



Praktische handleiding voor het FAIR maken van data

DCC-PO - mei 2024

Inhoud

| | |
|--|-----------|
| 1. INLEIDING | 3 |
| 1.1 Opdracht | |
| 1.2 Opbouw van het document | |
| 1.3 Doelgroep | |
| 2. BELANG VAN FAIR DATA BIJ ONDERZOEKSGEGEVENS | 4 |
| 2.1 Wat zijn onderzoeksdata? | |
| 2.2 Het belang van toepassen van FAIR bij onderzoeksdata | |
| 2.3 Voordelen van FAIR data | |
| 3. UITLEG FAIR PRINCIPES | 6 |
| 3.1 Findable | |
| 3.2 Accessible | |
| 3.3 Interoperable (uitwisselbaar) | |
| 3.4 Reusable (herbruikbaar) | |
| 4. FAIR PRINCIPES PER ONDERZOEKSFASE | 9 |
| 4.1 Voorafgaand aan het onderzoek | |
| 4.2 Tijdens het onderzoek | |
| 4.3 Na afronding van het onderzoek | |
| 5. BEGRIPPENLIJST | 15 |
| 6. ALGEMENE BRONNEN EN NOTEN | 16 |
| 7. BIJLAGEN | 17 |

1. Inleiding

1.1. Opdracht

Deliverable-1 van het werkpakket FAIR (onderdeel van het DCC-PO) omvatte de opdracht een praktische handleiding te schrijven voor het FAIR maken van data met speciale aandacht voor Interoperability en Reusability. Daarbij was het uiteraard zaak om die handleiding specifiek te richten op Praktijkgericht Onderzoek binnen hogescholen.

1.2 Opbouw van het document

In deze handleiding wordt beschreven waarom het belangrijk is dat opgeslagen en gearcheerde onderzoeksdata FAIR zijn. Daarnaast wordt per onderzoeksfase aangegeven wanneer de verschillende onderdelen van FAIR onder de aandacht gebracht moeten worden. FAIR is een acroniem van de Engelse termen; Findable (vindbaar), Accessible (toegankelijk), Interoperable (uitwisselbaar) en Reusable (herbruikbaar). Vanuit de behoefte om een aantal standaarden te ontwikkelen waarmee onderzoeksdata goed vindbaar, toegankelijk, en herbruikbaar te maken zijn, werden in 2016 de FAIR principes voor het eerst [beschreven](#).

1.3 Doelgroep

Deze handleiding richt zich op onderzoeksondersteuners (datastewards, research supporters) en de onderzoekers zelf. Om alle informatie voor deze doelgroepen overzichtelijk te presenteren, kan er bijvoorbeeld een LibGuide of andere webtool worden gemaakt.

2. Belang van FAIR data bij onderzoeksgegevens

In onderstaande paragrafen wordt het belang van FAIR data bij onderzoeksgegevens aangeduid.

2.1 Wat zijn onderzoeksdata?

Belangrijk is eerst vast te stellen wat in deze handleiding verstaan wordt onder de term 'onderzoeksdata'. Hierbij maken we gebruik van de onderstaande beschrijving die is geformuleerd aan de hand van omschrijvingen uit publicaties van diverse instellingen waaronder SURF, DCC-PO, NWO en RDNL:

"Onderzoeksdata kunnen worden omschreven als gegevens die in het kader van (wetenschappelijk en praktijkgericht) onderzoek worden verzameld voor het beantwoorden van een onderzoeksvraag of het testen van een hypothese. Een verzameling van feitelijke informatie of gegevens die worden gemeten, gecreëerd, gegenereerd, verzameld of gebruikt tijdens het uitvoeren van onderzoek. Deze gegevens omvatten diverse vormen van kwantitatieve en kwalitatieve informatie, zoals metingen, observaties, enquêteresultaten, interviews, experimentele resultaten, documenten, en andere relevante bronnen".

2.2 Het belang van toepassen van FAIR bij onderzoeksdata

Stakeholders, meer specifiek subsidieverstrekkingen en partners, stellen de laatste jaren steeds hogere eisen aan het zorgvuldig verzamelen, opslaan en toegankelijk maken van onderzoeksdata. FAIR helpt daarbij. Het is inmiddels een geaccepteerd begrip en opgenomen in vrijwel alle standaard data-managementplannen of -formats. Daarnaast heeft de Nederlandse overheid bepaald dat de resultaten van publiek gefinancierd onderzoek (zoveel mogelijk) openbaar gemaakt moeten worden (Open Science). Verder is de 'Wet implementatie Open datarichtlijn' in ontwikkeling en ook die heeft betrekking op onderzoeksdata van hogescholen. In de [Gedragscode Wetenschappelijke Integriteit](#) voor onderzoekers en instellingen (2018) zijn normen voor goede onderzoekspraktijken vastgesteld en ook daar worden de FAIR beginselen genoemd.

Bij het verzamelen en publiceren van data moet voldaan worden aan wetgeving (AVG/GDPR), veilige opslag en archivering door niet meer persoonsgegevens te verzamelen dan nodig zijn voor het beschrijven van de onderzoekspopulatie en/of het beantwoorden van de onderzoeksvraag. Ook mogen direct-identificerende persoonsgegevens nooit opgeslagen worden bij de onderzoeksdata, maar separaat ervan. Wanneer gevoelige persoonsgegevens worden verzameld of verwerkt dient er een [DPIA](#) (Data Protection Impact Assessment) te worden uitgevoerd. Daarmee breng je vooraf de privacy risico's bij het verzamelen en/of verwerken van persoonsgegevens in kaart om passende maatregelen te nemen om die risico's te verkleinen.

Het belang van FAIR data is bij het onderzoek naar Covid 19 gebleken. Ontwikkelingen en onderzoeksresultaten konden snel worden gepresenteerd doordat data in dit geval open gedeeld werden (zie bijvoorbeeld onderzoek bij [ZonMw](#)).

2.3 Voordelen van FAIR data

FAIR bevordert zorgvuldige publicatie, vindbaarheid en toegankelijkheid van onderzoeksdata. Ook datasets die niet toegankelijk zijn (met bijvoorbeeld privacygevoelige data) kunnen door het toevoegen van goede metadata gevonden worden. Wel zijn voorwaarden nodig die bepalen of toegang tot dergelijke onderzoeksdata mogelijk is. Een goede uitwisselbaarheid van data kan sneller tot nieuw onderzoek leiden. Andere onderzoekers moeten onderzoeksdata makkelijk kunnen controleren en die eventueel kunnen hergebruiken voor vervolgonderzoek. Dat biedt onderzoekers tal van nieuwe mogelijkheden en versnelt nieuw onderzoek. Persoonlijke voordelen voor onderzoekers met betrekking tot FAIR data zijn bijvoorbeeld: efficiënter werken met data, minder risico op dataverlies, 'Open data' kan zelfs leiden tot meer citaties van publicaties, zie bijvoorbeeld dit [artikel](#).

Vanuit het werkveld zal het vertrouwen in hogescholen als onderzoekspartner toenemen wanneer er duurzaam met data wordt omgegaan en gepubliceerd onderzoek makkelijk door externen kan worden ingezien en beoordeeld.

3. Uitleg FAIR principes

3.1 Findable

De F van Findable zegt “zorg dat data vindbaar zijn”.

Begin met het vindbaar maken van je onderzoek al in de planfase van het onderzoek en zorg voor vindbaarheid in alle onderzoeksfasen, dit voorkomt tijdrovende aanpassingen achteraf.

Om onderzoeksdata vindbaar te maken is het gebruik van een goede metadatering en Persistent Identifiers (PID) belangrijk. Het opnemen van de onderzoeksdata in internationaal erkende repositories (bijvoorbeeld Dimensions.ai) vergroot de vindbaarheid van onderzoek en onderzoeksdata.

Zie ook deze link over het FAIR maken van data, het [FAIRification proces](#)

Zie hier voor links met meer informatie over [Persistent Identifiers](#) en [metadatering](#).

Hier vind je informatie over [selectiecriteria](#) en informatie over erkende [repositories](#) voor onderzoeksoutput

3.2 Accessible

De A van Accessible beschrijft de toegangsvoorwaarden van de data.

Het motto van ‘accessible’ is “zorg dat mens en machine (lees: computer) weten onder welke voorwaarden ze bij de data kunnen”. (Meta)data kunnen worden opgehaald door hun identificatie met behulp van een gestandaardiseerd communicatieprotocol.

Een protocol is bijvoorbeeld een [https](#) of [FTP](#). Beiden hebben een veilige versleutelde uitwisseling van gegevens als doel en zijn zonder kosten open, vrij en universeel implementeerbaar. De eigenaar van de onderzoeksdata kan de gebruiks- en toegangsrechten specificeren; hoe de data te gebruiken en delen. Ook heeft de eigenaar de mogelijkheid, voor zover privacywetgeving dit toelaat, de mogelijkheid om bij te kunnen houden (loggen) wie de data hergebruikt/inziet en hoe vaak dit gebeurt. Het protocol maakt waar nodig een authenticatie- en autorisatieprocedure mogelijk. Denk daarbij aan inloggen met behulp van een extra inlogcode. Je maakt dan bijvoorbeeld gebruik van een [HMAC-authenticatie](#) methode waarbij wachtwoorden automatisch veranderen na ieder gebruik. HMAC zorgt ervoor dat berichten tijdens het verzenden niet kunnen worden aangepast. Ook wel 2FA of MFA (2 Factor/Multi Factor Authentication) genoemd. Steeds vaker is zo’n MFA vereist.

Bij gevoelige data is het ook FAIR om geen toegang te geven via een protocol maar om bij de metadata, de contactgegevens van de eigenaar van de data of contactpersoon voor het onderzoek te vermelden. Zo kunnen anderen de data dan wel inzien als ze dat willen. Het is belangrijk om vast te leggen wie de eigenaar(s) van de data zijn en wie toestemming mag geven voor inzage. Ook is het van belang om te zorgen dat die verantwoordelijkheid wordt overgenomen zodra contactpersonen niet meer in de organisatie werkzaam zijn. Uitgangspunt van het begrip ‘Accessible’ is dat iedereen met een computer en internetverbinding, altijd tenminste de metadata in kan zien die aan de data van een bepaald onderzoek zijn toegevoegd.

3.3 Interoperable (uitwisselbaar)

Bij interoperable is het belangrijk om met de onderzoeker het gesprek aan te gaan.

Over het algemeen zijn datastewards inhoudelijk niet op de hoogte van de datasets. Je stelt vragen aan de onderzoeker die helpend zijn in het interoperable maken van de data. Voorbeelden van vragen die je kunt stellen vind je in bijlage 4.

Een goede reden om ervoor te zorgen dat je data interoperable zijn is dat de door jou gecreëerde data ook interessant kunnen zijn voor iemand uit een andere discipline. Een voorbeeld hiervan is de Erfgoed sector. De afgelopen jaren hebben zij hun data gedigitaliseerd. Andere wetenschapsdisciplines zoals het historisch aardbevingsonderzoek maken hier dankbaar gebruik van. In [1904](#) zijn de eerste seismometers in Nederland geplaatst. Om meer te weten te komen over aardbevingen die vóór 1904 plaatsvonden raadplegen seismografen geschiedbronnen zoals brieven en geschriften.

Creëer een extra dimensie in je datasets door deze aan te vullen met bijvoorbeeld data van het CBS. Bij interoperable kun je onderscheid maken in data die door mensen begrepen worden en data die door computers begrepen worden.

Uitwisselbaarheid data - mensen

Mensen kunnen data interpreteren, maar ze hebben dan wel context nodig. Het is dus zaak dat men de betekenis van de variabelen van een dataset kan begrijpen. De naam én de betekenis van een variabele moet duidelijk zijn. Zo is de naam 'var1' op zich niet duidelijk, tenzij je gebruik maakt van een codeboek waarin de naam staat uitgelegd. Ook moet het duidelijk zijn in welke eenheid de variabele gemeten is. De variabele 'gewicht' kan bijvoorbeeld in gram maar ook in kilogram zijn genoteerd. In zogenaamde 'README- files' omschrijf je alle variabelen. Wanneer je kwantitatief onderzoek doet en je maakt gebruik van Excel kun je de omschrijving van de variabelen bovenaan in het document zetten.

Uitwisselbaarheid data - computers

Het type onderzoek en het vakgebied bepaalt welke tools/software je kunt gebruiken. Hoe specialis- tischer de tool, hoe moeilijker het is om de dataset over bijvoorbeeld tien jaar te kunnen openen en om de data te kunnen koppelen aan andere datasets of op te halen door andere systemen.

Data die door computers worden uitgewisseld moeten expliciet geformuleerd worden zodat er geen fouten kunnen ontstaan. Denk hierbij bijvoorbeeld aan de datumnotatie in de Verenigde Staten (mm-dd-yyyy) versus Nederland (dd-mm-yyyy).

Wanneer je kwantitatief onderzoek doet kun je uitzoeken of er al bestaande datasets zijn. Kijk welke variabelen gebruikt worden en hoe die geïnterpreteerd worden. Dit staat omschreven in de datadocu- mentatie. In bijlage 1 vind je een voorbeeld van datadocumentatie. Wanneer de betekenis van de variabelen overeenkomt met wat jij voor ogen hebt kun je de datasets mogelijk koppelen of samen- voegen.

Wanneer je kwalitatief onderzoek doet kun je uitzoeken of er al bestaande datasets zijn. Kijk of er bij de codering gebruik is gemaakt van een bestaand Vocabulaire. Kijk of de betekenis van de codes overeenkomt met wat jij voor ogen hebt. Het kan zijn dat een ander Vocabulaire beter bij je onderzoek past. Een overzicht van Vocabulaires vind je bij ['Linked Open Vocabulaires'](#). Op dit moment is er (nog) geen uitgewerkt voorbeeld van het gebruik van een Vocabulaire binnen praktijkgericht onderzoek beschikbaar.

Op de site van het RIVM staat een [voorbeeld](#) van een uitgewerkt document met datadocumentatie, de zogenaamde 'README-files'.

Een ander voorbeeld van interoperabiliteit is het onderzoeksproject 'Farm Data Train' van de WUR (zie bijlage 3)

Voor een uitgebreide training op het gebied van dataorganisatie, zie de [Social Science Workshop Overview \(datacarpentry.org\)](#)

3.4 Reusable (herbruikbaar)

Het is belangrijk de data zo te beschrijven en toegankelijk te maken dat andere onderzoekers onderzoeksgegevens en datasets makkelijk kunnen vinden, gebruiken en kunnen bewerken.

Bij herbruikbaarheid gaat het om:

1. het toekennen van duidelijke gebruiksvoorwaarden aan een dataset;
2. het geven van een gedetailleerde beschrijving van de data en hun herkomst;
3. het zoveel mogelijk gebruik maken van community standards bij metadatering.

Om hergebruik van de data mogelijk te maken is het van belang om een licentie mee te geven die duidelijk omschrijft onder welke (juridische) voorwaarden de data (her)gebruikt mogen worden. De afwezigheid van een licentie betekent niet dat de data voor iedereen vrij toegankelijk zijn te gebruiken, maar creëert juist juridische onzekerheid. De licentie is bij voorkeur 'machine-leesbaar'.

Een gedetailleerde beschrijving van de data en uitgebreide informatie over de context waarin de data zijn gecreëerd, maakt het mogelijk voor potentiële gebruikers om te bepalen of de data geschikt zijn om te gebruiken voor hun specifieke onderzoek.

Het moet voor mensen en computers duidelijk zijn waar de data vandaan komen, hoe ze zijn gegenereerd, wie ze heeft verzameld, welke verwerkingsstappen zijn gemaakt, en welke apparatuur, machines, meetinstrumenten, vragenlijsten, standard operating procedures, et cetera zijn gebruikt.

Om herbruikbaar te zijn, moeten de (meta-)data opgebouwd zijn volgens de geldende standaarden of best practices van een bepaalde onderzoeksgemeenschap, dat wil zeggen gebruik de standaard file formats en de metadatastandaard die gebruikelijk zijn voor het onderzoeksgebied en werk met een zo gestandaardiseerd mogelijke opbouw van de data. Als er geen community standard gebruikt wordt in een onderzoeksgebied, gebruik dan de algemene metadatastandaard die de repository hanteert, of kies dan zelf voor een gangbare standaard als Dublin Core of DataCite Metadata Schema.

4. FAIR Principes per onderzoeksfase

In dit hoofdstuk wordt per hoofdfase in het onderzoekstraject beschreven hoe de verschillende FAIR principes het best kunnen worden toegepast op onderzoeksdata.

4.1 Voorafgaand aan het onderzoek

Voordat het daadwerkelijke onderzoek überhaupt kan beginnen, moeten plannen worden gemaakt, onderzoeksvorstellen geschreven, financiering gezocht, datamanagementplannen opgesteld en afspraken gemaakt over samenwerking tussen onderzoekers binnen de hogeschool of met externe partners. Al in de beginfase moet worden nagedacht hoe om te gaan met het beheren van onderzoeksdata. Stakeholders verwachten dat een en ander in een Datamanagementplan (DMP) vooraf duidelijk beschreven wordt.

Findable (vindbaar)

De vindbaarheid van onderzoek in de planfase wordt beter door aandacht voor onderstaande punten.

- DataManagementPlan, DMP, hierin een duidelijke omschrijving van het onderzoek.
Denk hierbij aan welke data/tools/protocollen.
Maak voor het maken van een DMP gebruik van een template, bijvoorbeeld voorgeschreven door de subsidieverstrekker, eigen instelling of gebruik DMP-online.
- Pre-registration, (vooraf registreren).
Een DMP en preregistratie zijn tools om je onderzoek FAIR te maken
- Metadata. Voorzie je onderzoek van goede metadata om de vindbaarheid te vergroten.
Gebruik zoveel mogelijk gestandaardiseerde metadata, bijvoorbeeld Dublin Core of Datacite.
Belangrijk voor de vindbaarheid is ook de data te voorzien van een Persistent ID (PID)(bijvoorbeeld DOI), denk daar nu al over na, maar is in deze fase van het onderzoek nog niet essentieel. Neem de PID en de ORCID op in de metadata.
Zie deze link van de cursus [RDNL](#) voor aanvullende informatie over een DMP, DMP aan de tekentafel, ook aandacht voor preregistratie, preregistration.
[Klik hier](#) voor een voorbeeld en omschrijving over metadata, KU Leuven.
In de cursus van RDNL ook aandacht voor metadata [Klik hier](#) voor een voorbeeld en omschrijving.

Accessible (toegankelijk)

- Bepaal in het DMP-plan (samen met partners) welke onderzoeksdata openbaar/gesloten toegankelijk mogen zijn.
- Bepaal het eigenaarschap van de data en wie contactpersonen zijn om vragen met betrekking tot het onderzoek en de data te beantwoorden.
- Check het opslagmedium en de repository (waar je data opslaat tijdens en na het onderzoek) op de mogelijke typen en methoden van beveiliging en authenticatie.
- Bepaal voor de verschillende datasets wie er toegang en bewerkingsrechten hebben tijdens het onderzoek.
- Zet in het opslagmedium een heldere mappenstructuur op met duidelijke namen voor bestanden en mappen zodat data voor verschillende onderzoekers makkelijk vindbaar zijn. Op die manier ben je al voorbereid op het archiveren van de data na het onderzoek. Hier een [voorbeeld](#) mappenstructuur.

Interoperable (uitwisselbaar)

Wanneer je kwantitatief onderzoek doet kun je uitzoeken of er vanuit eerder onderzoek bruikbare datasets zijn om mee verder te werken. Deze vind je bijvoorbeeld in de digitale archieven van DANS (Easy) of bij 4TU. Kijk welke variabelen gebruikt worden en hoe die geïnterpreteerd worden. Dit staat omschreven in de datadocumentatie. Een uitgewerkt voorbeeld hiervan zie je in bijlage 1. Op de website van het [RIVM](#) vind je een voorbeeld van datadocumentatie.

Wanneer de betekenis van de variabelen overeenkomt met wat jij voor ogen hebt, kun je de datasets mogelijk koppelen of samenvoegen.

Ook als je kwalitatief onderzoek doet, kun je uitzoeken of er al bestaande datasets zijn. Kijk of er bij de codering gebruik is gemaakt van een bestaand Vocabulaire. Kijk of de betekenis van de codes overeenkomt met wat jij voor ogen hebt. Het kan zijn dat een ander Vocabulaire beter bij je onderzoek past. Een overzicht van Vocabulaires vind je bij '[Linked Open Vocabularies \(linkedata.es\)](#)'. Op dit moment is er (nog) geen uitgewerkt voorbeeld van het gebruik van een Vocabulaire binnen praktijkgericht onderzoek beschikbaar.

Check of alle data output van de door jou gebruikte tools/software in zogenaamde 'preferred formats' is bewaard. Duurzame bestandsformaten maken het mogelijk om de bestanden over tien jaar ook nog te kunnen openen en omgezet kunnen worden naar een duurzaam bestandsformaat om te voorkomen dat op termijn data verloren gaat.

Reusable (herbruikbaar)

- Bij het opstellen van het datamanagementplan moet al worden nagedacht over welke data opgeslagen gaan worden en hoe de data gedocumenteerd zullen worden. Hier kan de start gemaakt worden met bijvoorbeeld het codeboek of een logboek om je onderzoeksdata, methoden et cetera te beschrijven.
- Als in het onderzoek persoonsgegevens worden verwerkt neem dan in het toestemmingsformulier (informed consent) al de vraag op of de deelnemer toestemming geeft voor mogelijk toekomstig hergebruik van de verzamelde data.
- Werk in een onderzoeksprotocol je plan uit (en preregistreer dit, zie [RDNL](#)).
- In het DMP wordt ook gevraagd naar het gebruik van metadastandaarden. Hier kan al gekeken worden naar relevante community standards.

Zie de lijst met [community standards](#) van de RDA Metadata Standards Directory Working Group.

4.2 Tijdens het onderzoek

In deze fase worden data verzameld, onderling gedeeld en tijdelijk opgeslagen. Het is van belang om onderzoeksdata(sets) in deze fase al zo FAIR mogelijk op te slaan zodat onderzoekers hun data up-to-date houden, de data voor collega-onderzoekers goed toegankelijk zijn en zij met de juiste versies van bestanden werken.

Findable (vindbaar)

In de onderzoeksfase is de vindbaarheid van je onderzoek belangrijk; let daarom op onderstaande punten.

- Metadata (voorzie al je onderzoeksgegevens van goede metadata).
- Registratie van onderzoek.
- Datadocumentatie (workflow tijdens je onderzoek; vindbaar bijvoorbeeld door het delen en beschikbaar stellen van je labnotes, codes en aanpak (protocol)).

Links en informatie voor het vindbaar maken van je onderzoeksdata [Crossref](#), [Datacite](#) en [Datacite Netherlands](#), TU Delft.

Voorbeelden van workflow en datadocumentatie, zie bijvoorbeeld [Myexperiment](#)

Accessible (toegankelijk)

- Maak duidelijke README-files bij mappen en documenten waarin je uitlegt hoe documenten gebruikt kunnen worden waarom folders/documenten bijvoorbeeld niet of beperkt toegankelijk zijn. Op de website van het RIVM staat een [voorbeeld](#) van een uitgewerkt document met datadocumentatie, de zogenaamde 'README-files'. Ook geeft 4TU een heldere [uitleg](#).
- Check of bepaalde toegangsrechten nog steeds goed staan ingesteld.
- Bepaal welke data je tijdens je onderzoek (bijvoorbeeld tussenresultaten) al naar buiten wilt brengen?

Interoperable (uitwisselbaar)

Houd datadocumentatie bij. Hierin beschrijf je de (interpretatie van) variabelen, (interpretatie van) codes, gebruikte tools en software, redactie die je op je data gedaan hebt. Een voorbeeld hiervan vind je bij dit [RIVM-onderzoek](#).

Sla de definitieve bestanden die je na afloop van het onderzoek gaat archiveren/publiceren op in de eerder besproken 'preferred formats', zoals pdf/A, CSV, et cetera.

Tip: [Frictionless Data](#) biedt de tool 'DataPackage creator' die in bijvoorbeeld je csv-files onvolkomenheden zoals lege rijen kan opsporen. Zo verbeter je de kwaliteit van je dataset.

Reusable (herbruikbaar)

- Houd log books, code books, data dictionaries, etc. bij gedurende de data-workflow. Annoteer analyse-scripts/syntax met comments. Werk volgens 'standard operating procedures'. Documenteer versies, typenummers, etc. van gebruikte apparatuur, meetinstrumenten, software, etc.
- Meer informatie over datadocumentatie tijdens het onderzoek is te vinden in deze [Libgide](#) van Zuyd Hogeschool

4.3 Na afronding van het onderzoek

In deze fase worden onderzoeksresultaten verwerkt in conclusies, rapportages, publicaties en worden onderliggende data (sets) voor de langere termijn gearchiveerd. Het is van belang onderliggende onderzoeksdata zo FAIR mogelijk te archiveren. De data worden gearchiveerd in een institutionele of landelijke (certified) data repository zoals [4TU.datacentrum](#), [DANS EASY](#), [DataverseNL](#), [Figshare](#), [Zenodo](#).

Findable (vindbaar)

Na afronding van het onderzoek zijn de volgende punten van belang voor de vindbaarheid van onderzoeksdata, veel van deze punten zijn belangrijk in alle fasen van het onderzoek, en zijn daar ook al benoemd.

- Metadatering
- Persistent Identifiers, PID, toekennen, zie ook de informatie in de eerdere fasen van het onderzoek. Maak ook gebruik van toekenning/aanvraag van een ORCID om de vindbaarheid van de onderzoekers te vergroten. Neem de PID's op in de metadata. Toekenning van een DOI gebeurt vaak automatisch bij opname in een erkende data repository. Als je een DOI wilt aanmaken buiten een repository om, dan kan dat via een DOI Registration Agency
- Als de data zelf vertrouwelijk is, maak dan wel de metadata vindbaar en beschikbaar. In online repositories/publicatieplatforms (bv in Figshare is het mogelijk om alleen de metadata te publiceren, metadatafiles only) Het is ook mogelijk om onder embargo te publiceren, de onderzoeksdata zijn dan pas na het verstrijken van een bepaalde tijdsperiode onder voorwaarden toegankelijk.

Informatie over ORCID vind je [hier](#).

Informatie over DOI en DOI registration Agencies kun je vinden met deze [link](#) van de DOI-organisatie

Accessible (toegankelijk)

- Bepaal per dataset/folder welk toegangsprotocol toegepast moet worden. Zo'n protocol (bv https of ftp) kan waar nodig een authenticatie- en autorisatieprocedure mogelijk maken. Denk hierbij bv aan inloggen met behulp van een extra inlogcode. (HMAC-authenticatie)
- Ook kunnen gebruikersvoorwaarden worden toegevoegd zodat anderen weten onder welke voorwaarden de data gebruikt mogen/kunnen worden.
- Bepaal eventueel embargoperiodes.
- Check of beveiliging van data en protocollen goed zijn toegepast.
- In het geval van gevoelige data of data die niet extern gearhiveerd kunnen/mogen worden, kun je de beschrijving en metadata opslaan in een extern opslagmedium. Zo wordt het onderzoek zelf wel bekend, maar zijn de onderzoeksdata niet openbaar toegankelijk. Voeg gegevens van personen toe waarmee contact kan worden opgenomen als er vragen zijn over het onderzoek en de data. Regel ook de vervanging van contactpersonen als, voor het onderzoek relevante, personen de organisatie verlaten.

Informatie over toegang tot data bij de bekende archieven/opslagmedia:

DANS opslagmogelijkheden

[DataverseNL](#)

[DANS easy](#)

Surf opslagmogelijkheden

[Dataopslag en -beheer | SURF.nl](#)

[Langdurig data opslaan met Data Archive | SURF.nl](#)

[4TU](#)

Zenodo

[Policies](#)

[Zenodo | re3data.org](#)

Figshare

[figshare | re3data.org](#)

[Access policy](#)

[Copyright & License Policy](#)

[Registry of Research Data Repositories](#)

Interoperable (uitwisselbaar)

- Controleer of alle bestanden die moeten worden gearhiveerd/gepubliceerd in preferred formats zijn opgeslagen.
- Rond de datadocumentatie - README text file(s) - af en archiveer deze bij je datasets.

Tip:

- Gebruik bijvoorbeeld dit [document](#) van 4TU.nl met voorkeursformatie of de meer uitgebreide lijst van [DANS](#)
- 4TU biedt de mogelijkheid om je Github account te koppelen aan je 4TU account. Op deze manier kun je door jou ontwikkelde software duurzaam opslaan.

Reusable (herbruikbaar)

- Voeg aan te publiceren en/of archiveren (meta)data een duidelijke licentie toe waarin de gebruiksvoorwaarden voor hergebruik worden omschreven. Veel gebruikte datalicenties zijn de reeks [Creative Commons \(CC\)](#) licenties die duidelijk beschrijven hoe data wel en niet hergebruikt mogen worden.

De CC-licenties zijn onherroepelijk. Er zijn ook andere soorten licenties, onder meer voor software en code.

Meer informatie:

- [Datalicenties](#) in de Datamanagement Expert Guide van CESSDA
 - Informatie over [softwarelicenties](#)
 - Een [hulpmiddel](#) bij het kiezen van een licentie
- Voeg aan te publiceren en/of archiveren (meta)data een uitgebreide set met datadocumentatie toe. *Datadocumentatie omvat:*
 - Een beschrijving van de data (een overzicht van de bestanden met een beschrijving van de inhoud per bestand met informatie over onder andere bestandsformaat, gebruikte software om data te genereren, benodigde software om het bestand te kunnen openen, gebruikte codes en variabelen).
 - Een beschrijving van het dataverzamelingsproces en de gebruikte instrumenten (onder andere codeboek, logboek, vragenlijsten, handleidingen, meetinstrumenten).
 - Een omschrijving van de verandering van de dataset in de tijd (een historisch verslag van de bewerkingen van de dataset in de tijd).

Meer informatie:

[Datadocumentatie](#) in de Datamanagement Expert Guide van CESSDA

- Kies voor een gestandaardiseerde methode van metadateren, die past bij je onderzoeksdomein. Dat zorgt voor betere herbruikbaarheid van je onderzoek en dataset. Een data repository geeft vaak al een aantal algemene metadatatavelden op basis van Dublin Core of DataCite Metadata Schema. Meer specifieke metadata op basis van community standards zullen vaak extra moeten worden toegevoegd.
- Je datasets kun je checken via [F-UJI](#), een webservice om de FAIRness van datasets te beoordelen op basis van statistieken die zijn ontwikkeld door het FAIRsFAIR-project.

5. Begrippenlijst

Data Management Plan, DMP

Een schriftelijke overeenkomst waarin staat welke gegevens worden opgeslagen, hoe ze worden opgeslagen (bestandsformaat, versiebeheer, metadata), of gegevens worden ingediend bij een repository en onder welke voorwaarden (Bron: [LCRDM website](#), woordenlijst)

Digital Object Identifier (DOI)

Een type persistent identifier dat wordt gebruikt om objecten op unieke wijze te identificeren. Het DOI-systeem wordt met name gebruikt voor elektronische documenten zoals tijdschriftartikelen. Het DOI-systeem begon in 2000 en wordt beheerd door de International DOI Foundation. Zie ook: Persistent Identifier (PID) (Bron: [LCRDM website](#), woordenlijst)

Metadata

Gestandaardiseerde gestructureerde informatie die gegevensitems uitlegt, zoals, maar niet beperkt tot: doel, oorsprong, tijdreferenties, geografische locatie, maker, toegangsvoorwaarden en gebruiksvoorwaarden van een gegevensverzameling. Gegevens over gegevens. (Bron: [LCRDM website](#), woordenlijst)

Metadatastandaarden

Maak bij metadata zoveel mogelijk gebruik van standaarden die er al in de wereld zijn en die ook succesvol worden toegepast. Dat scheelt je veel werk én dat maakt de informatie beter vindbaar, herbruikbaar en relevant omdat ook andere organisaties van die standaarden gebruik maken. Denk bij metadatastandaarden aan [Dublin Core](#) of de Nederlandse [OWMS](#) die weer deels op de Dublin Core is gebaseerd.

ORCID

ORCID, wat staat voor Open Researcher en Contributor ID, is een wereldwijde, non-profit organisatie die wordt ondersteund door vergoedingen van [onze lidorganisaties](#). We zijn community-built en worden bestuurd door een [Raad van Bestuur](#) die representatief is voor ons lidmaatschap met een brede vertegenwoordiging van belanghebbenden. ORCID wordt ondersteund door een toegewijde en deskundige [professionele staf](#). (Bron : [ORCID](#))

Persistent Identifier (PID)

Een langdurige verwijzing naar een document, bestand, webpagina of ander object (echt of abstract). Zie ook: Digital Object Identifier (DOI) (Bron: [LCRDM website](#), woordenlijst)

Preregistratie

Bij preregistratie beoordelen peer reviewers de onderzoeksvraag en de onderzoeksmethodologie ([Center for Open Science, n.d.a.](#)). En dat doen ze nog voordat het onderzoek wordt uitgevoerd. Bij een positief oordeel leidt dit tot een voorlopige acceptatie van de toekomstige publicatie op voorwaarde dat de auteurs hun geregistreerde protocol volgen. Je kunt preregistratie zien als een soort van zelfbescherming tegen onze eigen denkfouten ([Begley, 2013](#)). (Bron: Cursus [RDNL](#))

6. Algemene bronnen en noten

<https://www.kuleuven.be/rdm/nl/metadata/metadata>

[SRS_poster_2019_FAIR-724x1024.jpg \(724x1024\)\(saxion.nl\)](#)

<https://www.lcrdm.nl/files/lcrdm/2020-10/Do-IPASS-for-FAIR.pdf>

[GO FAIR initiative: Make your data & services FAIR \(go-fair.org\)](#)

<https://www.nature.com/articles/sdata201618>

(bronartikel uit 2016 waarin FAIR initiatief wordt beschreven)

[FAIR principles - Onderzoekscyclus: 1.Planfase - LibGuides at Zuyd University](#)

<https://www.nwo.nl/research-datamanagement>

[The FAIR Data Principles | FORCE11](#)

https://lcrdm.nl/wp-content/uploads/2023/03/HOW-FAIR-IS-YOUR-DATA_flyer_2.pdf

[23 Things for research data management | Field guides to Research Data Management \(uu.nl\)](#)

[Three camps, one destination: the intersections of research data management, FAIR and Open \(uksg.org\)](#)

[FAIR-Aware \(knew.nl\)FAIR Principles - GO FAIR \(go-fair.org\)](#)

[RDM, open en FAIR | RDNL - Essentials 4 Data Support \(researchdata.nl\)](#)

[Steeds FAIRder, HVA rapport](#)

https://www.surf.nl/files/2020-03/landelijk-integraal-onderzoeksondersteuningmodel-hogescholen_0.pdf

7. Bijlagen

Bijlage 1: Format metadata/datadocumentatie

Dataset/resultaat - niveau

| Onderdeel | Omschrijving | Type |
|-----------------------------|---|---------------|
| Titel | De titel van het onderzoeksproject | Tekst |
| Doel | Het doel in één zin omschrijven | Meerkeuze (?) |
| Samenvatting / omschrijving | <ul style="list-style-type: none"> - Waarom het onderzoek - Korte methodiek - Kort resultaat - Betekenis van resultaten | Tekst |
| Status | Vorbereiding; lopend; stopgezet; afgerond | Meerkeuze |
| Contactpersoon | Naam van de contactpersoon | Tekst |
| Organisatie(s) | Organisatie die het uitgevoerd heeft | Tekst |
| Startdatum | De officiële startdatum van het project | Datum |
| Einddatum | De officiële / beoogde einddatum van het project | Datum |
| Thema/SDG | Bestaand thema vanuit de HBO kennisbank / SDG | Meerkeuze |
| Onderzoeksresultaat | Impact van het onderzoek | Tekst |
| Trefwoorden | 3 tot 5 passende trefwoorden voor het onderzoek | Tekst |
| Focus | Wetenschappelijk - onderwijs - combinatie | Meerkeuze |
| Projectnummer | Het nummer van het project | Tekst |
| Financiering | Organisatie die het gefinancierd heeft | |

Dataset/resultaat - niveau

| Onderdeel | Omschrijving | Type |
|----------------------------------|---|--------------|
| Titel | De titel van de dataset | Tekst |
| Beschrijving | - Waarom deze dataset van belang is - Korte methodiek totstandkoming | |
| Onderzoek / Projectnummer | Waar de dataset voor gebruikt is | Tekst |
| Contactpersoon | Vorbereiding; lopend; stopgezet; afgerond | Meerkeuze |
| Geografische extent | Voor welk gebied de data dekkend is | Tekst / tool |
| Datum uploaden | | |
| Interval van data | Periode waarvoor de data dekkend is | Datum |
| Gebruikersrechten / retourbeleid | Één van de zeven CC-typen | Meerkeuze |
| Producten | Dataset wordt gebruikt in andere applicatie | Meerkeuze |

Dataspécifiek - niveau

| Onderdeel | Omschrijving | Type |
|-----------------------|---|-----------|
| Attribuutomschrijving | Per opgenomen attribuut een omschrijving | |
| Inwinningsmethode | Moment van inwinnen | Datum |
| Verwerkingsmethode | Sensoren en tools die gebruikt zijn | Meerkeuze |
| | De stappen die zijn gezet om van de ruwe data informatie te maken | Tekst |
| Opmerkingen | Afrondingen, must-knows | Tekst |

Gebruiksrechten bij een IT project heeft in dit veld een andere waarde, bv Apache.

Bijlage 4

Vragen te stellen over interoperable:

Let op: interoperable gaat over hergebruik maar ook over het combineren van datasets.

[The Elements of FAIR - Interoperability - YouTube](#)



Technische interoperability

- Van welke bestandsformaten wordt gebruik gemaakt?
- Staan de bestandsformaten in de voorkeurslijst van DANS? Is die bij iedereen bekend? Worden er andere voorkeurslijsten gebruikt?
- Zijn de juiste metadata vastgelegd? Hierbij ook persistent identifiers zoals ORCID, DOI.
- Worden de variabelen op dezelfde manier geschreven (bijvoorbeeld datum)?

Semantic interoperability

- Bedoelen we allemaal hetzelfde met een specifiek woord? Bijv plant = groene blaadjes of fabriek, string = onderbroek of snaar
- Welke eenheden gebruik je?
- Maak je gebruik van specifieke vocabulaires?
- Heb je de afspraken of toelichting hierover vastgelegd in documentatie?
- Datadocumentatie waarin je beschrijft hoe je de data hebt verzameld, geïnterpreteerd, opgeschoond (bv onvolledig ingevulde vragenlijsten), gebruikte tools/software, etc.
- Heb je de variabelen (kwantitatief onderzoek) omschreven? Dit kun je bovenaan in het databestand (Excel file) doen.

Organisational interoperability

- Heb je het er over gehad met de consortium partners? Gebruiken werkveld en andere samenwerkingspartners hetzelfde jargon, staan zij achter de afspraken?

Legal interoperable

- Zijn er afspraken gemaakt over copyright?
- Zijn er afspraken gemaakt over licenties? Ligt dit vast in metadata?
- Hoe is omgegaan met GDPR en andere wet- en regelgeving?

Multidisciplinair

Kun je de data aanvullen met data uit andere bronnen waardoor je een extra dimensie krijgt? Bijvoorbeeld historisch aardbevingsonderzoek maakt ook gebruik van geschriften/brieven uit geschiedbronnen omdat seismologische metingen nog niet zolang gedaan worden.

Colofon

Digital Competence Center Praktijkgericht Onderzoek (DCC-PO), 2024
Deze uitgave is met zorg samengesteld met vermelding van relevante bronnen

Druk/versie

Versie 3.2

Auteursrecht

De tekstuele inhoud van deze publicatie valt onder de CC-BY 4.0 licentie creativecommons.org/licenses/by/4.0
Gebruikte logo's en huisstijl vallen onder het auteursrecht en mogen niet zonder toestemming hergebruikt worden.

Vragen?

Neem contact met ons op via info@dcc-po.nl

Volg DCC-PO

Via www.dcc-po.nl