

Financiële implicaties research datamanagement

8 september 2016

Werkgroep Financiële Aspecten RDM: Jacques Beursgens (Maastricht University), Magchiel Bijsterbosch (SURFsara) vz., Jan-Willem Boiten (Data4lifesciences), Jan-Willem Brock (Universiteit Leiden), Ingrid Dillo (DANS), Marlon Domingus (Erasmus Universiteit Rotterdam), Christina Elsenga (Rijksuniversiteit Groningen), Michelle van der Kooij (Technische Universiteit Delft), Reinout Raijmakers (Universiteit Utrecht), Marcel Ras (Nationale Coalitie Digitale Duurzaamheid), Ingeborg Verheul (SURFsara/LCRDM)

Inleiding

De financiële aspecten van onderzoeksdata zijn in de Roadmap van het Landelijk Coördinatiepunt Research Data Management (LCRDM) geïdentificeerd als één van de vraagstukken om op landelijk niveau aan te pakken.ⁱ Een werkgroep bestaande uit experts werkzaam binnen universiteiten en andere stakeholders binnen de levenswetenschappen en de lange termijn archivering heeft een analyse gemaakt van de voornaamste knelpunten die een duurzaam financieel kader voor het research datamanagementbeleid in de weg staan.ⁱⁱ Deze notitie bevat de voornaamste bevindingen van de werkgroep en een zestal concrete aanbevelingen ter verbetering van de huidige situatie.

Bevindingen

1. Meer kosten, maar weinig cijfers

Binnen universiteiten en andere onderzoeksinstellingen is sprake van een toenemende behoefte aan expertise en voorzieningen voor de opslag, de verwerking en het beheer van onderzoeksdata. Dit wordt veroorzaakt door de beleidseisen in het kader van verificerbaarheidⁱⁱⁱ en hergebruik van onderzoeksresultaten, en door een toename van reken- en dataintensief onderzoek in steeds meer disciplines en onderzoeksprojecten. De kosten voor expertise en voorzieningen nemen navenant toe. Tegelijkertijd blijven de onderzoeksbudgetten gelijk. Daarmee is er voor de onderzoeker geen stimulans om de datamanagementkosten apart te begroten: dit gaat immers ten koste van bijvoorbeeld personeel of instrumenten. Het gevolg daarvan is niet alleen dat onderzoeksdata op dit moment veelal niet voldoen aan de genoemde eisen van verificerbaarheid en hergebruik, maar ook dat er weinig inzicht is in de totale omvang van de datamanagementkosten binnen de onderzoeksinstellingen. Schattingen van de initiële kosten op basis van instellingsprogramma's en landelijke voorzieningen worden geraamd op ca. 10 M€ in 2016. De High Level Expert Group van de European Open Science Cloud neemt 5% van het onderzoeksbudget als uitgangspunt voor de dekking van structurele kosten van goed datamanagement. Uitgaande van de bekende cijfers over 2014 voor R&D-uitgaven in het hoger onderwijs, zou dit landelijk neerkomen op 213M€ voor research data management.^{iv}

Beter inzicht in de werkelijke omvang van de kosten, kan ontstaan door naast een inhoudelijke datamanagementparagraaf ook minimale eisen te stellen aan de projectbegroting. Een andere mogelijkheid is om datamanagementkosten op voorhand in subsidieprogramma's te oormerken. Een definitiekader en een benchmark met de belangrijkste kostencomponenten kunnen hierbij ondersteunen. Met de Curation Cost Exchange^v uit afgeronde 4C project en een lopend project binnen het werkprogramma Duurzaam van de *Nationale Strategie Digitaal Erfgoed*^{vi} wordt al gekeken naar de vergelijkbaarheid van kosten voor duurzame toegang. Met het raamwerk *Guide Research Data Management and Costs*^{vii} is een basis gelegd om concrete handvatten te bieden voor onderzoeksbegrotingen. Gezien de diversiteit in het onderzoek behoeft deze algemene aanzet wel nadere uitwerking met inbreng vanuit de verschillende instellingen, de disciplines, en de onderzoekers.

2. Kosten binnen projectperiode buiten de boot

In het datamanagementbeleid van bijvoorbeeld NWO^{viii} en H2020^{ix} worden kosten voor datamanagement subsidiabel gesteld. In de praktijk zijn de kosten van onder andere interne dienstverlening (zoals bijvoorbeeld kosten van centrale IT of juridische expertise) voor de onderzoeker dikwijls niet op te voeren in

zijn/haar projectbegroting. Dit wordt veroorzaakt door een combinatie van de interne boekhouding van universiteiten, subsidievoorwaarden van financiers m.b.t. indirecte kosten binnen de 2^e en 3^e geldstroom, en accountingprincipes. De onderzoeker wordt hiermee verleid tot het opzetten van puntoplossingen die wél subsidiabel zijn, maar die de robuustheid en schaalvoordelen van centrale voorzieningen missen (zie verder ook punt 4). Worden deze diensten wel geleverd, dan resteren de daaruit voortvloeiende kosten als bedrijfsvoeringskosten c.q. overhead voor de instelling en gaan deze - met gelijkblijvende budgetten - uiteindelijk ten koste van onderwijs en onderzoek binnen de 1^e geldstroom. De reeds bestaande *matchingsdruk* van universiteiten^x neemt hiermee verder toe.

Vanuit de gedachte dat met datamanagement ook daadwerkelijk waarde aan het onderzoek wordt toegevoegd, zou datamanagement zoveel mogelijk als direct aan het onderzoek gerelateerde kostenpost geregistreerd moeten worden. Dit maakt dat datamanagementkosten ook binnen de bestaande subsidievoorwaarden subsidiabel te maken zijn. Tegelijkertijd moet daarbij gewaakt worden voor hoge administratieve lasten en registratiekosten. Acceptatie van 'full-cost' voor datamanagement-gerelateerde kosten in subsidievoorwaarden van financiers zou daarbij voor een balans tussen verantwoordingsplicht en registratiekosten voor instellingen kunnen zorgen.

3. Geen aansluiting tussen projectfinanciering en langetermijn-ambities

Het beleid van instellingen (i.c. verifieerbaarheid) en onderzoeksfinanciers (i.c. herbruikbare data) ambiëert bewaartermijnen die het onderzoeksproject ruimschoots overstijgen. Een voorbeeld hiervan is de bewaartermijn van 10 jaar zoals bepaald in De Nederlandse Gedragscode Wetenschapsbeoefening.^{xi} Voor de bekostiging na afloop van de projectperiode is echter geen dekking. Hiermee bestaat een reële kans dat data verdwijnt en dat daarmee niet wordt voldaan aan voorwaarden van instellingen en onderzoeksfinanciers. Dit kan o.a. leiden tot reputatieschade voor zowel de betrokken onderzoekers als de instelling wanneer bij het vermoeden van wetenschapsfraude de beschuldiging niet weerlegd of bewezen kan worden zonder de oorspronkelijke onderzoeksdata.

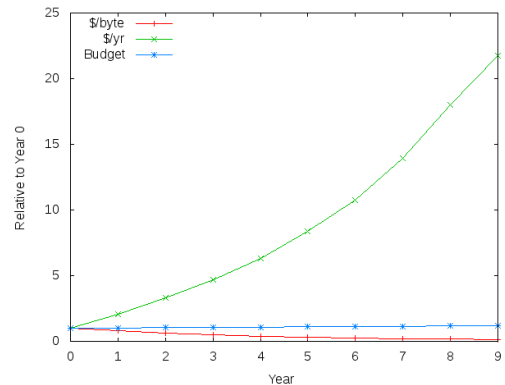
Door eenmalige bedragen voor de benodigde dienstverlening in rekening te brengen, kan aansluiting worden gemaakt met projectfinanciering. Dit kan op twee verschillende momenten in het onderzoekproces gebeuren. In de eerste variant betaalt de *producerende* onderzoeker op het moment van deponeren van de onderzoeksdata in het archief. Dit gebeurt in de vorm van *deposit fees* die ten laste komen van het projectbudget en de bekostiging van duurzame opslag is zo onderdeel van de afronding van het bestaande onderzoek. Als tweede variant is het daarnaast mogelijk om kosten van duurzame opslag als *access fees* in rekening te brengen aan de *consumerende* onderzoeker, op het moment dat deze eerder gedeponeerde data wil hergebruiken. Hoewel *access fees* in beginsel op gespannen voet staan met de *Open Access*-gedachte, kan het in sommige onderzoeksvelden of situaties toch opportuun zijn dit model te hanteren. Dit is bijvoorbeeld het geval wanneer specifieke toegevoegde waarde voor een onderzoeksproject wordt gecreëerd door data (her)bruikbaar te maken. In dit geval vindt de bekostiging plaats als onderdeel van nieuw onderzoek (vergelijkbaar met kosten voor dataverzameling of -generatie). Beide modellen blijven vanuit het oogpunt van duurzaamheid kwetsbaar vanwege onvoorspelbaarheid in de toekomstige vraag naar hergebruik van onderzoeksdata. Datasets die minder worden geraadpleegd of waarvan het op de lange termijn beschikbaar houden naar verhouding kostbaar is, zullen hiermee eroderen. Het complementeren van deze vraaggeoriënteerde bekostigingsmodellen met een aanbodgeoriënteerde financiering van dergelijke voorzieningen blijft noodzakelijk als we ook deze onderzoeksdata toegankelijk willen houden.

Universiteiten kunnen, eventueel in koepelverband, opdrachtgever zijn voor de langetermijn-voorziening waarin onderzoeksdata wordt gearchiveerd en toegankelijk gehouden. Daarnaast zullen met onderzoeksfinanciers afspraken moeten maken over een duurzaam financieringssysteem dat aansluit bij de langetermijn-ambities. In specifieke onderzoeksvelden kan ook naar oplossingen gezocht worden door internationale samenwerking.

4. Mogelijkheden tot kostenbeheersing

Het gezamenlijk organiseren van expertise en voorzieningen kan leiden tot aanzienlijke schaalvoordelen. Dienstverlening is (i.t.t. individuele investeringen) onder de huidige voorwaarden echter vaak niet subsidiabel. Het gezamenlijk inkopen wordt hiermee ontmoedigd en schaalvoordelen worden misgelopen.

Het volume en de complexiteit van data nemen sneller toe dan kostendaling van dataopslag of de besparingen uit schaalvoordelen kunnen compenseren voor de dataopslag sec. Laat staan dat er gecompenseerd kan worden voor het onderhoud ervan voor hergebruik. Niet voor alle data is het opportuun deze duurzaam tot in de eeuwigheid te bewaren en toegankelijk te houden. Kritische selectie aan de poort (“welke data moet echt bewaard worden”) en een actief deselectiebeleid (“moet data na een bepaalde periode nog steeds bewaard blijven”) blijven nodig om de totale omvang van data te beperken en daarmee kosten te beheersen. De te volgen strategie zou zoveel mogelijk bij aanvang van het onderzoek bepaald moeten worden. Dit is vanuit het oogpunt van wetenschappelijk integriteit cruciaal, om de schijn van datamanipulatie door het weglaten van onwelgevallige resultaten tegen te gaan. Daarnaast is vanuit economisch oogpunt selectie vooraf nagenoeg altijd goedkoper dan deselectie achteraf. Daarbij zal er een gedegen beleid moeten zijn waarop eventuele deselectie plaats kan vinden.



Figuur 1: ondanks de afname van de prijs per eenheid nemen de totale kosten voor opslag toe. Bron: Rosenthal, D.S. (2014, 18 november). Talk "Costs: Why Do We Care?"

Gezien de cumulatieve groei van gearchiveerde onderzoeksdata, is het gezamenlijk optrekken om schaalvoordelen te incasseren met name opportuun voor de langetermijn-opslag van onderzoeksdata.

5. Monitoring van beleidseffecten

De investering in research datamanagement zou zijn vruchten moeten afwerpen via hergebruik van bestaande kennis op de lange termijn. Blijft deze (maatschappelijke) impact echter uit, dan veroorzaakt (naast het belang van verifieerbaarheid en wetenschappelijke integriteit), datamanagement vanuit bedrijfseconomisch perspectief slechts een stijging van de overheadkosten voor het wetenschappelijk bedrijf die ten koste gaat van onderzoek (waaronder bijvoorbeeld promovendi). Tegelijkertijd zijn de baten van research datamanagement moeilijk meetbaar en niet enkel in financiële zin uit te drukken.

Zonder feiten en cijfers over de kosten en baten van research datamanagement is een objectieve basis voor bestuurlijke dialoog tussen universiteiten, financiers en rijksoverheid echter nauwelijks mogelijk. Een structurele monitoring van beleidseffecten kan de gewenste onderbouwing geven. Op basis van de waargenomen effecten kan beleid al dan niet worden bijgestuurd, bijvoorbeeld door vanuit de rijksoverheid alsnog te compenseren met aanvullende middelen of beleidseisen voor datamanagement bij te stellen. Om in de toekomst een dergelijke dialoog te kunnen voeren, moet zo snel mogelijk de aanvangssituatie in kaart worden gebracht.

Het Rathenau Instituut verzorgt met ‘Wetenschap in cijfers’ periodieke informatie over de financiering en wetenschappelijke output van Research & Development in Nederland.^{xii} Het Rathenau Instituut beschikt hiermee reeds over de kennis en expertise om het beleid van universiteiten en onderzoeksfinanciers aanvullend te informeren wat betreft investering in en resultaten van research datamanagement. Daarmee is zij in potentie een kandidaat voor de selectie van beleidsindicatoren en het uitvoeren van deze monitoring.

Aanbevelingen

De werkgroep komt aan de hand van bovenstaande bevindingen tot de volgende aanbevelingen die zij de Stuurgroep Onderzoek en Valorisatie ter overweging meegeeft:

1. *Stel minimale eisen aan de projectbegroting op het gebied van research datamanagement*
 - (alle) kosten voor research datamanagement die zijn gerelateerd aan het onderzoek moeten zichtbaar gemaakt worden in projectbegrotingen;
 - start vroegtijdig de discussie (tussen onderzoeker, instelling en betrokken financiers) over de dekking van datamanagementkosten van het onderzoeksproject én binnen de bewaartermijn na de projectperiode;
 - datamanagement wordt een expliciet onderdeel van de planning- & controlcyclus van universiteiten, zodat een basis voor beleidssturing ontstaat.
2. *Registreer datamanagement zoveel mogelijk als directe kosten voor het onderzoek*
 - kosten van datamanagement worden verder traceerbaar en eenvoudiger subsidiabel onder de huidige subsidievoorwaarden. Dit is verdedigbaar gezien de toegevoegde waarde die ondersteuners creëren in het onderzoek.
3. *Stimuleer interne en gezamenlijke inkoop van diensten*
 - binnen universiteiten, tussen universiteiten en in internationaal verband;
 - deel van investeringen en opstartkosten basis van gezamenlijke prioriteiten;
 - zoek naar schaalvoordelen, gezien de cumulatieve groei met name in de langetermijn-opslag van onderzoeksdata.
4. *Maak richtlijnen voor de kritische selectie van onderzoeksdata: welke data moet, voor welke termijn, met welk doel moeten worden bewaard?*
 - beperk de instroom van nieuwe data door selectie van data op basis van vooraf bepaalde criteria en daarmee de uiteindelijke kosten voor het toegankelijk houden;
 - zorg dat deze criteria om redenen van wetenschappelijk integriteit en economische voordelen zo vroeg mogelijk in het onderzoeksproces in overeenstemming met subsidievoorwaarden en beleidseisen worden bepaald;
 - datasets die nauwelijks meer relevant zijn of geraadpleegd worden, kunnen in de levensduur van de data actief (al dan niet automatisch) op basis van de vooraf bepaalde criteria gedeselecteerd worden.
5. *Stem met financiers subsidievoorwaarden af die ruimte bieden aan bovengenoemde maatregelen*
 - zorg voor een juiste balans tussen verantwoordingsplicht en registratiekosten;
 - maak kosten voor research datamanagement, inclusief gezamenlijke dienstverlening, subsidiabel;
 - zorg voor een duurzaam financieringsmodel voor onderzoeksdata na afloop van het project.
6. *Vraag het Rathenau Instituut zorg te dragen voor beleidseffectmeting*
 - door het ontwikkelen van deugdelijke beleidsindicatoren;
 - door het uitvoeren van een efficiënte monitoring.

ⁱ <https://www.lcrdm.nl/nieuws/2015/09/roadmap-rdm-met-focus-op-vijf-vraagstukken-uitgestippeld.html>

ⁱⁱ <https://www.lcrdm.nl/vraagstukken/financi%C3%ABle-aspecten>

ⁱⁱⁱ Koninklijke Nederlandse Akademie van Wetenschappen. (2012). Zorgvuldig en integer omgaan met wetenschappelijke onderzoeksgegevens.

<https://www.knaw.nl/nl/actueel/publicaties/zorgvuldig-en-integer-omgaan-met-wetenschappelijke-onderzoeksgegevens>

^{iv} Rathenau Instituut. (2016, 14 januari). *Financiering en uitvoering van R&D in Nederland*. <https://www.rathenau.nl/nl/page/financiering-en-uitvoering-van-rd-nederland>. 4262M€ totaal R&D voor Hoger Onderwijs.

^v <http://www.curationexchange.org/>

^{vi} <https://www.rijksoverheid.nl/documenten/publicaties/2015/03/09/nationale-strategie-digitaal-erfgoed>

^{vii} Westerhof, A. et al. (2015). *Research Data Management: Storage and Beyond / covering the Full Research Data Life Cycle*.

https://www.edugroepen.nl/sites/RDM_platform/Wiki%20Data%20Management%20Costs/Home.aspx

^{viii} <http://www.nwo.nl/beleid/open+science/datamanagement>

^{ix} http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

^x VSNU. (2014). *Matchingsdruk*. http://www.vsnu.nl/nl_NL/matchingsdruk.html

^{xi} VSNU. (2014). *De Nederlandse Gedragscode Wetenschapsbeoefening*.

http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Code_wetenschapsbeoefening_2004_%282014%29.pdf

^{xii} <https://www.rathenau.nl/nl/wetenschap-in-cijfers>