



ZOEKEN EN HERGEBRUIKEN VAN ONDERZOEKSDATA



INTRODUCTIE

- Deze handreiking is bedoeld voor onderzoekers en onderzoeksondersteuners, zoals data stewards en informatiespecialisten, die betrokken zijn bij het zoeken naar data of hierover advies willen geven aan onderzoekers
- Het is een wegwijzer voor databronnen, zoals data-archieven, zoekmachines en andere informatiebronnen. Daarnaast behandelt de handreiking zoekstrategieën voor data en biedt handvatten om kwaliteit en betrouwbaarheid van data te beoordelen
- Gebruikswijze: Klik op pijltjes om verder of terug te gaan en op de blauwe blokjes voor meer informatie over een onderwerp



INHOUD

Hergebruiken van data

Waarom zou je data hergebruiken?

Waarvoor kun je data(-documentatie) hergebruiken?

Zoekstrategie dataset

Beoordelen van data (FAIR, kwaliteit van dataset)

Documenteren van dataherkomst

Citeren van data

Licentie nieuwe dataset met hergebruikte data



HERGEBRUIKEN VAN DATA

Hergebruik van data doelt in deze context op het gebruiken van eerder verzamelde data uit onderzoek of andere bronnen voor een lopend onderzoeksproject. Voorbeelden van data die je kunt hergebruiken zijn:

- Datasets van eerder (eigen of niet eigen) onderzoek
- Beleidsonderzoek
- Patiëntendossiers
- Bedrijfsdata
- Overheidsdata
- Data van niet-academische instituten



WAAROM ZOU JE DATA HERGEBRUIKEN

- Geld besparend
- Tijd besparend
- Duurzaamheid
- Versnellen van wetenschappelijke ontwikkelingen
- Versterken van je eigen dataset door te combineren met bestaande data
- Specifieke doelgroepen worden minder overvraagd
- Ethische overweging (wanneer het onderzoek impact heeft op deelnemers, dieren of milieu, etc.)
- Mogelijkheid tot samenwerking vergroten (internationaal en interdisciplinair)





WAARVOOR KUN JE DATA(DOCUMENTATIE) HERGEBRUIKEN?

Bij onderzoeksdata:

- Om nieuwe onderzoeksvragen te beantwoorden
- Om te valideren (reproduceerbaarheid)
- Om nieuwe datamodellen te bouwen (Data Science)
- Om andere datasets als vergelijkingsmateriaal te gebruiken
- Om de vraagstelling te onderbouwen

Bij datadocumentatie:

- Om andermans methode over te nemen
- Om je te laten inspireren



ZOEKSTRATEGIE DATASET



Napkin. AI



WAT ZOEK IK

Bepalen van je zoekstrategie in stappen





WAAR ZOEK IK

Overzicht van mogelijke kanalen

Selectie van datarepositories

Zoekmachines

Nederlandse datarepositories

Voorbeelden van internationale repositories en platforms

Open data platforms (voorbeelden)

Datatijdschriften



WAAR ZOEK IK: OVERZICHT VAN MOGELIJKE KANALEN

- Online datarepository van je instituut
- Domeinspecifieke datarepository: selecteren op domein (Re3data, Openaire)
- Andere databanken: [CBS](#), [RIVM](#), [Kaggle](#), [Overheid.nl](#), [Centerdata](#), [OECD](#)
- Toegang via [Data Availability Statement](#) bij geselecteerde literatuur
- Datazoekmachines
- Vragen aan onderzoekspartners en je netwerk of een auteur van een relevante bron om onderliggende data te delen



WAAR ZOEK IK: SELECTIE VAN DATAREPOSITORIES

Datареpositories kun je bijvoorbeeld zoeken via een registry. In een Registry vind je geen datasets maar wel de repositories waar datasets in staan.

De belangrijkste datареpositories registries zijn:

Re3data.org

Open Access Directory

FAIRsharing.org



RE3DATA.ORG

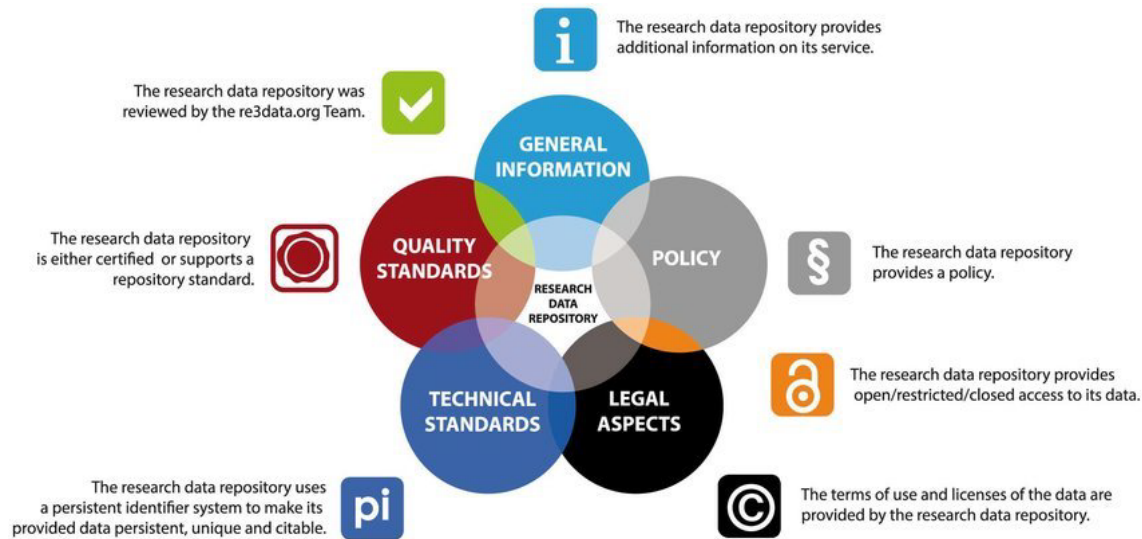
- re3data.org is een wereldwijd register van onderzoeksdatarepositories dat alle onderzoeksdisciplines dekt
- Je kunt zoeken naar repositories met verschillende soorten datasets: experimentele data, simulaties, afbeeldingen, geluid, video, enquêtes, fysieke monsters, softwarebronnen, genomische en geospatiale data, enz.
- Je kunt repositories filteren op basis van certificering, metadatastandaarden en verschillende andere criteria



re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

Search



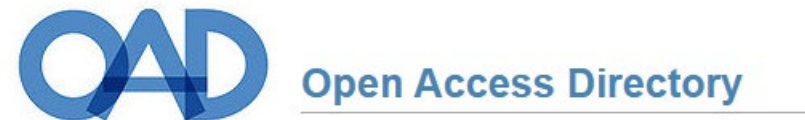
Information	Access	Licenses	Persistent Identifier	Certificates and Standards	Policy	Reviewed
			doi			
			urn			
			ark			
			hdl			
			purl			
			other			

Afbeeldingen: Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., & Dierolf, U. (2013). Making research data repositories visible: The re3data.org registry. PLoS ONE, 8(11), e78080. <https://doi.org/10.1371/journal.pone.0078080>



OPEN ACCESS DIRECTORY

- [Open Access Directory](#) biedt een alfabetische lijst van datarepositories, gesorteerd op onderwerpen
- Je kunt deze lijst raadplegen wanneer je op zoek bent naar domeinspecifieke datarepositories





FAIRSHARING.ORG

- [FAIRsharing](https://fairsharing.org) biedt een uitgebreide lijst van repositories in verschillende disciplines
- Je kunt repositories filteren op basis van criteria zoals discipline, landen, licentie
- FAIRsharing biedt gedetailleerde beschrijvingen van elke repository, inclusief het proces en voorwaarden van het deponeren van data, licenties en specifieke kenmerken, zoals datacuratie, data toegangsvoorwaarden, data versiebeheer, etc.





ZOEKMACHINES

Zoekmachines zijn online tools waarmee je meerdere repositories tegelijk kunt doorzoeken om een publicatie (van een dataset) te vinden. Dit zijn een aantal voorbeelden:

Datacite Commons

Odissei Portaal

Dataset Search

CESSDA Data Catalogue



- DataCite Commons is een organisatie die zich richt op vindbaarheid van data. Het platform monitort citaties van data en gebruik van repositories
- Het platform levert een DOI (voor bijvoorbeeld je dataset)
- Het host publicaties uit verschillende repositories
- Je kunt zoeken op soort publicatie (datasets, artikelen, events, collecties), auteur, organisatie en repository



ODISSEI PORTAAL



ODISSEI

Open Data Infrastructure for
Social Science and Economic Innovations

- ODISSEI staat voor *Open Data Infrastructure for Social Science and Economic Innovations* en brengt hogescholen, universiteiten, onderzoeksinstituten en publieke organisaties samen om grootschalige en innovatieve sociale data beschikbaar te maken
- Het ODISSEI Portal host data uit verschillende repositories (DANS data stations, Dataverse.nl, CBS, LIES pannel) van de sociale wetenschappen
- Je kunt met geavanceerde semantische zoekopdrachten zoeken naar datasets



DATASET SEARCH

Dataset Search

- Dataset Search is een zoekmachine voor datasets van o.a. onderzoeksinstellingen, universiteiten, overheidsportalen (bijv. [DATA.GOV](#), [European Data Portal](#)) en non-profit organisaties
- Dataset Search host metadata van repositories binnen verschillende domeinen, zoals geowetenschappen, sociale wetenschappen, chemie, biologie, landbouw, economie en overheidsdata
- Je kunt zoeken op titel, beschrijving van de inhoud, organisatie, publicatiedatum, bestandsformaten en licentie
- Je kunt filteren op type data, bestandsformaat, datum laatste aanpassing



CESSDA DATA CATALOGUE



- CESSDA staat voor Consortium of European Social Science Data Archives die sociale wetenschappers toegang geeft tot grootschalige en duurzame databronnen. Het doel is om samenwerking, open science en hergebruik van data in de sociale wetenschappen te bevorderen
- CESSDA Catalogue host metadata van nationale repositories van sociale wetenschappen uit 22 Europese landen
- Kwantitatieve, kwalitatieve of mixed data, cross-sectioneel, longitudinaal, recent en historische gegevens
- Maakt gebruik van gestructureerde metadata, zoals titel, beschrijving van de inhoud, organisatie, publicatiedatum, bestandsformaten en licentie
- Je kunt filteren op titel, trefwoorden, jaar van publicatie, land, uitgever



NEDERLANDSE DATAREPOSITORIES (VOORBEELDEN)

Datarepositories zijn in deze context systemen waarin datasets worden gepubliceerd.

Datarepository	Uitleg	Kenmerken volgens Re3data.org
DANS data stations	• Vier domeinen onderverdeeld in vier datastations. • Data stations: Archeologie, Social Sciences and Humanities • Life Sciences, Physical and Technical Sciences	 
DataverseNL	• Alleen datasets/datadocumentatie • Alle domeinen bij elkaar (eventueel uitgesplitst door instituten met lidmaatschap)	
4TU.ResearchData	• Data en software in het technische domein	

INTERNATIONALE REPOSITORIES EN PLATFORMS (VOORBEELDEN)

Datarepository	Uitleg	Kenmerken volgens Re3data.org
Dryad	• open access repository voor data van medische wetenschappen • Alle datasets hebben CC0 licentie • Alledataset hebben een Read.me bestand	
OSF	• De OSF is een generalistisch archief, wat betekent dat het niet beperkt is tot één enkele discipline.	
Zenodo	• Naast datasets staan er ook artikelen, presentaties, en andere vormen van digitale objecten.	



OPEN DATA PLATFORMS (VOORBEELDEN)

Categorie	Organisatie	Website	Inhoud
Overheid	Overheid.nl	https://data.overheid.nl/	Datasets van overheidsinstellingen
	RIVM	https://www.rivm.nl/	Dataset van de domeinen: gezondheid, natuur & milieu, geo wetenschappelijke data, etc.
	CBS	Statline	Economie & maatschappij. CBS verzamelt daarnaast Microdata. Deze data zijn niet open)
	Gemeentes, b.v. Rotterdam	https://data.rotterdam.nl/pages/home/	Overheidsdata gericht op een gemeente
Onderzoeksinstituten	CentreData Research Institute	https://www.centerdata.nl/	Maatschappij
Internationale organisaties	OECD	https://www.oecd.org/en/data.html	Statistische data betreffende landen die lid zijn van OECD.
	Europese Unie: Eurostat	https://ec.europa.eu/eurostat/web/main/data/database	Data betreffende de Europese Unie, verschillende domeinen.
	The Platform for Experimental, Collaborative Ethnography	https://worldpece.org/	Etnografie
	World Bank	https://data.worldbank.org/	World Development Indicators (WDI), statistische data, maatschappij
Research Communities	Kaggle	https://www.kaggle.com/	Data science platform

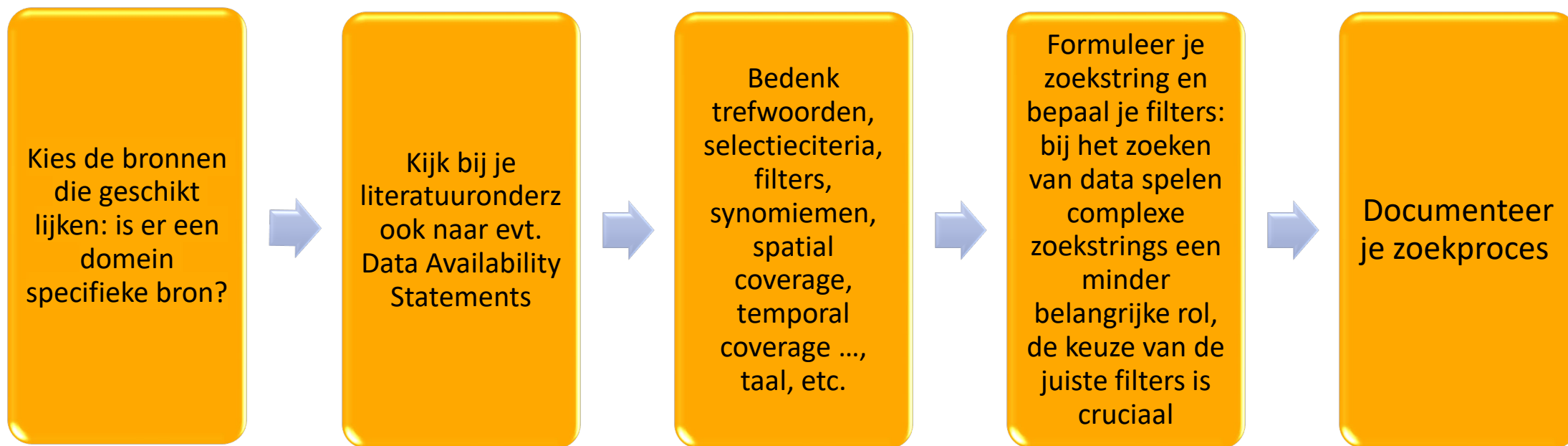


DATATIJDSCRIFTEN

- In een datatijdschrift worden geheel of gedeeltelijk datagerichte artikelen gepubliceerd
- Lijst van datatijdschriften (voorbeelden):
 - <https://zenodo.org/records/7082126>
 - <https://www.gbif.org/data-papers> (tijdschriften over biodiversiteit)



HOE ZOEK IK?





DOCUMENTEREN ZOEKPROCES

Documenteer je zoekproces in bijvoorbeeld een logboek. De omschrijving van het zoekproces kunnen bestaan uit de volgende punten:

- De gebruikte repository en/of zoekmachine (noteer informatie over bruikbaarheid, functionaliteiten en relevantie)
- Gebruikte zoekwoorden en filters
- Aantal gevonden resultaten
- Evaluatie van je zoekproces en eventuele verbeteringen die je hebt aangebracht



WAT HEB IK? BEOORDELEN VAN DATA

Maak een selectie uit de zoekresultaten op grond van de volgende criteria:

FAIRness van data

Kwaliteit



FAIRNESS VAN DATA

Datasets - FAIR principles:

Findable ✓

Accessible

Interoperable

Reusable



ACCESSIBLE

Wat is de toegankelijkheid van de dataset?

- Metadata is toegankelijk
- Toegangsvoorwaarden en licenties zijn duidelijk vermeld
- Toegangsvoorwaarden van de dataleveranciers of onderzoekspartners



INTEROPERABLE

Kan je deze dataset combineren of uitwisselen?
Hierbij kun je letten op de volgende aandachtspunten:

- Metadatastandaarden
- Omschrijving variabelen (legenda)
- Consistentie in omschrijvingen
- Codering antwoordopties
- Meeteenheden



REUSABLE

Wat heb ik nodig om deze dataset te begrijpen en te kunnen gebruiken?

- Datadocumentatie
- Gebruikslicentie: <https://creativecommons.nl/uitleg/>
- Bestandsformaten (heb ik speciale software nodig?)



KWALITEIT

Helaas worden onderzoeksdata nog in hele beperkte mate gepeerreviewed. De volgende punten kunnen een indicatie geven of de onderzoeksdata van goede kwaliteit zijn:

- De (meta)data en datadocumentatie zijn goed
- De wetgeving is nageleefd
- De repository heeft een certificaat (b.v. Core Trust Seal) of is goed aangeschreven
- De hoofdonderzoeker heeft ook artikelen gepubliceerd
- De hoofdonderzoeker heeft een ORCID
- Het onderzoeksinstituut waar de hoofdonderzoeker bij aangesloten zit is bekend
- Van de dataset is ook een artikel gepubliceerd



DOCUMENTEREN VAN DATAHERKOMST

De omschrijving van de herkomst van gekozen datasets kan bestaan uit de volgende punten:

- De database/repository/organisatie
- Toegankelijkheid (open, onder embargo, beperkt, gesloten)
- Persistent identifier(s) (bijv. DOI)
- Datum van wanneer je de dataset hebt gevonden

Je kunt de datasets toevoegen aan je reference manager



CITEREN VAN DATA

Gebruik bij het citeren van een dataset in een artikel de citatiestijl die door de redacteur/uitgever wordt voorgeschreven. Als er geen stijl wordt aanbevolen voor datasets, neem dan eventueel een standaard citatiestijl aangeboden door een data repository en pas deze aan de stijl van je artikel aan.

- Meer over het citeren van data:

[How to Cite Datasets and Link to Publications | DCC](#)



LICENTIE NIEUWE DATASET MET HERBRUIKTE DATA

1. Wanneer je een bestaande dataset hebt bewerkt en wilt publiceren dan moet je rekening houden met de licentie die op de bestaande dataset staat.

Publiceren van bewerkte dataset

2. Wanneer je bestaande datasets gaat samenvoegen en publiceren moet je rekening houden met de licenties op de bestaande datasets.

Publiceren van samengevoegde bestaande datasets



PUBLICEREN VAN BEWERKTE DATASET

Wanneer je een bestaande dataset bewerkt en deze wilt publiceren, ben je bij de keuze voor een nieuwe licentie afhankelijk van de licentie die op de originele dataset staat. Het volgende schema geeft aan welke keuze je hebt:

- **Groen:** Licentie die je mag toekennen op bewerkte data met een bepaalde oorspronkelijke licentie.
- **Geel :** juridisch gezien mag het maar wordt afgeraden.
- **Grijs:** licenties die je niet mag gebruiken

Adapter's license chart		Adapter's license						
		BY	BY-NC	BY-NC-ND	BY-NC-SA	BY-ND	BY-SA	PD
Status of original work	PD							
	BY							
	BY-NC							
	BY-NC-ND							
	BY-NC-SA							
	BY-ND							
	BY-SA							

Bron: <https://creativecommons.org/faq/#when-is-my-use-considered-an-adaptation>



PUBLICEREN VAN SAMENGEVOEGDE BESTAANDE DATASETS

Wanneer je een nieuwe dataset maakt van bestaande datasets moet je er op letten dat de licenties van de datasets compatibel zijn.

- Vinkje: datasets mogen gecombineerd worden
- Kruisje: datasets mogen niet gecombineerd worden.

Bron: <https://creativecommons.org/faq/#when-is-my-use-considered-an-adaptation>



	PUBLIC DOMAIN	PUBLIC DOMAIN	CC BY	CC BY SA	CC BY NC	CC BY ND	CC BY NC SA	CC BY NC ND
PUBLIC DOMAIN	✓	✓	✓	✓	✓	✗	✓	✗
PUBLIC DOMAIN	✓	✓	✓	✓	✓	✗	✓	✗
CC BY	✓	✓	✓	✓	✓	✗	✓	✗
CC BY SA	✓	✓	✓	✓	✗	✗	✗	✗
CC BY NC	✓	✓	✓	✗	✓	✗	✓	✗
CC BY ND	✗	✗	✗	✗	✗	✗	✗	✗
CC BY NC SA	✓	✓	✓	✗	✓	✗	✓	✗
CC BY NC ND	✗	✗	✗	✗	✗	✗	✗	✗



AUTEURS EN CONTACTGEGEVENS

- Auteurs: Inge Boot, Anne-Marie Borgers-van Boeckel, Esther Goedegebuure, Sylvia Huwaë, Marta Kargól, Renate Mattiszik
- Contactgegevens DCC-PO: secretariaat@dcc-po.nl
- Licentie: CC BY



Inge Boot, Anne-Marie Borgers-van Boeckel, Esther Goedegebuure, Sylvia Huwaë, Marta Kargól, Renate Mattiszik, Zoeken en hergebruiken van onderzoeksdata.

Handreiking DCC-PO, 2026. Zenodo: [10.5281/zenodo.22271821](https://zenodo.org/record/22271821). Dit werk is gelicenseerd onder de licentie CC BY 4.0.