workshop zijn de volgende vragen beantwoord.

Wat is de praktische bruikbaarheid van het ontwikkelde concept-datamanagementplan?

In de workshop is uitgebreid gesproken over de praktische bruikbaarheid van de items uit de uitgebreide checklist DMP. De meeste items bleken duidelijk te zijn en goed te gebruiken om een datamanagementplan mee te maken. Ook is aangegeven dat de checklist bijdraagt aan de bewustwording van de activiteiten die ondernomen moeten worden bij het beschikbaar stellen van data. Enkele items zijn naar aanleiding van het commentaar van de projectleiders aangepast of toegevoegd.

Op welke onderdelen voldoen huidige projecten niet aan de elementen uit het datamanagementplan?

Tijdens de workshop, de interviews en discussies met de experts/projectleiders werd duidelijk dat nog lang niet alle projectleiders en onderzoekers ingevoerd zijn in datamanagement. De vindbaarheid op internet en de metadatering van onderzoeksdata laten veelal nog te wensen over. Er zijn ook een aantal uitzonderingen te noemen, koplopers op het gebied van metadatering en vindbaarheid van data zoals het [Genoom van Nederland](#), [LifeLines](#), [BBMRI](#) en [CTMM TraIT](#), [Profiles registry](#) en de catalogus van het [Parelsnoer instituut \(PSI\)](#).

Er lijkt een tweedeling te zijn ontstaan tussen de grotere projecten en kleine projecten. Grote projecten zoals deze koplopers hebben aparte datamanagementcentra met grote opslagfaciliteiten voor files en databases, evenals ter zake kundige medewerkers. Terwijl bij kleine projecten vaak geen specifieke opslag of datadeskundigheid voorhanden is. Daarbij is er zichtbaar beweging bij de instituten om datamanagementservices te realiseren, maar dit is nog niet voldoende. Bij de koplopers zijn experts op het gebied van datamanagement betrokken. Dit lijkt dus ook een voorwaarde te zijn voor goed datamanagement, al is het nog lastig om te bepalen wanneer een datamanager voldoende gekwalificeerd is. Certificering van deze expertise voert momenteel nog te ver, maar het is belangrijk dat men bekend raak met aanbod zoals de [cursus Data Intelligence](#) van Research Data Netherlands en het [Witboek Dataprofessionals](#) uit 2011.

Waarom voldoen projecten nog niet aan het DMP en in hoeverre is het haalbaar dat projecten in de toekomst hier wel aan voldoen?

De voor de hand liggende verklaring is dat het tot nu toe nog niet van onderzoekers gevraagd werd, althans niet door ZonMw; wel door sommige wetenschappelijke tijdschriften. Dit blijkt ook uit interviews met onderzoekers. Het op de juiste manier beschikbaar maken en houden van data kost bovendien veel tijd en wordt meestal gezien als een lastig 'sluitstuk' van een onderzoek. Deze datamanagementinspanning is voor onderzoekers niet aantrekkelijk omdat er niet voldoende beloning tegenover staat. Carrières binnen de wetenschappelijke wereld worden immers nog niet gemaakt of bestendigd door data beschikbaar te stellen. Erkenning voor het genereren en beschikbaar stellen van waardevolle onderzoeksgegevens, vergelijkbaar met de Hirsch-index voor wetenschappelijke publicaties, zou een grote stap voorwaarts zijn. Naarmate meer financiers, wetenschappelijke uitgevers en kennisinstellingen de zichtbaarheid van data bevorderen – door enerzijds hun voorwaarden en anderzijds steun voor een goede citatie-index voor data en voor de data-infrastructuur – zal dit veranderen.

Overige conclusies uit de workshop

Steun voor gedegen en gestructureerd datamanagement

Op basis van de workshop en de expert meeting (zie hoofdstuk 5) is de conclusie gerechtvaardigd dat de projectleiders en andere aanwezigen zich achter het idee scharen van gedegen en gestructureerd datamanagement. Onderzoekers onderschrijven dat zij verantwoordelijk zijn voor het vindbaar, toegankelijk en uitwisselbaar maken van hun onderzoeksdata. Er is in het veld weinig bezwaar tegen het verplicht stellen van een datamanagementplan, mits alle randvoorwaarden duidelijk zijn en aanvragers van subsidie geholpen worden bij het opzetten van goed datamanagement.

Maak duidelijke afspraken over eigenaarschap, hergebruik en metadatering

Uit de discussies mag geconcludeerd worden dat het van groot belang is om in een vroeg stadium duidelijke afspraken te maken over rechten op de data, hergebruik en de metadatering van de data waarvoor subsidie verleend is; dit geldt m.n. ook voor bewerkte data van derden. In dezen zijn de financiers van onderzoek aan zet. De weg die ZonMw is ingeslagen om het vindbaar, beschikbaar en uitwisselbaar maken van gegevens te verbinden aan de subsidie is dus een heel belangrijke stap in de goede richting.

Metadata en data nog erg versnipperd

Hoewel er veel positiefs te melden valt over een groot aantal belangrijke projecten, instituten en andere initiatieven als het gaat om het beschrijven en beschikbaar stellen van 'hun' data, is er nog wel het probleem dat het om uiteenlopende metadatabases gaat op verschillende plekken. Een onderzoeker moet in feite het internet 'afstruinen' voordat hij dankzij metadata relevante data vindt. Centrale opslag van data is niet nodig, maar een overkoepelende online catalogus of webportaal met metadata uit verschillende bronnen is zeer wenselijk. Het NARCIS-portaal van de KNAW kan een deel van deze functie op zich nemen (zie ook paragraaf 5.1).

Datamanagement veelal sluitstuk project

Datamanagement wordt nog te vaak beschouwd als een sluitstuk van een onderzoek. Er wordt bijvoorbeeld niet van meet af aan gewerkt aan goede metadatering, eenduidige bestandsnamen, het bijhouden van logboeken en andere manieren om het (verloop en de tussenresultaten) van een onderzoek te beschrijven. Veelal eindigt een onderzoek met het zo goed mogelijk opslaan van de (bewerkte) data met enige metadata, en wel binnen de eigen gelederen. In het slechtste geval staan de gegevens op een harddisk of ander medium dat niet geback-up't wordt en niet vindbaar is voor derden.

Behoeft aan ondersteuning

Er is duidelijk behoefte aan ondersteuning bij het invullen van de checklist en bij het voldoen aan de voorwaarden van ZonMw, bijvoorbeeld hulp bij het duurzaam opslaan van data. Daarbij valt in eerste instantie te denken aan *best practices*, sjablonen, loketten (binnen het eigen instituut) en dergelijke. Velen zijn van mening dat er (vooral) op instituutsniveau faciliteiten moeten komen voor datamanagement. De individuele onderzoeker heeft namelijk veelal te weinig expertise en tijd om voor een specifieke projectaanvraag de gevraagde data-informatie te verzamelen.

De projectleiders zijn van mening dat ZonMw hen in dit opzicht zou moeten ondersteunen bij het plannen van datamanagement. Verder worden er knelpunten voorzien met betrekking tot de financiering en de toetsing van het datamanagementplan. Budgettair wordt een knelpunt voorzien, omdat het lastig is om van te voren te bepalen hoe hoog de kosten worden.

Desgevraagd gaven de projectleiders nog de volgende adviezen aan ZonMw:

1. faciliteer en financier datamanagement zo veel mogelijk;
2. stimuleer dat alle onderzoeksinstituten een datamanagementloket hebben;
3. stimuleer nationale coördinatie van die loketten bijvoorbeeld via DTL/DISC;
4. stimuleer dat onderzoekers experts inschakelen en niet alles zelf doen;
5. controleer databeloftes en verbind consequenties aan het niet nakomen van beloftes;
6. houd onderzoekers aan hun verantwoordingsplicht;
7. stimuleer datamanagement met voorbeeldprojecten en voorbeelddataparagrafen;
8. laat het datamanagementplan in de aanvraag opnemen, dus voordat subsidie toegekend is (toelichting: dan kunnen bijvoorbeeld de benodigde ICT-voorzieningen eerder begroot worden);
9. verschaf duidelijkheid over de rol van de dataparagraaf in de beoordeling van de aanvraag.

5 Management van onderzoeksdata

In dit hoofdstuk geven we antwoord op vijf vragen inzake het beter vindbaar, beschikbaar en uitwisselbaar maken van onderzoeksdata:

1. Hoe kunnen bestaande databronnen beter vindbaar gemaakt worden?
2. Hoe kan de duurzame opslag van data gefaciliteerd worden met bijbehorende metadata volgens bestaande (internationale) standaarden?
3. Wat is het beste financieringsmodel voor de (duurzame) opslag van data?
4. Aan welke minimale eisen rondom privacy en gebruiksvoorwaarden dienen aanvragen te voldoen?
5. Wat moet er specifiek voor biobanken en beeldmateriaal toegevoegd worden?

Verder beschrijven we de expert meeting op 5 september 2013 die tot doel had om te achterhalen hoe verschillende experts aankijken tegen datamanagement en de plannen van ZonMw.

5.1 Hoe kunnen bestaande databronnen beter vindbaar gemaakt worden?

Het meest eenvoudige antwoord op deze vraag is: door de beschrijving van de beschikbare databestanden, de metadata, op een gestructureerde manier op het internet te zetten, makkelijk vindbaar voor veelgebruikte zoekmachines zoals Google. Zowel uit dit project als uit de voorstudie voor “Goed Gebruik Geneesmiddelen” (Grootveld et al., 2012) blijkt dat onderzoekers metadata toekennen, maar hierbij weinig gebruik maken van internationale standaarden zoals de [Dublin Core standaard](#), [SNOMED CT](#) en het [Data Documentation Initiative](#). Deze laatste standaard wordt bijvoorbeeld aanbevolen door de Britse Medical Research Council. Het gebruik van standaarden, zoals de data dictionaries van het [Parelsnoer Instituut](#) of de zogenoemde Minimum Information About- oftewel [MIA-standaarden](#) in de biowetenschappen, neemt wel degelijk toe, omdat het naast vindbaarheid ook voordelen biedt bij het aansluiten op andere dataverzamelingen (zie voor overzicht <http://biosharing.org/standards>). De standaarden zijn echter niet allemaal vrij toegankelijk en zijn vaak multi-interpretabel. Het beeld is dat de meer complexe metadatastandaarden nog niet tot standaard in het gezondheidsonderzoeksveld kunnen worden verheven. Het is niettemin belangrijk om data goed gestructureerd op te slaan en te voorzien van adequate metadatering.

Dataportals vormen een belangrijk instrument in de vindbaarheid van data doordat ze metadata vergaren en doorzoekbaar aanbieden. Repositories en portals werken met verschillende soorten metadata, maar een gemene deler is het generieke Dublin Core metadataschema. Het is aan te raden deze beperkte set aan metadata voor studies op te leggen, omdat deze set gebruikt wordt in uitwisselingsprotocollen ter bevordering van de vindbaarheid van data, zoals in het [OAI-PMH-protocol](#). Het bekende portal voor onderzoeksinformatie, [NARCIS](#), kan metadata “oogsten” uit onderliggende repositories die hun metadata aanbieden volgens OAI-PMH. Dit geldt zowel voor publicaties als ook voor data. Ook andere portals, bijvoorbeeld disciplinespecifieke, kunnen via dit protocol metadata “oogsten”. Verschillende portals kunnen ook verschillende gradaties – eenvoudiger of rijkere – aan metadata oogsten en presenteren, wanneer die rijkere metadata ten minste zijn toegekend.

Websites zoals www.zorggegevens.nl zijn thematisch ingedeeld en zorgen er in ieder geval voor dat gegevens over de volksgezondheid en zorg op één centraal punt te vinden zijn. De website geeft voldoende overzicht en inzicht om bijvoorbeeld een specifieke dataset te vinden en een contactpersoon. Echter, meer gedetailleerde onderzoeksvragen over beschikbare data kunnen niet met deze website beantwoord worden. Daarvoor zijn de beschrijvingen niet gedetailleerd genoeg.

Het is zaak om als onderzoeksfinancier de bestaande metadatatportals en -tools te gebruiken en wellicht zelfs financieel te ondersteunen. Om de metadata enigszins centraal te houden is het een optie om een keuze te maken voor een specifiek metadatatportal (per thema) en de onderzoeker te verplichten hierbij passende metadata toe te kennen. In ieder geval is het belangrijk dat de metadata zelf voldoen aan internationaal geldende criteria.

5.2 Hoe kan duurzame opslag van data gefaciliteerd worden met bijbehorende metadata volgens bestaande (internationale) standaarden?

Duurzame archivering van onderzoeksgegevens is essentieel voor hergebruik op de langere termijn. Voor zover gegevens digitaal zijn of gedigitaliseerd kunnen worden, komen hiervoor zogenoemde Trusted Digital Repositories (TDR) in aanmerking. Dit zijn data-archieven en repositories met een lange-termijnambitie die aan bepaalde criteria voldoen. De eenvoudigste set criteria is die van het internationale [Data Seal of Approval](#). Deze criteria en het bijbehorende keurmerk zijn onafhankelijk van het vakgebied. Een repository kan het Data Seal of Approval (DSA) verwerven wanneer het zelf aan alle richtlijnen voldoet of wanneer het een deel hiervan heeft uitbesteed aan een ander DSA-gecertificeerd repository.

ZonMw kan verlangen dat onderzoekers data in een dergelijk Trusted Digital Repository archiveren. Specifieke zorgdatarepositories met het DSA zijn er nog nauwelijks. Een onderzoeker kan echter bij bestaande brede repositories terecht (zie de [lijst met Seals](#)). Daarnaast kunnen instellingen die zorgdatarepositories onderhouden en het DSA hiervoor willen verwerven, deze intentie kenbaar maken bij ZonMw en een peildatum afspreken waarop de instelling het keurmerk dient te hebben. Hiervoor valt te denken aan een periode van twee à drie jaar.

Met een dergelijk groeitraject versterkt ZonMw de infrastructuur voor onderzoeksgegevens. Een knelpunt bij duurzame beschikbaarheid van data, ongeacht de discipline, is immers dat opslagvoorzieningen met regelmaat bekostigd (moeten) worden uit projectfinanciering, die per definitie eindig is. ZonMw zou zich hierbij niet moeten beperken tot uitvoerders van door ZonMw gesubsidieerde projecten, maar zou langetermijnarchivering ook kunnen stimuleren op het niveau van instellingen en grote projecten zoals het Parelsnoer Instituut en CTMM-TraIT.

Als onderzoekers gewend zijn hun gegevens op te slaan in internationale repositories, zouden ze dit moeten continueren, zeker indien tijdschriften of (co-)financiers dit eisen, zoals bijvoorbeeld in [ArrayExpress](#) voor gen-expressie-experimenten en [EGA](#) voor genotype-phenotype associate en next-generation sequencing experimenten van het European Bioinformatics Institute en zusterdatabases [GEO](#) en [dbGaP](#) bij de Amerikaanse National Institutes of Health. Hierbij wordt veelal gebruik gemaakt van vakgebiedspecifieke normatieve standaarden (die bepalen welke informatie moet worden gerapporteerd voor hergebruik, zoals [MIAME](#)) en formatteringsstandaarden (die bepalen hoe de informatie moet worden geformatteerd, zoals [MAGE-TAB](#)). Het is zeer onwenselijk als onderzoekers de data eveneens in een (Nederlands) trusted digital repository zouden moeten opslaan; bovendien zijn deze internationale voorzieningen al geoptimaliseerd voor de veelal grote datasets. DSA-certificering (of DIN, ISO, WDS/ICSU) van internationale repositories is zeker wenselijk, maar dit vergt een lange adem en lijkt niet direct de rol van ZonMw. In plaats hiervan verdient het de voorkeur dat (Nederlandse) metadatatportals gaan verwijzen naar deze opslagfaciliteiten om hergebruik te bevorderen.

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

1. de ruwe data (indien de onderzoeker bestaande data hergebruikt, is er in het onderzoek wellicht geen sprake van “ruwe” data; in dit geval moeten de bestaande data duurzaam gearchiveerd zijn of alsnog gearchiveerd worden);
2. de data die ten grondslag liggen aan publicaties;
3. documentatie over de gehanteerde onderzoeksmethodologie (zoals codeboeken, metadatering, 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.

Dit geheel wordt wel aangeduid als het “replication package”. Een goede uitwerking van dit minimum, 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. Het verdient aanbeveling voor ZonMw om deze lijst te laten aanvullen voor andere soorten onderzoek en de resulterende tabel toe te voegen aan de online toelichting op de dataparagraaf en datamanagementplan.

Verschillende codes en richtlijnen hanteren verschillende bewaartermijnen. De [Nederlandse Gedragscode Wetenschapsbeoefening](#) stelt “De bewaartermijn van ruwe onderzoeksgegevens is minimaal 5 jaar”. Het ligt voor de hand dat ZonMw bij deze algemeen geldende gedragscode aansluit, maar ook verder gaat. De genoemde [NFU-publicatie](#) bevat namelijk diverse langere bewaartermijnen voor uiteenlopend materiaal. Het verdient aanbeveling om draagvlak te zoeken en te ontwikkelen voor een langere minimale bewaartermijn voor zorggegevens dan vijf jaar. Een praktische aanbeveling is daarnaast om een dergelijk overzicht van uiteenlopende richtlijnen toe te voegen aan de online toelichting op de checklist. Overigens kunnen er ook goede redenen zijn om data niet te bewaren, bijvoorbeeld wanneer reproductie voordeliger is dan archivering of als data van derden niet bewaard mag worden na afloop van het project.

5.3 Wat is het beste financieringsmodel voor (duurzame) opslag van data?

Binnen het gezondheidsonderzoek wordt er gewerkt met zeer uiteenlopende soorten data. Die diversiteit is terug te zien in de verschillende repositories. Binnen onderzoeksprojecten is de lange-termijnarchivering is niet altijd even doordacht. Het verdient aanbeveling om hier in de projectbegroting expliciet aandacht voor te vragen en daar een bedrag voor te reserveren. Wat zeker ook kan helpen is om dit bedrag pas uit te keren nadat de data duurzaam zijn gearchiveerd.

Diverse partijen werken aan kostenmodellen voor duurzame dataopslag. Ter illustratie: het Europese FP7-project “[Collaboration to Clarify the Cost of Curation](#)”, dat tot een kostenmodel voor digitale data-archivering wil komen, betreft in zijn analyse al meer dan tien serieuze modellen-in-ontwikkeling. Het laatste woord is hier voorlopig nog niet over gesproken en archieven bepalen hun eigen tariefbeleid. Een concreet financieringsmodel voor opslag van medische data valt dan ook niet aan te wijzen. Een subsidieaanvrager/onderzoeker kan de kosten (in termen van inspanningen) beter schatten naarmate hij meer ervaring heeft met datamanagement, maar in het algemeen is die ervaring weinig concreet, ook bij instellingen die datadiensten verlenen, onder meer door de genoemde diversiteit. De eenvoudigste optie voor ZonMw is om dezelfde aanpak te kiezen als voor communicatie en implementatie, namelijk een vast percentage van het projectbudget ervoor te reserveren en in de begroting op te nemen. Bij gebrek aan concrete gegevens zou dit ook hier 5% kunnen zijn, deels te besteden tijdens het onderzoek aan metadatering en documentatie, deels te betalen aan het lange-termijnarchief. Belangrijke factoren zijn in het algemeen:

1. de benodigde capaciteit voor opslag en back-up, waarbij dit voor beelddata uiteraard sneller oploopt dan voor beschrijvingen;

2. de mate van handwerk bij het toekennen van metadata en het opstellen van overige documentatie zoals codeboeken en gebruikte queries in het statistische pakket;
3. de benodigde mate van beveiliging van de data;
4. het inhuren van externe (datamanagement-)expertise.

5.4 Aan welke minimale eisen rondom privacy en gebruiksvoorwaarden dienen aanvragen te voldoen?

Organisaties borgen de privacy van respondenten en proefpersonen op uiteenlopende manieren: met ICT-maatregelen, door beperkt toegang te verlenen tot identificerende gegevens, door het anonimiseren en pseudonimiseren van privacygevoelige gegevens (voor hergebruik van de data) en met behulp van informed consent. De ideeën van de deskundigen variëren als het gaat om de vraag wanneer de waarborging van de privacy van de respondenten geregeld dient te zijn: al binnen het onderzoeksproject of pas bij disseminatie.

Procedures en contracten voor hergebruik van onderzoeksgegevens lijken te bestaan. Echter, de voorwaarden worden vaak ad hoc gesteld en er is nog geen consensus over de minimale eisen die gesteld kunnen worden aan gebruiksvoorwaarden. De uitdaging is dan ook met name om meer consistentie in de procedures te creëren. Een goed voorbeeld voor datatoegang lijkt het model van DANS, waarbij enkele algemene typen voorwaarden worden gehanteerd waaruit elk onderzoek kan kiezen, namelijk Open Access (dit vergt slechts registratie in het repository) of Restricted Access (waarbij de deponerende van de data beslist over de aanvraag tot hergebruik). Beslissingen over de goedkeuring van gegevensaanvragen worden wel eens door de projectleider of hoofdonderzoeker genomen of door een stuurgroep. Dit laatste is bijvoorbeeld het geval bij grotere projecten zoals [Lifelines](#).

Voorwaarden voor hergebruik zijn verbonden met de uitwisselbaarheid van data: mogen de gevraagde data gerelateerd worden aan andere data? Naast de vraag of het “mag” is er eveneens de vraag of het “kan”. Er moet immers een gemeenschappelijk element in de databronnen zijn. Net als hiervoor besproken bij metadata hanteren onderzoekers ook voor uitwisselbaarheid idealiter standaarden. In de genoemde voorstudie voor “Goed Gebruik Geneesmiddelen” (Grootveld et al., 2012) wordt hier aandacht aan besteed. [Informatie over zulke standaarden en ICT-voorzieningen](#) is ook te vinden op de website van het programma “Toegang tot data”.

5.5 Wat moet specifiek voor biobanken en beeldmateriaal toegevoegd worden?

Biobanken hebben vrijwel altijd een eigen datamanagementinfrastructuur en je zou deze datainfrastructuren kunnen beschouwen als lange-termijnarchieven. Om de vindbaarheid van biobankstudies te verbeteren heeft de Nederlandse Biobanking and Biomolecular Research Infrastructure (BBMRI-NL) een nationale [catalogus](#) van de ongeveer 200 Nederlandse biobanken gerealiseerd die nu verder wordt uitgewerkt in een Europese catalogus in BBMRI-ERIC verband en die dan ook zonder wachtwoord openbaar beschikbaar moet komen. Daarnaast heeft BBMRI-NL ook een centrale datamanagementinfrastructuur ingericht voor grote data-analyse in genomics, ter ondersteuning van grote biobankanalyses, dat bestaat uit 500 TB aan opslag en 1000 cpu aan rekenkracht opgesteld bij UMCG en CIT Groningen ([www.bbmri.nl](#) and [www.nlgenome.nl](#)). Daarnaast zijn er in BBMRI-NL-context een tweetal biobankinfrastructuren aanwezig met nationale uitstraling in het Parelinoer Instituut (waarin 8 UMC's data delen) en de Lifelines biobank (met herbruikbare grootschalige sample- en dataopslag gedreven door een cohort van 165.000 participanten). In de aanpalende CTMM TraIT infrastructuur gericht op translationele studies is een nationaal archief voor beeldmateriaal opgericht ([Bio-Medical Imaging Archive](#), BMIA) voor veilige en duurzame opslag van

(radiologische) DICOM beelden uit onderzoek. Ook in TraIT is er een faciliteit voor de ondersteuning van kleinschalige 'investigator initiated studies' die zelf geen eigen dataopslagfaciliteit hebben met behulp van [OpenClinica](#) programmatuur. Ook is het Nationaal Initiatief Hersenen en Cognitie, regiegegaan van NWO, bezig met het project "Sharing NeuroImaging Data", dat binnenkort een white paper uitbrengt over het delen van visuele onderzoeksdata. Tenslotte zijn er voor de grote genetische/moleculaire studies EU-archieven (zoals hiervoor genoemde ArrayExpress, EGA, GEO en dbGaP). Het is wenselijk dat alle metadata uit deze catalogi worden toegevoegd aan andere metadataportals om onderzoekers een centraal loket te kunnen bieden.

5.6 Overige informatie: de expert meeting

De hieronder staande informatie is gebruikt als input voor het ontwikkelen van de checklist Datamanagement en de checklist dataparagraaf. Op 5 september 2013 is er een expert meeting gehouden, waarbij de hieronder vermelde vragen aan de experts zijn voorgelegd.

Vraag 1: Wat zijn de specifieke uitdagingen van zorggegevens?

Is privacy specifiek voor zorggegevens?

Waarschijnlijk niet. Ook andere disciplines, zoals criminologie, kennen dit aspect. Waarschijnlijk de meest specifieke uitdaging bij zorggegevens ligt bij het gebruik ervan, vooral het eigenaarschap. De sleutel tot het gebruik van de 'eigen' zorggegevens zou in handen moeten liggen van de patiënt. Ben je als patiënt eigenaar? En hoe zit het met de databeveiliging? Herleidbaarheid is eveneens een issue bij zorggegevens. Anonimiseren wordt steeds moeilijker; het beschermen van datasets daarmee ook. Dit komt mede doordat respondenten/burgers/patiënten steeds meer data in de openbaarheid brengen via social media en initiatieven als 'patients like me'. Deze data gaan zich overigens zelf ook steeds beter lenen voor onderzoek, dus er zitten twee kanten aan dit verhaal.

Het diffuse karakter van zorggegevens?

Een ander specifiek punt is het diffuse karakter van zorggegevens: denk bijvoorbeeld aan EPD-gegevens, EEG-data, röntgenfoto's en MRI's. Er wordt geopperd dat er geen goede metadatastandaard is voor zorggegevens, noch een goede standaard voor het uitwisselen van zorggegevens. Uitwisselstandaard HL7 is zeer moeilijk te implementeren, niet open en niet gratis. Uitwisselstandaarden helpen vaak niet bij het doel van hergebruik en uitwisselbaarheid. Een goede basisregistratie is veel belangrijker. Deze moet zo zijn ingericht dat er op verschillende – en steeds veranderende – standaarden kan worden aangesloten. Standaarden die afkomstig zijn uit de verschillende vakgebieden zelf en met een hoge mate van flexibiliteit zijn vaak de beste katalysatoren voor hergebruik en uitwisselbaarheid. Als onderzoeksfinanciers het gebruik van standaarden (let wel: niet één standaard) stimuleren of afdwingen, helpt dit wel bij de onderhoudbaarheid van de data en metadata.

Maatschappelijk nut?

Het maatschappelijk nut van zorggerelateerd onderzoek wordt ook als specifiek voor zorggegevens gezien. Behalve uiterste voorzichtigheid op het gebied van privacy van patiënten ligt er ook een duidelijke maatschappelijke verantwoordelijkheid als het gaat om het zo optimaal mogelijk benutten van de mogelijkheden die er zijn, het juist wel gebruiken van persoonsgebonden data zoals zorggegevens. Patiënten willen veelal juist meewerken aan onderzoek en hun data ter beschikking stellen wanneer patiëntbehandeling en wetenschappelijk onderzoek daarin voorop staan.

Samenvattend: geen eenduidig antwoord

Er is geen eenduidig antwoord te geven op de vraag hoe 'speciaal' zorggegevens zijn en het is dus aan te bevelen om ook naar andere disciplines te kijken en om van anderen te leren. Er zijn geen steekhoudende argumenten te vinden waarom zorggegevens een echt bijzondere 'positie' innemen

en niet gedeeld zouden hoeven worden, mits de privacy goed geborgd is. Het lijkt sterk aan te bevelen om de patiënt eigenaar te maken van zijn of haar zorggegevens en dit op een goede manier te faciliteren, naar analogie van het bankpasje of de donorregistratie.

Vraag 2: Wat kan of mag ZonMw eisen van onderzoekers?

Tijdens de bijeenkomst werd duidelijk dat onderzoeksfinanciers best iets mogen eisen van een wetenschapper; het gaat immers om maatschappelijk geld. Onderzoekers mogen gestimuleerd/gedwongen worden om efficiënt met dat geld om te gaan. Daarnaast mag er uit kwaliteits- en uit integriteitsoogpunt reproduceerbaarheid geëist worden. De subsidiegever moet financiële ruimte bieden voor goed databeheer. Er mag hier dus geld voor gebudgetteerd worden.

Wie is de eigenaar van de gegevens?

De vraag komt opnieuw naar voren wie nu eigenlijk de eigenaar is van de data en de verantwoordelijkheid voor deze data draagt. De onderzoeker (en de achterliggende organisatie) heeft een bepaalde verantwoordelijkheid en beheert de data gedurende een bepaalde tijd. Eigenaarschap is volgens de groep van experts een moeilijk en heikel punt. De financier is (of vindt zichzelf) co-eigenaar, om te kunnen voorkomen dat de onderzoekinstelling de data strategisch achter slot en grendel houdt. Onduidelijk is wat dit co-eigenaarschap aan verantwoordelijkheden en bevoegdheden met zich meebrengt. In dit kader is het beter om terug te vallen op de [databankenwet](#). De onderzoeker is in dat kader de producent van de databank. De financier kan naar voren treden als co-producent. Gebaseerd op deze wet en rolverdeling kunnen beter gefundeerde afspraken gemaakt worden die zowel praktisch als juridisch hout snijden.

Het koppelen van nieuwe en bestaande data brengt aanvullende vragen met zich mee. Feit is dat bij hergebruik van data de eerste efficiëntiewinst al binnen is (een van de doelstellingen van instanties als ZonMw). Over de nieuwe dataset die dan ontstaat, moeten eigenaren ook afspraken maken aangaande de toegankelijkheid, herbruikbaarheid, uitwisselbaarheid en controleerbaarheid.

Benadruk de positieve kanten en onderwijs!

Eisen stellen aan onderzoekers is één ding. Het vertellen van succesverhalen, positieve berichtgeving, is iets anders en een niet te overschatten factor bij het beschikbaar krijgen van data. Het gaat om een cultuurverandering die zich als een olievlek zal moeten verspreiden. Het gaat hier in eerste instantie niet om een technische uitdaging, maar om het toepassen van *best practices*. Het delen van ervaringen is hierin heel belangrijk. Het delen van data resulteert in nieuwe contacten en nieuwe inzichten en kan uiteindelijk leiden tot betere netwerken van wetenschappers: *“Alleen ga je sneller, maar samen kom je verder!”*. Databeheer zal ook zijn plek moeten vinden in de wetenschappelijke opleidingen, bijvoorbeeld voor promovendi, zoals al onderschreven werd door de commissie-Schuyt.

Vraag 3: Wat is er nodig aan ondersteunende diensten en instrumenten?

Er is een duidelijk gevoelde behoefte aan onafhankelijke organisaties die kunnen helpen bij alle aspecten van datamanagement (zie ook paragraaf 4.3.4). Uit de discussies komt naar voren dat vrijwel alle instituten momenteel worstelen met datamanagementvraagstukken en zich hierop nog aan het oriënteren zijn. De wetenschapper zal zichzelf ook als dienstverlener moeten gaan zien. Veel datasets kunnen niet zonder de uitleg van de initiële onderzoeker/onderzoeksgroep.

De bestaande instituten zoals DANS, 3TU.Datacentrum, SURF, CentERdata, CIT Groningen en DTL/DISC kunnen op meerdere manieren helpen en hebben ieder hun eigen specialiteiten en interessegebieden. Andere archieven, universiteitsbibliotheken, trusted digital repositories en trusted third parties kunnen dit eveneens. Deze partijen bestaan al, maar er zijn er meer nodig en

hun dienstenpakket moet verbeterd worden en meer bekendheid krijgen. Een overzicht van organisaties die zich bezig houden met zorggegevens zou veel kunnen toevoegen.

Onderzoeker vs. onderzoeksinstituut

Commitment vanuit het management van een onderzoeksinstituut is noodzakelijk, mede in verband met tijd en geld voor datamanagement. Hier wordt niet alleen van de subsidieverstrekker iets verwacht, maar eveneens van het bestuur van de instituten en van organisaties die zich bezig houden met data. Er is mede daarom behoefte aan een organisatiebeleid op het gebied van datamanagement: eenvoudige richtlijnen op hoog niveau waar onderzoekers en onderzoeksgroepen bij kunnen aanhaken voor het plannen en uitvoeren van databeheer. Deels is hierin al voorzien, bijvoorbeeld door de VSNU-richtlijnen, maar op andere onderdelen moet er nog aanvulling komen.

Conclusies uit de expert meeting

De vraag of datamanagement belangrijk is voor goede wetenschapsuitoefening wordt unaniem bevestigend beantwoord. Over de precieze invulling van bijvoorbeeld de eisen die aan onderzoekers gesteld mochten worden, verschilt men wel enigszins van mening, maar men is het erover eens dat er vrij vergaande eisen gesteld mogen worden. Het belangrijkste argument hiervoor is dat wat tot stand komt met publiek geld, ook beschikbaar moet komen voor de gemeenschap, het brede publiek. Uiteraard dient dit te gebeuren op een manier die de privacy en veiligheid van personen en instituten waarborgt.

De conclusie is dus dat onderzoeksfinanciers duidelijke en stringente eisen aan onderzoekers mogen en moeten stellen, maar dat dit ook verplichtingen voor hen met zich meebrengt, vooral in financiële zin. Idealiter kan de subsidieverstrekker ook helpen in immateriële zin, bijvoorbeeld door te wijzen op de faciliteiten en de infrastructuren die al beschikbaar zijn met betrekking tot datamanagement en door goede wegwijzers te ontwikkelen en informatie te verschaffen over zaken zoals de juridische aspecten van het delen van data: wat mag wel en wat niet? Vergeet daarnaast niet om te benadrukken dat het delen van data veel oplevert, voor de maatschappij maar ook voor de individuele onderzoekers. Een positieve benadering werkt wellicht beter dan het al te veel hameren op plichten en plichtsbesef. Er liggen duidelijk nieuwe kansen die op een goede manier voor het voetlicht gebracht moeten worden.

6 Conclusies en adviezen

ZonMw is voornemens in subsidieaanvragen een dataparagraaf toe te voegen en vervolgens, na toekenning van een subsidie, een datamanagementplan te verlangen van de onderzoekers. Met deze middelen stimuleert ze aandacht voor en planning van databeheer, in het bijzonder met het doel om onderzoeksgegevens beter vindbaar, toegankelijk en uitwisselbaar te maken. Om onderzoekers hierbij te helpen, heeft het projectteam twee checklists inclusief toelichtingen ontwikkeld: een korte voor de dataparagraaf en een uitgebreidere voor het datamanagementplan. Daarnaast zijn diverse aspecten van datamanagement inzichtelijk gemaakt om ZonMw in staat te stellen om te zijner tijd de ingevulde documenten te beoordelen. Dit rapport, inclusief de checklists, is het resultaat van bijeenkomsten en interviews met vele deskundigen uit het veld, expertise van de projectpartijen en literatuurstudie.

Een belangrijke conclusie uit de twee georganiseerde bijeenkomsten en de afgenomen interviews is dat over het algemeen de houding ten opzichte van het beschikbaar en uitwisselbaar maken van onderzoeksgegevens positief is. Ook is men positief over de ontwikkelde checklist voor het datamanagementplan. Men ziet het vooral als een mooi instrument dat potentiële subsidieaanvragers bewuster maakt van het proces dat doorlopen moet worden voordat onderzoeksgegevens beschikbaar kunnen worden voor een breder publiek.

Een andere conclusie is dat het hele traject van het beschikbaar en uitwisselbaar maken van onderzoeksgegevens een groeipad is. Hierin worden nu de eerste stappen gezet en daarmee zijn we er nog niet. De huidige checklist voor een datamanagementplan is vrij ambitieus en vernieuwend in opzet. Consequentie is dat de overgang van de huidige situatie (waarbij onderzoekers geen werkwijze ten aanzien van onderwerpen als privacy, datakwaliteit, datadelen hoeven te rapporteren) naar de implementatie van de checklist vrij groot is. Het risico is aanwezig dat niet alle criteria uit de huidige checklist door alle onderzoekers ingevuld kunnen worden en dat ondersteuning noodzakelijk is. Om het opstellen van een datamanagementplan (aan de hand van de checklist) te faciliteren hebben wij de volgende adviezen.

Adviezen voor ZonMw:

1. Algemeen: benadruk vooral de positieve effecten van het delen van data, er is nog veel te winnen.
2. Focus: zowel uit de voorbeelden in hoofdstuk 2 als in dit adviestraject blijkt dat toegang tot data veel wensen en verwachtingen oproept aan “de” data-infrastructuur. Er is onder meer behoefte aan expertise, ICT-infrastructuur en afstemming over definities, regelgeving en verantwoordelijkheden. Dit zijn complexe zaken, die tijd vergen. De werkgroep herkent en onderschrijft deze behoefte, maar adviseert ZonMw om bij de dataparagraaf en het datamanagementplan de blik te richten op praktisch haalbare en zichtbare resultaten. Hiernaast kan ZonMw zich in diverse gremia en in programma’s zoals de Nationale roadmap grootschalige onderzoeksfaciliteiten inzetten voor het gaandeweg versterken van de data-infrastructuur. Diverse hiernavolgende adviezen houden verband met deze complexere aspecten, bijvoorbeeld 6 (duidelijke bevoegdheden), 9 (metadatastandaarden), 10 (meer trusted digital repositories), 11 (minimale bewaartermijn voor data) en de adviezen voor de kennisinstellingen.
3. Tijdens de workshop, de interviews en discussies met de experts en projectleiders werd duidelijk dat nog lang niet alle projectleiders en onderzoekers ingevoerd zijn in datamanagement. Er is een overzicht nodig waarin wordt weergegeven welke partijen expertise hebben op de relevante terreinen, met bijbehorende links naar de websites en/of contactpersonen van deze partijen. De werkgroep heeft een aanzet voor een dergelijk overzicht gemaakt (opgenomen in bijlage 1). Het

advies aan ZonMw is om dit overzichtsdocument om te zetten naar een website zodat gebruikers direct kunnen doorklikken naar experts die ze kunnen raadplegen of inschakelen bij het schrijven van een datamanagementplan.

4. Geef voorbeelden van best practices en probeer een netwerk(-website) van experts en aanbieders op te zetten op het gebied van datamanagement, voortbouwend op ‘emerging’ netwerken zoals DTL/DISC, lopende infrastructures als CTMM/TraIT en BBMRI-NL, en bestaande instituten zoals DANS en CentERdata. Dus niet alleen een overzicht maar ook het faciliteren van contact (bijvoorbeeld via de sociale media) en het ‘uitdragen’ van goede voorbeelden van datamanagement in een wiki-achtige opzet. Idealiter vinden experts het interessant om zichtbaar deel van dit netwerk te zijn, zodat zij zelf hun informatie en expertise/diensten actueel houden en de inspanning voor ZonMw beperkt blijft.
5. Maak van tevoren duidelijke afspraken over het hergebruik en de metadatering van de data die uit het gefinancierde onderzoek voortkomen en/of in het onderzoek gebruikt zijn (afkomstig van derden, al dan niet bewerkt).
6. Volgens artikel 20 is ZonMw co-producent van de data. Dat is aan te bevelen maar het is nog onduidelijk wat dit precies gaat betekenen in de praktijk. Onduidelijk is bijvoorbeeld wat dit co-producentschap aan verantwoordelijkheden en bevoegdheden met zich meebrengt. Dit zal nader bekeken moeten worden door zowel juristen, projectleiders als andere experts. Vermijd al te veel papierwerk, probeer ‘high trust high penalty’ toe te passen.
7. Met het oog op hergebruik van onderzoeksgegevens en de mogelijke replicatie van het onderzoek adviseren wij dat een onderzoeker minimaal de volgende data en documentatie moet archiveren:
 - a. de ruwe data (indien de onderzoeker bestaande data hergebruikt, is er in het onderzoek wellicht geen sprake van “ruwe” data; in dit geval moeten de bestaande data duurzaam gearchiveerd zijn of alsnog gearchiveerd worden);
 - b. de data die ten grondslag liggen aan publicaties;
 - c. documentatie over de gehanteerde onderzoeksmethodologie (zoals codeboeken, metadatering, 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.
8. Data moeten gestructureerd worden opgeslagen maar er is geen algemene ‘gouden standaard’. Dit moet per project bepaald worden. De standaarden die binnen specifieke vakgebieden worden gebruikt moeten leidend zijn. De onderzoeker moet in het eindverslag voor ZonMw opnemen in welk repository de data zijn opgeslagen en via welke catalogus of catalogi de data te vinden zijn.
9. Voor metadata is aan te bevelen om het generieke metadataschema van de Dublin Core Standard op te leggen als minimum op het niveau van de dataset. NARCIS, het portal voor onderzoeksinformatie, kan zulke metadata “oogsten” bij repositories die gebruik maken van het OAI-PMH-protocol, wat de vindbaarheid bevordert. Daarnaast zijn op diverse onderzoeksgebieden rijkere metadataschema’s gangbaar om de data fijnkorreliger te beschrijven en daarmee hergebruik te vereenvoudigen; ZonMw kan het gebruik hiervan stimuleren. Een rolverdeling tussen NARCIS – voor metadata op hoog niveau van alle door ZonMw gefinancierde onderzoeken – en specifiekere catalogi lijkt in het Nederlandse landschap het best te passen.

10. Neem langetermijnarchivering mee als vaste kostenpost in de projectbegroting, bijvoorbeeld als 5% van het totale budget, dit is slechts een eerste schatting. Houd dit bedrag achter en keer het pas uit nadat de data duurzaam zijn opgeslagen bij een daarvoor geschikt trusted data repository, zoals een archief met een DSA. Instellingen die zorgdatarepositories onderhouden en het DSA hiervoor willen verwerven, kunnen deze intentie kenbaar maken bij ZonMw en een peildatum afspreken waarop de instelling het keurmerk dient te hebben.
11. Het verdient aanbeveling om draagvlak te zoeken en te ontwikkelen voor een langere minimale bewaartermijn voor zorggegevens dan vijf jaar. Een praktische aanbeveling is daarnaast om een overzicht van bestaande richtlijnen met uiteenlopende minimumbewaartermijnen toe te voegen aan de online toelichting op de checklist. Wellicht is het praktisch gezien het meest efficiënt en haalbaar om per ZonMw Programma aanbevelingen te doen over/eisen te stellen aan de minimale bewaartermijn van de data. Hier kan dan per project een beslissing over genomen worden.

Adviezen in relatie tot de checklist voor het datamanagementplan:

12. Het is belangrijk om te beseffen dat de huidige checklist een eerste versie is, die in dit bewegende veld de komende jaren aangepast zal moeten worden. Het advies aan ZonMw is om het gebruik van de checklist op korte termijn als een pilot te beschouwen: voer de checklist in de komende programma's in, houd commentaren van onderzoekers en referenten bij en analyseer die in samenhang met de ingevulde datamanagementplannen. Kijk of er aanleiding is om de checklist, de toelichting en/of de vorm te verbeteren, en overweeg of er projecttypes te identificeren zijn waarvoor de checklist te zwaar blijkt, zoals kleine projecten of projecten die relatief weinig data opleveren (het NWO-DANS-[datacontract](#) kan dan een alternatief zijn). Kijk ook of het mogelijk is om de checklist te vereenvoudigen met formuleringen als "ZonMw is akkoord met de keuze 'Ja' bij de volgende X items; andere keuzes worden nader bekeken". Het is aan te raden om ook na die verbeterslag periodiek, bijvoorbeeld om de drie jaar, te bezien of er aanleiding is voor aanpassingen. M.n. geldt dit advies ook voor de kleine checklist voor de dataparagraaf. Dit komt overeen met de klassieke *plan-do-check-act-kwaliteitscyclus*.
13. Wij raden ZonMw aan om na te denken over het gebruik van de checklist: is het een formulier en is met het invullen van het formulier een datamanagementplan geschreven of wil ZonMw een apart datamanagementplan waarin de criteria uit de checklist worden beschreven? De werkgroep acht het eerste voorstel het meest realiseerbaar voor de onderzoekers en vindt dat ZonMw dit kan eisen. Er zullen projecten zijn die (eventueel in aanvulling op de ingevulde checklist) een apart en gedetailleerder datamanagementplan zullen maken. Dit moet te allen tijde aangemoedigd worden, en het is zeker niet uitgesloten dat dit binnen afzienbare tijd de norm wordt.
14. Zoals hiervoor geconcludeerd staan de geraadpleegde onderzoekers positief tegenover datamanagement en de checklists. Er is echter wel onduidelijkheid over de status van een datamanagementplan. Zo vraagt men zich af of een dergelijk plan getoetst wordt door ZonMw en zo ja, wie dit gaat doen. Ook vraagt men zich af of er (financiële) consequenties zijn als partijen geen datamanagementplan schrijven of hun data uiteindelijk niet beschikbaar stellen. Men is van mening dat ZonMw haar eigen eisen goed moet toetsen en consequenties dient te formuleren voor onderzoekers wier databeheer in gebreke blijft. De werkgroep onderschrijft dit.

Adviezen voor kennisinstellingen:

15. Een groot deel van de checklist kan op instituutsniveau worden ingevuld. Universiteiten, onderzoeksinstituten en andere organisaties die subsidie aanvragen bij ZonMw wordt geadviseerd zoveel mogelijk criteria van de checklist op instituutsniveau in te vullen en aan hun

medewerkers beschikbaar te stellen. Dit verlaagt de inspanning en benodigde kennis om een datamanagementplan te schrijven. Ook is het advies aan deze organisaties om een centraal “loket” in te richten waar onderzoekers vragen kunnen stellen bij het. Dit wordt urgenter naarmate meer onderzoeksfinanciers voorwaarden gaan stellen aan datamanagement. Een dergelijk frontoffice op locatie past goed in de federatieve data-infrastructuur zoals beschreven in paragraaf 2.3.

16. Commitment van het bestuur van kennisinstellingen, universiteiten en andere organisaties voor het opzetten en onderhouden van datamanagement(faciliteiten) binnen de instelling of organisatie is van groot belang. Bevorder dit commitment en benadruk het algemene belang van het delen van data. Er is veel te winnen met het beter beschikbaar, vindbaar en uitwisselbaar maken van data en het is belangrijk om dit bij bestuurders onder de aandacht te brengen.

Literatuur

DANS: Data Management Plan voor wetenschappelijk onderzoek: Wat komt daarbij kijken? Den Haag, 2012. Beschikbaar via <http://www.dans.knaw.nl/content/data-management-plan>.

Grootveld, M., Polhout, M., Tjalsma, H., Visser, S., Bruin, W. de, Nugteren, R., Weda, M., Verheij, R., Dijk, L. van: Inventory patient registries in the Netherlands. Utrecht, 2012. Beschikbaar via <http://www.nivel.nl/sites/default/files/bestanden/Rapport-Inventory-Patient-Registries-Netherlands.pdf>.

Raad voor Gezondheidsonderzoek (RGO): Van gegevens verzekerd. Kennis over de volksgezondheid in Nederland nu en in de toekomst. Den Haag, 2008. Beschikbaar via <http://www.gezondheidsraad.nl/nl/adviezen/innovatie-en-kennisinfrastructuur/van-gegevens-verzekerd-kennis-over-de-volksgezondheid-ned>.

VSNU: De Nederlandse Gedragscode Wetenschapsbeoefening, den Haag, 2004, herziening 2012. Beschikbaar via [http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Code_wetenschapsbeoefening_2004_\(2012\).pdf](http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Code_wetenschapsbeoefening_2004_(2012).pdf).

Vollaard, H., van de Bovenkamp, H.M., Vrangbæk, K.: The emerging EU quality of care policy: from sharing information to enforcement. In Health Policy, 2013 Aug; 111(3): 226033. Doi:10.1016/j.healthpol.2013.05.004. Beschikbaar via <http://www.sciencedirect.com/science/article/pii/S0168851013001292>.

Bijlage 1: Aanzet kennisnetwerk datamanagement

Hieronder is een eerste aanzet gemaakt voor het in kaart brengen van organisaties die diensten en/of kennis hebben in relatie tot datamanagement. Deze aanzet zal worden voorgelegd aan alle partijen die betrokken zijn geweest bij de totstandkoming van dit rapport. Het eindresultaat zal apart worden opgeleverd en beschikbaar worden gesteld via www.ZonMw.nl.

KENNISNETWERK DATAMANAGEMENT

Privacy

Medisch-ethische toetsingscommissie bij de eigen kennisinstelling

ZorgTTP <https://www.zorgttp.nl>

Custodix <https://www.custodix.com>

Mondriaan <http://www.projectmondriaan.nl/>

NIVEL <http://www.nivel.nl/>

Medlawconsult <http://www.medlaw.nl/>

Vektis <http://www.vektis.nl/>

Parelsnoer Instituut <http://catalogus.parelsnoer.org/catalogus/>

Koppeling

Centraal Bureau voor de Statistiek, CBS <http://www.cbs.nl/>

ZorgTTP <https://www.zorgttp.nl>

Mondriaan <http://www.projectmondriaan.nl/>

NIVEL <http://www.nivel.nl/>

Vektis <http://www.vektis.nl/>

Parelsnoer Instituut <http://catalogus.parelsnoer.org/catalogus/>

Dataopslag tijdens het project

ICT-afdeling van de eigen kennisinstelling

Dutch Dataverse Network <https://www.dataverse.nl/dvn/>

SURFsara (grootschalig) <https://www.surfsara.nl/>

Target (grootschalig) <http://www.rug.nl/science-and-society/target/>

Parelsnoer Instituut <http://catalogus.parelsnoer.org/catalogus/>

Langetermijnarchivering³⁹

DANS⁴⁰ <http://dans.knaw.nl/>

3TU.Datacentrum (voor o.a. NetCDF) <http://datacentrum.3tu.nl/>

³⁹ N.B. Alle hier genoemde repositories hebben een duidelijke ambitie om onderzoeksgegevens voor de lange termijn te archiveren en er toegang toe te verschaffen. Ze zijn nog niet alle duurzaam in de zin dat zij bijvoorbeeld het Data Seal of Approval hebben verworven.

⁴⁰ voor o.a. tekst, databases, spreadsheets, audio, audiovisuele data, XML.

CentERdata (m.n. surveydata)	http://www.centerdata.nl/
Nederlands Instituut voor Beeld en Geluid	http://www.beeldengeluid.nl
SURFsara (grootschalig)	https://www.surfsara.nl/
Target (grootschalig)	http://www.rug.nl/science-and-society/target/

Metadata

CentERdata (DDI)	http://www.centerdata.nl/
DANS ⁴¹	http://dans.knaw.nl/
DDI, Data Documentation Initiative	http://www.ddialliance.org/what
Dublin Core	http://dublincore.org/specifications/
Metadatastandaarden voor onderzoek	http://www.dcc.ac.uk/resources/metadata-standards

Data beschikbaar stellen (zie ook de organisaties bij “Langetermijnarchivering”)

Zorggegevens.nl	http://www.zorggegevens.nl/
BioSharing (internationaal)	http://www.biosharing.org/
Durrer Center for cardiogenetic research	http://www.durrercenter.nl/
Parelsnoer Instituut	http://catalogus.parelsnoer.org/catalogus/
NARCIS.nl, tabblad “Datasets”	http://www.narcis.nl/
NIVEL	http://www.nivel.nl/
Centraal Bureau voor de Statistiek, CBS	http://www.cbs.nl/

⁴¹ Dublin Core, DDI. Voor datamanagers en beheerders van repositories: zichtbaarheid van data in Narcis.nl