

Rapport hergebruik onderzoeksdata van Nederlandse universiteiten in kaart

Versie 31 januari 2017

De getallen waarop resultaten in dit rapport gebaseerd zijn, zijn te vinden in het zip-bestand 'UKB_RD_reusereport2016DATA.zip'.

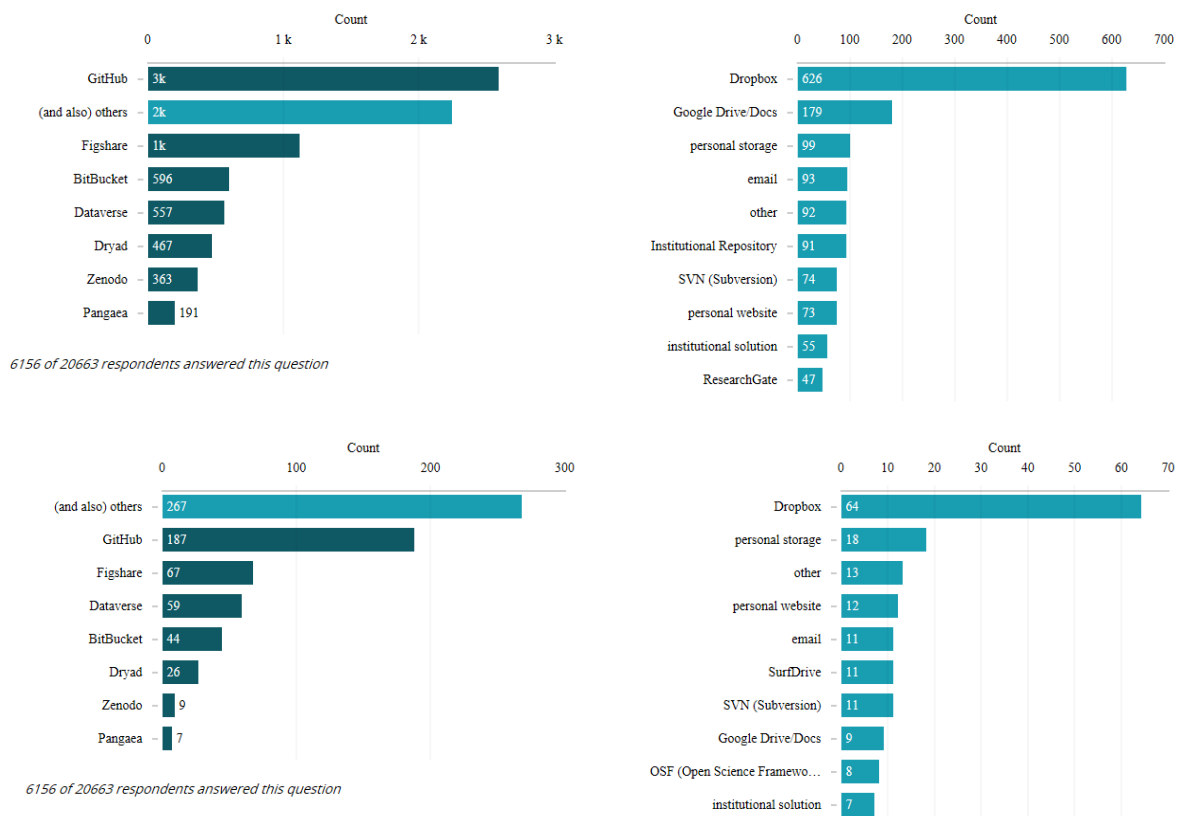
Introductie

Deze inventarisatie werd uitgevoerd door de UKB werkgroep Research Data voor het jaarplan 2016. Tessa Pronk (UU), Ellen Fest (WUR), Leon Osinski (TU/e), Petra Ploeg (TiU), Paul Plaatsman (EUR). Met dank aan: Christina Elsenga (RUG).

Hergebruik van onderzoeksdata zorgt voor mogelijk nieuw onderzoek, en efficiënter onderzoek. Meerwaarde dus voor de onderzoeksdata. Dat is één van de redenen dat er een beweging is naar het ter beschikking stellen van onderzoeksdata voor anderen. Het delen van onderzoeksdata is ook positief voor de delende onderzoeker (citaties, erkenning).

In dit rapport geven we een indicatie van hoe groot het hergebruik van onderzoeksdata is *van en door* onderzoekers van de Nederlandse universiteiten. Het gaat slechts om een indicatie, omdat:

- het delen van data voor hergebruik veelal nog op een informele manier gaat (Figuur 1);
- de gedeelde sets niet centraal geregistreerd staan;
- hergebruik niet standaard geregistreerd wordt;
- datacitatie nog niet op grote schaal plaatsvindt.



Figuur 1. Antwoord op de vraag: 'What tools/sites do you use to archive/share your data and code?' van het 101 scholarly innovations dashboard <http://dashboard101innovations.silk.co/page/Archive-share-data-%26-code>. In Nederland wordt relatief vaak de optie '(and also) others' aangegeven, wat veelal informeel delen inhoudt. Bovenste plaatje: situatie wereldwijd. Onderste plaatje: situatie Nederland. Plaatjes linkerkant: gegeven standaard opties bij de vraag. Plaatjes rechterkant: de genoemde 'others'.

Dit rapport bevat de beschrijving van drie inventarisaties, gedaan in de periode maart tot mei 2016, rond hergebruik van onderzoeksdata van en door onderzoekers die werkzaam zijn aan een Nederlandse onderzoeksinstituten (voor het gemak genoemd 'Nederlandse onderzoekers'):

1. Deponeren van datasets in drie Nederlandse data repositories (1a)

2. Hergebruik door Nederlandse onderzoekers (2a-b)

3. Hergebruik van Nederlandse datasets (3a-d)

De resultaten moeten geïnterpreteerd worden met het voorbehoud dat registratie van deze gegevens veelal niet op affiliatie wordt gedaan, en dat werkelijke cijfers daarom mogelijk anders zijn.

Daarnaast beschrijft het rapport een vierde inventarisatie rond:

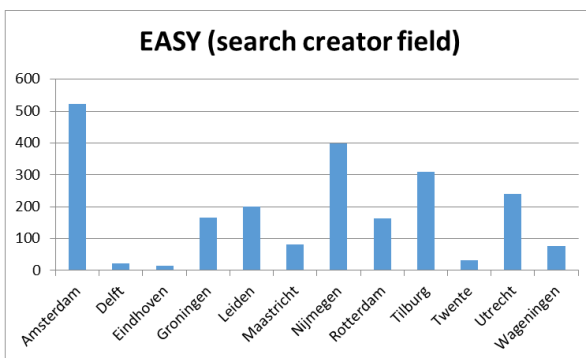
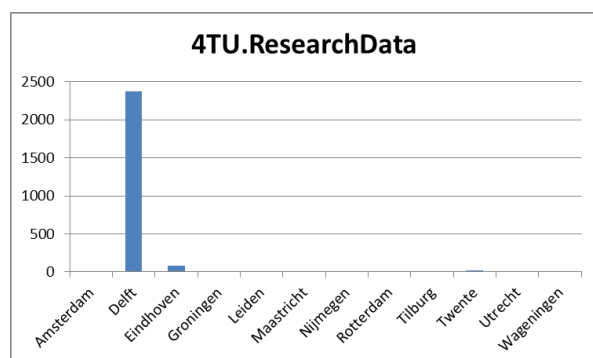
4. Metadata voor het inhoudelijk beschrijven van datasets in repositories (4a)

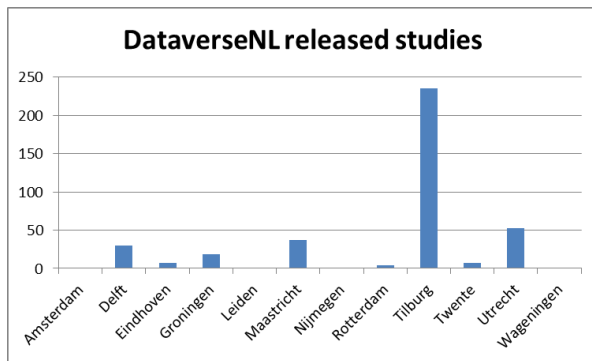
Als laatste geeft het rapport een korte conclusie, gevolgd door een aantal case studies waarin wordt bijgedragen aan hergebruik, gevolgd door een opsomming van mogelijke belemmeringen voor hergebruik van onderzoeksdata, met suggesties voor maatregelen om het hergebruik van onderzoeksdata te vergroten.

1. Deponeren van datasets in drie Nederlandse data repositories

Inventarisatie 1a: Datasets/records in de data repositories EASY, DataverseNL, 4TU.ResearchData, per universiteit.

Methode 1a: Voor een brede opbrengst, zoeken we als benadering voor de affiliatie van deponerende op universiteitsstad. Er is hiermee geen onderscheid meer tussen UvA en VU. In EASY zoeken we in het veld 'creator' datasets met daar de naam van de universiteitsstad. Voor DataverseNL tellen we handmatig de studies. Voor 4TU.ResearchData krijgen we een overzicht van de repository zelf.





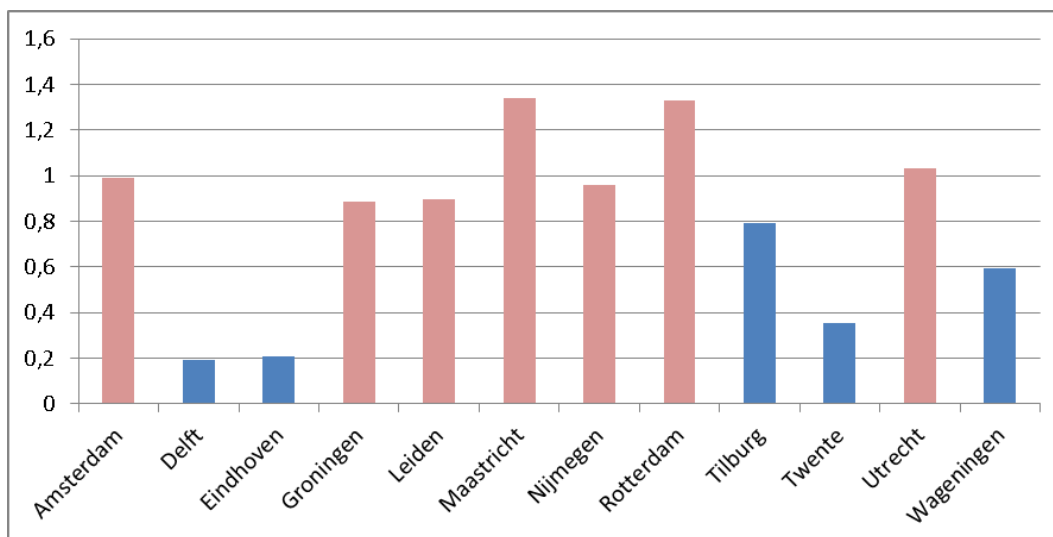
Figuur 2. Het aantal gedeponeerde datasets in Nederlandse repositories, per universiteitsstad.

Resultaten 1a: In EASY staan de meeste sets van diverse universiteiten. De minste sets hier hebben de vier technische universiteiten Delft, Eindhoven, Twente, Wageningen. Delft heeft veel 'records' in 4TU.ResearchData, dit zijn veelal automatische periodieke uploads van lopende experimenten. Tilburg heeft veel studies in DataverseNL omdat ook Tilburgse datasets die elders zijn opgeslagen in DataverseNL worden geregistreerd, bijv. die van CentERdata.

2. Hergebruik door Nederlandse onderzoekers

Inventarisatie 2a: het hergebruik door Nederlandse onderzoekers per universiteit: op indicatie van hergebruik in titel/abstract van publicaties.

Methode 2a: We zoeken per affiliatie (naam universiteitsstad) publicaties welke een indicatie geven dat er data hergebruikt is in het paper. Dit delen we door het totaal aantal papers met de affiliatie, waardoor we een percentage krijgen. De zoekopdracht is in Scopus: (TITLE-ABS-KEY (meta analysis OR reanalysis OR secondary analysis OR integration OR reuse OR cleaning) AND TITLE-ABS-KEY (data) AND AFFIL (*invullen*)).



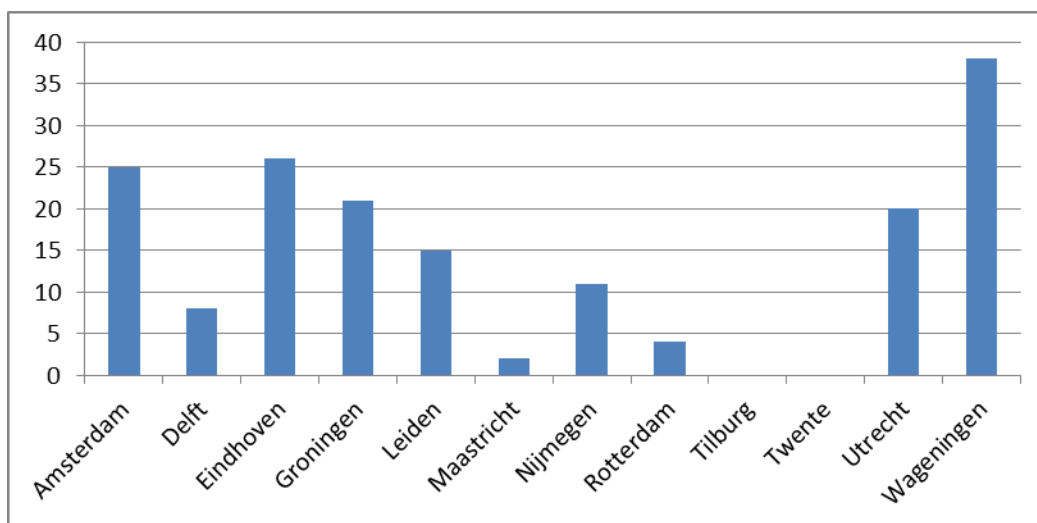
Figuur 3. Het procentuele aantal papers met een hint naar hergebruik van data in de titel/abstract, per affiliatie van een Nederlandse Universiteitsstad. In roze de universiteitssteden met een UMC.

Resultaat 2a: Het aantal titel/abstracts met een indicatie van hergebruik ligt voor alle universiteiten onder de 1.4 procent.

Inventarisatie 2b: Per universiteit referenties naar datasets in publieke repositories

Methode 2b: We kijken naar het aantal papers per affiliatie (naam universiteitsstad) met een referentie naar persistente identifiers van zeven veelgebruikte (generieke) repositories Zenodo, Dryad, Figshare, EASY, 4TU.ResearchData, DataverseNL, B2Share. In Scopus is de zoekactie: (AFFIL (*invullen*) AND REF ("net/10411" OR "net/11304" OR "10.4121/uuid" OR "10.17026/dans" OR "10.5281/zenodo" OR "10.5061/dryad" OR "m9.figshare")).

Resultaat 2b: Er is verschil tussen de universiteiten in de citatie van datasets. De getallen zijn echter over het geheel laag, met een maximum van 37 citaties van datasets in alle publicaties van Wageningen.



Figuur 4. Het aantal papers per affiliatie van een Nederlandse universiteitsstad met in de referentielijst een verwijzing naar een dataset in onderzoeksdata repositories, als indicatie van hergebruik/citatie van datasets door onderzoekers van Nederlandse Universiteiten.

3. Hergebruik van Nederlandse datasets

Inventarisatie 3a: Hoe worden datasets uit Nederlandse repositories 4TU.ResearchData, DataverseNL, EASY geciteerd in wetenschappelijke publicaties. Het aantal papers in totaal in Scopus met daarin referenties naar persistente identifiers van de drie generieke Nederlandse repositories.

Methode 3a: In Scopus zoekstring invoeren: (REF ("net/10411") OR REF ("10.17026/dans") OR REF ("10.4121/uuid")) dit zijn: DataverseNL OF Easy OF 4TU.ResearchData.

Search Alerts Lists My Scopus

(REF ("net10411") OR REF ("10.17026/dans") OR REF ("10.4121/uuid"))
 Show results for: (REF ("net10411") OR REF ("10.17026/dans") OR REF ("10.4121/uuid"))

49 document results

Search within results: [input type="text"]

Refine: [Limit to] [Exclude]

Year: 2016 (6), 2015 (21), 2014 (14), 2013 (8)

Author Name: Falstad, D. (7), Van Der Aalst, W.M.P. (7), Verbeek, H.M.W. (6), Loemans, S.J.J. (5), Maggi, F.M. (4)

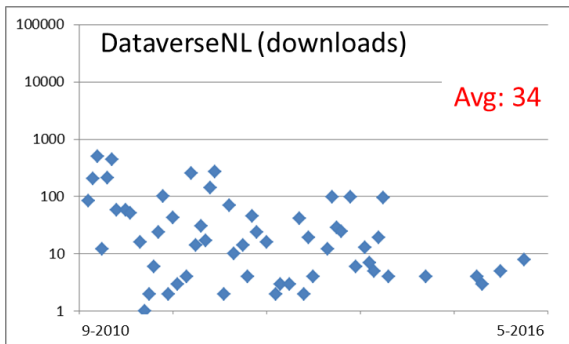
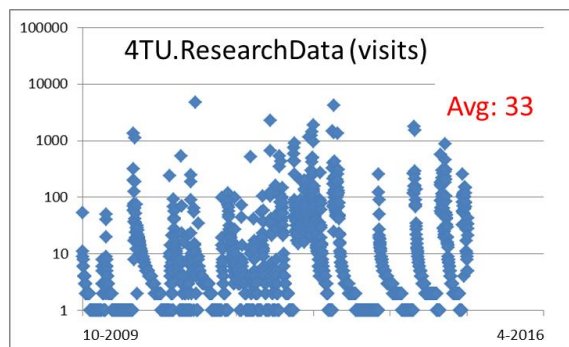
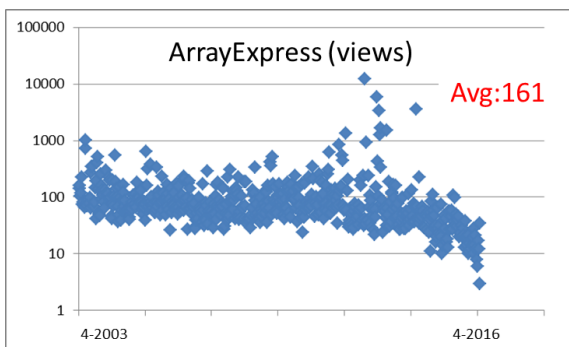
Subject Area: Computer Science (37), Engineering (19), Mathematics (19)

<input type="checkbox"/> Detecting careless respondents in web-based questionnaires: Which method to use?	Niessen, A.S.M., Meijer, R.R., Tendeiro, J.N.	2016	Journal of Research in Personality	0
<input type="checkbox"/> Rush hour commuting in the Netherlands: Gender-specific household activities and personal attitudes towards responsibility sharing	Oakil, A.T.M., Nijland, L., Dijkstra, M.	2016	Travel Behaviour and Society	0
<input type="checkbox"/> Global analysis of seagrass restoration: The importance of large-scale planting	van Katwijk, M.M., Thorhaug, A., Marbà, N., (...), Ransijn, A.M.A., Verduin, J.J.	2016	Journal of Applied Ecology	0
<input type="checkbox"/> Efficient discovery of Target-Branched Declare constraints	Di Cicco, C., Maggi, F.M., Mendling, J.	2016	Information Systems	0
<input type="checkbox"/> Households, consumption and the development of medical care in the Netherlands, 1650-1900	Deneweth, H., Wallis, P.	2016	Journal of Social History	1

Resultaat 3a: in totaal zijn er 49 citaties in Scopus naar de Nederlandse repositories 4TU.ResearchData, Easy, DataverseNL.

Inventarisatie 3b: Voor alle Nederlandse universiteiten samen: interesse voor Nederlandse datasets uit enkele repositories als maat voor hergebruik. Hiervoor bekijken we het aantal views van datasets met een contactpersoon van een Nederlandse universiteit van een discipline-specifieke repository (ArrayExpress), als maat voor mogelijk hergebruik. Het aantal downloads van datasets in DataverseNL. Het aantal visits aan datasets in 4TU.ResearchData.

Methode 3b: Voor arrayexpress selecteerden we Nederlandse datasets door te zoeken op "eur.nl OR tue.nl OR uvt.nl OR wur.nl OR vu.nl OR uva.nl OR tudelft.nl OR maastrichtuniversity.nl OR uu.nl OR rug.nl OR utwente.nl OR ru.nl OR leidenuniv.nl" in het standaard zoekveld. Daarna konden de views hiervan gekopieerd worden. Voor DataverseNL kopieerden we de downloadstatistieken van alle dataverses van de website. Voor 4TU.ResearchData kregen we de data van de repository.



Figuur 5. Op de y-as het aantal views van datasets in ArrayExpress met een contactadres van een auteur van een Nederlandse universiteit, als maat van hergebruik. Het aantal downloads per dataverse, in DataverseNL, als maat van hergebruik. Het aantal bezoeken (visits) aan de datasets in 4TU.ResearchData. Het tijdstip van deponeren staat op de x-as.

Resultaat 3b: Het is onduidelijk in hoeverre de views, visits, downloads uiteindelijk leiden tot hergebruik van de data. N.B. voor EASY was het gemiddeld aantal downloads voor datasets tien (hier waren geen gegevens per dataset beschikbaar).

Inventarisatie 3c: Hoe vaak worden datasets met een Nederlandse affiliatie geciteerd volgens de Web Of Science Data Citation Index (DCI).

Methode 3c: in DCI gezocht op 'Netherlands' in de affiliatie.

Auteurs	Titel	Bron	Adres	Citaties	year
Marchand, Miquelle	LISS panel > Personality > Wave 1	LISS panel data	CentERdata	6	2009
Van Den Brink, Nico	10 year trend of levels of organochlorine pollutants in Antarctic seabirds	Australian Antarctic Data Centre	Wageningen	3	2003
Feenstra, Robert C; Inklaar, Robert; Timmer, Marcel P	Penn World Table 8.1	Penn World Table	RU Groningen	1	2015
Agami, Reuven	Transcriptome-profiling (RNA-seq) and Ribosome-profiling (Ribo-seq) in proliferation, quiescence, senescence and transformed states.	European Nucleotide Archive	Netherlands Cancer Institute	1	2013
Agami, Reuven	Transcriptome-profiling (RNA-seq) and Ribosome-profiling (Ribo-seq) of BJ cells treated with Nutlin-3a, an MDM2 inhibitor, which induces p53.	European Nucleotide Archive	Netherlands Cancer Institute	1	2013
Feenstra, Robert C; Inklaar, Robert; Timmer, Marcel P	Penn World Table 8.0	Penn World Table	RU Groningen	1	2013
Neefjes, Jacques	Epigenetic changes due to histone eviction induced by anthracycline doxorubicin and aclarubicin.	European Nucleotide Archive	Netherlands Cancer Institute	1	2013
Neefjes, Jacques	Genome-wide analysis of histone eviction by FAIRE-Seq.	European Nucleotide Archive	Netherlands Cancer Institute	1	2013
Stunnenberg	Dynamic binding of RBPJ is determined by Notch signalling status.	European Nucleotide Archive	Radboud U	1	2013
Agami, Reuven	Wide-scale analysis of alternative polyadenylation (APA) associated with proliferation and transformation using 3'-Seq.	European Nucleotide Archive	Netherlands Cancer Institute	1	2012
Eishout, Suzan	Initial Questionnaire > 2010	LISS panel data	CentERdata	1	2012
Van Den Toorn, Henk WP	Quantitative erythrocyte membrane proteome analysis with Blue-Native SDS PAGE	PRIDE	U Utrecht	1	2009
Helbig, Andreas	Blue native electrophoresis combined with stable isotope labeling shows that Saccharomyces cerevisiae mitochondrial respiratory chain supercomplexes are still present under anaerobic conditions	PRIDE	U Utrecht	1	2008
Berns, Katrien; Hijmans, E Marielle; Mullenders, Jasper; Brummelkamp, Thijn R; Velds, Arno; Heimerikx, Mike; Kerkhoven, Ron M; Madredjo, Mandy; Nijkamp, Wouter; Weigelt, Britta; Agami, Reuven; Ge, Wei; Cavet, Guy; Linsley, Peter S; Beijersbergen, Roderick L; Bernards, Rene	p53 pathway components	GenomeRNAi			2004
Connell, Dave	Macrobenthos of the North Sea - World Biodiversity Database CD-ROM Series	Australian Antarctic Data Centre	Amsterdam	1	2003
Van Der Putten, Peter	Insurance Company Benchmark (COIL 2000)	UCI Machine Learning Repository	Amsterdam	1	2000
Bajpe, Prashanth Kumar	GSM169312: SD3.23_UT_2	Gene Expression Omnibus	Netherlands Cancer Institute	0	2016
Bajpe, Prashanth Kumar; Heynen, Guus J; Nevedomskaya, Ekaterina; Schlicker, Andreas; Zwart, Wilbert; Bernards, Rene	GSM169313: SD3.23_RA_1	Gene Expression Omnibus	Netherlands Cancer Institute	0	2016
Bajpe, Prashanth Kumar	GSM169310: RAR alpha ChIP-seq-UT-rep2	Gene Expression Omnibus	Netherlands Cancer Institute	0	2016
Bajpe, Prashanth Kumar; Heynen, Guus J; Nevedomskaya, Ekaterina; Schlicker, Andreas; Zwart, Wilbert; Bernards, Rene	GSM169312: SD3.23_UT_2	Gene Expression Omnibus	Netherlands Cancer Institute	0	2016
Bajpe, Prashanth Kumar; Heynen, Guus J; Nevedomskaya, Ekaterina; Schlicker, Andreas; Zwart, Wilbert; Bernards, Rene	GSM169311: SKNSH-PLX-GFP_UT_2	Gene Expression Omnibus	Netherlands Cancer Institute	0	2016
Bajpe, Prashanth Kumar; Heynen, Guus J; Nevedomskaya, Ekaterina; Schlicker, Andreas; Zwart, Wilbert; Bernards, Rene	GSM169314: SD3.23_RA_2	Gene Expression Omnibus	Netherlands Cancer Institute	0	2016
Bajpe, Prashanth Kumar; Heynen, Guus J; Nevedomskaya, Ekaterina; Schlicker, Andreas; Zwart, Wilbert; Bernards, Rene	GSM169310: Input-UT	Gene Expression Omnibus	Netherlands Cancer Institute	0	2016
Bajpe, Prashanth Kumar	GSM169310: MAML3 ChIP-seq-RA-rep2	Gene Expression Omnibus	Netherlands Cancer Institute	0	2016

Resultaat 3c: de hoogst geciteerde dataset wordt zes keer geciteerd. De affiliatie is niet bij alle datasets als metadata genoteerd. Bij navraag bij DCI komen deze cijfers niet van citaties in artikelen, maar de bronnen zijn de citaties die zelf zijn verzameld door 325 datacentrums.

Inventarisatie 3d: Zijn er, als kleine steekproef, Utrechtse datasets in Dryad met een altmetrics score?

Methode 3d: Zoeken in scopus op affiliatie 'utrecht' en referenties '10.5061', vervolgens handmatige selectie van datasets met auteurs met Utrechtse affiliatie.

Tabel 1. Datasets gevonden in Dryad met affiliatie Utrecht, citaties en Altmetrics score.

Zoeken in scopus op affiliatie 'utrecht' en referenties '10.5061', daarna handmatige controle: 7 hits UU (voornamelijk ecologie)	Citaties	Altmetrics
http://datadryad.org/resource/doi:10.5061/dryad.mv13v	-	-
http://datadryad.org/resource/doi:10.5061/dryad.p5v4g	-	1 tweet
http://datadryad.org/resource/doi:10.5061/dryad.3f5j9	-	1 tweet
http://datadryad.org/resource/doi:10.5061/dryad.dr491	-	1 tweet
http://datadryad.org/resource/doi:10.5061/dryad.h8p4n	-	-
http://datadryad.org/resource/doi:10.5061/dryad.4472d	-	-
http://datadryad.org/resource/doi:10.5061/dryad.564sc	-	-

Resultaat 3d: van de tien geselecteerde sets zijn er geen geciteerd, en drie hebben een Altmetrics score (tweet van de Dryad repository). Er is geen bewijs dat datasets hergebruikt worden en dat dit hier met Altmetrics gemeten kan worden.



Metadata voor het inhoudelijk beschrijven van datasets in repositories

Inventarisatie 4a: Een van de actiepunten van de werkgroep was een vergelijking te maken van Nederlandse repositories op basis van hun metadata. En dit zo mogelijk te linken met het hergebruik.

Methode 4a: Van verschillende repositories Zenodo, Figshare, Dryad, 4TU.ResearchData, EASY, DataverseNL, B2Share en ArrayExpress hebben we de metadatavelden verzameld, en daar de metadata ter beschrijving van de inhoud van de dataset uitgehaald, ter vergelijking.

Resultaat 4a: Wij hebben niet kunnen vaststellen dat er een verband is tussen hergebruik en geboden metadata. Dit wordt mede veroorzaakt doordat de repositories verschillende maten geven om gebruik van datasets aan te geven; visits, downloads, views, of niets.

Een aanbeveling hiervoor in het algemeen, om hergebruik te faciliteren, is om met het bepalen van in te vullen metadata niet alleen rekening te houden met wat de repository-manager wil weten en de depositor wil meegeven, maar ook met wat een hergebruiker zou willen zien. Met name zal dit inhoudelijke informatie zijn over de dataset. Zie ook de [powerpoint van een studie van het DIPIR project](#). 'A Context-driven Approach to Data Curation for Reuse' (2016).

Tabel 2. Een overzicht van metadata over de inhoud van de dataset in Nederlandse en internationale repositories.

	Zenodo	Figshare	Dryad	4TU.RD	Easy	Dataverse NL	B2 Share	Array Express
Title	x	x	x	x	x	x	x	x
Description	x	x	x	x	x	x	x	x
Keyword	x	x	x	x	x	x	x	
Topic Classification/subject	x	x		x	x	x	x	
Time Period Covered			x	x	x	x		
Date of Collection				x		x		
Country/ Nation						x		
Geographic Coverage			x	x	x	x		
Universe						x		
Kind of Data	x	x				x	x	x
Audience	x				x		x	
Source					x			
Remarks	x		x		x			x
Taxonomic names			x				x	x
Experimental Design								x
Experimental Factor Name								x
Experimental Factor Type								x
Quality Control Type								x
Replicate Type								x
Normalization Type								x
Protocol Type								x
Protocol Description								x
Protocol Parameters								x
metadata sheet on individual measurements								x
Average downloads		1,5	32.3		10	34		
Average views		22						161
Average visits				33				
Sources: Deposit interfaces of the repositories. Numbers on Figshare website: total uploads 350.000 and 8 million total page views 500.000 downloads. Numbers on Dryad website: 49756 files, 1607086 downloads								

Conclusie

Wat kan er nu geconcludeerd worden uit het beeld dat naar voren komt bij het in kaart brengen van hergebruik?

Reuse is not widespread It cannot easily be quantified



Hergebruik is wel zo belangrijk voor het nuttig ter beschikking stellen van onderzoeksdata, dat dit ten volle gestimuleerd en ondersteund zou moeten worden. Hieronder presenteren we drie case studies gevolgd door een tabel met een overzicht van hordes voor hergebruik, en mogelijke maatregelen om deze hordes te verminderen.

Stimuleren van hergebruik

Case studies:

Met deze case studies willen we wat inspirerende voorbeelden geven voor het stimuleren van hergebruik. Bij elke case study wordt specifiek genoemd hoe de universiteitsbibliotheek (UB) hierbij een rol kan spelen.

Cohen, M. Z., Thompson, C. B., Yates, B., Zimmerman, L., & Pullen, C. H. (2015, APRIL) Implementing common data elements across studies to advance research. *Nursing Outlook*, 63(2), 181-188. <http://dx.doi.org/10.1016/j.outlook.2014.11.006>.

This paper is about implementing the collection and reporting of common data elements to facilitate the combining of data from separate studies. This is important for example to enable overcoming small sample sizes and representing more populations, thereby generalizing findings. It will enable to create a body of research rather than conducting isolated studies with perhaps little chance to translate them into actual practice. In a project a database with recommended measures on concepts can be made (i.e. 891 standardized measures of 343 concepts) (i.e. concept 'quality of life' standardized measure: EuroQol quality of life scale, and for instance measures of disease outcomes at standardized time intervals). Needed for implementation according to the paper:

- Tight cooperation in deciding on measures to fit the (multiple-) theoretical framework(s) of different projects
- communicating the purpose of the common measures to increase acceptance
- teamwork to conform to the concepts and measures
- managing large(r) data sets in the end
- good preparation as adding a concept and especially a new measurement tool to a study after the original design phase is time-consuming and challenging
- unintended consequences: respondent burden as possibly more measures are done than needed for the individual projects, and more time asked of principal investigators to decide a beforehand on the concepts and measurements

Hoe kan de UB helpen?

De UB kan de databases van concepten en metingen gecentraliseerd hosten. De UB kan voorlichting bieden over het gebruik van common data elements. De UB kan begeleiding bieden bij het opstellen van common data elements database en opzet in de startup fase van grote projecten.

Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gotzsche PC, Krumholz HM, Gherzi D, Van derWorp HB. 2014. Increasing value and reducing waste: addressing inaccessible research. *Lancet* 383:257–266 DOI 10.1016/S0140-6736(13)62296-5.

This paper is about how much of the health research suffers from incomplete and selective reporting. The paper states: When full information about studies is inaccessible, billions of dollars in investment are wasted, bias is introduced, and research and care of patients are detrimentally affected. The paper states three recommendations. First, academic institutions and funders should *reward* investigators who fully disseminate their research protocols, reports, and participant-level datasets. Second, *standards* for the content of protocols and full study reports and for data sharing practices should be rigorously developed and adopted for all types of health research. Finally, journals, funders, sponsors, research ethics committees, regulators, and legislators should endorse and enforce *policies* supporting study registration and wide availability of journal reports, full study reports, and participant-level datasets.

Hoe kan de UB helpen?

Het mede ontwikkelen en promoten van standaarden voor inhoud van protocollen, studie rapportages en het proces van data delen. Het bijhouden van een gebruikte standaarden catalogus. Een embedded librarian.

Van Tuyl S, Whitmire AL (2016) *Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia*. PLoS ONE 11(2): e0147942. doi:10.1371/journal.pone.0147942

The paper defines 'data sharing' and presents a protocol for assessing whether data have been shared and how effective the sharing was. Data sharing is defined as addressing four criteria: DATA

- **Discoverability:** using direct and persistent links
- **Accessibility:** data should be shared via an open platform with minimal barriers to access
- **Transparency:** data should have collocated documentation that describes it sufficiently for an expert to use
- **Actionability:** one should be able to use the data in analytical platforms with minimal reprocessing or reformatting, e.g. no figure in article or table in pdf

The authors evaluate the data sharing practices of researchers funded by the NSF at Oregon State University in two ways: by attempting to discover project-level research data using the associated DMP as a starting point, and by examining data sharing associated with journal articles that acknowledge NSF support. Sharing at both the project level and the journal article level was not carried out in the majority of cases, and when sharing was accomplished, the shared data were often of questionable usability due to access, documentation, and formatting issues. The article offers recommendations for how data producers, journal publishers, data repositories, and funding agencies can facilitate the process of sharing data in a meaningful way:

<p>Recommendations data producers</p> <ul style="list-style-type: none"> - citation of datasets should be treated much the same as citation of journal articles (cite dataset not database; cite data in references not elsewhere) - consider sharing your data through an open repository or journal - engage with research data service providers either at the local level (often part of their academic library) or at the national level (as sharing can be difficult and time consuming) 	<p>Recommendations journal publishers</p> <ul style="list-style-type: none"> - set a policy for data sharing - guidelines for how to identify what data were used in the research (including data citation guidance) - when data are shared through a journal article, it is important for the journal to take some responsibility for ensuring the quality of the shared data (even if the data are not shared via the journal but through an external repository)
<p>Recommendations data repositories</p> <ul style="list-style-type: none"> - take some responsibility for ensuring the quality of the data and documentation deposited therein (does the dataset deposited have sufficient metadata, and are the data in a format that is actionable and open?) 	<p>Recommendations funding agencies</p> <ul style="list-style-type: none"> - engage more fully with grantees, journals, repositories, and data services providers to help ensure robust and practical standards for data sharing

Hoe kan de UB helpen?

Direct: kwaliteit van data en documentatie van data in eigen repository garanderen. Onderzoekers meer bewust maken hoe en waar ze hun data effectief kunnen delen. Onderzoekers ondersteunen bij hoe, waar en wanneer ze hun eigen datasets moeten citeren en hoe ze effectief metadata kunnen ontwikkelen.

Indirect: uitgevers en subsidieverstrekking stimuleren de genoemde aanbevelingen uit te voeren of verbeteren.

Tabel 3. Een suggestie voor onderdelen van het proces van hergebruiken, en de hordes hierin, met mogelijke maatregelen ter verbetering.



Hurdle		Measures to improve
Concern about ethical reuse of data [1]	●	<ul style="list-style-type: none"> • Have a data transfer agreement if personal data is exchanged between legal entities • Have an ethical committee judge the validity of the research before handing out the personal data • Only share fully anonymized data • Have written arrangements on the exact goal and scope of the use
Lack of available (suitable) data[1]	●	<ul style="list-style-type: none"> • Actively lobby at relevant parties for making data available • Organize and host common data elements for disciplines to make more study designs align [3] • Promote sharing by a top down policy from an institute, funder, or journal • Promote sharing bottom up by offering education on the <i>benefits of sharing</i>
Data is hard to find [1][2]	●	<ul style="list-style-type: none"> • Harvest databases through data portals to reduce 'scattering' of datasets • Standardization of relevant searchable metadata [3] • Advanced community and project-specific databases • Library assistance in finding and using appropriate datasets • Create extensive catalogues with better search functionality
Data is inaccessible (restricted or faulty)	●	<ul style="list-style-type: none"> • Use persistent identifiers • Use preferred formats • Have a 'notify' option if the data is not there or cannot be opened, let a data manager amend the study • Make deposits 'open' as a default, 'restricted' on request
Much time is spent to appraise unfamiliar data [1][2]	●	<ul style="list-style-type: none"> • Standardization of metadata, vocabulary, and documentation [3] • Have a data manager coordinate and/or curate documentation • Make use of tools to request and secure metadata and documentation • Increase incentives to provide rich documentation at deposit (i.e. quality seal) • Explicitly note flaws so these data can be omitted by users ('comment' field in data repositories) • Offer techniques or tools for easy assessment of dataset quality i.e. 'arrayanalysis.org'
Quality of data is poor	●	<ul style="list-style-type: none"> • Have a quality assessment before admittance to a repository • Educate researchers on data appraisal before sharing • Improve data quality, i.e. by the use of peer review or data journals [4]
Unfamiliarity with methods of secondary analysis [1]	●	<ul style="list-style-type: none"> • Provide courses on the methods for secondary analysis • Offer tools for pre-processing or data cleaning (i.e., 'OpenRefine' or 'R statistical language')

[1] Sage et al. (2014) Managing and sharing research data.

[2] Pronk et al. (2015) A game theoretic analysis of research data sharing. PeerJ 3:e1242. DOI 10.7717/peerj.1242

[3] Cohen et al. (2015) Implementing common data elements across studies to advance research. Nursing Outlook, 63(2), 181-188. DOI 10.1016/j.outlook.2014.11.006.

[4] Gorgolewski et al. (2013) Making data sharing count: a publication-based solution. Frontiers in Neuroscience 7(9), 1-7. DOI 10.3389/fnins.2013.00009.