

FAQ Metadata

(Versie 28/6/2022)

Wat is metadata?

Metadata bestaat uit informatie die data kenmerkt. Metadata wordt gebruikt om documentatie voor dataproducten te leveren. In wezen beantwoordt metadata wie, wat, wanneer, waar, waarom en hoe zit het met elk facet van de gegevens die worden gedocumenteerd.

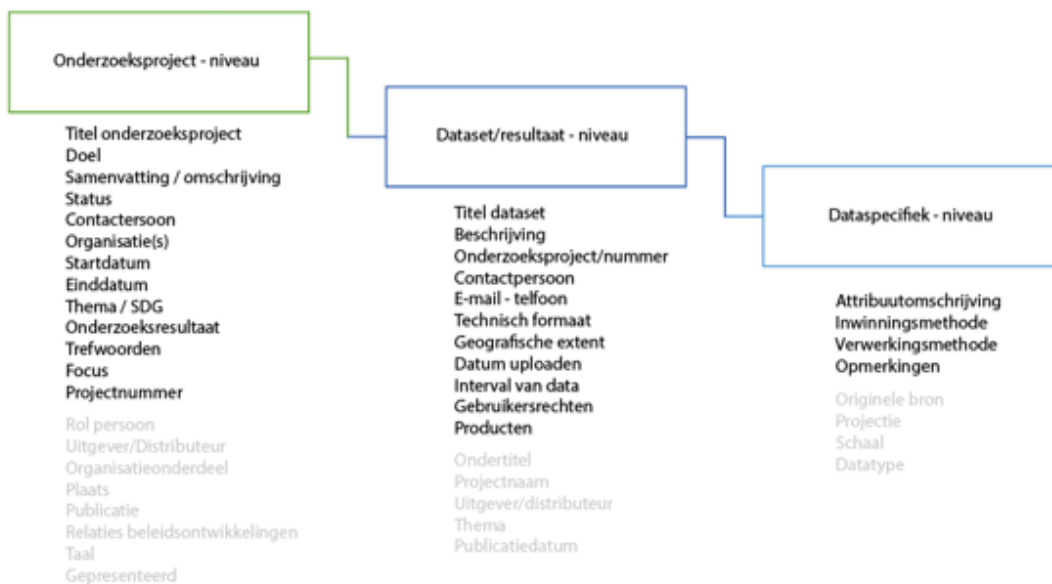
Waarom is het creëren van metadata belangrijk?

Metadata zorgt voor een duidelijke publicatie van uw project, onderzoek en data. Het helpt derden bij het begrijpen van het onderzoek. Daarnaast biedt het handvatten om het onderzoek te reproduceren.

Welke metadata standaard kan ik gebruiken?

Dit hangt af van het domein. Soms kan het ook afhangen van de repository (opslagplaats) die gekozen wordt. Een minimale metadata set is afgebeeld hieronder:

Bron: Acceptatieformulier, NPPO (metadatumodel),



Een voorbeeld op dataspécifiek niveau is

employee_id	first_name	last_name	nin	department_id
44	Simon	Martinez	HH 45 09 73 D	1
45	Thomas	Goldstein	SA 75 35 42 B	2
46	Eugene	Comelsen	NE 22 63 82	2
47	Andrew	Petulescu	XY 29 87 61 A	1
48	Ruth	Stadick	MA 12 89 36 A	15
49	Barry	Scardelis	AT 20 73 18	2
50	Sidney	Hunter	HW 12 94 21 C	6
51	Jeffrey	Evans	LX 13 26 39 B	6
52	Doris	Bemdt	YA 49 88 11 A	3
53	Diane	Eaton	BE 08 74 68 A	1
54	Bonnie	Hall	WW 53 77 68 A	15
55	Taylor	Li	ZE 55 22 80 B	1

Data

Column	Data Type	Description
employee_id	int	Primary key of a table
first_name	nvarchar(50)	Employee first name
last_name	nvarchar(50)	Employee last name
nin	nvarchar(15)	National Identification Number
position	nvarchar(50)	Current position title. e.g. Secretary
department_id	int	Employee departmnet. Ref: Departmetns
gender	char(1)	M = Male, F = Female, Null = unknown
employment_start_date	date	Start date of employment in organization.
employment_end_date	date	Employment end date. Null if employee st

Bron figuur: <https://dataedo.com/kb/data-glossary/what-is-metadata>

Een algemeen overzicht is te vinden op:

<https://www.dcc.ac.uk/guidance/standards/metadata/list>

Wanneer je niet een specifiek vakgericht metadata schema wil gebruiken is de [Dublin Core](#) een goede mogelijkheid.

Een paar specifieke metadata schema's:

Geo: [ISO 19115](#) Bruikbaar voor het beschrijven van geo-informatie en services. Let ook op eventuele specifieke uitbreidingen

Agri: [AgMES](#) Agricultural Metadata Element Set, onderhoud door FAO

Food: <http://internet-of-food.org/projects/metadata-for-food/> Vanuit Internetsociety Internet of Food SIG

Leefomgeving: Een belangrijk overzicht van metadata is te zien bij [GeoNovum](#).

Hoe kan ik bepalen welke metadata van belang is?

Als provider (degene die data maakt) helpt het om te denken aan een potentiële gebruiker(s). Wat zou de gebruiker willen weten? Locatie, inhoud, verwijzing naar project, verwijzing naar publicatie, ouderdom, gebruik etc.

Wat is een repository?

Een repository is een plek in de cloud waar de data opgeslagen kan worden. Diverse organisaties bieden diverse mogelijkheden (tegen kosten, afhankelijk van hoeveelheid en toegang). Een repository wordt gebruikt als zoeklocatie. Om tijdens het zoeken de juiste data te vinden is het van belang om de metadata in te vullen. Soms is dit specifiek, soms is dit gekoppeld aan de gebruikte metadata standaard.

Welke repository zal ik kiezen om mijn publicatie en data op te slaan?

HAS Hogeschool heeft nog geen standaard repository gekozen. Op dit moment is de voorkeur voor publicaties op Green-I met een link (bij voorkeur een PID) naar de data op een repository.

Wat is Green-I?

Dit de [online bibliotheek](#) van en voor het groene hbo. Binnen de HAS kan meer info gegeven worden door m.willems@has.nl of G.vdBoer@has.nl.

Welke repository wordt veel gebruikt?

De keuze wordt mede bepaald door het domein, die ook de mogelijk mede gebruikers gebruikt wordt. Dus waar je zelf zoekt, kan een goede plek zijn.

Onderstaande repositories zijn bekeken in juni 2022.

[Dans EASY](#) Verbonden aan de KNAW. Onderwerpen op Social Sciences and Humanities; Archeology; Life, Health and Medical Sciences; Physical and Technical Sciences

[DataverseNL](#) Tb.v. langdurige opslag en het aanmaken van persistent identifiers. Naast 4TU. ResearchData zijn de volgende organisaties aangesloten: NIOO, Trimbos instituut, Utrecht University, VU, Hanzehogeschool Groningen, Universiteit Leiden, Avans Hogeschool, Tilburg University, PThU, Maastricht University, Rijksuniversiteit Groningen, UMC Utrecht, Rotterdam University of Applied Sciences, Fontys, Rijksdienst Cultureel Erfgoed.

[4TU.ResearchData](#) (Tu Delft, University of Twente, Eindhoven university of Technology, Wageningen University and Research.

Onderwerpen: Mathematics, Physics, Chemistry, Earth Sciences, Atmospheric Sciences, Physical Geography and Environmental GeoScience, Geomatic Engineering, Environmental Sciences, Biology, Agricultural and Veterinary Sciences, Information and Computing Sciences, Nanotechnology, Aerospace Engineering, Civil Engineering, Electrical and Electronic Engineering, Built Environment and Design, Materials Engineering, Mechanical Engineering, Medical and Health Sciences, Business and Management.

[SURF](#) is de ict-coöperatie van onderwijs en onderzoek. Voor meer informatie (o.a. Surf research drive en SURF Sharekit) zie <https://www.surf.nl/portfolio-datadiensten-welke-dienst-heb-je-nodig-voor-beheren-en-opslaan-van-je-data> SURF Research drive wordt vooral gebruikt tijdens het onderzoek.

Een overzicht van diverse repositories kan gevonden worden op [Re3data.org](#) (Registry of research repositories)

Wat is een PID (Persistent Identifier).

Hiermee kan een permanente link worden gemaakt naar de data. Vergelijk DOI of ISBN voor publicaties. De PID worden geregeld via bij [ePIC](#) (Persistent Identifiers for eResearch). Het specifieke nummer wordt in Nederland door [SURFsara](#) als dienst aangeboden.

Moet ik onderscheid maken tussen metadata voor ruwe, onbewerkte data en metadata voor bewerkte, geanalyseerde data?

In principe niet, behalve dat voor de bewerkingen en analyse ook vastgelegd moeten worden in het kader van transparantie. Let wel op dat in het kader van de AVG de oorspronkelijke data geanonimiseerd moeten worden.

Hoe hangt het FAIR principe samen met metadata?

Het [Fair principe](#) Findable is alleen mogelijk als de metadata waarop gezocht wordt, ingevuld is. Zonder beschrijving is het zeker dat de data niet gevonden kan worden.

Wanneer begin ik met het metadata?

Als onderzoeker is het van belang om al tijdens het project te starten met het bijhouden van metadata. Een aantal velden in de gekozen metadata standaard kunnen al tijdens het onderzoek aangevuld worden. Het is belangrijk om aan te sluiten bij de metadata standaard die veel gebruikt wordt binnen het vakgebied. Eerdere publicaties van vergelijkbare data kunnen daarbij helpen.

Kan ik metadata het beste aan het einde van het onderzoeksproject maken, omdat ik dan beter weet wie potentiële gebruikers zijn?

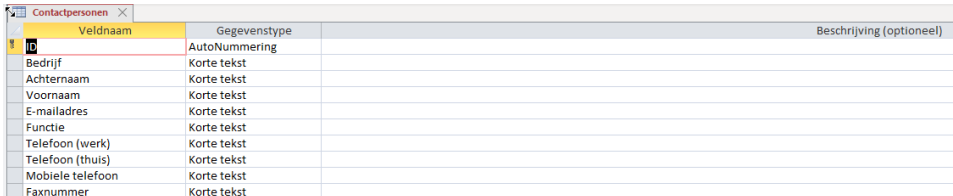
In principe is het aan te raden z.s.m. met het invullen van metadata te beginnen. Zodoende is het nog bij te houden en kan bij vragen snel met de betrokkenen nog extra informatie opgehaald worden. In principe is de metadata onafhankelijk van de gebruiker. Wel kan gekeken worden of sommige metadata velden belangrijker of essentiëler zijn voor de gebruiker. Vroeg bijhouden levert de volgende voordelen. Ten eerste zorgt het ervoor dat je nadenkt over je project. Ten tweede voorkomt het dat je achteraf niet meer weet wat je gedaan hebt. Ten derde kan het voor een team van onderzoekers tijdens het project voor iedereen helpen actueel te blijven. Als laatste kan het ook het gericht voeren van discussies over het proces ondersteunen.

Wie moet metadata creëren?

Sommige data-inwin tools doen dit automatisch. Meestal moet je dit echter zelf doen. Als onderzoeker is het van belang om metadata te creëren. Jij weet immers welke processen de data doorloopt en wat de betekenis is van de data. Echter is het van belang dat iedereen in het project metadata aanvult.

Hoe maak ik metadata?

Er zijn verschillende oplossingen. Soms zit het al gekoppeld aan de software, of zit daarin een specifieke mogelijkheid (b.v. [ArcGIS Pro](#)). Maar meer voorkomend is het gebruik van README-files (zie <https://www.makeareadme.com/>). Het kan zijn dat je aangeeft welke velden tot de minimale metadata set behoren. Ook is het mogelijk in een spreadsheet een eerste tabblad te gebruiken door alle velden nader te beschrijven. In een database kan elk veld beschreven worden in de ontwerpweergave van een tabel (zie voorbeeld).



Veldnaam	Gegevenstype	Beschrijving (optioneel)
Bedrijf	AutoNummering	
Achternaam	Korte tekst	
Voornaam	Korte tekst	
E-mailadres	Korte tekst	
Functie	Korte tekst	
Telefoon (werk)	Korte tekst	
Telefoon (thuis)	Korte tekst	
Mobiele telefoon	Korte tekst	
Faxnummer	Korte tekst	

Hoelang duurt het creëren van metadata?

Afhankelijk van je ervaring duurt het 2 uur om een opzet te maken. Daarna is het verstandig om elke week een half uur uit te trekken om de metadata in te vullen. Vooral bij het opzetten van databestanden moet gelijk aandacht gegeven worden aan de opzet van de metadata. Wanneer veel data als tussenproducten gemaakt worden is de keuze van de naam van het bestand essentieel. Zodoende geeft dat al zicht op inhoud van het bestand. Denk hier ook aan het reproduceren van de analyse.

Waar sla ik mijn metadatabestand op?

In principe tijdens het project op het hoogste niveau (afhankelijk van de gekozen map indeling), zodat een nieuwkomer in het project gelijk de relevante informatie over de data kan vinden. Eventueel kan ook een aparte map met metadatabestanden worden gemaakt. Na afloop van een project moet de metadata samen met de data opgeslagen worden.

Wat is de rol van een datasteward t.o.v. metadata?

Een datasteward heeft de taak om ondersteuning en advisering te geven aan onderzoekers op het gebied van Research Data management. Daarbij hoort dat de datasteward de onderzoekers kan informeren en motiveren over metadata.

Wat kan een datasteward of onderzoekondersteuner, bezig met publicatie van data, doen als de metadata niet volledig ingevuld is bij publicatie?

In principe weigeren, en zorgen dat de metadata aangevuld wordt. Het kan zijn dat het minder belangrijke velden niet zijn ingevuld, dan kan het proces doorgang vinden. Maar in eerste instantie is het van belang de onderzoeker op te wijzen dat goed onderzoek eist dat de metadata van de data op orde is. Hierdoor kan (her)gebruik van data vergemakkelijkt worden.