

Richtlijnen Datamanagement Instituut voor Geschiedenis

Introductie

Het hier voorliggende document is de uitwerking van de Regeling Datamanagement Universiteit Leiden voor onderzoek dat plaatsvindt binnen het Leidse Instituut voor Geschiedenis. Het is bedoeld voor onderzoekers die binnen het Instituut voor Geschiedenis aan een onderzoeksproject beginnen waarvoor mogelijk een datamanagementplan wordt opgesteld en/of bij externe financiers een voorstel gaan indienen waarbij datamanagement mogelijk een rol speelt. Dit document is bindend voor projecten die beginnen na de datum waarop een definitieve versie van deze uitwerking is vastgesteld door het MT. In gevallen waarin deze uitwerking niet voorziet is de Wetenschappelijk Directeur van het Instituut voor Geschiedenis beslissingsbevoegd.

Regeling Datamanagement Universiteit Leiden (RDUL)

Het CvB heeft per 19 april 2016 een 'Regeling Datamanagement Universiteit Leiden' vastgesteld waarin de kaders geschetst worden voor universitair datamanagementbeleid. De Universiteit heeft de Faculteiten opgedragen om vast te stellen of de nadere uitwerking van de Regeling Datamanagement op facultair niveau plaatsvindt of op het niveau van de instituten. De faculteit der Geesteswetenschappen heeft besloten dat het Datamanagementbeleid binnen de faculteit op instituutniveau wordt opgestart.

De RDUL definieert 'data' zoals NWO dat doet: data is zowel verzamelde, onbewerkte data, alsook door analyse gegenereerde data, en het gaat daarbij zowel om digitale als om niet-digitale onderzoeksgegevens. Het CvB stelt dat op het niveau van de discipline vastgesteld moet worden welke data moeten worden opgeslagen.

De RDUL bepaalt dat **ieder onderzoek** aan de universiteit dat data verzamelt of genereert, moet werken met een van tevoren opgesteld **datamanagementplan**, waarin het datamanagementbeleid van de betreffende faculteit of het betreffende instituut wordt uitgewerkt voor het onderzoeksproject in kwestie. Dit datamanagementplan moet garanderen dat:

- Onderzoeksdata gedurende het onderzoek veilig worden opgeslagen, waarbij de integriteit, beschikbaarheid en vertrouwelijkheid van de data gegarandeerd blijven.
- Onderzoeksdata na afloop van het onderzoek toegankelijk, begrijpelijk, herbruikbaar en duurzaam opgeslagen kunnen worden, in combinatie met metadata, documentatie en voor hergebruik benodigde software.
- Dataopslag plaatsvindt volgens internationaal geldende richtlijnen, liefst digitaal (bijv. via een Trusted Digital Repository) en minstens voor tien jaar.

Het CvB is verantwoordelijk voor het bieden van adequate voorzieningen en ondersteuning om verantwoord databeheer mogelijk te maken. De WD van het instituut is verantwoordelijk voor de uitvoering van het beleid; op universitair niveau is er een data protection officer die toeziet op de naleving van het beleid omtrent de veiligheid van onderzoeksgegevens.

De RDUL stelt een aantal kaders aan de uitwerking van de universitaire regeling op decentraal niveau. De uitwerking moet aansluiten bij de normen die gangbaar zijn in de discipline, en daarnaast vastleggen:

- Hoe de verantwoordelijkheden m.b.t. datamanagement binnen het instituut zijn belegd;
- Wat de relevante data zijn om op te slaan, gezien het type onderzoek binnen het instituut.
- Beleid t.a.v. data voortkomend uit BA, MA en RMA-onderwijs (en scripties)
- Hoe tijdens het onderzoek integriteit, beschikbaarheid en vertrouwelijkheid van data gegarandeerd blijven
- Hoe na het onderzoek de data vindbaar, toegankelijk, begrijpelijk en herbruikbaar duurzaam worden opgeslagen.
- De maximale bewaartermijn van onderzoeksgegevens en DMP's.

Datamanagement binnen het Instituut voor Geschiedenis

Het onderzoek binnen het instituut kenmerkt zich door een grote diversiteit en, deels, in een snelle ontwikkeling in de manieren waarop met gegevens wordt gewerkt. Het meeste onderzoek in het instituut is in grote mate individueel van aard, maar binnen projecten wordt ook in toenemende mate in teamverband met onderzoeksgegevens gewerkt.

Niet al het onderzoek binnen het IvG valt onder de RDUL. Twee typen onderzoek zijn geheel of gedeeltelijk vrijgesteld van de in de RDUL opgelegde bepalingen:

1. Voor een deel van de onderzoekers zal datamanagement geen rol in de onderzoekspraktijk besteden, omdat zij data verzamelen noch genereren. Dit geldt specifiek voor historisch onderzoek op basis van secundaire literatuur, en voor kwalitatief onderzoek op basis van uitgegeven teksten of archiefstukken die bibliografisch traceerbaar en publiekelijk toegankelijk zijn. Voor dit type onderzoek is een DMP onnodig.
2. Daarnaast is er onderzoek waarbij de data als appendices gepubliceerd worden. Het instituut stelt dat onderzoeksgegevens die als appendix bij een wetenschappelijke publicatie openbaar zijn gemaakt alleen *tijdens* het onderzoek vallen onder de RDUL. Het is onnodig om hen ook na afloop nog apart als gegevens op te slaan. Voor dit type onderzoek is echter wel een DMP nodig.

Voor een deel van het onderzoek geldt echter dat er wel degelijk data worden verzameld of gegenereerd die in principe volledig vallen onder de RDUL. Dit betreft in ieder geval:

3. onderzoeksprojecten waarin gedurende het project meerdere onderzoekers tegelijkertijd met dezelfde gegevensbestanden werken.
4. onderzoeksprojecten waarin gegevens worden samengebracht en geanalyseerd in databases: door onderzoek gegenereerde databases vallen per definitie onder de RDUL.
5. onderzoeksprojecten waarin gegevens worden verzameld via interviews, veldwerk of onderzoek in niet-openbare archieven en vastgelegd in aantekeningen, op foto's, via geluidsopnames of op film; hier geldt dat zowel de (ruwe) aantekeningen als de geproduceerde mediabestanden onder de RDUL vallen.
6. onderzoeksprojecten waarin gegevens uit online databases worden verzameld: de resultaten van een zoekopdracht moeten gezien worden als door onderzoek gegenereerde gegevens indien de zoekopdracht niet reproduceerbaar is met (exact) dezelfde versie van de database die voor de oorspronkelijke zoekopdracht gebruikt werd.

Onderzoeksgegevens

Uit bovenstaande volgt dat de onderzoeksgegevens waarop de RDUL betrekking heeft binnen het IvG uiteenvallen in meerdere categorieën. De voornaamste hiervan zijn:

- Databases met unieke gegevensverzamelingen voortkomend uit veldwerk of archiefstudie, of verzameld via onderzoek in één of meerdere online databases.
- Mediabestanden met unieke inhoud voortkomend uit veldwerk, archiefstudie of interviews, zoals afbeeldingen, audiobestanden en video-opnames.
- Aantekeningen van replicerbare observaties voortkomend uit veldwerk of archiefstudie. Het kan daarbij gaan om beschrijvingen van objecten, situaties, en documenten, en om afschriften van niet-openbare archiefstukken.
- Aantekeningen van niet-repliceerbare observaties, met name voortkomend uit interviews. Het kan hier gaan om geaccordeerde of niet-geaccordeerde aantekeningen.

Van deze vier typen gegevens zal tijdens het onderzoek de integriteit, en na het onderzoek de toegankelijkheid en herbruikbaarheid gegarandeerd moeten worden. Voor gegevensbestanden die voortkomen uit interviews of uit studie in besloten archieven dient indien nodig tevens de vertrouwelijkheid gewaarborgd te worden. Ieder datamanagementplan van onderzoek waarin levende personen een rol spelen dient uit te werken hoe zal worden voldaan aan de Algemene Verordening Gegevensbescherming (AVG). Ondersteuning daarbij is beschikbaar op de universiteit (privacysupport@hum.leidenuniv.nl).

Externe samenwerking

Indien er sprake is van samenwerkingsverbanden tussen onderzoekers van het IvG in Leiden met onderzoekers die elders werkzaam zijn, geldt dat:

- Onderzoekers van het IvG dragen, bij uitwisseling van bestaande gegevenssets voor onderzoeksdoeleinden, geen directe verantwoordelijkheid voor de kwaliteit van de opslag van onderzoeksgegevens die elders verzameld worden, maar zijn verplicht schriftelijke informatie op te vragen over de manier waarop gegevens verzameld zijn, en de hoe zij veilig, toegankelijk en herbruikbaar opgeslagen zullen worden.
- Bij projectmatige samenwerkingsverbanden waarbij nieuwe gegevens verzameld zullen worden, geldt de in deze richtlijnen gedefinieerde werkwijze als bindende leidraad, ook als de te verzamelen gegevens gedurende het project niet primair in Leiden worden beheerd.

Onderwijs: gegevens voortkomend uit BA, MA en RMA-cursussen en scripties

Voor zover het onderwijs van de aan het IvG verbonden onderzoekers en docenten werkt met primaire onderzoeksgegevens is het gebonden aan de RDUL en de in deze uitwerking gegeven richtlijnen daar waar het onderwijs tot een publiekelijk zichtbaar resultaat kan leiden. Hierbij geldt:

1. Het is de verantwoordelijkheid van de docent of scriptiebegeleider om erop toe te zien dat, indien nodig, door studenten een DMP opgesteld wordt, en dat onderzoeksgegevens na afloop van het project duurzaam worden opgeslagen.
2. Het is de verantwoordelijkheid van de opleidingen om te bepalen in hoeverre zij onderzoeksprojecten van studenten waarvoor een DMP nodig is de ruimte geven binnen het opleidingstraject.
3. Docenten zijn niet verplicht vakken aan te bieden waarin gewerkt wordt met nieuw te verzamelen onderzoeksdata die onder de RDUL vallen.

In de praktijk zal het bij projecten waarvoor een DMP nodig is in de meeste gevallen gaan om een MA-scriptie, aangezien deze normaliter via het repository openbaar toegankelijk worden gemaakt. In uitzonderlijke gevallen kan het gaan om BA-eindwerkstukken, BA Seminars of MA Research seminars waaruit wetenschappelijke publicaties voortvloeien. Het verdient de aanbeveling om bij het inrichten van onderwijs waarin studenten gehouden zijn een DMP op te zetten rekening te houden met de werkdruk voor de docent en de studeerbaarheid voor studenten.

Procedure Datamanagementplan

Procedureel maakt het Instituut voor Geschiedenis onderscheid tussen twee scenario's:

- Lopend, niet projectgebonden onderzoek dat binnen de reguliere aanstelling plaatsvindt. Hiervoor geldt dat onderzoekers en hun leidinggevendenden op periodieke basis evalueren in hoeverre voor lopend onderzoek een DMP nodig is. Hierbij geldt:
 - Het het IvG hanteert een *'nee, tenzij'*-beleid: veruit het meeste van dit onderzoek valt onder de hierboven omschreven benadering (1).
 - Waar dat niet het geval is wordt, voor specifieke, nieuw te vergaren datasets van substantiële omvang, een DMP opgesteld.
 - Niet-projectgebonden gegevensbestanden hoeven pas publiekelijk gearchiveerd te worden als het onderzoek volledig is afgerond. Hiervoor hoeft in het DMP geen termijn te worden gespecificeerd, zolang voorzien is in een periodieke evaluatie.
- Projectgebonden onderzoek met een beperkte looptijd waarvoor financiering is gekregen bij externe geldschieters. Hiervoor geldt:
 - Bij het indienen van een financieringsaanvraag, wordt vastgesteld of, en op welke gronden, het opstellen van een DMP noodzakelijk is, en welke middelen daartoe benodigd zijn.
 - Na aanvang van het onderzoeksproject wordt binnen een maand een definitief datamanagementplan opgesteld. Dit gebeurt op basis van het universitaire template Data Management Plan.
 - Gedurende het project wordt het datamanagementplan beheerd en indien nodig bijgewerkt door de onderzoeker, maar heeft de WD van het IvG altijd direct toegang tot de meest recente versie van het plan. Bij wijzigingen wordt versienummering toegepast, en oudere versies van het DMP worden gearchiveerd.
 - Na afloop van het project is het IvG verantwoordelijk voor de archivering van het datamanagementplan voor de in het RDUL genoemde termijn van twintig jaar.¹

Datamanagement tijdens het project: integriteit, beschikbaarheid en vertrouwelijkheid

Binnen het onderzoek zoals dat op het IvG gebeurt is er voor datamanagement tijdens het onderzoeksproces een cruciaal onderscheid tussen projecten waarbij meerdere personen aan één en dezelfde dataset werken, en projecten waarbij dat niet (structureel) gebeurt.

- Als meerdere onderzoekers aan dezelfde dataset werken, moet gebruik gemaakt worden van een gedeelde werkomgeving die voor alle medewerkers van het project vanaf hun werkplek toegankelijk is.
- Als dit niet het geval is, volstaat in principe de opslag op de eigen werkplek via het ULCN-account.

¹ Zolang er geen universitair archief voor DMP's is.

Voor een gedeelde werkomgeving wordt in principe gebruik gemaakt van de faciliteiten en servers die de universiteit intern aanbiedt. Er wordt in principe geen gebruik gemaakt van de diensten van commerciële providers. Het Centre for Digital Scholarship is als universitair aanspreekpunt voor datamanagement de eerste aangewezen partner als het gaat om faciliteiten die de standaardwerkplek overstijgen.

Duurzame opslag: metadata, software en documentatie

De RDUL stelt dat onderzoeksdata moeten worden opgeslagen in combinatie met ‘metadata, overige documentatie en eventueel de software(versie) die noodzakelijk zijn voor potentieel hergebruik van de data’.

Metadata

Voor de gegevenstypen waarmee onderzoekers aan het IvG werken geldt standaard dat in een dataset minimaal als meta-informatie wordt aangegeven:

- De datum waarop de gegevens gecreëerd zijn
- De datum waarop de gegevens voor het laatst bewerkt zijn
- In het geval van mediabestanden: plaats waar de opname is gemaakt en naam van de maker.
- In geval van mediabestanden: de omvang (in pixels) en/of de duur.
- In het geval van interviews: naam van de geïnterviewde (evt. te anonimiseren).
- Eventuele toegangsrestricties op de individuele gegevens

Bestandsformaten en software

Met het oog op eventueel hergebruik (RDUL, artikel 7) ligt het voor de hand dat bestanden worden gearchiveerd in het bestandsformaat waarin ze binnen het project gebruikt zijn. Voor de gegevenstypen waarmee onderzoekers aan het IvG werken geldt dat het hierbij veelal gaat om veel voorkomende bestandsformaten waarvan binnen de in de RDUL gegeven termijn de leesbaarheid niet in het geding zal zijn.

Het DMP specificeert de bestandsformaten die een rol spelen binnen het project. Voor bestandsformaten die uitgelezen kunnen worden met goed gearchiveerde open-source software, of met standaard softwarepakketten als Microsoft Office en Adobe Creative Suite geldt dat geen nadere toelichting noodzakelijk is. Voor bestandstypen die niet met deze software uitgelezen kunnen worden geldt dat in het DMP aangegeven moet worden hoe de toegankelijkheid gegarandeerd is, of dat de benodigde software mee gearchiveerd moet worden.

Overige documentatie

Behoudens de data, en eventueel software, vereist de RDUL dat aanvullende documentatie moet worden opgeslagen. Het IvG verstaat hieronder documentatie die betrekking heeft op de herkomst, onderzoekscontext en gebruiksgeschiedenis van de dataset als geheel. Het gaat daarbij in ieder geval om:

- Titel en looptijd van het onderzoeksproject
- Naam (en contactgegevens?) van de PI
- Namen van onderzoekers die meegewerkt hebben aan de totstandkoming van de dataset
- Beschrijving van de inhoud en aard van de gegevens, inclusief technische details
- Beschrijving van de herkomst van de gegevens en van de gevolgde verzamelmethode
- Bibliografische details van de publicaties die gebaseerd zijn op de dataset.

Duurzame opslag na afloop van het project

Na afloop van het project moeten onderzoeksgegevens worden opgeslagen in een Trusted Digital Repository. Het IvG hanteert, met NWO en ERC, de norm dat een Trusted Digital Repository voorzien moet zijn van een 'Data Seal of Approval' of een Nestor Seal of moet voldoen aan de RAC-standaarden (ISO 16363). Het Centre for Digital Scholarship beheert als universitair aanspreekpunt voor datamanagement het overzicht van repositoria die aan deze standaarden voldoen.

Het IvG ziet erop toe dat de voorzieningen voor verantwoord databeheer die het CvB in het kader van RDUL art. 15 ter beschikking stelt aan of toegankelijk maakt voor onderzoekers voldoen aan de uiteenlopende behoeften van de onderzoekers die binnen het instituut actief zijn, met name met het oog op onderzoeksgegevens die vanuit de eerste geldstroom worden gegenereerd.