

Duurzame data-archivering bij Research Data Netherlands

Vertegenwoordigt u een onderzoeksinstelling of onderzoeksfinancier, of bent u zelf onderzoeker? En gunt u onderzoeksdata een langer leven, zodat we meer vragen uit samenleving en wetenschap kunnen beantwoorden? Dan zijn de duurzame dataservices van Research Data Netherlands mogelijk relevant. Research Data Netherlands biedt langetermijnopslag in Trustworthy Digital Repositories (TDR) en zorgt dat uw onderzoeksdata vindbaar blijft via Persistent Identifiers. Zo kunt u bestaande data hergebruiken en uw eigen data veilig en langdurig archiveren.

USE CASE 1: voor instellingen, front-offices, research support office

Een universiteit wil de onderzoeksdata van haar onderzoekers die nu in een lokaal datamanagementsysteem is opgeslagen, preserven voor de lange termijn. Via een koppeling van het institutionele systeem naar een langetermijn-archief worden data en metadata rechtstreeks overgezet, zodat de onderzoeker de data niet nogmaals hoeft te deponeren. De instelling is verantwoordelijk voor de naleving van de deponer-eisen van het duurzame archief (zoals preferred formats, metadata). Vervolgens kan de universiteit de Persistent Identifier die het duurzame archief toekent, in het institutionele systeem opnemen.

Zie "Data archiveren bij de RDNL-partners" en "Betrouwbaarheid bij de RDNL-partners".

USE CASE 2: voor onderzoekers die hun gegevens goed willen archiveren

Een onderzoeker zoekt veilige opslag voor zijn data om deze beschikbaar te maken voor hergebruik en verificatie. Bovendien wil de onderzoeksfinancier in de eindrapportage een Persistent Identifier van de data hebben, om eraan te kunnen refereren.

De onderzoeker informeert eerst bij zijn front-office (research support office) of er binnen de eigen instelling dan wel het vakgebied afspraken zijn met duurzamearchieven.

Kan uw front-office u niet verder helpen, zie dan "Data archiveren bij de RDNL-partners" en "Betrouwbaarheid bij de RDNL-partners".

USE CASE 3: voor onderzoekers die geschikte data zoeken

Een onderzoeker wil haar bevindingen vergelijken met die van anderen. Binnen haar eigen vakgroep kan ze voortbouwen op onderzoek van collega's, maar hoe zit het daarbuiten? Via het onderzoeksportal NARCIS.nl komt ze terecht bij een data-archief dat data op haar vakgebied heeft. Een deel ervan is direct open toegankelijk; voor andere data krijgt ze op verzoek toestemming van de oorspronkelijke onderzoekers om de data te gebruiken, op voorwaarde dat ze haar bewerking van de data eveneens duurzaam toegankelijk maakt.

Zie: "Bestaande data bij de RDNL-partners gebruiken".



Langetermijnopslag en -toegang

Als instelling, onderzoeksgemeenschap of onderzoeker kunt u (selecties van) uw onderzoeksdata voor de lange termijn archiveren in een Trustworthy Digital Repository (TDR). De RDNL-partners bieden deze gestandaardiseerde en goed gedocumenteerde opslag, zodat de vindbaarheid en de herbruikbaarheid van de data zijn gewaarborgd. Financiers eisen steeds vaker dat u uw onderzoeksdata in een TDR deponert.

1. Bestaande data bij RDNL-partners gebruiken

<p>Open Access, Restricted Access, embargo: Open Science, Open Access en Open Data worden steeds belangrijker. RDNL ondersteunt dit vanzelfsprekend, en erkent tegelijkertijd dat sommige data niet met iedereen gedeeld kunnen worden; hiervoor bestaan bijvoorbeeld een tijdelijk embargo en Restricted Access.</p>	<p>3TU.Datacentrum-archief: Alle data zijn Open Access (mandatory). Daarnaast is een embargo met een maximum van twee jaar mogelijk.</p>	<p>DANS EASY: Er zijn verschillende opties beschikbaar, waaronder Open Access CC0 Waiver (zie ook http://www.dans.knaw.nl/nl/actueel/nieuws/dans-maakt-open-access-mogelijk). Daarnaast is een embargo met een maximum van twee jaar mogelijk.</p>	<p>SURFsara TDR: Alle data zijn standaard Open Access, maar de toegang kan beperkt worden tot specifieke gebruikers of groepen. Embargo's zijn door gebruikers zelf in te stellen.</p>
<p>Licenties voor datagebruik: de onderzoeker of instelling kiest bij het deponeren van een dataset voor Open of Restricted Access e.d. Deze keuze bepaalt wat een gebruiker met de dataset mag doen, en een gebruikslicentie maakt dit expliciet.</p>	<p>3TU.Datacentrum-archief: Zie http://datacentrum.3tu.nl/onderzoek-publiceren/data-uploaden/</p>	<p>DANS EASY: Zie http://www.dans.knaw.nl/nl/over/organisatie-beleid/juridische-informatie/juridische-informatie</p>	<p>SURFsara TDR: Licentie wordt bij het deponeren van de data gekozen door de gebruiker.</p>
<p>Voornaamste vakgebieden: archieven verschillen in de vakgebieden waarop zij zich (primair) richten. Dit hoeft overigens geen belemmering te zijn om 'disciplinevreemde' data te deponeren, bijvoorbeeld als het een bestandsformaat betreft waarmee het betreffende archief veel ervaring heeft.</p>	<p>3TU.Datacentrum-archief: Exacte en technische wetenschappen</p>	<p>DANS EASY: Sociale wetenschappen en geesteswetenschappen</p>	<p>SURFsara TDR: Generiek</p>

<p>Datagebruik in andere systemen: voor onderzoekers die data uit een archief willen analyseren of bewerken in een applicatie, is het praktisch als het archief met die applicatie te koppelen valt.</p>	<p>3TU.Datacentrum-archief: Voor bepaalde datatypen (bijvoorbeeld NetCDF) bieden we toegang via het OpenDAP protocol waarmee directe toegang vanuit programmeertalen (waaronder Python) mogelijk is.</p>	<p>DANS EASY: EASY biedt geen rechtstreekse koppelingen aan naar specifieke applicaties. Wel wordt waar mogelijk doorverwezen naar externe partijen die deze functionaliteit bieden (bijvoorbeeld CentERdata voor panel-en surveydata).</p>	<p>SURFsara TDR: Voor de toegang tot de data worden verschillende transportprotocollen ondersteund die geschikt zijn voor toepassingen onder zowel Linux-, OSX- als Windows-omgevingen. Daarnaast is er uitstekende toegang naar andere SURFsara-faciliteiten waaronder HPC Cloud, Supercomputer en Grid.</p>
---	--	---	---

2. Data archiveren bij de RDNL-partners

<p>Licenties bij datadeponering: wie een dataset deponeert (archiveert), sluit een licentie met een archief, waarin bijvoorbeeld staat wat het archief met de data mag en moet doen, en welke rechten de deponerder op de data houdt.</p>	<p>3TU.Datacentrum-archief: zie http://datacentrum.3tu.nl/onderzoek-publiceren/data-uploaden/</p>	<p>DANS EASY: zie http://www.dans.knaw.nl/nl/over/organisatie-beleid/juridische-informatie/juridische-informatie</p>	<p>SURFsara TDR: De deponerlicentie bij deponeren van dataset dient geaccepteerd te worden. De deponerder blijft eigenaar van de data.</p>
<p>Data bij een archief deponeren (ingest):archieven bieden de onderzoeker of instelling verschillende manieren om data te deponeren, bijvoorbeeld afhankelijk van de omvang van de data of op basis van een deponeringsovereenkomst. Voor 'deponeren' wordt ook de term 'ingest' gebruikt.</p>	<p>3TU.Datacentrum-archief:</p> <ul style="list-style-type: none"> • Webtoegang voor individuele deponerders • Bulk (via data engineer) 	<p>DANS EASY:</p> <ul style="list-style-type: none"> • Webtoegang voor individuele deponerders • API voor instellingen of communities (SWORD interface, op contractbasis) • Bulk (op contractbasis) 	<p>SURFsara TDR:</p> <ul style="list-style-type: none"> • Webtoegang voor individuele deponerders • API voor instellingen of communities (op contractbasis) • Bulk (data ingest service, op contractbasis) • Accessing and placing data: via internet, using SSH, (HPN)SCP, SFTP, rsync, GridFTP; via iRODS federations
<p>Bestandsformaten voor archivering (preferred formats): om de ambitie van langdurige toegang tot data waar te kunnen maken, concentreren duurzame archieven hun diensten en</p>	<p>3TU.Datacentrum-archief: Zie http://datacentrum.3tu.nl/onderzoek-publiceren/databeschrijving-en-formaten/</p>	<p>DANS EASY: Zie http://dans.knaw.nl/nl/deponeren/toelichting-data-deponeren/DANSpreferredformatsNL.</p>	<p>SURFsara TDR: Geen voorkeur voor data- of bestandsformaat. Bij gebruik van bepaalde servicecomponenten kunnen hier wel restricties op zitten.</p>

<p>expertise vaak op een - brede - selectie van bestandsformaten: de zogenoemde voorkeursformaten. Deponerders wordt sterk aangeraden alle data in een voorkeursformaat te deponeren.</p>		<p>pdf</p>	
<p>Gebruikte standaard voor de beschrijvende metadata: informatie waarmee de deponerder een dataset beschrijft, helpt om de dataset te vinden en om een eerste indruk te krijgen van de bruikbaarheid voor vervolgonderzoek. Archieven hanteren vaak een standaardschema voor zulke beschrijvende metadata, zo ook de RDNL-archieven.</p>	<p>3TU.Datacentrum-archief: Dublin Core aangevuld met specifieke metadata-elementen. Exportformaten: oai-ore resource maps (rdf), datacite xml.</p>	<p>DANS EASY: Dublin Core aangevuld met specifieke metadata-elementen. Zie ook: https://easy.dans.knaw.nl/schemas</p>	<p>SURFsara TDR: Dublin Core</p>
<p>Toegekende Persistent Identifier(s): in een duurzaam archief krijgt elke dataset een unieke, duurzame identifier, zoals een boek een ISBN krijgt. De Persistent Identifier (PID) is nodig om ook in de verre toekomst datasets te vinden en te citeren. Er bestaan verschillende 'smaken' PID's.</p>	<p>3TU.Datacentrum-archief: DOI (Digital Object Identifier)</p>	<p>DANS EASY: DOI (Digital Object Identifier), URN:NBN</p>	<p>SURFsara TDR: EPIC</p>
<p>Metadata in zoekportals: archieven stellen metadata en de Persistent Identifier niet alleen beschikbaar in hun eigen catalogus, maar ook via zoekmachines en portals. Voor deze uitwisseling gebruiken ze de OAI-PMH-standaard. Onderzoekers en instellingen hoeven hier niets voor te doen (maar kunnen ook niet kiezen welke zoekmachines e.d. de metadata komen 'harvesten').</p>	<p>3TU.Datacentrum-archief: Ja, OAI-PMH</p>	<p>DANS EASY: Ja, OAI-PMH</p>	<p>SURFsara TDR: Ja, OAI-PMH</p>

3. Betrouwbaarheid bij de RDNL-partners

<p>Bewaartermijn van de data: de langetermijnarchieven van de RDNL-partners houden de data en metadata ook toegankelijk na de minimale bewaartermijn van tien jaar (Nederlandse Gedragscode voor Wetenschapsbeoefening).</p>	<p>3TU.Datacentrum-archief: Long-term preservation met garantie voor minimaal vijftien jaar</p>	<p>DANS EASY: Long-term preservation</p>	<p>SURFsara TDR: Long-term preservation</p>
<p>Toegangscontrole, autorisatie: bij de deponering is vastgelegd wie toegang tot (welk deel van) de data hebben, bijvoorbeeld individuen of bepaalde groepen. Het archief regelt dat alleen geautoriseerde personen en/of groepen toegang krijgen. Toegang hebben tot data is iets anders dan gebruik mogen maken van de data; gebruik wordt geregeld via licenties. De beschrijvende metadata zijn altijd voor iedereen zichtbaar.</p>	<p>3TU.Datacentrum-archief: Datasets die zijn geüpload via het webformulier zijn toegankelijk na gebruikersregistratie (standaardmethode). Datasets die zijn gearchiveerd met tussenkomst van een data librarian (vanwege speciaal bestandsformaat of bestandsgrootte) zijn toegankelijk zonder gebruikersregistratie.</p>	<p>DANS EASY: De datasets met een CC0 Waiver licentie zijn zonder registratie toegankelijk voor iedereen. Voor toegang tot de overige datasets geldt ten minste een gebruikersregistratie in EASY.</p>	<p>SURFsara TDR: Standaard zijn datasets toegankelijk voor iedereen; de gebruiker kan dit instellen bij deponering of naderhand. Bij downloaden is registratie vereist.</p>
<p>Preservation strategy: duurzame archieven hanteren verschillende technieken om de data toegankelijk te houden. De bits en bytes moeten intact blijven, harde schijven en tapes moeten worden vervangen, en soms moet een bestandsformaat geconverteerd worden naar een nieuw formaat (file format migration).</p>	<p>3TU.Datacentrum-archief: file format migration, zie http://datacentrum.3tu.nl/onderzoek-publiceren/lange-termijn-opslag/</p>	<p>DANS EASY: file format migration, zie http://dans.knaw.nl/nl/deponeren/toelichting-data-deponeren/DANSpreservationpolicyUK.pdf (Engels)</p>	<p>SURFsara TDR: Bit preservation; curatie mogelijk in een later ontwikkelingsstadium van de dienst.</p>
<p>Certificering: Er bestaan verschillende certificeringen voor 'Trustworthy Digital Repositories'. De duurzame archieven van RDNL hebben minimaal het Data Seal of Approval. Ook bestaan er bijvoorbeeld certificeringen voor informatiebeveiliging.</p>	<p>3TU.Datacentrum-archief: Data Seal of Approval (DSA)</p>	<p>DANS EASY: Data Seal of Approval (DSA), World Data System (WDS-ICSU)</p>	

4. Contact en meer informatie over de RDNL-partners

<p>Voor de kosten die mogelijk aan een dienst verbonden zijn, hanteren de partners elk een eigen model.</p> <p>We helpen u graag als u meer wilt weten over een van onze diensten. Op de websites vindt u veel informatie, maar neem ook gerust contact met ons op.</p>	<p>Informatie over het 3TU.Datacentrum-archief: http://datacentrum.3tu.nl/onderzoek-publiceren/data-uploaden/</p> <p>Archief https://data.3tu.nl/</p> <p>Contact via datacentrum@3tu.nl</p>	<p>Informatie over DANS EASY: http://dans.knaw.nl/nl/deponeren/toelichting-data-deponeren/praktische-informatie-data-deponeren</p> <p>Archief http://easy.dans.knaw.nl/</p> <p>Contact via info@dans.knaw.nl</p>	<p>Informatie over SURFsara TDR: Contact via info@surfsara.nl</p>
---	--	--	---