

Richtlijn archivering wetenschappelijk onderzoek voor Nederlandse faculteiten Maatschappij- en Gedragwetenschappen Versie 2.1, april 2018

1. Preambule

In de discussies over wetenschappelijke integriteit staan transparantie en reproduceerbaarheid van onderzoek centraal. Beiden veronderstellen toegang tot kwalitatieve of kwantitatieve onderzoeksdata, een gedetailleerde beschrijving van onderzoeksmaterialen en -aanpak, en een weergave van het proces van dataverwerking en van publicatie. In dit document staat een richtlijn voor de archivering van wetenschappelijke publicaties alsook informatie die nodig is om de resultaten zoals vermeld in de publicatie te kunnen reproduceren. Het document heeft dus betrekking op de archivering van gepubliceerd wetenschappelijk onderzoek; het is geen richtlijn voor datamanagement, bewerkers overeenkomsten en privacyaspecten. Het document is te zien als een initiatief dat onderdeel uitmaakt van een bredere onderneming om wetenschappelijke integriteit te bevorderen onder kwantitatief en kwalitatief georiënteerde onderzoekers bij Nederlandse faculteiten Maatschappij- en Gedragwetenschappen. Het is niet bedoeld als een strak keurslijf, maar het geeft wel een duidelijke richting aan, die naar lokale facultaire omstandigheid verder kan worden ingevuld. Hierbij geldt het adagium "Pas toe, of leg uit!".

1.1 Doel van deze richtlijn

Deze *Richtlijn archivering wetenschappelijk onderzoek* formuleert randvoorwaarden ten aanzien van de archivering van data, materialen en informatie die ten grondslag liggen aan publicaties. Dat wil zeggen beschrijving van deze data, materialen en informatie om onderzoeksresultaten te reproduceren, en de opslag hiervan. De richtlijn heeft betrekking op de data, materialen en informatie van publicaties die vanaf 1 juni 2018 in definitieve vorm verschijnen.ⁱ De richtlijn is gebaseerd op verantwoording achteraf, d.w.z. nadat een publicatie is verschenen. Uitgangspunt van deze richtlijn is dat elke onderzoeker verantwoordelijk is voor een verantwoorde en transparante archivering van data, materialen en informatie en daarop gebaseerde publicaties. Mocht deze richtlijn in voorkomende gevallen geen sluitend voorschrift bevatten, dan wordt de onderzoeker geacht om in de geest van de richtlijn te handelen.

Faculteiten worden geacht de landelijke richtlijn te hanteren. Om onderzoekers de gelegenheid te geven vertrouwd te raken met de voorwaarden in de richtlijn, kunnen faculteiten de richtlijn gefaseerd invoeren.

1.2 Bestaande kaders waarbinnen deze richtlijn functioneert

- Wet bescherming persoonsgegevens en de vanaf 25 mei 2018 geldende Algemene Verordening Gegevensbescherming (AVG)
- Auteurswet, Rijksoctrooiwet, Databankenwet
- Medisch-ethische onderzoeksprotocol eisen voortvloeiend uit de wet Medisch-wetenschappelijk onderzoek met mensen (WMO) en die wordt getoetst door een METC
- CAO Nederlandse universiteiten
- Gedragscode VSNU
- Universitaire beleidskaders van de eigen universiteit

1.3 Op wie is deze richtlijn van toepassing?

Deze richtlijn is van toepassing op alle medewerkers van de faculteit die onderzoek verrichten in een tijdelijk of vast dienstverband, op alle promovendi die onder begeleiding van hoogleraren promotieonderzoek doen, en op alle researchmaster studenten. De richtlijn is niet van toepassing op bachelor- en eenjarige master studenten, tenzij het onderzoek van deze studenten leidt tot een wetenschappelijke publicatie. Het onderzoek van bachelor- en eenjarige masterstudenten valt voornamelijk onder de formele verantwoordelijkheid van de begeleiders.

Alle onderzoekers van de faculteit onderschrijven *The Netherlands Code of Conduct for Scientific Practice*.ⁱⁱ Deze richtlijn geeft een concrete uitwerking van het aspect "controleerbaarheid", zoals opgenomen in de *code of conduct* van de VSNU.

2. Richtlijn inzake publicatie package

Deze richtlijn heeft betrekking op alle onderzoekspublicaties die in het wetenschappelijk jaarverslag van de faculteit worden vermeld. Voor de transparantie en reproduceerbaarheid van kwalitatief en kwantitatief empirisch onderzoek is het van belang dat alle informatie wordt gearchiveerd (in het Engels), die nodig is om de resultaten te kunnen reproduceren. Deze informatie wordt opgeslagen in een zogenaamde 'publicatie package'.

2.1 Wat moet er in een publicatie package worden opgeslagen

Voor elke gepubliceerde empirische studie (d.w.z. artikel, bundel, hoofdstuk in boek, hoofdstuk in proefschrift, research master thesis, opvraagbaar intern rapport, etc.) worden de volgende materialen opgeslagen:

1. Het gepubliceerde (of geaccepteerde) manuscript of publicatie.
2. Een korte beschrijving van probleemstelling, onderzoeksdesign, conceptueel kader, dataverzameling (steekproeftrekking, selectie en representativiteit van informanten) en toegepaste methoden. Als regel volstaat hier een elektronische versie van het gepubliceerde manuscript.
3. De instructies, procedures, het ontwerp van het experiment en stimulusmaterialen (topiclijst, interviewguide, vragenlijsten) die redelijkerwijs nodig zijn voor een reproductie van het onderzoek. De materialen dienen beschikbaar te zijn in de taal waarin het onderzoek is uitgevoerd. Het publicatie package is Engelstalig.
4. Bij gebruik van primaire gegevens, de geanonimiseerde ruwe datafiles (die de meest directe registratie vormen van het gedrag dan wel reactie van proefpersoon/respondent, bijvoorbeeld: een ongefilterde export file van een online survey of ruwe tijdreeksen bij een EEG meting, e-dat bestanden bij een E-Prime gedragsexperiment, opnames of transcripten van gesprekken, beschrijvingen van observaties, archief- en ander bronnen- en/of mediamateriaal). Indien de ruwe datafiles in een extern archief (zoals opslagfaciliteiten bij DANS) zijn opgeslagen en toegankelijk zijn, dan volstaat een bronverwijzing naar de files in dit archief. Deze extern gearchiveerde ruwe data kunnen primaire of secundaire gegevens zijn. Ruwe data worden niet gewijzigd nadat deze voor het eerst digitaal beschikbaar zijn gekomen.
5. Computercode (bijvoorbeeld Atlas.ti, SPSS-syntaxfile, MATLAB analysescripts, code in R) die de stappen beschrijft die zijn genomen om de ruwe data te verwerken tot analyse data. Dit met inbegrip van korte verklaringen van de stappen in het Engels, bijvoorbeeld een korte beschrijving van de stappen die zijn doorlopen bij de kwalitatieve analyse van de primaire onderzoeksgegevens, d.w.z. thema's, domeinen, taxonomieën, componenten.
6. (Al dan niet verwerkte) datafiles die uiteindelijk voor de totstandkoming van het artikel zijn geanalyseerd (dus bijvoorbeeld een SPSS-datafile na transformatie van variabelen, na toegepaste selecties, enzovoort). Dit laatste is niet nodig indien de ruwe datafile direct is geanalyseerd.
7. Computercode (bijvoorbeeld een syntaxfiles van SPSS, Atlas.ti, Matlab, R; syntaxen van maatwerk software) die de stappen beschrijft om van de analysedata naar de resultaten in het manuscript te komen. Dit met inbegrip van korte verklaringen van de stappen in het Engels.
8. Een readme file (metadata) die beschrijft waar welke documenten en/of files te vinden zijn en hoe de documenten en/of files geïnterpreteerd dienen te worden. De readme-file bevat tevens de volgende informatie:
 - a. Naam van de persoon die de documenten en/of files heeft opgeslagen.
 - b. Rolverdeling van auteurs, waarbij i.i.g. is aangegeven wie de data heeft geanalyseerd.
 - c. Datum waarop het manuscript is geaccepteerd, inclusief referentie.
 - d. Datum/periode waarop de data zijn verzameld.
 - e. Namen van de personen die de data hebben verzameld.
 - f. (Indien van toepassing) Adressen van de locatie in het veld waar de data zijn verzameld en de (eventuele) contactpersonen.
 - g. Of er sprake is geweest van Ethische Toetsing voorafgaande aan het onderzoek, en eventuele uitspraken daarover van de Ethische Toetsingscommissie.
9. De readme-file moet voldoende helder zijn. Een relevante collega onderzoeker moet op basis van de onderdelen van het publicatie package de resultaten zoals vermeld in de publicatie kunnen reproduceren.
10. Indien aanwezig, documenten die betrekking hebben op de ethische goedkeuring.

2.2 Wanneer moet een publicatie package worden opgeslagen?

Een publicatie package wordt opgeslagen uiterlijk 1 maand na definitieve publicatie van het manuscript. Voor iedere research master student geldt dat, bij indiening van de master thesis, er een publicatie package is opgeslagen. Voor iedere promovendus geldt dat, bij indiening van het manuscript bij de leescommissie, er voor ieder empirisch hoofdstuk uit het proefschrift een publicatie package is opgeslagen (of één publicatie package in geval van een monografie).

Nadat het publicatie package is opgeslagen, wordt het gefixeerd, en kan het niet meer worden gewijzigd (*read only*).

2.3 Wie draagt de verantwoordelijkheid voor de opslag van een publicatie package?

Als de eerste auteur werkzaam is bij één van de faculteiten maatschappij- of gedragswetenschappen, dan is hij of zij altijd verantwoordelijk voor de archivering van het publicatie package, d.w.z. de opslag van ruwe en bewerkte data, syntax en materialen, en additionele informatie over het publicatieproces zoals hierboven geformuleerd. Als de tweede of latere auteur werkzaam is bij één van de faculteiten maatschappij- of gedragswetenschappen, dan is hij/zij op de hoogte dat de data zorgvuldig zijn opgeslagen en van de wijze waarop de opslag is geregeld. Dit geldt met name wanneer de eerste auteur niet werkzaam is bij één van de faculteiten maatschappij- of gedragswetenschappen.

Indien de eerste auteur werkzaam is bij één van de faculteiten maatschappij- of gedragswetenschappen, dan mag de tweede of latere auteur er van uit gaan dat de eerste auteur de regelgeving van de eigen universiteit volgt en hoeft de tweede of latere auteur geen publicatie package te vervaardigen.

Bij promovendi en research masterstudenten is de promotor respectievelijk dagelijks begeleider verantwoordelijk voor de opslag van een publicatie package. Hij/zij kan de uitvoering delegeren, maar dit laat onverlet dat hij/zij eindverantwoordelijk is.

2.4 Wie heeft toegang tot het publicatie package?

De eerste auteur heeft leesrechten. Indien een faculteit werkt met een *co-pilot* systeem, waarbij de *co-pilot* de analyses controleert, dan heeft ook deze *co-pilot* leesrechten. Met het oog op eventuele audits van publicatie packages kan het bestuur van de faculteit een functionaris aanwijzen die namens haar leesrechten heeft. Te denken valt aan de coördinator van een onderzoeksprogramma of een lid van een commissie wetenschappelijke integriteit.

2.5 Bewaartermijn en centrale opslag van publicatie packages

Het publicatie package wordt gedurende een periode van minimaal 10 jaar na verschijnen van de publicatie centraal opgeslagen op een beveiligde server voorziening van de faculteit. De leesrechten zijn bepaald in 2.4.

3. Richtlijn inzake opslag ruwe data

In het kader van transparantie en reproduceerbaarheid van onderzoek is het uiteraard van belang dat ruwe data worden opgeslagen. Ruwe data zijn de onbewerkte gegevens die in het kader van een onderzoek zijn verzameld. Daaronder vallen bijvoorbeeld

- Registraties afkomstig uit experimenteel onderzoek;
- Survey data afkomstig van ingevulde vragenlijsten, in het kader van (al dan niet longitudinaal) onderzoek, zelf verzameld of door een extern veldwerk organisaties;
- Transcripties van videomaterialen verzameld in kwalitatief onderzoek (open interviews, observaties);
- Notities die in het kader van kwalitatief onderzoek of onderzoek met bronnenmateriaal zijn verkregen.

De ruwe data worden bij voorkeur digitaal opgeslagen. Opgeslagen ruwe data zijn altijd geanonimiseerd en daarmee niet direct tot personen of groepen personen herleidbaar.^{iii,iv} Direct tot de persoon herleidbare data zijn persoonsdata. Men kan hier denken aan NAW gegevens, maar ook aan foto's en aan videomateriaal. De ruwe data en de persoonsdata vormen samen de onderzoeksdata.

3.1 Minimum bewaartermijn onderzoeksdata

Ten aanzien van de termijn van het bewaren van de ruwe data volgt de faculteit de VSNU-richtlijnen en de APA guidelines: 'minimum of 10 years after the publication of this research'. De opgeslagen ruwe data (die dus niet tot de persoon herleidbaar zijn) hebben geen maximale bewaartermijn en mogen langer bewaard worden dan nodig voor het doel waarvoor ze zijn verzameld of verwerkt (conform artikel 10 lid 2 WBP).^v

Een uitzondering is het bewaren van persoonsgegevens voor een medisch dossier en dus ook data afkomstig uit WMO-plichtig onderzoek. Deze data dienen minimaal 15 jaar bewaard te worden (conform artikel 454 lid 3 Wet op de geneeskundige behandelingsovereenkomst (WGBO)).

3.2 Maximum bewaartermijn onderzoeksdata

De tot de persoon herleidbare gegevens mogen in principe *niet* koppelbaar zijn aan onderzoeksdata als dit niet meer nodig is voor het doel van het onderzoek. Deze tot de persoon herleidbare gegevens moeten vernietigd worden als het doel waarvoor ze zijn verzameld is bereikt (conform artikel 10 lid 1 Wet Bescherming Persoonsgegevens). Voor bepaalde onderzoeken is het nodig de tot de persoon herleidbare gegevens te bewaren, bijvoorbeeld voor het doen van vervolgonderzoek of bij longitudinale studies. In dat geval mogen de tot de persoon herleidbare gegevens bewaard blijven (conform artikel 10 lid 2 WBP).

De verantwoordelijkheid voor het toezicht op de vernietiging van de onderzoeksdata op de geëigende datum ligt bij het hoofd van de betreffende afdeling. De formele eindverantwoordelijkheid ligt bij de decaan.

Een complicerende factor is de wens om tot de persoon herleidbare gegevens te bewaren voor onderzoek naar de integriteit van het onderzoek, bijvoorbeeld om te kunnen controleren of de participanten daadwerkelijk hebben deelgenomen aan het onderzoek. Indien dergelijk integriteitsonderzoek wordt beschouwd als een onderdeel van het onderzoek waarvan de integriteit wordt onderzocht, lijkt de wet de mogelijkheid te bieden voor dit doel tot de persoon herleidbare gegevens te bewaren. Bij publicatie van onderzoek worden deze persoonsdata *niet* bewaard in het publicatie package maar blijven zij gescheiden opgeslagen.

3.3 Hoe is de opslag en archivering van onderzoeksdata geregeld?

De geanonimiseerde ruwe data worden opgeslagen op een eerste facultaire server die voldoet aan de eisen die aan de opslag van data gesteld kunnen worden in termen van veiligheid, robuustheid en voorziening voor automatische back-up. Het is aan te bevelen de ruwe data *read only* op te slaan en wel voordat de gegevens beschikbaar komen voor bewerking. Bij deze opslag worden de ruwe data gefixeerd, d.w.z. onderzoekers kunnen deze niet meer modificeren.

De tot persoon herleidbare gegevens worden gescheiden van de ruwe data opgeslagen op een tweede facultaire server, die fysiek gescheiden is van de eerste facultaire server. Indien een sleutel nodig is om de geanonimiseerde ruwe data te koppelen aan de tot de persoon herleidbare gegevens, dan wordt deze sleutel opgeslagen op de tweede facultaire server.

Externe opslag van ruwe data zoals in nationale of internationale data-archieven, zoals DANS – waarmee de data publiekelijk beschikbaar, vindbaar, en citeerbaar wordt - is aan te bevelen en in een aantal gevallen verplicht, bijvoorbeeld wanneer NWO dit contractueel eist, maar ontslaat de onderzoeker niet van interne opslag op de eerste facultaire server.

Individuele opslag op eigen HD, USB en opslag 'in the cloud' zoals Dropbox voldoet niet. Data die verzameld worden in het kader van promotieonderzoek^{vi} of postdoc-onderzoek worden zo gearhiveerd dat continuïteit gegarandeerd is wanneer een promovendus of postdoc de faculteit verlaat.

Deze eisen van opslag gelden niet voor (die delen van de ruwe) data die onder beheer van externe organisaties staan. De onderzoeker die gebruik maakt van data van externe organisaties dient zich ervan te vergewissen dat deze organisatie haar data opslaat volgens een protocol dat overeenkomt met de eisen van deze facultaire richtlijn.

3.4 Wie draagt de verantwoordelijkheid voor de opslag van onderzoeksdata?

Eerstverantwoordelijk voor het naleven van deze richtlijn inzake opslag ruwe data is de onderzoeker. Bij promovendi en research master studenten zijn respectievelijk promotor en dagelijks begeleider verantwoordelijk. Eindverantwoordelijke is degene die het onderzoeksprogramma coördineert waaronder de publicaties vallen (leerstoelhouder, programmaleider, voorzitter departement, dit varieert per faculteit). Toezicht op het naleven van de richtlijn is onderdeel van de functionerings- en beoordelingsgesprekken. De decaan draagt formele eindverantwoordelijkheid.

4. Nader te bepalen facultair beleid

Faculteiten kunnen kiezen om de hierboven genoemde randvoorwaarden inzake publicatie package en inzake opslag ruwe data aan te vullen m.b.t. de volgende zaken:

- (i) Een faculteit kan al dan niet beslissen dat de richtlijn ook van toepassing is op data die verzameld zijn in het kader van eenjarige master-onderzoek en bachelor-onderzoek. Men kan daar de begeleider verantwoordelijk voor maken.
- (ii) Een faculteit kan beslissen deze richtlijn uit te breiden naar de opslag van alle data, ook van niet gepubliceerd onderzoek. Dit valt onder het opstellen van data management plannen.
- (iii) Een faculteit kan regels opstellen m.b.t. het eigendom van de data, bijvoorbeeld dat het eigendom van data niet verandert door opslag van de data in een publicatie package.
- (iv) Een faculteit kan beslissen al dan niet regelmatig steekproefsgewijze controles uit te voeren om de aanwezigheid en kwaliteit van publicatie packages te controleren.
- (v) Een faculteit kan andere termijnen hanteren, en bijvoorbeeld aangeven dat een publicatie package na acceptatie (in plaats van publicatie) van een manuscript gearhiveerd moeten worden.
- (vi) Een faculteit kan de landelijke richtlijn gefaseerd invoeren.
- (vii) Een faculteit kan beslissen dat in het manuscript moet worden vermeld waar de data zijn opgeslagen (een data statement) en wat de rolverdeling van de auteurs is geweest.

ⁱ Dit document is tot stand gekomen door werkzaamheden van een hiertoe door het DSW ingestelde commissie bestaande uit Marc van Veldhoven (UvT, later vervangen door Jelte Wicherts), Rob Eisinga (RU), Rosanne Janssen (UM) en Peter van der Heijden (UU).

ⁱⁱhttp://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/The_Netherlands_Code_of_Conduct_for_Scientific_Practice_2012.pdf

ⁱⁱⁱ Persoonsgegevens zijn 'alle informatie betreffende een geïdentificeerde of identificeerbare natuurlijke persoon' (Artikel 1 sub a *Wet Bescherming Persoonsgegevens*).

^{iv} Artikel 10 *Wet Bescherming Persoonsgegevens*

1. Persoonsgegevens worden niet langer bewaard in een vorm die het mogelijk maakt de betrokkene te identificeren, dan noodzakelijk is voor de verwerking van de doeleinden waarvoor zij worden verzameld of vervolgens worden verwerkt.

2. Persoonsgegevens mogen langer worden bewaard dan bepaald in het eerste lid voor zover ze voor historische, statistische of wetenschappelijke doeleinden worden bewaard, en de verantwoordelijke de nodige voorzieningen heeft getroffen ten einde te verzekeren dat de desbetreffende gegevens uitsluitend voor deze specifieke doeleinden worden gebruikt.

Artikel 11

1. Persoonsgegevens worden slechts verwerkt voor zover zij, gelet op de doeleinden waarvoor zij worden verzameld of vervolgens worden verwerkt, toereikend, ter zake dienend en niet bovenmatig zijn.

2. De verantwoordelijke treft de nodige maatregelen opdat persoonsgegevens, gelet op de doeleinden waarvoor zij worden verzameld of vervolgens worden verwerkt, juist en nauwkeurig zijn.

^v Wel is er sprake van een *Vrijstellingsbesluit Wet Bescherming persoonsgegevens*, Artikel 30. Wetenschappelijk onderzoek en statistiek

1. Artikel 27 van de wet is niet van toepassing op verwerkingen van organisaties voor wetenschappelijk onderzoek of statistiek die uitsluitend ten dienste staan van door hen te verrichten of verricht onderzoek, voor zover deze verwerkingen voldoen aan de in dit artikel vermelde eisen.

2. De verwerking geschiedt slechts voor het verzamelen, verwerken en controleren van de gegevens ten behoeve van een bepaald onderzoek of een bepaalde statistiek.

3. Geen andere persoonsgegevens worden verwerkt dan:

a. naam, voornamen, voorletters, titulatuur, geslacht, geboortedatum, adres, postcode, woonplaats, telefoonnummer en soortgelijke voor communicatie benodigde gegevens, alsmede bank- en girorekeningnummer van de betrokkene;

b. een informatieloos administratienummer;

c. andere dan de onder a en b bedoelde gegevens, die ten behoeve van een bepaald onderzoek of een bepaalde statistiek zijn verkregen.

4. De persoonsgegevens worden slechts verstrekt aan:

a. degenen, waaronder begrepen derden, die zijn belast met of leiding geven aan de in het tweede lid bedoelde activiteiten of die daarbij noodzakelijk zijn betrokken;

b. anderen, in de gevallen bedoeld in artikel 8, onder a, c en d, en artikel 9, derde lid, van de wet.

5. De in het derde lid onder a bedoelde persoonsgegevens, met uitzondering van geslacht, woonplaats en geboortjaar, worden verwijderd uiterlijk zes maanden nadat de in dat lid onder c bedoelde gegevens omtrent de betrokkene zijn verkregen.

^{vi} De onderdelen van een dissertatie (of de dissertatie als geheel) gelden ten formele als een publicatie ook als deze niet één op één in een tijdschrift zijn gepubliceerd.