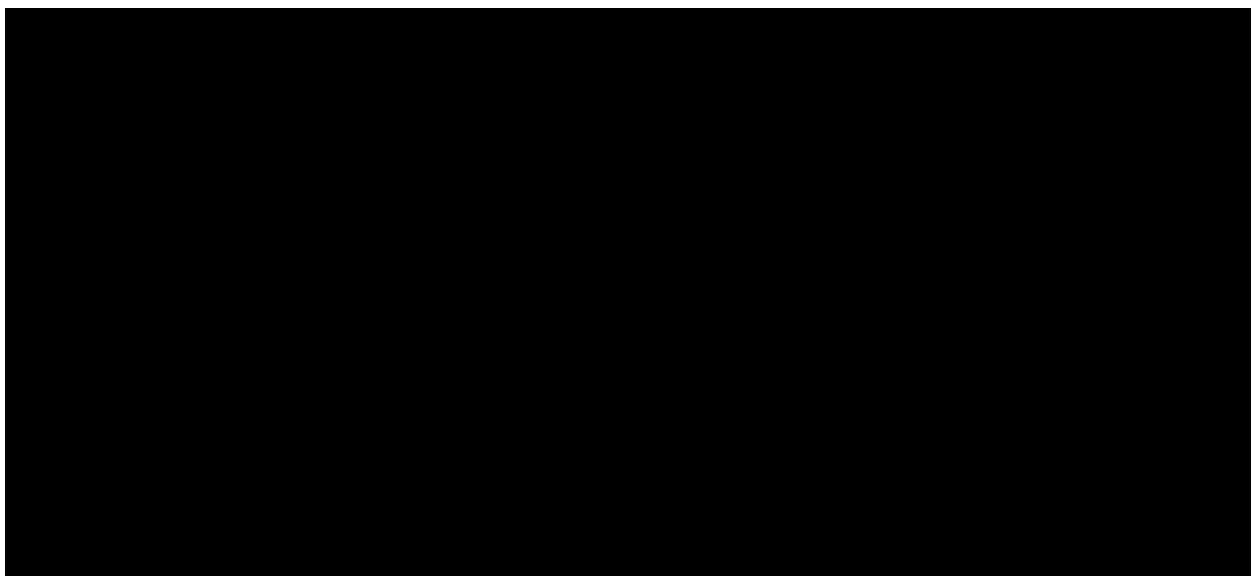




Handleiding data deponeren in EASY

– algemeen

juni 2016





Handleiding data deponeren in EASY - algemeen

U staat op het punt om data te deponeren bij DANS in het online archiveringssysteem EASY. Daarmee zorgt u ervoor dat uw data duurzaam worden gearcheveerd en door anderen gevonden en hergebruikt kunnen worden.

In EASY deponeert u uw data als een dataset. Een dataset bestaat uit de volgende onderdelen:

- Aeen beschrijving in EASY, de metadata. Er wordt automatisch een Persistent Identifier toegekend, een unieke identificatie met een duurzame link die altijd naar de dataset zal verwijzen.
- Ade documentatiebestanden die het onderzoek en de bestanden beschrijven.
- Ade databestanden.

Deze handleiding geeft u handvatten voor alle drie de onderdelen. U krijgt antwoord op vragen als:

- Awelke keuzes heeft u binnen EASY? Hoe vult u de schermen in?
- Awelke documentatie heeft een hergebruiker nodig om uw dataset te begrijpen?
- Awelke gegevens kan DANS wel en niet opslaan?

Er zijn ook enkele extra aanwijzingen voor de disciplines:

- AMaatschappij- en gedragswetenschappen
- AHistorische Wetenschappen
- ALevenswetenschappen en geneeskunde
- ATaal- en Literatuurwetenschappen

Er zijn afzonderlijke handleidingen voor Archeologie en Oral History. Alle handleidingen staan op de DANS-website met de toelichting op data deponeren.¹

Uw data worden gearcheveerd volgens de richtlijnen van het internationaal erkende Data Seal of Approval² het ICSU-WDS³ en het NESTOR seal for trustworthy digital archives⁴. De gebruikte documentatievelden (metadata) zijn in overeenstemming met de richtlijnen van de Dublin Core standaard.⁵

¹ <http://dans.knaw.nl/nl/deponeren/toelichting-data-deponeren>

² <http://www.datasealofapproval.org>

³ <https://www.icsu-wds.org/>

⁴ http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.html

⁵ <http://dublincore.org>



1: Data voorbereiden

Voor de daadwerkelijke deponering is het belangrijk dat data goed omschreven zijn.

Data- en documentatiebestanden

Bij de voorbereiding van het deponeren moet u die data- en documentatiebestanden van uw onderzoek selecteren die zinvol zijn voor mogelijk hergebruik.

Als u een groot aantal bestanden deponeert, ontvangen wij daar graag een bestandenlijst van. Deze lijst bevat een opsomming van de bestandsnamen, evenals een beschrijving van de inhoud van en de mogelijke samenhang tussen de afzonderlijke bestanden.

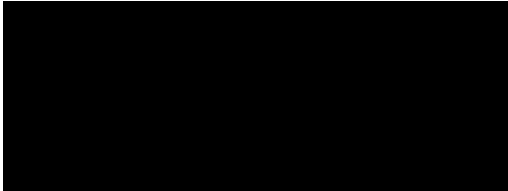
Wanneer andere onderzoekers uw data gebruiken, is het noodzakelijk dat er informatie is over de onderzoeksopzet, de manier waarop de data zijn verzameld en de codering van de data, d.w.z. een verklaring van de variabelen, afkortingen en terminologie binnen het onderzoek. In veel gevallen zal een codebook en een publicatie over het onderzoek voldoende zijn. Zorg ervoor dat uw dataset deze informatie bevat.

Voor data uit **Historische Wetenschappen** is het belangrijk dat u een beschrijving van de (archivalische) bronnen, de gebruikte selectieprocedure en de manier waarop de bronnen zijn gebruikt, aan de dataset toevoegt. Vermeld ook of de informatie gestandaardiseerd is, en zo ja, op welke manier. Indien u gebruik hebt gemaakt van bepaalde standaarden of classificatiesystemen zoals HISCO, vermeld dit dan ook in de beschrijving.

Voor **Maatschappij- en gedragswetenschappen** geldt daarnaast: De databestanden moeten volledige en duidelijke variabele- en valuelabels bevatten.

Essentiële documentatiebestanden zijn:

- A de vragenlijst(en) of andere onderzoeksinstrumenten
- A de veldwerkrapportage (indien beschikbaar)
- A een codebook, of een beschrijving van de variabelen en informatie over:
 - oA de populatie
 - oA soort data (units of observation / analysis)
 - oA de steekproef en steekproefprocedure
 - oA de (non)response
 - oA de methode van data verzamelen
 - oA wegingsvariabelen
 - oA geconstrueerde en/of afgeleide variabelen
 - oA informatie over anonimisering
- A publicaties gebaseerd op de data (indien beschikbaar) of bibliografische beschrijvingen hiervan



Meer informatie vindt u ook in de publicatie 'Preparing data for sharing'.⁶

Bij Taal- en Literatuurwetenschappen: CLARIN-metadata

Heeft u subsidie van CLARIN ontvangen voor uw onderzoek? In dat geval moet uw dataset bij voorkeur één, of indien noodzakelijk meerdere, metadatabestanden bevatten in het zogenaamde Component MetaData Infrastructure (CMDI) formaat (een XML-formaat). Ook wanneer u subsidie heeft ontvangen die vereist dat u uw resultaten CLARIN-conform aanlevert, dienen CMDI-metadata deel van de dataset uit te maken. Voor meer informatie over CMDI zie: <http://www.clarin.eu/cmdi>

Om EASY in staat te stellen uw CMDI-metadata beschikbaar te stellen, bijvoorbeeld aan externe zoekmachines, moet u in EASY de passende optie selecteren. Meer informatie staat onder 4: Documentatie en toegangsniveau, Scherm 1: Required elements - verplichte velden.

Anonimisering

Voordat u de bestanden uploadt, is het nodig om na te gaan of ze privacygevoelige informatie bevatten zoals omschreven in de Wet Bescherming Persoonsgegevens. Als u de data beschikbaar wilt stellen, dienen de bestanden en de beschrijving volledig geanonimiseerd te worden.

Voor het deponeren van interviewdata gelden specifieke voorschriften. Uitgebreide informatie leest u in de desbetreffende instructies en in H. Tjalsma, Rapport authenticiteit en juridische aspecten kerncollectie Getuigen Verhalen.⁷

Daarnaast biedt DANS de mogelijkheid om het oorspronkelijke, niet-anonieme bestand op een voor buitenstaanders niet toegankelijke server te bewaren. Als u hiervan gebruik wilt maken, neem dan contact op met DANS. Hou er rekening mee dat de metadata in EASY altijd openbaar is en ook gebruikt wordt door zoekmachines.

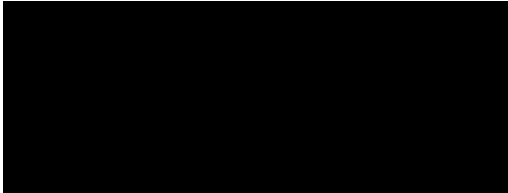
Voor Maatschappij- en gedragswetenschappen geldt daarnaast: Geanonimiseerde databestanden mogen geen variabelen meer bevatten die kunnen leiden tot de identificatie van individuele respondenten. Voorbeelden hiervan zijn:

- A naam van respondent
- A adres van respondent
- A telefoonnummer van respondent
- A burgerservicenummer of sociaalfiscaal nummer (BSN / SoFinummer) van respondent

Onderstaande variabelen moeten worden gehercodeerd:

⁶ <http://www.dans.knaw.nl/nl/over/organisatie-beleid/publicaties/DANSpreparingdataforsharing.pdf>

⁷ <http://dans.knaw.nl/nl/deponeren/toelichting-data-deponeren>

- 
- A geboortedatum (moet worden gehercodeerd naar geboortjaar)
 - A postcode (moet worden gehercodeerd naar alleen de vier cijfers van de postcode)

Ook exacte namen van beroepen dient u te hercoderen naar een beroepenclassificatie. U kunt hiervoor de Standaard Beroepenclassificatie (SBC)⁸ van het Centraal Bureau voor de Statistiek gebruiken of de International Standard Classification of Occupations (ISCO)⁹ van de International Labour Organization (ILO).

Bestandsformaten

DANS hanteert een lijst met voorkeursformaten (preferred formats)¹⁰ voor de te deponeren bestanden. Van data die in een preferred format worden aangeleverd, kan DANS garanderen dat deze langdurig bewaard kunnen worden en toegankelijk blijven. Omwille van de duurzaamheid ontvangen we bij voorkeur data in een voorkeursformaat. Indien u hiervan wilt afwijken, neem voor deponering contact op met een datamanager van DANS.

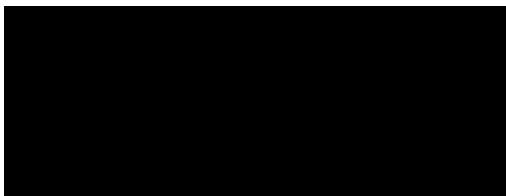
Grootte van bestanden

Bestanden tot 100 MB kunt u eenvoudig uploaden in EASY. Als uw bestanden groter zijn dan 100 MB kunt u overwegen om het bestand te zippen tot een .zipbestand. Hierover staat meer informatie bij [5: Databestanden uploaden](#). Bij twijfel kunt u het beste eerst overleggen met een datamanager van DANS.

⁸ <http://www.cbs.nl/nl-NL/menu/methoden/classificaties/overzicht/sbc/2010/default.htm>

⁹ <http://www.ilo.org/public/english/bureau/stat/isco/index.htm>

¹⁰ <http://www.dans.knaw.nl/nl/deponeren/toelichting-data-deponeren>

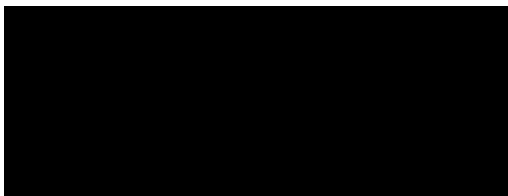


2: Naar EASY

Ga naar <https://easy.dans.knaw.nl> en log in. Nieuwe gebruikers maken eerst een account aan via *Register*.

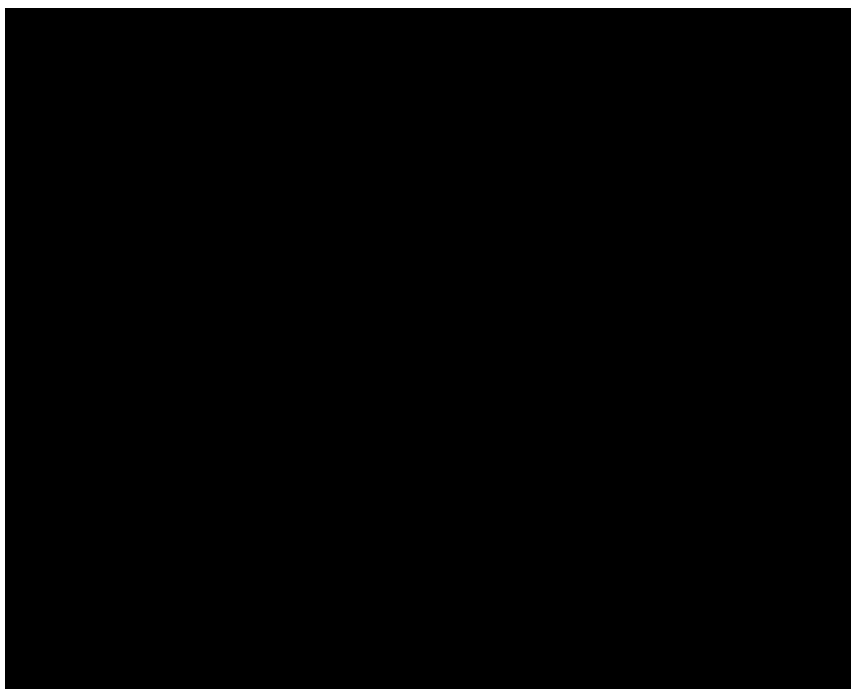
EASY is Engelstalig om data ook voor onderzoekers uit het buitenland te kunnen ontsluiten. U heeft echter de keuze om uw data in het Nederlands of het Engels te beschrijven. Zie ook het kopje **Taal** onder [4: Documentatie en toegangsniveau](#).

Afbeelding 1: Website van EASY



3: Start het deponeren

Ga in EASY naar *DEPOSIT YOUR DATA*. Vervolgens selecteert u de juiste discipline en klikt u op de knop *Start deposit*.



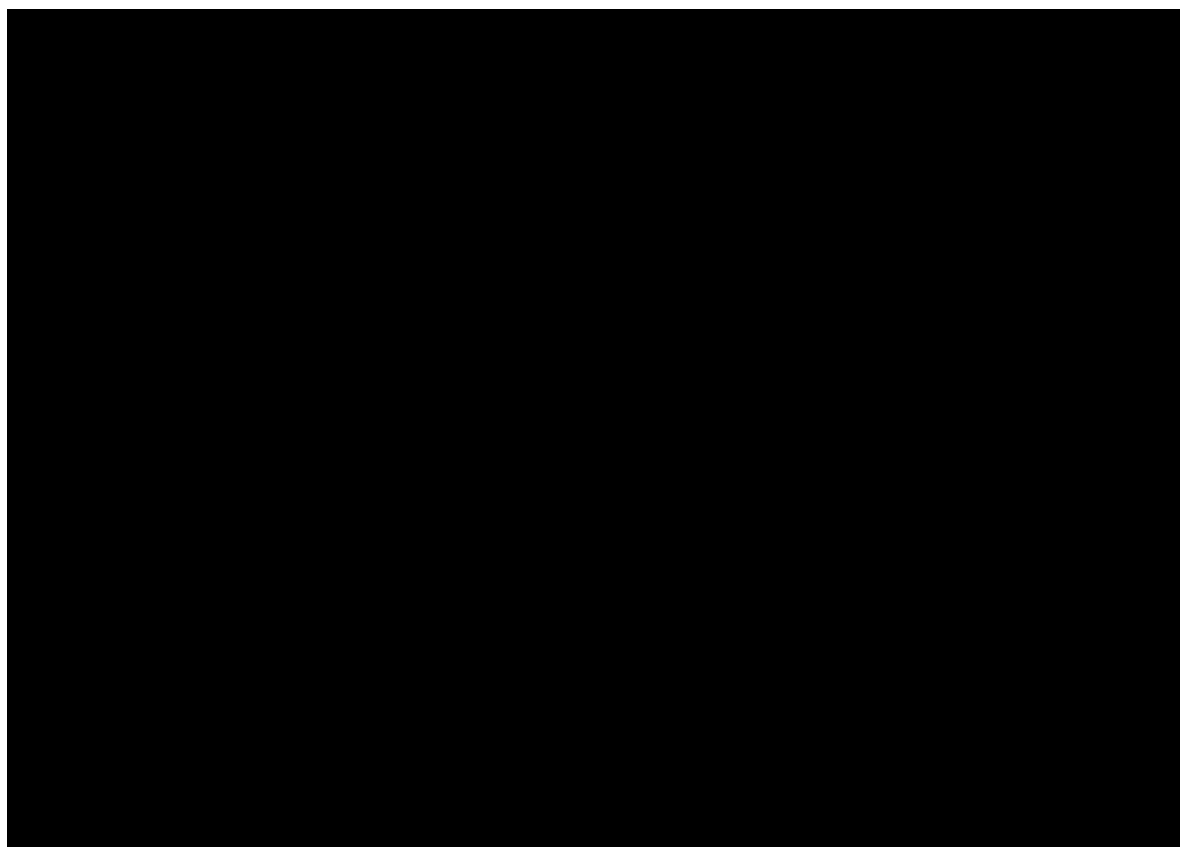
Afbeelding 2: Kiezen van de discipline



4: Documentatie en toegangsniveau

Nadat u op *Start deposit* heeft geklikt, ziet u de volgende vijf schermen:

1. Required elements – verplichte velden
2. Upload files - uploaden bestanden
3. Recommended elements – aanbevolen velden
4. Additional elements – aanvullende velden
5. Overview and submitting – overzicht en data versturen



Afbeelding 3: Metadatavelden– sommige verplicht, sommige optioneel (*optional*)

Binnen de vijf schermen zijn acht velden verplicht om in te vullen. De niet-verplichte velden herkent u aan de aanduiding (*optional*) achter de veldnaam, zoals *Date created (optional)* hierboven. Echter, hoe meer velden u invult, hoe gemakkelijker het voor anderen is om uw data te vinden, te gebruiken en te citeren.

Taal

U bent in EASY vrij in uw taalkeuze, maar het advies is om de taal te gebruiken waarin de gegevens zijn vastgelegd. Is dit een andere taal dan Nederlands of Engels, bijvoorbeeld Duits of Frans, voeg dan naast de originele beschrijving ook een samenvatting en trefwoorden in het



Nederlands of Engels toe. Voor data uit de **Maatschappij- en gedragswetenschappen** is ook de taal van de 'variable labels' en 'value labels' belangrijk: zijn deze in het Engels, beschrijf de dataset dan ook in het Engels. Neem bij twijfel contact op met DANS.

Tussentijds opslaan

Onderaan elk scherm vindt u de volgende knoppen:



: Sla tussentijdse wijzigingen op



of



: Ga naar het vorige of volgende scherm

Na iedere stap in de procedure wordt uw input opgeslagen. Als u uw input tussentijds wilt opslaan, klikt u op *Save draft*. U kunt nu uitloggen bij EASY, op een later tijdstip weer inloggen en verder gaan waar u gebleven was. De eerder ingevoerde metadataset vindt u na inloggen als 'draft' onder het kopje *My datasets*.

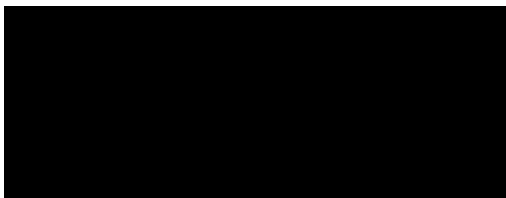
Op het laatste scherm, **5: Overview and submitting**, vindt u de knop



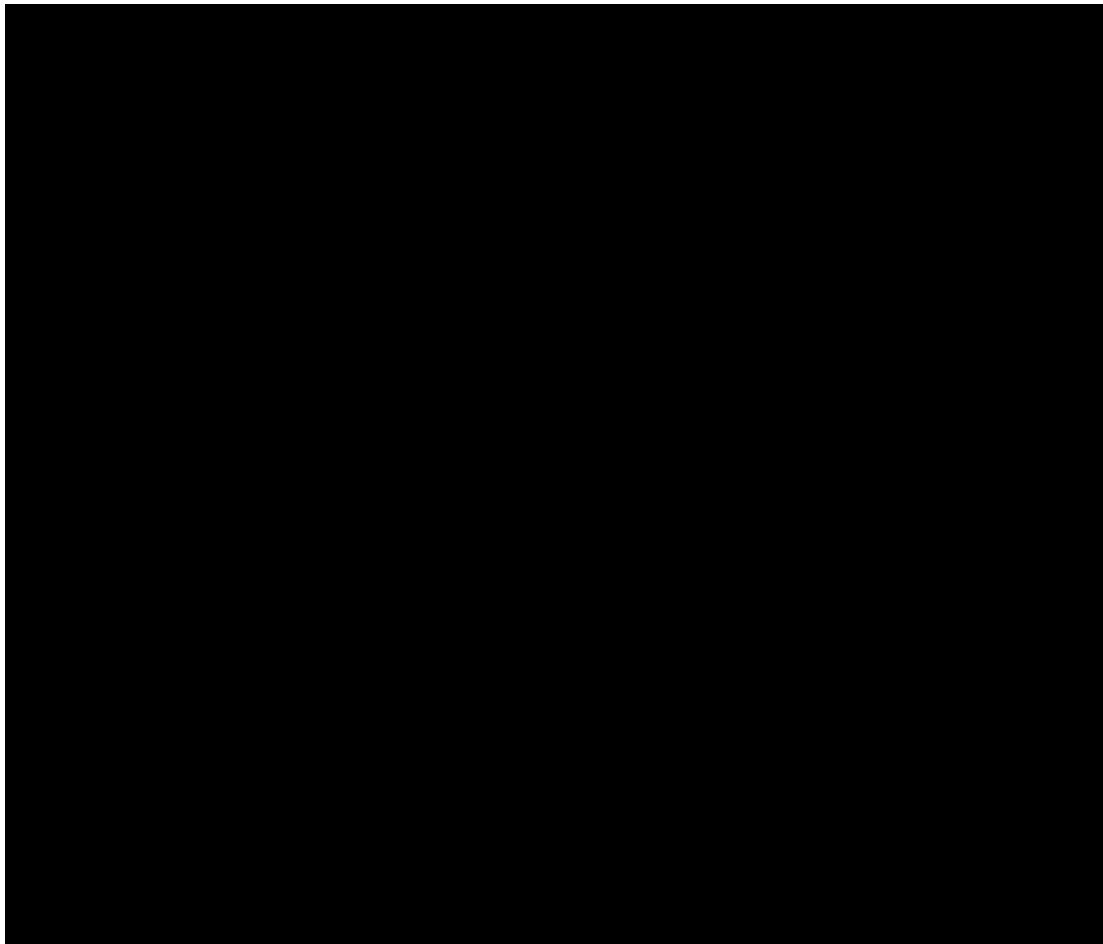
om de dataset in te dienen.



Wanneer u op de informatieknop klikt, verschijnt er een help-scherm met aanvullende informatie over het desbetreffende metadataveld. Door op het plusteken te klikken, kunt u een extra veld voor hetzelfde item aanmaken.



Scherf 1: Required elements - verplichte velden



Afbeelding 4: Scherm 1 – Required elements met default Open access for registered users

De meeste velden op scherm 1 zijn verplicht.

Access rights

De toegang tot uw data regelt u in het veld *Access rights*. In dit veld kunt u aangeven onder welke voorwaarden uw data toegankelijk worden.

Standaard is aangevinkt:

Open access – CCO Waiver (direct toegankelijk voor iedereen zonder in te hoeven loggen in EASY volgens de Creative Commons Zero Waiver licentie.¹¹ Een van de voorwaarden van de CCO-licentie is dat u afstand doet van al uw rechten op de data).

¹¹ <http://creativecommons.org/publicdomain/zero/1.0>




Andere opties zijn:

Open access for registered users (direct toegankelijk voor ingelogde gebruikers van EASY conform de licentieovereenkomst van DANS).¹²

Restricted access: request permission (pas na uw toestemming mag een gebruiker de data inzien en downloaden conform de licentieovereenkomst van DANS).¹³

Other access (de data zijn niet via EASY, maar via een andere repository toegankelijk). De categorie Other access is enkel te gebruiken wanneer u een contract heeft afgesloten met DANS.

Meer informatie over bovengenoemde categorieën vindt u onder de informatieknop  op het scherm. Indien u dat wenst, kan een datamanager binnen 1 dataset enkele bestanden op aanvraag zetten, en andere bestanden op Open Access. Vermeld in het veld *Remarks* dan welke bestanden op welke manier beschikbaar zijn, dan past de datamanager dat voor u aan voor hij / zij de dataset publiceert. De metadata van uw dataset (na publicatie te vinden onder het tabblad *Description* in EASY) zijn beschikbaar als open access; hier rusten geen rechten op.

Date available

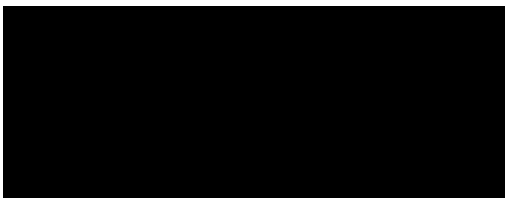
In het veld *Date available* kunt u aangeven of u een tijdelijk embargo op de data wilt instellen van maximaal twee jaar. Zodra het embargo verlopen is, wordt de dataset volgens de eerder door u gekozen toegangscategorie in *Access rights* opengesteld. Meer informatie vindt u onder de informatieknop op het scherm. Wilt u het embargo verlengen, neem dan contact op met DANS.

Publisher

De deponeeringang *Life sciences and medicine* voor de [Levenswetenschappen en geneeskunde](#) biedt het veld *Publisher*. Voor data van de Wageningen UR is afgesproken dat hier 'Wageningen UR' wordt ingevuld.

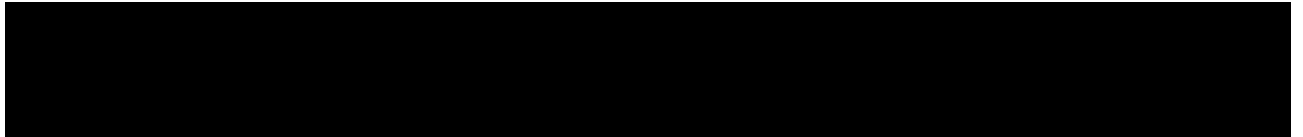
¹² <http://www.dans.knaw.nl/nl/over/organisatie-beleid/juridische-informatie>

¹³ <http://www.dans.knaw.nl/nl/over/organisatie-beleid/juridische-informatie>



Extra metadata

Voor **Taal- en literatuurwetenschappen** is een extra veld beschikbaar, het veld *Extra metadata*. Indien uw onderzoek gefinancierd is door de CLARIN Research Infrastructure, dient u ook de CMDI-metadata als XML-bestand te uploaden. Vink in dit veld alvast aan dat uw dataset CLARIN metadata bevat, en upload het CMDI-bestand samen met uw databestanden via **scherm 2: Upload files**.

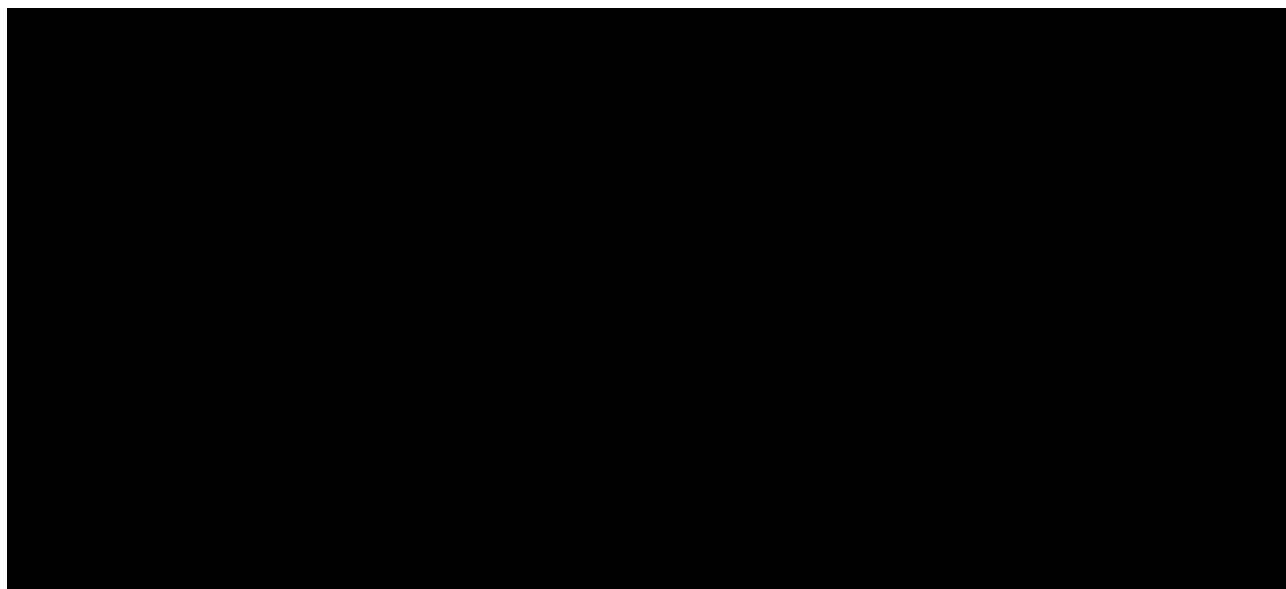


Afbeelding 5: CLARIN metadata, ja of nee



5: Bestanden uploaden

Via **scherm 2: Upload files** kunt u uw bestanden uploaden, zowel de data- als de documentatiebestanden.



Afbeelding 6: Scherm 2 - Upload files

Gebruik de *Browse*-knop achter het metadataveld *Upload dataset* om uw bestand te zoeken, selecteer het en klik op *Upload* om de dataoverdracht te starten.

ZIP-bestanden in EASY

Bij een omvangrijke dataset adviseren we om eerst alle bestanden samen te bundelen tot één ZIP-bestand. Bij het uploaden van het ZIP-bestand zal EASY dit gecomprimeerde bestand zelfstandig weer uitpakken en alle oorspronkelijke directories en bestanden in één keer aanmaken.

Enkel ZIP-bestanden eindigend op .zip worden door EASY ondersteund en / of uitgepakt, andere ZIP-formaten zoals 7zip (.7z) niet. Omdat onzeker is of ZIP-bestanden op de lange termijn nog geopend kunnen worden, accepteren wij geen dubbelgezipte bestanden zonder meer.

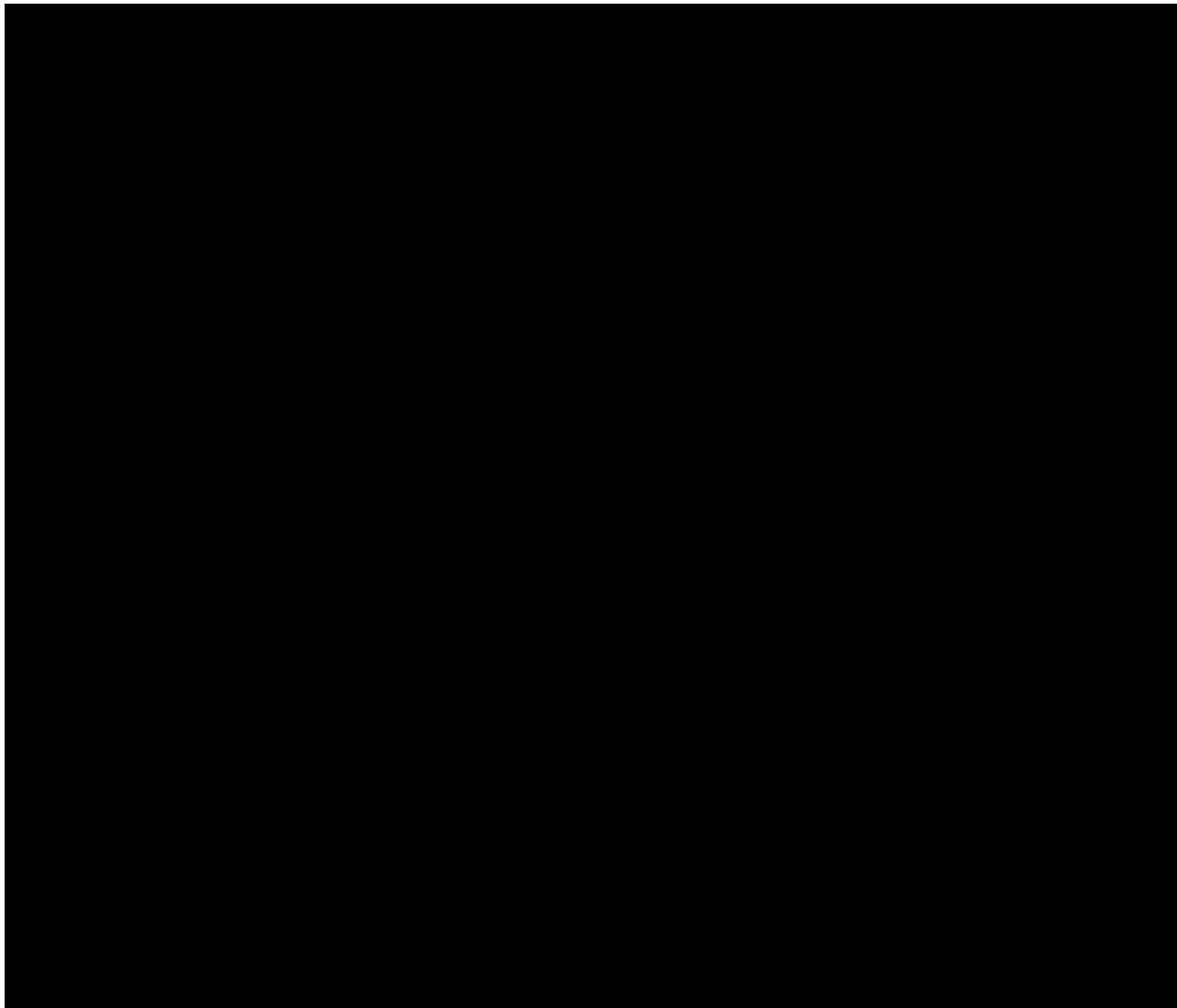
Bij zeer omvangrijke datasets (groter dan 100 MB) vragen we u om wel alle beschrijvende gegevens van een onderzoeksproject in te vullen, maar het uploaden achterwege te laten. Vermeld in het *Remarks*-veld dat de bestanden te groot zijn. De datamanager die uw dataset bekijkt, zal contact met u opnemen voor een alternatieve aanlevering van de bestanden.

“ ~



Documentatie en toegangsniveau - vervolg

Scherf 3: Recommended elements - aanbevolen velden

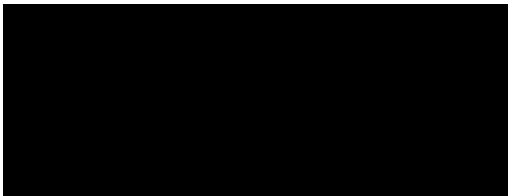


Afbeelding 7: Scherm 3 – Recommended elements

Scherf 3 bevat een aantal niet-verplichte metadatavelden. We raden u echter aan om het veld *Subject* en, indien van toepassing, ook de velden *Spatial coverage* en *Temporal coverage* in te vullen. In het veld *Subject* noemt u trefwoorden waardoor de dataset makkelijker gevonden kan worden.

Spatial coverage

In *Spatial coverage* noemt u het gebied waarover de dataset gaat, bijvoorbeeld 'The Netherlands' of 'Groningen'.



Naast het tekstveld *Spatial coverage* biedt de deponeeringang *Life sciences and medicine* voor de [Levenswetenschappen en geneeskunde](#) de velden *Spatial point* of *Spatial box*. Indien u uw onderzoek vindbaar wilt maken op geografische coördinaten, kunt u die in deze velden invullen.

Temporal coverage

In *Temporal coverage* noemt u de periode waarover het onderzoek gaat.

Identifier

EASY wijst automatisch een Persistent Identifier toe aan uw dataset zodra de deponeerprocedure is afgerond. Als uw dataset al een of meer identifiers van een andere organisatie heeft, noem deze hier dan, inclusief de desbetreffende organisatie, bijvoorbeeld:

WODC Projectnummer: 1433

Synthesgra projectnummer S130087

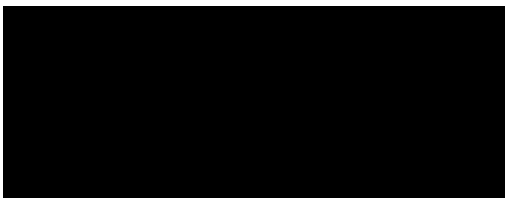
Als u de dataset deponeert als onderdeel van een NWO-DANS datacontract of een ARVODI-overeenkomst, vul dan bij Identifier het NWO dossiernummer of het ARVODI-nummer in, bijvoorbeeld:

NWO-Middelgroot 480.06.001

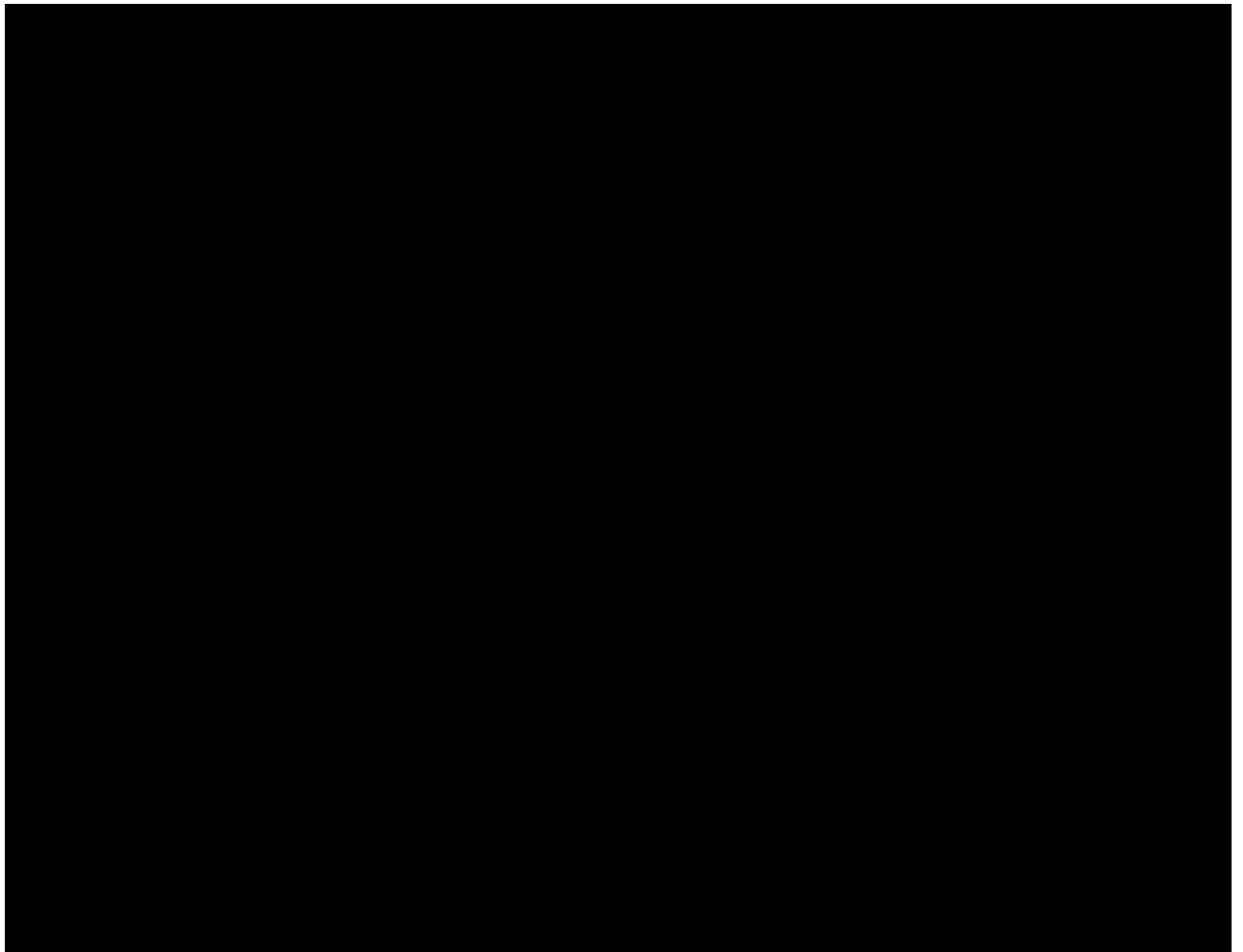
NWO-Veni 016.135.161

NWO-Veni 016.115.26

ARVODI - Projectnummer 2828281



Scherf 4: Additional elements - aanvullende velden



Afbeelding 8: Scherm 4 – Additional elements

Scherf 4 bevat een aantal niet-verplichte metadatavelden.

Relation

In het veld *Relation* kunt u verwijzen naar een gerelateerde dataset, publicatie of tijdschriftartikel, bij voorkeur met vermelding van de DOI (Digital Object Identifier). Is het artikel nog niet gepubliceerd, meld dan in het veld *Remarks* waar en wanneer het artikel naar verwachting zal verschijnen.

U kunt in dit veld ook refereren aan een gerelateerde website. Vul bij *Title* de naam van de website in en bij *URL* het volledige webadres, inclusief <http://>.

Door *Emphasize this relation* aan te vinken, maakt u de *Relation* zichtbaar op de eerste pagina van de dataset nadat deze gepubliceerd is, het tabblad *Overview*.



Remarks

Het metadataveld *Remarks* (Opmerkingen) kunt u gebruiken om aanvullende informatie over de dataset te geven die u elders niet kwijt kunt.

U kunt dit veld ook gebruiken als een communicatiescherm om advies te geven over het gebruik van de data door toekomstige gebruikers, of om een verzoek te doen aan de datamanager bij DANS. Dit is vooral van toepassing wanneer u een dataset deponeert die uiteindelijk zowel Open access als Restricted access: request permission bestanden zal bevatten. In het veld *Remarks* geeft u dan aan welke bestanden volgens welke toegangscategorie beschikbaar gesteld moeten worden.

Open Access tijdschriften

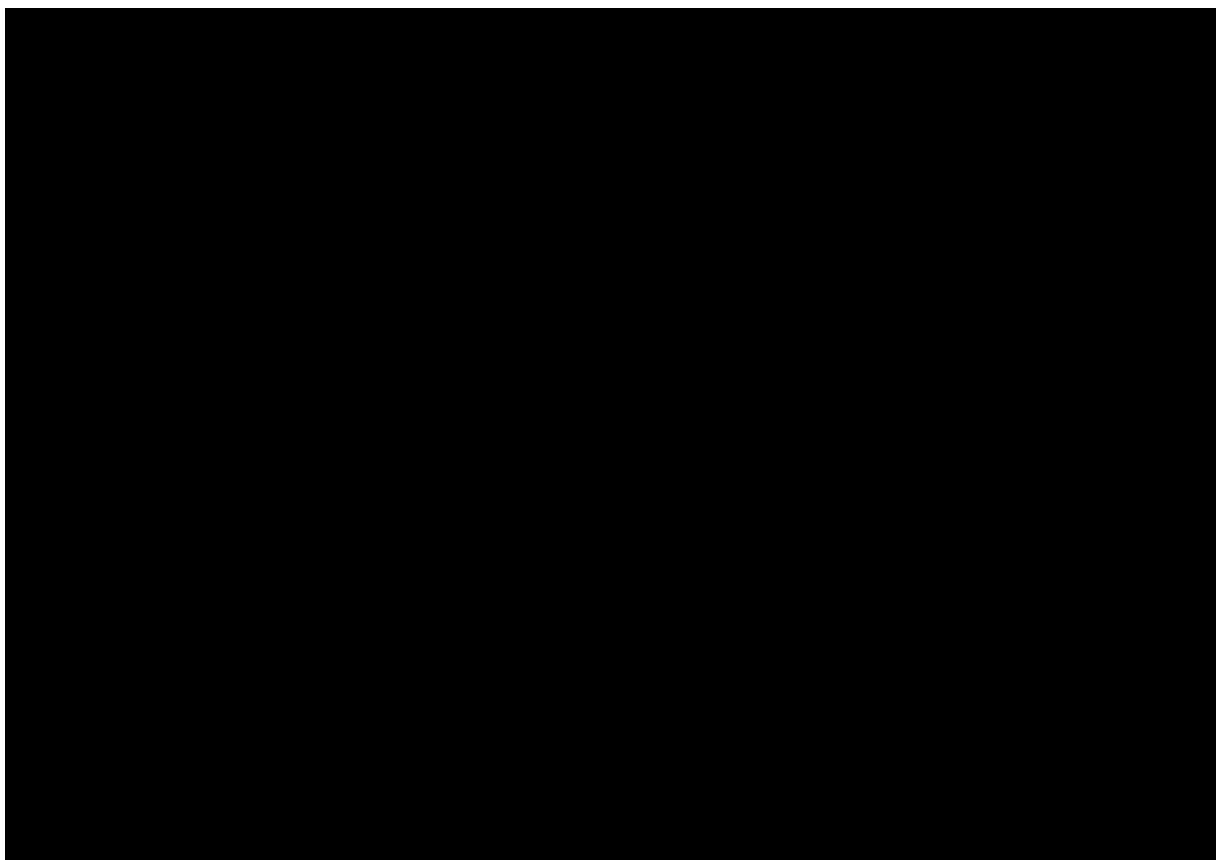
Heeft u een artikel geschreven voor een Open Access-tijdschrift zoals *PlosOne*, maar moet u eerst de data deponeren voor het artikel wordt gepubliceerd?

U kunt de data en het documentatiebestand alvast in EASY deponeren, maar vermeld er dan wel bij voor welk tijdschrift het is en wanneer het artikel ongeveer zal verschijnen. Het documentatiebestand bevat o.a. een beschrijving van de structuur van de dataset met een beschrijving van evt. variabelen en de samenhang tussen de bestanden. Ten slotte ontvangen we achteraf graag de DOI van het artikel. Wij kunnen deze dan toevoegen aan de beschrijving van de dataset, zodat niet alleen vanuit het artikel naar de data wordt verwezen, maar ook vanuit de dataset naar het artikel. Op deze manier vergroten we ook de aandacht voor het artikel.

Betreft het een artikel wat niet openbaar toegankelijk is (niet online beschikbaar en/of geen Persistent Identifier hebbend en/of niet gratis), voeg dan een preprint van het artikel en/of een codebook toe aan de dataset voor een goed begrip van de data.



6: Bestanden versturen



Afbeelding 9: Scherm 5 – Overview and submitting

Op **Scherf 5: Overview and submitting - overzicht en versturen** ziet u een overzicht van alle metadata die u heeft ingevuld. Controleer de metadata een laatste keer en geef aan akkoord te zijn met de DANS Licence Agreement door het vakje aan te vinken. Daarna kunt u op de knop *Submit* klikken en de deponering afronden. De dataset wordt nu naar DANS gestuurd voor controle en publicatie.

Nadat u op *Submit* hebt geklikt, kunt u de dataset niet meer zelf wijzigen of updaten. Als u toch wijzigingen wilt (laten) doorvoeren, neemt u dan contact op met een datamanager van DANS.

Als het u niet lukt om de data te 'submittten', kan het zijn dat een verplicht veld niet is ingevuld. EASY geeft dan aan om welk veld het gaat. Zodra u dit alsnog heeft ingevuld, klikt u nogmaals op *Submit*.

Zodra de dataset is ingediend, ontvangt u een e-mail met de licentieovereenkomst en de Persistent Identifier van de dataset. Deze Persistent Identifier wordt actief na publicatie van de dataset door DANS.



7: Publicatie door DANS

Nadat u de dataset gedeponereerd heeft, verwerkt een datamanager van DANS deze volgens een vastgesteld protocol.¹⁴ Hij / zij kan u vragen om de dataset aan te passen. Als aan alle voorwaarden is voldaan, maakt hij of zij de dataset publiek toegankelijk volgens het toegangsniveau dat u heeft aangegeven en een eventueel embargo.

Uw dataset is nu duurzaam gearchiveerd en kan door anderen gevonden en hergebruikt worden.

Meer informatie

Veel gezochte informatie over het deponeren van data kunt u vinden op de DANS-website.¹⁵

Bij vragen en problemen kunt u contact opnemen met een datamanager van DANS via [info at dans.knaw.nl](mailto:info@dans.knaw.nl) of 070 34 94 450.

¹⁴ <http://dans.knaw.nl/nl/deponeren/toelichting-data-deponeren>

¹⁵ <http://www.dans.knaw.nl/nl/deponeren/toelichting-data-deponeren/praktische-informatie-data-deponeren>