



DANS Data Stations - Data Processing Team Handleiding

Please note that this Data Processing Team Manual is currently only available in Dutch. An English translation will be added soon.

Omgang met persoonsgegevens	2
Deponering	3
Curatie	5
Inventarisatie	5
Controle	5
Basis Curatie (Metadata)	6
Bestanden	9
Publiceren	12
BIJLAGE 1 Autorisatiematrix Data Stations	14



Omgang met persoonsgegevens

Voor de verwerking van persoonsgegevens gelden de volgende afspraken voor het Data Processing Team:

Toegang beperken tot gegevens tijdens de verwerking

Voor het doel waartoe persoonsgegevens worden verwerkt, is het alleen nodig om bepaalde personen toegang tot deze persoonsgegevens te geven en om deze bepaalde handelingen te kunnen laten verrichten.

Maatregel: DANS hanteert een autorisatiestructuur waarmee de rollen en rechten zijn vastgelegd, toegang verloopt altijd conform deze structuur. Zie bijlage 1.

Up to date houden van gegevens (juistheid - "Accuracy")

Conform de [Terms Of Use](#), mag DANS metadata corrigeren en bestanden omzetten ten behoeve van duurzame archivering en/of ten behoeve van de wensen van de eindgebruiker. Het is van belang dat persoonsgegevens niet onbedoeld wijzigen tijdens deze handelingen.

Daarnaast geldt op basis van de Terms Of Use dat DANS bestanden niet inhoudelijk mag bewerken. Het anonimiseren van gegevens die naar oordeel van DANS onbedoeld zijn opgenomen is daarmee op die basis dus ook niet toegestaan.

Maatregel:

- Wanneer DANS conversies doet op de data betreft dit enkel de vorm/het bestandsformaat, ter bevordering van de lange-termijn duurzaamheid en/of toegankelijkheid. Door geconverteerde bestanden achteraf te openen (of een representatieve steekproef bij een groot aantal) controleert de datamanager of de inhoud inderdaad ongewijzigd is gebleven.
- Het is geen onderdeel van het werkproces om databestanden inhoudelijk te bewerken, de aandacht gaat uit naar de vorm en het bestandsformaat. Bij wijze van uitzondering kan DANS een databestand inhoudelijk bewerken wanneer een deposant/rechthebbende hiertoe toestemming geeft. Het bestand mag dan alleen op de afgesproken manier bewerkt worden.

Bewaring en verwijdering van gegevens

Gegevens mogen niet langer bewaard worden dan nodig. Dit verkleint daarnaast ook de kans op een datalek. Het Data Processing Team bewaart data op verschillende media.

Lokaal:

- Laptops/PC: (lokaal beheerd)
- DANS server: (lokaal beheerd)

- Microsoft Office (lokaal beheerd)
- Open Office (lokaal beheerd)
- Libre Office (lokaal beheerd)
- StatTransfer (lokaal beheerd op licentiebasis)
- Python scripts (lokaal beheerd)
- FFMPEG (lokaal beheerd)
- Adobe Suite (lokaal beheerd)

Extern

- Vancis (conform Terms of Use, oneindig indien mogelijk)
- Microsoft 365 (op termijn): conform bewaarbeleid KNAW (wordt nog opgesteld)
- SURFfilesender: SURF hanteert een standaard bewaartermijn van 4 maanden

Maatregel

- Bij datasets met persoonsgegevens: zo snel als mogelijk en uiterlijk binnen een maand na publicatie van de datasets moet de lokale opslag zijn opgeschoond.
- SURFfilesender: de gegevens worden door SURF automatisch verwijderd na 4 maanden. Deze bewaartermijn is passend voor DANS om zo de ruimte te hebben om de data te kunnen downloaden.

Passende bescherming bij gebruik SURFfilesender

Voor het versturen van persoonsgegevens of andere 'restricted' data van gevoelige aard moet de encryptie-optie van SURF filesender worden gebruikt.

Maatregel: Bij het opvragen van gegevens via SURFfilesender wijst de Datamanager de deposant altijd op het verzorgen van extra bescherming middels encryptie bij het toesturen van persoonsgegevens.

Deponering

Stappen die de klant onderneemt voor het deponeren van een dataset: de curator moet dit proces goed kennen.

Zie ook [Handleiding Data Deponeren](#).

- De klant logt in
- De klant klikt op 'Add Data' => 'New Dataset'
- Er wordt een formulier geopend voor het invullen van de metadata. Dit formulier is in elkaar gezet op basis van een 'template' - bovenaan het formulier is de Host Dataverse en het gebruikte Dataset Template te zien. Per Data Station heeft DANS een standaard Template als 'default' staan.
- De klant beschrijft de dataset in de metadata-velden en heeft hier ook de optie om bestanden toe te voegen.

- Bij het invullen van de metadata geeft de klant zelf aan of er persoonsgegevens in de dataset zitten ('Personal Data In Dataset?'), met een selectiemenu met als opties: Yes/No/Unknown. Deze indicatie wordt nagekeken bij de curatie (zie Curatie). De deposant blijft de eindverantwoordelijke voor de dataset.
- Na 'Save Dataset' kan de klant een andere licentie kiezen bij het tabblad 'Terms' => 'Edit Terms Requirements'
- Na 'Save Dataset' kan de klant voor elk geüpload bestand beschrijvingen toevoegen of de beschikbaarheid aanpassen via File Options (de drie puntjes rechts van het bestand). Hier is het mogelijk om het bestand op 'Restrict' te zetten, met of zonder 'Enable access request'. Als de klant het vinkje voor 'Enable access request' uitzet is deze verplicht om *Terms of Access for Restricted Files* op te geven. DANS is er om het hergebruik van onderzoeksdata te bevorderen: het aanzetten van de mogelijkheid voor het doen van 'access requests' is in principe standaard, hier mag enkel van af worden geweken op afspraak en na instemming van de Data Station Manager.
- De File Options geven ook de mogelijkheid om een embargo aan bestanden toe te kennen. Dit kan per bestand apart ingesteld worden. DANS raadt een maximum embargo van twee jaar aan, maar dit is niet technisch afgedwongen.
- Na 'Save Dataset' kan de klant verdere bewerkingen aan de metadata doen (Add + Edit Metadata) en/of meer bestanden uploaden (Upload Files).
- De klant dient de dataset in met 'Submit for review'. Hierna kan de klant zelf geen aanpassingen aan de dataset meer doen.

Aandachtspunten

- Het toevoegen van een dataset aan een collectie kan door in de metadata bij 'Relation Metadata' - 'Collection' via autofill de juiste collectie te zoeken. Deze functionaliteit was niet in EASY aanwezig: het vervangt de *Thematische Collecties* (dit waren meestal lege datasets met een HTML voorpagina met lijsten van collecties en links naar de desbetreffende datasets). Het is iets waar wij gebruikers op zullen moeten wijzen, waar men aan zal moeten wennen en waar bij de curatie op moeten blijven letten: zijn de datasets hier al toegewezen aan de juiste collectie, of moeten wij dit nog doen?
- Bij het toevoegen van bestanden valt met 'Edit' per bestand (of voor alle bestanden, d.m.v. aanvinken) in te stellen of dit bestand 'Restricted' is of niet. Edit geeft de opties: Delete/Restrict/Unrestrict. Wanneer je hier voor Restrict kiest, moet je nog apart aangeven of voor het bestand een toegangsverzoek in te dienen is of niet: 'Request Access' - 'Enable access request' (aanvinkvakje). Restricted data zonder deze optie is effectief Closed Access: bestanden zijn zichtbaar maar kunnen niet gedownload worden. Het is *niet* mogelijk in Dataverse om bestanden onzichtbaar te maken (visibility:NONE, zoals we dit in EASY hadden).
- Wanneer meerdere bestanden op Restricted worden gezet, is het *niet* mogelijk om voor een deel van deze bestanden een toegangsverzoek mogelijk te maken en voor een deel niet. **Wanneer de optie 'Enable access request' wordt aangevinkt voor ook maar**

één Restricted bestand, wordt deze keuze doorgevoerd voor *alle* Restricted bestanden in *alle* versies van de dataset.

Dit maakt dat datamanagers goed moeten nakijken of de juiste keuzes zijn gemaakt. Staan de bestanden die op Restricted Access zouden moeten staan inderdaad op Restricted en zijn deze opvraagbaar? Restricted Access data zou standaard opvraagbaar moeten zijn. Voor uitzonderingen waarvoor het toch wenselijk zou zijn om enkele Restricted Access bestanden *niet* opvraagbaar te maken (denk aan een 'informed consent' bestand), zie *Curatie - Restricted Access*.

- Het zal voor niet iedereen even duidelijk zijn hoe de licentie van een dataset aangepast moet worden; dat dit ná 'Save metadata' bij 'Terms' ingesteld moet worden. Dit moeten de datamanagers wel goed weten om hier begeleiding bij te kunnen leveren en dit te kunnen controleren.

Curatie

Zie ook de [Data Stations Policy](#). Datamanagers verzorgen beoordeling en in ieder geval basiscuratie van alle handmatig gedeponeerde datasets die worden ingediend voor beoordeling. Na inloggen in een Data Station ziet de datamanager deze datasets door te filteren op **Publication Status - In Review**. Alle 'in review' datasets hebben ook de status 'Draft' totdat deze door ons worden gepubliceerd. Geheel nieuwe deponeringen hebben daarnaast ook de status 'Unpublished'; ingediende versies van bestaande datasets hebben *niet* de status Unpublished en moeten ook door de datamanager worden nagekeken, gecureerd en gepubliceerd. Datasets met de status 'Draft' die *niet* de status 'In Review' hebben, zijn nog in bewerking door de deposant. Deze zal de deponering eerst moeten indienen alvorens de datamanager ermee aan de slag kan gaan.

Inventarisatie

Voor elk Data Station zijn interne overzichten beschikbaar waar je de datasets in kan registreren.

Controle

Een dataset moet beoordeeld worden op de algemene inhoud, op die manier wordt bepaald of de dataset te accepteren is voor archivering. Zie ook het DANS selectiebeleid.

Wetenschappelijk onderzoek

Een dataset hoort het gevolg te zijn van wetenschappelijk onderzoek of onderzoek ten behoeve van beleid. De beoordeling gebeurt in principe door de datamanagers. Bij twijfel wordt overlegd met de Data Processing Team Leader.

Nieuwe inhoud

Het kan voorkomen dat een deposant niet doorheeft dat een dataset al bij DANS is gearchiveerd, maar dat de datamanager dit wel in de gaten krijgt. Een tweede, aparte deponering van een dataset met identieke data wordt niet geaccepteerd. Het is wel mogelijk om een nieuwe versie op een bestaande dataset te maken.

Licentie

Controleer de [licentie](#) bij License/Data Use Agreement. Als het [CC0](#) is, kijk na in het desbetreffende inventarisatie spreadsheet of dit de standaard keuze is van de organisatie. Doe navraag bij de depositor als je het niet zeker weet. Deposanten hebben niet altijd door hoe de licentie omgezet moet worden, sommige deposanten willen het wel eens vergeten.

Als de dataset Restricted Files bevat, staat bij het tabblad Terms wat de Terms of Access zijn en daarbij of de optie voor het kunnen doen van Access Requests is aangevinkt. Als er restricted files zijn en er staat 'Users may **not** request access to files', kijk na of een goede reden is opgegeven bij 'Terms of Access for Restricted Files', bespreek eventueel met de Data Station Manager of dit acceptabel is.

Doe enig inzicht op van het onderwerp en de inhoud van de dataset aan de hand van de beschrijvingen en het nakijken van documenten en andere databestanden waar eventueel persoonsdata in zou kunnen zitten. Als het vermoeden bestaat dat de dataset persoonsgegevens bevat maar er is voor een open licentie gekozen, neem dan contact op met de deposant om dit na te vragen.

Basis Curatie (Metadata)

Vóór publiceren kunnen aanpassingen aan de metadata worden gedaan door naar het tabblad Metadata te gaan en de knop 'Add + Edit Metadata' te gebruiken.

- Controleer of de ingevulde metadata correct en volledig is. Kleine aanvullingen kunnen worden aangebracht indien dit de FAIRness van de dataset ten goede komt.

Citation Metadata

- Controleer of alle (co-) **Authors** zijn ingevuld, vul aan wat ontbreekt.
- De notatie van namen moet als volgt: **initialen tussenvoegsel achternaam**
- Controleer ORCIDs en andere Author Identifiers: deze moeten niet als links worden ingevuld (<https://orcid.org/ID>) maar alleen als ID.
- Controleer de **Description**: kan een andere onderzoeker op basis hiervan begrijpen waar de dataset over gaat? Ga hier alleen aanpassingen doen (of navraag bij de deposant) als het echt té beknopt of té omslachtig is. Het is gebruikelijk om hier de samenvatting of de

conclusies van de rapportage te plaatsen, het streven blijft naar een beknopte omschrijving.

- Bevat de tekst in de **Description** ‘harde returns’? Zo ja, dan kan dit opgeschoond worden: kopieer de tekst even naar een programma zoals Notepad++, haal de harde returns er uit, kopieer de tekst dan weer terug.
- Let op het type onderzoek en op de conclusies die in de **Description** worden vermeld: hier zijn mogelijk Keywords uit te halen of gestandaardiseerde termen uit vocabulaires (*Archaeology-Specific Metadata; Social Sciences and Humanities Metadata*)
- Is een **Date** ingevuld onder de Description? Dat zou alleen van toepassing moeten zijn als er meerdere descriptions bij zijn. Zo niet, dan kan dit verwijderd worden (of verplaatst naar Production Date).
- Is **Language** ingevuld? Zo niet: doe dit alsnog. Dit betreft de taal van de data.
- Is **Production Date** ingevoerd? Zo niet: doe dit alsnog. Deze datum kan gelijk gesteld worden aan de publicatiedatum van de eindpublicatie / de opleverdatum van de dataset.
- Is er een **Distribution Date** ingevoerd die in de toekomst ligt? Zo ja, dan is dit waarschijnlijk bedoeld als embargodatum. Echter, de file settings worden hier niet automatisch op aangepast. Controleer daarom in dit geval of er een embargo op de files staat gelijk aan de Distribution Date.

Rights Metadata

- Is hier aangegeven dat er persoonsdata in de dataset zit? Zo ja: klopt dat, gaat het echt om persoonsgegevens conform de AVG? Zie ook het aparte kopje [Persoonsgegevens](#) op dit punt. Namen van auteurs, medewerkers en opdrachtgevers gelden ter verantwoording van de dataset en kunnen gewoon in de metadata worden opgenomen, dat is geen reden om dit veld op ‘Yes’ te zetten. Als het echt klopt: raadpleeg de juridisch adviseur.
- Als de deposant geen taal heeft ingevuld bij “Language Of Metadata”, voeg deze dan toe. Dit betreft de taal van de metadata, als meerdere talen worden gebruikt in de omschrijving kun je meerdere talen toevoegen.

Relation Metadata

- Controleer of de dataset aan de juiste **Collection** is toegekend, zo niet, pas dit aan.
- Mocht je in de gaten hebben dat er gerelateerde datasets bestaan, dan kunnen deze in het veld **Relation** worden aangegeven.

Temporal and Spatial Coverage

- Zorg er voor dat, waar beschikbaar, coördinaten in de metadata staan bij **Spatial Point** (centrumcoördinaat) of **Spatial Box** (uiterste N-O-Z-W punten). Voor de meeste archeologische projecten worden coördinaten in het Rijksdriehoekstelsel opgegeven (scheme = RD), voluit geschreven nummers op de meter, zonder tussenpunt. De X zit tussen de 0 en de 300000, de Y tussen de 300000 en 600000.

- Mochten Latitude/Longitude coördinaten worden opgegeven, let er dan op dat het formulier eerst om een X-coördinaat en dan om een Y-coördinaat vraagt, maar dat is niet hoe Lat/Lon normaal wordt genoteerd: Latitude = Y en Longitude = X.
- Wij zien graag gebruik van **Spatial Coverage (free text)** voor de invoer van plaatsbepalende gegevens. Vul elke noemer in een eigen veld in (+ om het veld te herhalen): toponiem (straat; plek) / stad / gemeente / provincie. Ter onderscheid van de gemeente t.o.v. steden zetten we het woord Gemeente voor de gemeentenaam.

Archaeology-Specific Metadata

- Belangrijk voor archeologie: zorg er voor dat het **Archis Zaakidentificatie** is ingevuld. Voor de meeste deponeringen zal een ZaakID bestaan. Elk archeologisch onderzoek wordt in het systeem Archis van de RCE geregistreerd en krijgt een dergelijk nummer. Het huidige systeem Archis 3 (sinds ca. 2015) kent een 10-cijferig nummer. Voor oudere onderzoeken bestaat het Archis Onderzoeksmeldingsnummer, dat meestal 5 cijfers (soms 4) heeft.
- Het veld **Archis ID** is een plek voor identifiers uit het Archis systeem anders dan de Zaakidentificatie, een invoer gaat samen met de juist typering. Het kan gaan om een waarneming, vondstmelding, monument of onderzoek. ('Onderzoeknummer' is anders dan een 'Onderzoeksmeldingsnummer': bij aanmelding van een onderzoek in Archis 2 kreeg het project een onderzoeksmeldingsnummer, bij voltooien en afmelden van het onderzoek kreeg men een apart onderzoeksnummer).
- Controleer het gebruik van de ABR vocabulaire velden, vul aan indien mogelijk: **Report (ABR Rapporten)** met bijbehorend **Rapportnummer**; **Methods of Recovery (ABR Verwervingswijzen)** (zie hiervoor meestal de subtitel van het rapport of het begin van de samenvatting/Description); **Subject (ABR Complextypen) / Artefact (ABR Artefacten) / Temporal (ABR Periodes)** - alleen als met zekerheid iets is aangetoond, zie hiervoor de conclusies/de Description.

Social Sciences metadata

- Controleer de classificatie van de onderwerpen met een (of meerdere) term(en) uit de European Language Social Science Thesaurus ([ELSST](#)) en een term van de CESSDA topic classification, een [lijst van onderwerpen](#) samengesteld door CESSDA. Als deze velden leeg zijn maar op basis van de omschrijving of gegeven keywords duidelijk is welke term(en) uit de ELSST en de CESSDA Topic Classification relevant is, voeg deze termen dan toe.
- Neem bij vragen contact op met de deponeerder (Contact Owner)

OH-Smart

OH-Smart is een losstaande deposit module voor Oral History datasets. Deponeringen van OH-Smart datasets komen in het Data Station Social Sciences & Humanities terecht. Datasets gedeponerd met de OH-Smart deposit module bevatten vijf bestanden met OH-specifieke metadata, die niet in de DS metadata te passen valt: 1 JSON LD bestand met alle metadata, 1

JSON LD met alle restricted metadata, 1 JSON LD met alle openbare metadata, 1 TXT bestand met alle metadata, 1 TXT bestand met alle openbare metadata. Hier zijn deze datasets ook goed aan te herkennen. Bij deze datasets is het volgende van belang:

- Er zijn 3 metadata velden die niet door SWORD kunnen worden overgebracht: subtitle; grant agency; grant ID - deze worden wel vastgelegd, maar moeten nog handmatig in de DS metadata worden gezet.
- Kijk goed na of de bestanden die restricted access zouden moeten zijn, dit ook zijn.

Bestanden

- Controleer de beschikbaarheid van de bestanden: een groen slotje op het bestandsicoontje betekent dat het bestand op Restricted is gezet. Bij meerdere bestanden valt via het filter te controleren of/welke bestanden op Restricted zijn gezet. Komt dit overeen met de licentiekeuze/bedoeling? Bij de licentiekeuze DANS Licence moeten de bestanden op Restricted staan. Ook moet 'Access Request' dan enabled zijn, anders kan niemand een aanvraag doen. Let op: voor elk bestand kan apart worden besloten of deze op Restricted staat of niet, maar het vinkje voor 'Access Request enabled' heeft effect op alle bestanden van de dataset die op Restricted staan.

Tenzij op expliciet verzoek van de deposant, mogen datamanagers voor publiceren géén files aanpassen. Indien er op file niveau iets aangepast moet worden, stuurt de datamanager de dataset terug: 'Return to Author'.

Compleetheid documentatiebestanden

De dataset moet de benodigde documentatie bevatten om hergebruik door derden mogelijk te maken. Er is sprake van twee typen documentatie, de onderzoeksdocumentatie en de datadocumentatie.

Onderzoeksdocumentatie

Onderzoeksdocumentatie is de documentatie over het onderzoek, met de eventuele gebruikte instrumenten, en documentatie over wat de dataset aan bestanden bevat. Dit laatste is alleen nodig bij datasets met grote aantallen bestanden, of complexe structuur. De documentatie is noodzakelijk voor een goede interpretatie van de dataset.

Het volgende moet onderdeel zijn van de onderzoeksdocumentatie:

- Algemene informatie over de context van het onderzoek
- Informatie over de onderzoeksofzet en methode van dataverzameling
- (Indien van toepassing) Gebruikte instrumenten

- (Indien nodig) Informatie over de bestanden en/of structuur: aantallen bestanden, relaties onderling etc. Dit wordt aangeleverd in de vorm van een bestandslijst

Een (gepubliceerd) artikel kan als onderzoeksdocumentatie dienen, maar ook een informeel document zoals een readme-bestand. De documentatie kan uit één of meerdere bestanden bestaan.

Daarbij moet de documentatie ook inhoudelijk compleet zijn (bijvoorbeeld: alle nodige hoofdstukken en bijlagen in een rapport) en definitief (geen conceptversies).

De documentatie moet als bestand(en) bij de dataset worden opgenomen. Op die manier kan ook deze documentatie duurzaam gearriveerd worden.

Datadocumentatie

Datadocumentatie is documentatie op het niveau van de data. Het gaat om bijvoorbeeld een verklaring van begrippen en variabelen die voorkomen in een databestand. Deze verklaring kan in een databestand staan (bijvoorbeeld als labels in een SPSS-bestand) of als apart bestand worden meegeleverd. Het is noodzakelijk dat deze documentatie aanwezig is zodat alle waarden en codes op de juiste manier geïnterpreteerd kunnen worden.

Bij onjuiste of ontbrekende documentatiebestanden

Een datamanager neemt contact op met de deponerder om de benodigde documentatie toe te voegen, te vervangen of onjuiste informatie te verwijderen. In overleg met de deponerder kan de datamanager dit ook zelf uitvoeren.

Compleetheid databestanden

Het aantal databestanden moet compleet zijn. Alle delen moeten aanwezig zijn die worden verwacht op basis van de documentatie (bijvoorbeeld: bestand A t/m D bij een datasets waar de documentatie bestand A t/m D noemt). Is dit niet mogelijk dan moet dit duidelijk worden verantwoord. Bij twijfel wordt contact opgenomen met de deponerder.

Bij onjuiste of ontbrekende databestanden

Een datamanager neemt contact op met de deposant om de aanvullingen en/of aanpassingen te regelen vóór publicatie van de dataset.

Bij aanlevering buiten het Datastation om

In overleg met de deposant kan de datamanager het compleet maken van de dataset ook zelf uitvoeren. De deposant levert de nodige bestanden dan buiten het Data station om aan. Bij kleine bestanden kan dit als bijlage per mail, indien het grotere of veel bestanden betreft is SURF Filesender beter geschikt.

Bestandscuratie en Enhanced Curation

- Controleer of formaten zijn geleverd die DANS niet als voorkeursformaat/preferred format heeft aangemerkt in onze [richtlijnen](#). Zo ja, bekijk of wij deze zelf kunnen converteren (*zie aparte protocollen - momenteel te vinden in interne documentatie voor Student Assistenten van het Data Processing Team*) of dat wij contact willen opnemen met de deposant om de vorm van de data te overleggen.
- Controleer bij aanlevering van een groot aantal formaten of deze overzichtelijk zijn gestructureerd. Zo niet, dan kunnen wij besluiten tot het aanbrengen van een (betere) folderstructuur.
- Controleer of - in het geval dat de onderzoeksdata persoonsgegevens bevatten, of lijken te bevatten - de metadata, de licentie en de toegankelijkheid daarbij aansluiten. Neem bij twijfel contact op met de deposant. Neem in het algemeen contact op met de deposant indien het vermoeden ontstaat dat er onbedoeld persoonsgegevens in de dataset zijn opgenomen (metadata en bestanden). Voor dit laatste is geen actieve controle nodig, het gaat om de situatie wanneer dit vermoeden per toeval ontstaat.
- Controleer of alle aangeleverde data relevant is: geen conceptversies, tijdelijke werkbestanden, systeembestanden, ...
- Controleer of er geen ZIP- of andere compressiebestanden zijn aangeleverd. ZIPs willen we bij voorkeur niet in het archief, enkel als uitleverformaat indien dit de beste wijze voor disseminatie is. In dat geval willen wij de uitgepakte data óók bij de dataset plaatsen.
- Controleer of bestandsnamen kloppen en geen persoonsgegevens bevatten.

Als het nodig wordt geacht om conversies van bestanden naar voorkeursformaten te verzorgen, mogen de nieuwe formaten *niet* toegevoegd worden aan de gedeponeerde versie van de dataset. **In dit geval moet de dataset éérs gepubliceerd worden, vervolgens maken wij een nieuwe versie aan.**

Het toevoegen van conversies van bestanden in voorkeursformaten of het aanpassen van de structurering van de dataset heet **Enhanced Curation**.

Geef in de inventarisatie aan of Enhanced Curation nodig wordt geacht.

Herstructurering van de dataset kan door de bestanden te selecteren op het Files tabblad en vervolgens bij Edit Files => Metadata het 'File Path' te bewerken. Als de dataset zoveel bestanden bevat dat het niet efficiënt werkt om dit via de interface te doen, kan de dataset gedownload worden; lokaal in folders worden gestructureerd; een nieuwe versie van de dataset te beginnen door de huidige data te verwijderen en vervolgens de data in de mappenstructuur opnieuw te uploaden door de gestructureerde dataset in een ZIP-bestand te verpakken en deze ZIP te uploaden. Dit ZIP-bestand wordt automatisch uitgepakt.

Als wij bestanden converteren naar voorkeursformaten downloaden we de originele data lokaal en zetten wij deze data om conform protocollen (*zie aparte protocollen - momenteel te vinden in interne documentatie voor Student Assistenten van het Data Processing Team*). Maak een nieuwe versie van de dataset aan door het originele bestand te verwijderen en het geconverteerde bestand te uploaden.

Let op!!! Bij het opnieuw toevoegen van data moet heel goed worden gekeken naar de beschikbaarstelling van de bestanden in de originele Version 1 dataset. Alle data die 'Restricted' zijn, moeten bij de nieuwe upload óók op Restricted worden gezet. Voor alle data waar een embargo op rust, moet dit embargo ook worden ingesteld voor de nieuwe upload.

Publiceren

Publiceer de dataset middels de knop 'Publish Dataset'.

Als wijzigingen worden aangebracht aan een gepubliceerde dataset, resulteert dit in een nieuwe versie.

In Dataverse bestaan 'minor versies' (bijvoorbeeld versie 1.1) en 'major versies' (bijvoorbeeld versie 2). Een Superuser (zie autorisatiematrix, bijlage 1) heeft de mogelijkheid om wijzigingen aan te brengen en de dataset te overschrijven zonder dat dit tot een nieuwe versie leidt.

Onze standaard:

- Gepubliceerde dataset, wijzigingen in de metadata = Minor revision
- Gepubliceerde dataset, wijzigingen in de bestanden = Major revision
- Aanpassingen aan de licentie of de toegankelijkheid van de data = Major revision. Let er bij dergelijke wijzigingen op of eerdere versies van de dataset niet op een onwenselijke instelling blijven staan, zo ja: dan moet(en) deze eerdere versie(s) *deaccessioned* worden.

Deposanten mogen nieuwe versies indienen waarin de data meer open is dan voorheen. Andersom is dit anders: in principe kan na publicatie geen meer gesloten licentie worden gekozen (je kan immers niet met terugwerkende kracht aanvullende gebruikscondities opleggen) en het minder toegankelijk maken van data gaat in tegen de doelstelling van DANS om onderzoeksdata herbruikbaar te maken. Over dergelijke verzoeken moet eerst overlegd worden.

Let op wanneer een deposant een nieuwe versie indient, of inderdaad een Major versie is aangegeven als het gaat om wijzigingen in de bestanden of in de licentie. Zo niet, dan moet deze nieuwe versie worden afgekeurd.

**Datasets met Restricted files
wanneer Access Request wordt toegestaan**

Omdat DANS deze dataset publiceert komen eventuele Access Requests ook bij DANS binnen. Door de depositor een extra rol met permissie 'ManageDatasetPermissions' te geven, krijgt de klant de Access Request ook en kan vervolgens honoreren of afwijzen.

Dit mogen we uitsluitend doen wanneer het is vastgesteld dat de klant gerechtigd is om de Access Requests te behandelen.

wanneer Access Request wordt toegestaan, maar bepaalde bestanden echt niet gepubliceerd kunnen worden

Zoals hierboven ook is genoemd: Dataverse kent geen combinatie van Restricted files waarbij enkele bestanden wel opvraagbaar zijn en enkele niet. Wanneer de optie 'Enable access request' wordt aangevinkt voor ook maar één Restricted bestand, wordt deze keuze doorgevoerd voor *alle* Restricted bestanden in *alle* versies van de dataset.

Mocht het toch zo zijn dat bepaalde bestanden echt niet opvraagbaar mogen worden (bijvoorbeeld wanneer een origineel bestand persoonsdata bevat en enkel geanonimiseerde afgeleiden van dit bestand opvraagbaar mogen zijn), dan moet de eerste versie van de dataset op Deaccessioned gezet worden.

Edit Dataset => Deaccession Dataset => kies de eerste versie. Standaard reden: 'Legal issue or Data Usage Agreement'.

Accounts en rechten

Het kan voorkomen dat een account rechten op een dataset moet krijgen. Bijvoorbeeld wanneer iemand niet de depositor is, maar wel gerechtigd is om een nieuwe versie van de dataset te verzorgen. Of wanneer een dataset is gemaakt op een verouderd account van de depositor en deze met een nieuw account toegang tot de dataset wil krijgen. Of wanneer een aanvullend persoon toestemmingsaanvragen voor Restricted Data moet gaan beoordelen.

Wij kunnen deze rechten toekennen via Edit dataset=>Permissions=>Dataset, Assign Roles to Users/Groups, een account op te zoeken en deze 'Contributor with ManageFilePermissions' rechten te geven.

Mocht het gaan om een scenario waarbij de dataset wordt 'overgenomen' van de originele depositor door een nieuw account, let dan ook op het account dat genoemd staat bij *Dataset Contact*, pas dit aan wanneer nodig.

Verschillen in beleid tussen Data Stations en EASY

Er zijn een aantal verschillen tussen de contracten en het beleid van de data stations en EASY. De verschillen tussen contracten staan [hier](#) op de website.

Datasets die in EASY gedeponereerd zijn blijven de voorwaardes van EASY volgen, tenzij de depositors het expliciet overdragen naar de terms of use van de Data Stations. Er wordt nog gekeken of dat mogelijk is.

Dataset met persoonsgegevens moest in EASY restricted access, dezelfde voorwaarden blijven gelden dus dit mag niet aangepast worden.

BIJLAGE 1 Autorisatiematrix Data Stations

Rol \ Permissie (op object)	Anonieme bezoeker	Ingelogde bezoeker	Datamanager	Functioneel beheerder	Applicatiebeheerder	ICT-support beheerder
Metadata inzien (gepubliceerde datasets)	X	X	X	X	X	X
Dataset maken en indienen		X	X	X	X	
Dataset publiceren			X	X	X	
Datasetversie overschrijven				X	X	
Dataset deaccesssionen			X	X	X	
Non-restricted files downloaden (gepubliceerde datasets)	X	X	X	X	X	
Restricted files downloaden (alle datasets)		X*	X	X	X	
Toegang verlenen tot restricted files (zelf gedeponeerde datasets)		X				
Toegang verlenen tot restricted files (alle datasets)			X	X	X	
Deaccession (dataset(versie))			X	X	X	
Accountgegevens inzien (alle accounts)				X	X	
Secure Shell toegang met beperkte rechten (data station server)			X		X	X
Secure Shell toegang met onbeperkte rechten (data station server)						X

* Na verlening van toegang door geautoriseerde gebruiker (typisch na een access request, maar kan ook proactief verleend worden).

DANS