



# Uitgeversbeleid en Onderzoeksdata

Rapportage in het kader van het Jaarplan 2016

UKB Werkgroep Research Data

Mei 2017 [v2 oktober 2017 ivm literatuur update]

Rutger de Jong, Universiteit Leiden

Henk van den Hoogen, Universiteit Maastricht

Fieke Schoots, Universiteit Leiden

Mariëtte van Selm, Universiteit Amsterdam

Madeleine de Smaele, TU Delft

Rapportage op basis van inventarisatie: PublishersDataPolicies\_UKBwg2016.xlsx

## Inhoud

0. Samenvatting.....	3
1. Inleiding.....	3
2. Datajournals .....	4
3. Data Availability Policy's.....	6
3.1. Opzet en algemene bevindingen .....	6
3.2. Beleid t.a.v. onderzoeksdata .....	6
3.3. Data delen per discipline .....	7
3.4. Locatie van de data .....	8
3.5 Open Access.....	8
3.6. Eigendomsrecht / copyright .....	9
3.7. Conclusies m.b.t. Data Availability Policy's .....	9
4. Voorbeelden van policy's .....	10
Elsevier.....	10
Wiley.....	10
Springer / Nature.....	11
PLoS One.....	12
5. Data archieven en data citatie .....	12
6. Conclusie .....	12
7. Bronnen .....	14
a. Data journals.....	14
b. Data Availability Policy's.....	14
c. Onderzoeken naar (niet) beschikbaarheid data van journals met een DAP: .....	16
d. data policies en data archieven.....	17
e. data policies en de rol van bibliotheken / instellingen .....	17
f. Over (on)mogelijkheid data policy registry:.....	18
g. overzichten van uitgeversbeleid t.a.v. onderzoeksdata.....	18
Bijlage I : vragen bij inventarisatie data availability policies & data repositories .....	19

## 0. Samenvatting

Wetenschappelijke uitgevers kunnen een belangrijke rol vervullen bij het duurzaam toegankelijk maken van onderzoeksdata. Uit een steekproef van tijdschriften waarin veel wordt gepubliceerd door Nederlandse onderzoekers blijkt echter dat maar een klein deel expliciet eisen stelt aan de beschikbaarheid van data behorende bij de artikelen, in de vorm van een zogeheten Data Availability Policy (DAP). Ook 'data journals', die het publiceren over datasets mogelijk maken, vragen niet altijd om toegang tot de data. Uitgevers bieden vaak wel de mogelijkheid om onderzoeksdata via het eigen platform aan te bieden, als 'supplementary materials', maar dat bemoeilijkt de vindbaarheid en herbruikbaarheid van de gegevens, op z'n minst voor degenen die niet geabonneerd zijn op het tijdschrift. Dus hoewel we in ons beknopte onderzoek geen expliciete aanwijzingen hebben gevonden dat uitgevers het eigendomsrecht op onderzoeksdata claimen, verdwijnen onderzoeksdata regelmatig achter de 'paywall' van het tijdschrift. In dit rapport bevelen we instellingen, meer specifiek de universiteitsbibliotheken, aan om hun onderzoekers actief te informeren over het beleid van uitgevers en te bevorderen dat data behorende bij publicaties in duurzame, gecertificeerde repositoria wordt gedeponereerd in plaats van overgedragen aan de uitgever. De data archieven kunnen het linken van data en publicatie bevorderen door het vergemakkelijken van datacitatie en peer review. Uitgevers, tot slot, moeten in de eerste plaats streven naar duidelijke richtlijnen en harmonisatie van het beleid tav onderzoeksdata. Dat beleid moet gericht zijn op de vindbaarheid en herbruikbaarheid van de onderzoeksdata voor alle geïnteresseerden.

## 1. Inleiding

Aanleiding voor deze rapportage is de constatering dat bij het publiceren van wetenschappelijke artikelen uitgevers steeds vaker ook om de achterliggende dataset vragen. Voorkomen moet worden dat er onduidelijkheden ontstaan over de rechten op de data en dat er op termijn gelijke discussies met uitgevers moeten worden gevoerd als nu regelmatig het geval is m.b.t. Open Access van artikelen. Ook is er de laatste jaren regelmatig gepubliceerd over de slechte vindbaarheid van data bij publicaties. Binnen de UKB werkgroep Research Data hebben we daarom besloten, als één van de thema's voor het jaarplan 2016, meer helderheid te verschaffen over de eisen die uitgevers aan hun auteurs stellen. Ook willen we een advies uitbrengen aan de UKB over de rol die UB's kunnen spelen ten aanzien van het beleid van uitgevers. Een deel van onze bevindingen hebben we gepresenteerd bij het seminar 'The Making of Research Data Management Policy'.<sup>1</sup>

De volgende deelvragen stonden aan de basis van ons onderzoek:

1. Wat zijn de ontwikkelingen ten aanzien van Data Journals en Data Papers?
2. Wat zijn de Data Availability Policy's van uitgevers?
3. In hoeverre ondersteunen repositoria de opslag en de citatie van data behorende bij een publicatie? Wat zijn de samenwerkingsverbanden tussen uitgevers en repositoria?
4. Hoe zit het met de rechten van data behorende bij een publicatie waarvan de rechten worden overgedragen aan de uitgever? Hoe zit het met de rechten van data bij een open access publicatie?

---

<sup>1</sup> Seminar georganiseerd door UKB werkgroep Research Data en Landelijk Coördinatiepunt RDM, Wageningen, December 1, 2016 (<https://www.surf.nl/en/agenda/2016/12/seminar-the-making-of.-research-data-management-policy/index.html>, bekeken op 3 mei 2015)

Om antwoord te kunnen geven op bovenstaande vragen maken we om te beginnen onderscheid tussen ‘reguliere’ wetenschappelijke tijdschriften en de relatief nieuwe data journals die ‘data papers’ of ‘data descriptors’ publiceren. De ontwikkelingen t.a.v. de datajournals worden beschreven in hoofdstuk 2.

In hoofdstuk 3 inventariseren we de Data Availability Policy’s van ‘reguliere’ wetenschappelijke tijdschriften. Wat houden deze richtlijnen in en zijn er verschillen in discipline? Wat zeggen de richtlijnen over de rechten op de data? In hoofdstuk 4 geven we voorbeelden van de Data Availability Policy’s van een aantal uitgevers.

In hoofdstuk 5 kijken we naar een aantal data repositoria die worden genoemd in Data Availability Policy’s: hoe faciliteren zij het linken van data aan publicaties, werken zij samen met bepaalde uitgevers, zijn er citatierichtlijnen, is peer review op de data mogelijk etc.?

Hoofdstuk 6, de conclusie, bevat enkele aanbevelingen en komt terug op de wenselijkheid van een nationaal overzicht van uitgeversbeleid.

Tot slot volgt een lijst met bronnen over uitgevers en databeleid.

Enkele opmerkingen vooraf :

- I. Ontwikkelingen in databeleid gaan snel en we zien dat uitgevers hun beleid regelmatig aanpassen. Onze inventarisatie is gedaan in de eerste helft van 2016 en afgesloten in september 2016.
- II. We hebben ons beperkt tot – desktop – research en openbare informatie op de websites van de uitgevers. We hebben geen goed beeld van aanvullende eisen t.a.v. bijvoorbeeld peer review van data als een artikel is ingediend.
- III. Er is de afgelopen jaren veel gepubliceerd over de (niet) beschikbaarheid van onderzoeksdata behorende bij publicaties, datacitatie, verrijkte publicaties, modelbeleid, etc. (zie bronnenlijst). We hebben hier geen uitgebreid literatuuronderzoek willen doen, maar de focus beperkt tot een praktisch onderzoek naar de consequenties van het beleid van uitgevers voor onderzoekers aan Nederlandse instellingen.

## 2. Datajournals

In 2013 werd binnen het project Peer Review for Publication and Accreditation of Research Data in the Earth Sciences (PREPARDE) een overzicht van data journals en hun richtlijnen opgesteld.<sup>2</sup> In 2014 werd deze lijst door Katherine Akers (University of Michigan) aangevuld tot een lijst van 73 tijdschriften die “either exclusively publish data (or database, software, or model) papers or that publish data papers as one article type.”<sup>3</sup> In 2015 werd door Candela et al. een nieuwe lijst gepubliceerd.<sup>4</sup> Deze vier lijsten zijn samengevoegd en ontdaan van titels die niet meer bestaan of met/tot andere titels zijn gefuseerd. Vervolgens zijn alle gemengde tijdschriften, waarin zowel reguliere als data papers gepubliceerd worden, buiten beschouwing gelaten – alleen het beleid van dertien ‘pure’ data journals is geanalyseerd.

Aquatic Data

<http://www.journals.elsevier.com/aquatic-data/>

<sup>2</sup> <http://proj.badc.rl.ac.uk/preparde/blog/DataJournalsList> (bekeken op 3 mei 2017)

<sup>3</sup> <https://mlibrarydata.wordpress.com/2014/05/09/data-journals/> (bekeken op 3 mei 2017)

<sup>4</sup> Candela, L., Castelli, D., Manghi, P. and Tani, A. (2015), Data journals: A survey. *Journal of the Association for Information Science and Technology*, 66: 1747–1762.

Atomic Data and Nuclear Data Tables	<a href="http://www.journals.elsevier.com/atomic-data-and-nuclear-data-tables/">http://www.journals.elsevier.com/atomic-data-and-nuclear-data-tables/</a>
Biodiversity Data Journal (BDJ)	<a href="http://biodiversitydatajournal.com/">http://biodiversitydatajournal.com/</a>
Data in Brief	<a href="http://www.journals.elsevier.com/data-in-brief/">http://www.journals.elsevier.com/data-in-brief/</a>
Dataset Papers in Science	<a href="http://www.hindawi.com/journals/dpis/">http://www.hindawi.com/journals/dpis/</a>
Earth System Science Data	<a href="http://www.earth-system-science-data.net/">http://www.earth-system-science-data.net/</a>
Geoscience Data Journal	<a href="http://www.geosciencedata.com">http://www.geosciencedata.com</a>
Journal of Open Archaeology Data	<a href="http://openarchaeologydata.metajnl.com/">http://openarchaeologydata.metajnl.com/</a>
Journal of Open Health Data	<a href="http://openhealthdata.metajnl.com/">http://openhealthdata.metajnl.com/</a>
Journal of Open Psychology Data	<a href="http://openpsychologydata.metajnl.com/">http://openpsychologydata.metajnl.com/</a>
Journal of Open Research Software	<a href="http://openresearchsoftware.metajnl.com/">http://openresearchsoftware.metajnl.com/</a>
Scientific Data	<a href="http://www.nature.com/scientificdata/">http://www.nature.com/scientificdata/</a>
Research Data Journal for the Humanities and Social Sciences (RDJ)	<a href="http://www.brill.com/rdj">www.brill.com/rdj</a>

Tabel: geanalyseerde datajournals

Juist deze journals waren als eerste onderwerp van analyse, omdat we bij data journals de helderste voorschriften ten aanzien van de beschikbaarheid van data (data availability policy's) verwachtten.

Een vergelijking tussen de dertien tijdschriften levert een aantal opvallende zaken op:

- **Data papers** (artikelen in *data journals*) zijn in overgrote meerderheid **open toegankelijk**: van de dertien tijdschriften zijn er twaalf Open Access en is er één hybride.
- Bij acht van de dertien tijdschriften dienen ook de **data** die in de *papers* beschreven worden **open toegankelijk** te zijn. De andere tijdschriften moedigen het delen van de data wel sterk aan, maar stellen het niet als voorwaarde voor publicatie in het tijdschrift.
- Het delen van de **code** die voor het verkrijgen van de data geschreven is, wordt bij één van de tijdschriften benoemd en bij één ander tijdschrift verplicht gesteld. Om de herkomst van de data die in een data paper beschreven wordt, goed te kunnen duiden, zou meer aandacht voor het (verplicht) beschikbaar stellen van code verwacht mogen worden.
- Het verschaffen van **supplementary materials** bij het data paper wordt door twee titels, beide van Elsevier, aangemoedigd. De andere tijdschriften maken hiervan geen melding.
- Voor de **opslag van onderzoeksdata** geven de meeste tijdschriften suggesties door middel van een niet-verplichtende lijst van *trusted repositories*. Eén tijdschrift, Biodiversity Data Journal, verwijst voor grote datasets expliciet naar specifieke repositories en biedt de mogelijkheid om kleine datasets via de website van het tijdschrift te uploaden.
- Aan het **eigenaarschap** van de data wordt in geen van de dertien tijdschriften aandacht besteed. Dit zou positief kunnen worden geduid: de uitgever van het tijdschrift werpt zich dus niet op als (aspirant-)eigenaar van de data die ten grondslag liggen aan de artikelen in het tijdschrift. Of de uitgevers van de dertien tijdschriften zich stilzwijgend rechten toeëigenen, is in een beknopt onderzoek als verricht is, niet na te gaan.
- Het belang van **datacitatie** wordt bij slechts drie van de dertien titels expliciet benoemd. Dat kan erop duiden dat de uitgevers het belangrijker vinden dat het data paper wordt geciteerd dan dat naar de beschreven dataset zelf wordt verwezen.

## 3. Data Availability Policy's

### 3.1. Opzet en algemene bevindingen

De leden van de werkgroep hebben een selectie gemaakt van de tien wetenschappelijke tijdschriften waarin onderzoekers van de eigen instelling volgens de respectievelijke CRIS systemen het meest hebben gepubliceerd voor acht vakgebieden, in de periode 2014 – 2015<sup>5</sup>. Na ontdebbling leverde dit een lijst op van in totaal 188 tijdschriften. Doordat we beperkt waren in de tijd, hebben we niet alle tijdschriften uit deze selectie kunnen beoordelen. Uiteindelijk hebben we voor 128 tijdschriften vragen beantwoord als 'is er een data sharing policy', 'geldt deze voor alle journals van de uitgever of alleen voor dit journal', 'is er een eis ten aanzien van de locatie waar de data worden opgeslagen, 'zijn er richtlijnen voor supplementary data', 'zijn er richtlijnen voor de citatie van data', 'zijn er bepalingen mbt de rechten op de data' etc<sup>6</sup>. Het bestand met de resultaten van deze inventarisatie worden bij deze rapportage gevoegd<sup>7</sup>. In de conclusie komen we terug op nut en noodzaak van het verder verspreiden en actualiseren van het bestand.

We spreken in dit rapport van een Data Availability Policy als een tijdschrift duidelijke richtlijnen heeft ten aanzien van het beschikbaar maken van onderzoeksdata, al dan niet onder de benaming 'Data Availability Policy'. Deze richtlijnen zijn veelal onderdeel van de 'author' of 'submission guidelines'. Uit onze inventarisatie kwam naar voren dat een kleiner deel dan verwacht, namelijk 15%, van de door ons onderzochte tijdschriften expliciete richtlijnen heeft voor onderzoeksdata. In 23% van de gevallen wordt het delen van data aangemoedigd, in 15 % wordt het verondersteld, bijv. doordat data worden beschouwd als 'supplementary materials' die al dan niet bij het artikel kunnen worden gevoegd. Ook bleek dat de richtlijnen in het overgrote deel van de gevallen alleen gelden voor een specifiek tijdschrift en niet voor tijdschriften van een uitgever. Soms heeft de uitgever wel een overkoepelend beleid maar geven tijdschriftredacties daar een invulling aan die aansluit bij het vakgebied. Hieronder volgen de meest opvallende bevindingen.

### 3.2. Beleid t.a.v. onderzoeksdata

Van de 128 tijdschriften in onze steekproef:

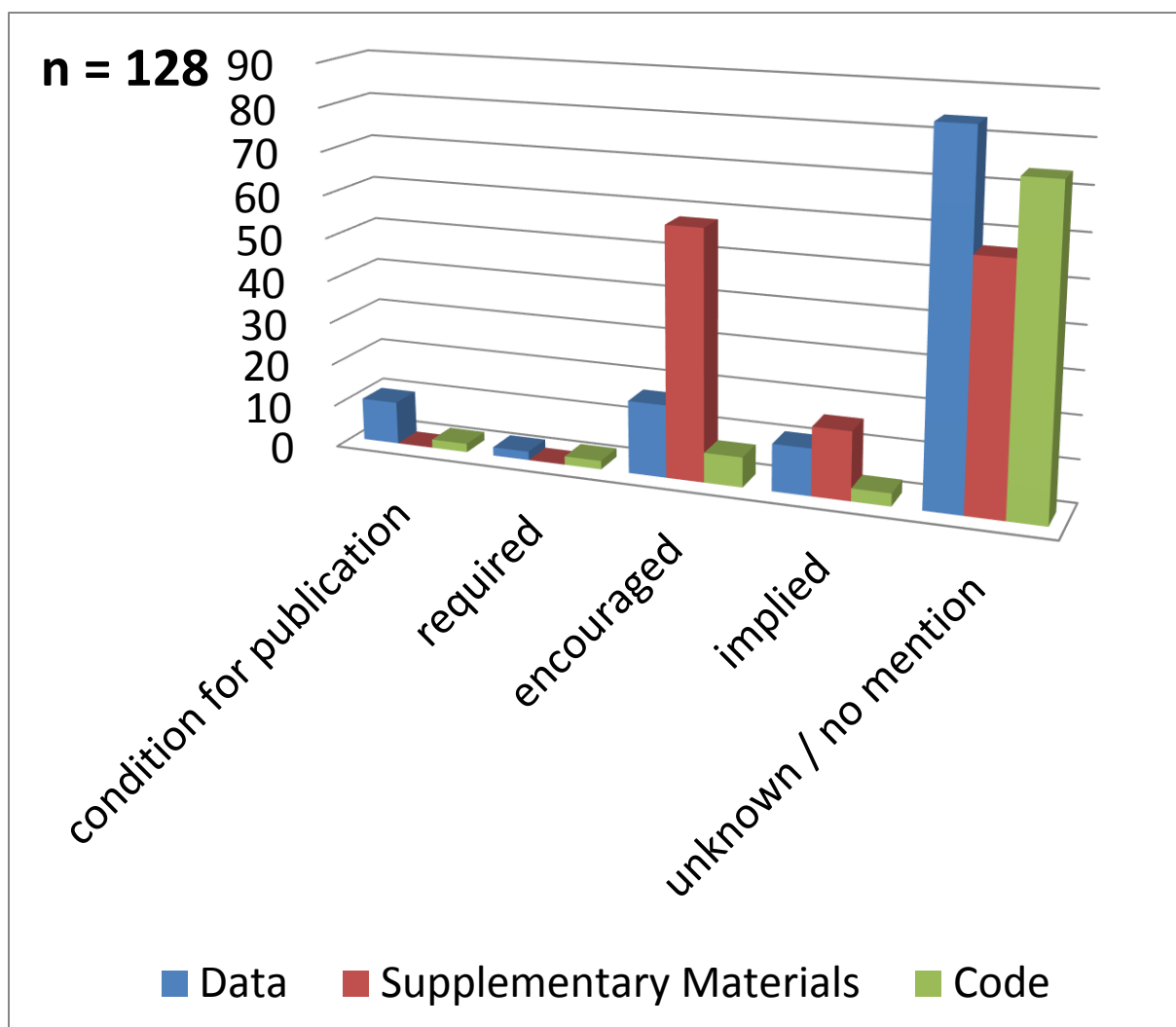
- hebben 10 tijdschriften een DAP die inhoudt dat een artikel niet wordt gepubliceerd zonder bijbehorende data [uitgevers: Wiley, Elsevier, Frontiers, Copernicus, NPG];
- hebben 2 tijdschriften wel beleid maar verbinden zij geen duidelijke consequentie aan het ontbreken van data (Angewandte Chemie, Astronomy and Astrophysics).
- moedigen 18 tijdschriften het delen van data aan zonder dit verplicht te stellen (o.a. Wiley, Sage, Elsevier, Taylor & Francis, IOP, Nature);
- wordt bij 10 tijdschriften het delen van data 'verondersteld', bijv. doordat een 'data availability statement' wordt gevraagd zonder dat er duidelijke eisen worden geformuleerd;
- worden bij 87 tijdschriften onderzoeksdata niet genoemd in de richtlijnen voor auteurs.

---

<sup>5</sup> Universiteit Leiden, Universiteit Maastricht, Universiteit van Amsterdam, TU Delft. Vakgebieden: Archeologie, Bestuurskunde en Economie, Geesteswetenschappen, Geneeskunde, Rechtsgeleerdheid, Sociale wetenschappen, Techniek, Science.

<sup>6</sup> Zie bijlage I voor een compleet overzicht. De vragen zijn gebaseerd op de lijst uit Stodden, V., Guo, P., & Ma, Z. (2013). Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals. *PLoS ONE*, 8(6), 2–9. <http://doi.org/10.1371/journal.pone.0067111>.

<sup>7</sup> PublishersDataPolicies\_UKBwg2016.xlsx (laatste toevoeging 1 september 2016)



Figuur 1 Sharing is...

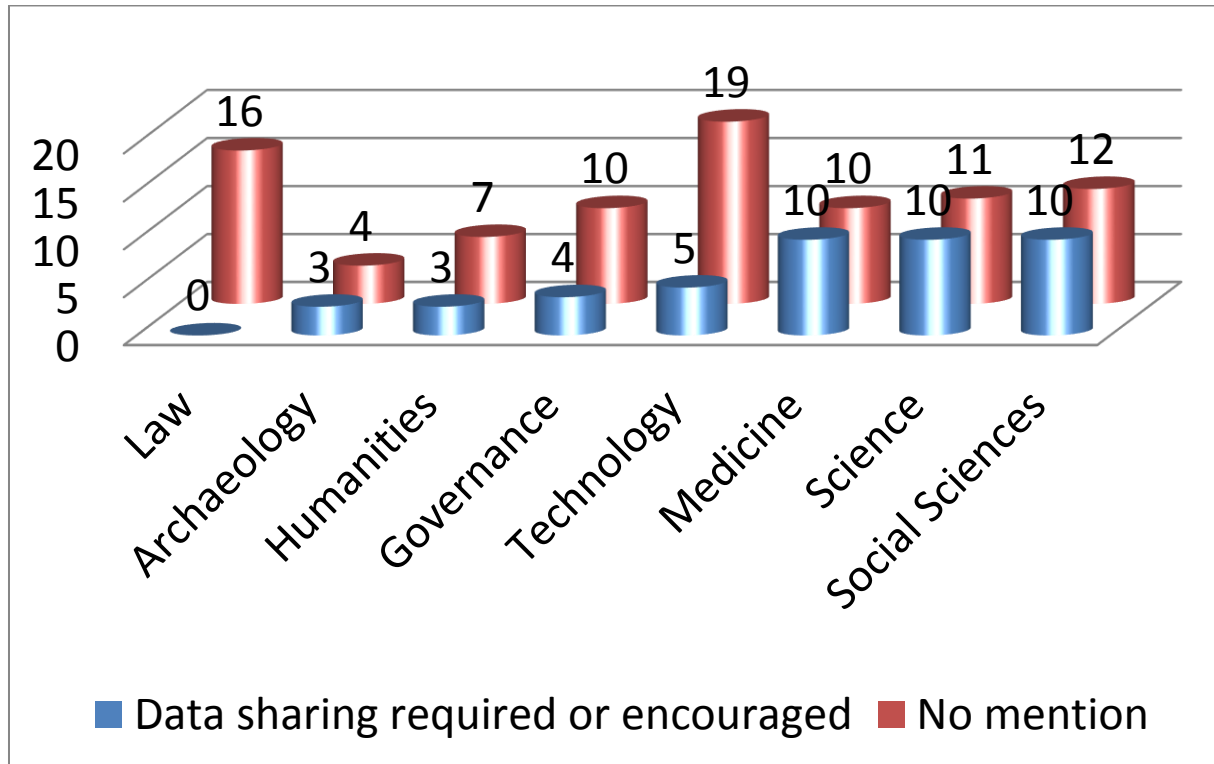
Voor *supplementary materials* is het beeld heel anders. Supplementary materials stammen uit de tijd dat veel tijdschriften ook of alleen in gedrukte vorm verschenen. Het is het materiaal dat ten grondslag ligt aan, of zeer relevant is voor de publicatie zelf, maar te omvangrijk is om af te drukken en daarom online wordt aangeboden, veelal op de website van de uitgever. Het zijn van oudsher datasets onderliggend aan tabellen of extra figuren, maar tegenwoordig ook powerpoints, videomateriaal of extra visualisaties. 58 tijdschriften uit onze steekproef (74%) moedigen het uploaden van 'supplementary materials' bij het artikel aan, bij 16 tijdschriften wordt de mogelijkheid geboden zonder dat dit verplicht is of aanbevolen wordt.

Met de aandacht voor onderzoeksdata is het besef gegroeid dat *code* ook bewaard en gedeeld kan worden. In onze steekproef zijn er vier tijdschriften die expliciete aanwijzingen geven voor het delen van code, zeven tijdschriften moedigen het delen van code aan, bij drie wordt het verondersteld.

### 3.3. Data delen per discipline

Als we voor de tijdschriften in onze steekproef inzoomen naar de aanwezigheid van Data Availability Policy's bij de verschillende disciplines valt op dat het aandeel van tijdschriften mét DAP voor de Sociale Wetenschappen, Science, Geneeskunde en de Archeologie rond de 50% ligt. Bij Geesteswetenschappen en Bestuurskunde ligt het aandeel van tijdschriften met DAP rond de 30%,

voor de Technische wetenschappen rond de 20%. Overigens heeft ca. 80 % van de tijdschriften op het gebied van de Technische wetenschappen wel richtlijnen voor supplementary materials. Geen van de Rechten tijdschriften heeft beleid voor data.



Figuur 2 Beleid voor het delen van data per discipline

### 3.4. Locatie van de data

Een aantal van de grotere uitgevers geeft als onderdeel van de richtlijnen voor onderzoeksdata of voor ‘supplementary materials’ een lijst met repositoria waar de data verplicht moeten worden gedeponeerd. Dit geldt voor 13 tijdschriften uit de steekproef (m.n. van Elsevier, Taylor & Francis, American Society for Hematology, BioMed Central, Springer, Nature, Wiley). Vaak gaat het om specifieke disciplinegebonden archieven (European Nucleotide Archive, NCBI trace and short-read archive, ENA's Sequence Read Archive, GEO, ArrayExpress, Genbank, Human Brain Mapping database, BrainMap etc. of voor source code : bioinformatics.org, sourceforge.net)<sup>8</sup>. Meer algemene voorzieningen die worden genoemd zijn Figshare, Dryad en Science Direct (Elsevier).

Bij 13 andere tijdschriften staat geen verplichte lijst maar worden o.a. bovenstaande repositoria als voorbeeld genoemd, zie bijv. het beleid van PlosOne.<sup>9</sup>

### 3.5 Open Access

Gaandeweg ons onderzoek rees de vraag of tijdschriften met een Open Access beleid vaker beleid hebben voor (open) data. Het blijkt niet zo te zijn dat alle Open Access tijdschriften een DAP hebben.

<sup>8</sup> Zie bijv.: <http://www.nature.com/sdata/policies/repositories>; <https://www.elsevier.com/books-and-journals/content-innovation/data-base-linking/supported-data-repositories> (bekeken op 3 mei 2017)

<sup>9</sup> <http://journals.plos.org/plosone/s/data-availability> (bekeken op 3 mei 2017)



Wel is het zo dat van de 41 tijdschriften die een of andere vorm van beleid hebben voor het delen van data (incl. 'implied') iets minder dan de helft (18 tijdschriften) ook Open Access publiceren mogelijk maakt. Hetzelfde beeld zien we bij de 70 tijdschriften die een of andere vorm van beleid hebben voor supplementary materials. Hier gaat het in 32 gevallen om Open Access tijdschriften. Om verdere conclusies te trekken, zouden we in onze steekproef eerst een duidelijkere definitie moeten hanteren van 'open access'.

### ***3.6. Eigendomsrecht / copyright***

Hoewel ons onderzoek is ingegeven door de vraag hoe uitgevers om zullen gaan met het eigendomsrecht van data, wordt dit onderwerp nauwelijks genoemd in de richtlijnen voor onderzoeksdata. Dit ligt anders voor de supplementary materials die voor sommige tijdschriften expliciet onder het copyright van de uitgever vallen. Om hergebruik van data te stimuleren is het belangrijk om te weten bij wie de rechten liggen. Onduidelijkheden over eigendomsrecht of het afstaan van eigendomsrecht maakt dat data onbruikbaar zijn voor bijvoorbeeld overkoepelende databases of data mining.

### ***3.7. Conclusies m.b.t. Data Availability Policy's***

In zijn algemeenheid constateren we dat er relatief nog weinig DAP's zijn en dat zij bovendien weinig concreet zijn. Zelfs als er richtlijnen zijn, zijn er onduidelijkheden over:

- wat te deponeren (data vs supplementary materials);
- waar;
- verwachtingen tav toegang;
- wanneer (bij indienen artikel ; voor peer review ; bij publicatie of later)<sup>10</sup>;
- peer review van de data;
- evt. sancties bij het niet opvolgen van DAP.

Veelal worden onderzoeksdata als 'traditionele' supplementary materials behandeld. Deze benadering wordt vanuit het oogpunt van duurzame toegang problematisch beschouwd omdat:

- data, of delen van de data, niet beschikbaar zijn;
- data achter betaalmuren verdwijnen;
- het format waarin de data worden gepubliceerd (pdf, word) hergebruik van de data verhindert;
- er geen persistente link naar de data is;
- de data kunnen niet worden geciteerd.

Anderzijds biedt de bekendheid met het publiceren van 'supplementary materials' ook mogelijkheden. Dat er al een cultuur is waarin data gepubliceerd worden, kan een goede insteek bieden om een overgang naar duurzamere repositories te bewerkstelligen.

---

<sup>10</sup> Zie ook "Table 1. Summary of main points discovered from survey of journal data policies" in: Sturges, P., Bamkin, M., Anders, J. H.S., Hubbard, B., Hussain, A. and Heeley, M. (2015), Research data sharing: Developing a stakeholder-driven model for journal policies. *J Assn Inf Sci Tec*, 66: 2445–2455.

[Http://doi.org/10.1002/asi.23336](http://doi.org/10.1002/asi.23336)

## 4. Voorbeelden van policy's

### *Elsevier<sup>11</sup>*

Elsevier heeft twee manieren om het uitgangspunt te ondersteunen dat ruwe data voor alle onderzoekers beschikbaar moeten zijn en dat auteurs vrij moeten zijn om hun data publiek te maken.

- Open data pilot : Data worden als supplementary materials geupload naar Science Direct en zijn dan beschikbaar voor downloaden onder een CC BY licentie.

- Mendeley data: auteurs uploaden hun data in Mendeley, krijgen een DOI en kiezen zelf een licentie. Voor tijdschriften die niet de mogelijkheid bieden data bij het artikel te publiceren, geeft Elsevier een lijst van vakspecifieke repositoria waarin data via een speciale 'database linking tool' aan de publicatie kunnen worden gelinkt. Elsevier moedigt verder het schrijven van 'data profiles' aan waarin auteurs kunnen specificeren waar de data beschikbaar zijn of kunnen uitleggen waarom dat niet mogelijk is.

### *Wiley<sup>12</sup>*

Wiley heeft een workflow uitgewerkt in vier stappen. Eerst dient de auteur de 'data sharing policy' van het tijdschrift waarin hij wil publiceren te bestuderen. Vervolgens worden de data ingediend, samen met het artikel. Als de data gelijktijdig worden aangeleverd, zijn deze beschikbaar voor peer review. Als het artikel wordt geaccepteerd, worden de datafiles automatisch gedeponeerd in de Figshare data repository van het tijdschrift en bij verschijnen van het artikel publiek toegankelijk gemaakt met een CC0 licentie in Figshare. Artikel en data verwijzen naar elkaar d.m.v. links en er wordt een data availability statement gegenereerd.

---

<sup>11</sup> <https://www.elsevier.com/books-and-journals/enrichments/open-data>, (bekeken op 10 februari 2017)

<sup>12</sup> Zie [http://media.wiley.com/assets/7307/97/Data\\_Sharing\\_Service.pdf](http://media.wiley.com/assets/7307/97/Data_Sharing_Service.pdf) en <https://authorservices.wiley.com/author-resources/Journal-Authors/licensing-and-open-access/open-access/data-sharing.html> (bekeken op 10 februari 2017)



Figuur 3: Wiley workflow voor het delen van data in Figshare

## Springer / Nature<sup>13</sup>

Springer / Nature heeft onlangs, vanuit het besef dat het research data policy landschap te complex is geworden en in een poging om te streven naar meer harmonisatie, vier data policy typen gelanceerd. Afhankelijk van de mate waarin de eigen onderzoeksgemeenschap het delen van data omarmt, kunnen tijdschriftredacties kiezen welk type het beste past bij de eigen gemeenschap.

- type 1: data delen en data citatie wordt aangemoedigd
- type 2: data delen en bewijs van data delen wordt aangemoedigd
- type 3: data delen wordt aangemoedigd en 'data availability statements' zijn verplicht
- type 4: data delen, bewijs van data delen en peer review van de data zijn verplicht.

## Policy Types



Figuur 4: Springer / Nature policy types

<sup>13</sup> <http://blogs.nature.com/ofschemesandmemes/2016/07/05/promoting-research-data-sharing-at-springer-nature> (bekeken op 3 mei 2017)

## *PLoS One*<sup>14</sup>

Megajournal *Public Library of Science* is begonnen als Open Access-tijdschrift en heeft snel een hoge impact weten op te bouwen, waardoor het een voorbeeld is binnen de OA-gemeenschap. Onderdeel van het open beleid is ook een streng beleid op het gebied van data: “PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.” PLoS accepteert geen artikelen waarvan de data niet toegankelijk zijn omwille van octrooien, omwille van verder onderzoek of als de conclusies puur zijn gebaseerd op met eigendomsrechten beschermde data (van de farmaceutische industrie bijvoorbeeld). Wel bestaat de optie om data die om ethische of praktische redenen niet in een repository gedeeld kunnen worden, beschikbaar te maken op aanvraag. Auteurs leveren bij hun artikel een Data Availability Statement.

De policy vermeldt duidelijk dat indien data niet beschikbaar zijn na publicatie, het artikel een correctie of zelfs retractie kan krijgen en dat contact met auteurs, instituut en financier wordt opgenomen. De beschikbaarheid van de data lijkt echter niet standaard door PLoS gecontroleerd te worden, zoals ook peer review van data niet standaard is, ondanks dat er in de policy over peer review van data wordt gesproken.

De data policy van PLoS is erg uitgebreid, van aanwijzingen voor het anonimiseren van data tot een lijst met voorbeeldrepositories naar onderwerp. PLoS heeft een Data Repository Integration Partner Program om samenwerking met repositories te bevorderen. Daarnaast kunnen data ook als supplementary material aangeleverd worden. Kortom, de uitgebreide en heldere beschrijving van PLoS is een voorbeeld als het om data availability gaat, waarin alleen standaard peer review van data ontbreekt.

## 5. Data archieven en data citatie

We bekijken van een aantal repositoria die door uitgevers worden ‘aanbevolen’ wat zij doen om citatie van datasets mogelijk te maken, wat hun beleid is t.a.v. supplementary materials, copyright e.d.. De inventarisatie staat op een apart tabblad in ons bestand.

Van de 10 repositoria of data archieven waarvan de gegevens zijn verzameld, zien we geen opmerkelijke verschillen tussen de generieke en de discipline-specifieke data repositoria. Ongeveer de helft van de repositoria verstrekt informatie over hoe de dataset geciteerd moet worden, en enkelen verstrekken ook het aantal citaties. Het merendeel maakt het linken van de dataset naar de gerelateerde publicatie mogelijk. Om data beter in het publicatieproces te integreren zou een review van data, vergelijkbaar met de zgn. peer-review van artikelen, een belangrijke bijdrage kunnen leveren. Echter, door geen van de bekeken repositoria wordt deze mogelijkheid geboden, of wordt vermeld dat dit in de publicatieworkflow is opgenomen.

## 6. Conclusie

Uit onze inventarisatie van databeleid van de uitgevers blijkt dat er veel diversiteit is met betrekking tot de richtlijnen en dat, om de termen van Springer/ Nature te gebruiken, te veel tijdschriften nog

---

<sup>14</sup> <http://journals.plos.org/plosone/s/data-availability> (bekeken op 3 mei 2017)

een 'type zero' beleid hebben, d.w.z. dat zij geen beleid hebben voor data of dat zij data beschouwen als traditionele 'supplementary materials'. Dit heeft tot gevolg, zoals ook uit meerdere studies in onze literatuurlijst blijkt, dat data niet of slecht beschikbaar zijn voor replicatie, validatie of hergebruik. We formuleren daarom een aantal aanbevelingen voor de verschillende partijen die betrokken zijn bij de wetenschappelijke communicatie.

- Uitgevers kunnen:
  - eisen dat data behorende bij publicaties publiek toegankelijk is in duurzame formaten die geschikt zijn voor hergebruik;
  - het gebruik van gecertificeerde, duurzame archieven promoten;
  - in artikelen het citeren van data bevorderen;
  - peer review van data organiseren;
  - op z'n minst policy type 3, maar bij voorkeur type 4 van de Springer / Nature DAP indeling hanteren;
  - gezamenlijk naar harmonisatie van richtlijnen streven.
- Data archieven kunnen:
  - peer review van de data mogelijk maken vóór publicatie van een studie;
  - het citeren van data vergemakkelijken;
  - het linken naar de publicatie vergemakkelijken;
  - het aantal citaties van een datasets bijhouden.
- Bibliotheken (of onderzoeksinstituten) kunnen:
  - onderzoekers informeren over Data Availability Policy's en richtlijnen in hun vakgebied;
  - onderzoekers adviseren over de beste manier om data onderliggend aan hun publicatie beschikbaar te maken;
  - onderzoekers helpen om data te deponeren in een duurzaam archief;
  - bijdragen aan een (internationale) tool voor de registratie van Data Availability Policies (zodra de wenselijkheid daarvan is vastgesteld, zie hieronder).

Aanvankelijk was het plan om, na een eerste steekproef, onze inventarisatie uit te breiden naar alle voor UKB instellingen relevante tijdschriften en uitgevers en de informatie gezamenlijk up-to-date te houden in een gemeenschappelijke tool. Dit zou bij voorkeur opgezet worden in samenwerking met het Landelijke Coördinatiepunt RDM. De tool zou een instrument zijn voor publicatie advies en onderzoekers en ondersteuners helpen om snel een overzicht te krijgen van de eisen die de diverse tijdschriften. Een vergelijkbare tool is in de UK al eens ontwikkeld voor sociaal-wetenschappelijke tijdschriften in het project JORD<sup>15</sup>. Dit project is echter beëindigd bij gebrek aan financiering. Tijdens onze inventarisatie hebben we gezien dat de ontwikkelingen snel gaan en informatie snel verouderd. Een dergelijke tool vereist dus een solide organisatie voor het beheer. Daar komt bij dat uitgevers veelal internationaal opereren en het wellicht beter is een dergelijke tool in internationale samenwerking op te zetten. Een uitbreiding van Sherpa Romeo met data policy's ligt voor de hand. Uit een recente discussie op de door JISC gefaciliteerde Research Dataman discussielijst blijkt dat internationaal behoefte is aan meer inzicht in het databeleid van uitgevers en de invloed van dat beleid op de replicerbaarheid van onderzoek<sup>16</sup>. De studie van L. Naughton & D. Kernohan (2016) , werpt echter de vraag op of het niet beter is eerst te streven naar meer harmonisatie en

---

<sup>15</sup> <https://jordproject.wordpress.com/reports-and-article/journals-and-their-policies-on-research-data-sharing/>

<sup>16</sup> <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=RESEARCH-DATAMAN>

standaardisatie voordat het zin heeft om een 'data policy registry' in het leven te roepen. ANDS geeft een aanzet tot harmonisatie met een 'Research data for journal editors' gids<sup>17</sup>. Een ander voorbeeld is Dryad dat tijdschriftredacties de elementen aanlevert die in een 'journal data policy' zouden moeten staan<sup>18</sup>. Springer / Nature geeft concreet invulling aan dit streven met het lanceren van vier standaard policy's. ANDS en Springer behoren tot de initiatiefnemers voor een Interest Group van de Research Data Alliance over 'Data policy standardisation and implementation'<sup>19</sup>.

Op dit moment, en met de huidige middelen, lijkt het ons als UKB werkgroep Research Data niet doenlijk om een accurate, gezamenlijke tool te ontwikkelen en te onderhouden. Wel stellen we het excelbestand met onze bevindingen samen met deze rapportage voor iedereen ter beschikking. Het kan interessant zijn om onze inventarisatie over een aantal jaar te herhalen om te zien óf en hoe het beleid van uitgevers zich ontwikkelt. De UKB werkgroep Research Data heeft zich dan ook voor 2017 tot doel gesteld de bevindingen uit deze rapportage breder onder de aandacht te brengen en de vinger aan de pols te houden bij de ontwikkelingen in de wereld van de wetenschappelijke communicatie.

## 7. Bronnen

De afgelopen jaren is er veel geschreven en gepubliceerd over uitgevers(beleid) en data. Hieronder volgt een selectie van bronnen.

### a. Data journals

- Candela, L., Castelli, D., Manghi, P. and Tani, A. (2015), Data journals: A survey. J Assn Inf Sci Tec, 66: 1747–1762. <http://doi.org/10.1002/asi.23358>

"In this study of more than 100 currently existing data journals, we describe the approaches they promote for data set description, availability, citation, quality, and open access. We close by identifying ways to expand and strengthen the data journals approach as a means to promote data set access and exploitation."

- Projectblog: Preparate: <http://www2.le.ac.uk/projects/preparate>

List of Data Journals: <http://proj.badc.rl.ac.uk/preparate/blog/DataJournalsList>

**Peer REview for Publication & Accreditation of Research Data in the Earth sciences (PREPARDE)** is a UK JISC funded international project that brings together a wide range of experts in research, academic publishing and data management to produce data publication guidelines applicable across a range of research disciplines and data types.

- Blogpost: Data@MLibrary, Katherine Akers: A Growing list of Data Journals, May 9, 2014, <https://mlibrarydata.wordpress.com/2014/05/09/data-journals/>

### b. Data Availability Policy's

- Vasilevsky NA, Minnier J, Haendel MA, Champieux RE. (2017) Reproducible and reusable research: are journal data sharing policies meeting the mark? PeerJ 5:e3208

<https://doi.org/10.7717/peerj.3208>

---

<sup>17</sup> <http://www.ands.org.au/guides/data-policies-and-journals>

<sup>18</sup> [http://wiki.datadryad.org/Journal\\_instructions](http://wiki.datadryad.org/Journal_instructions)

<sup>19</sup> <https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation>, 2 mei 2017.

“Our study confirmed earlier investigations which observed that only a minority of biomedical journals require data sharing, and a significant association between higher Impact Factors and journals with a data sharing requirement. Moreover, while 65.7% of the journals in our study that required data sharing addressed the concept of reproducibility, as with earlier investigations, we found that most data sharing policies did not provide specific guidance on the practices that ensure data is maximally available and reusable.”

- Sturges, P., Bamkin, M., Anders, J. H.S., Hubbard, B., Hussain, A. and Heeley, M. (2015), Research data sharing: Developing a stakeholder-driven model for journal policies. *J Assn Inf Sci Tec*, 66: 2445–2455. <http://doi.org/10.1002/asi.23336>

“In the absence of a consolidated infrastructure to share data easily, a model journal data sharing policy was developed by comparing quantitative information from analyzing existing journal data policies with qualitative data collected from stakeholders. This article summarizes and outlines the process by which the model was developed and presents the model journal data sharing policy.”

- Recommendations for the Role of Publishers in Access to Data, *Plos Biology*, October 28, 2014

Jennifer Lin, Carly Strasser <http://dx.doi.org/10.1371/journal.pbio.1001975>

As appeals for public access of research data continue to proliferate, many scholarly publishers— alongside funders, institutions, and libraries—are expanding their role to address this need. Here we outline eight recommendations and a set of suggested action items for publishers to promote and contribute to increasing access to data.

- Stodden V, Guo P, Ma Z (2013) Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals. *PLoS ONE* 8(6): e67111.

doi:10.1371/journal.pone.0067111

This article extends the literature by studying journal data sharing policies by year (for both 2011 and 2012) for a referent set of 170 journals. [...]We build a predictive model of open data and code policy adoption as a function of impact factor and publisher and find higher impact journals more likely to have open data and code policies and scientific societies more likely to have open data and code policies than commercial publishers. We also find open data policies tend to lead open code policies, and we find no relationship between open data and code policies and either supplemental material policies or open access journal status. Of the journals in this study, 38% had a data policy, 22% had a code policy, and 66% had a supplemental materials policy as of June 2012. This reflects a striking one year increase of 16% in the number of data policies, a 30% increase in code policies, and a 7% increase in the number of supplemental materials policies. We introduce a new dataset to the community that categorizes data and code sharing, supplemental materials, and open access policies in 2011 and 2012 for these 170 journals.

- Piwowar, H. A., & Chapman, W. W. (2008). Identifying Data Sharing in Biomedical Literature. *AMIA Annual Symposium Proceedings*, 2008, 596–600. Retrieved from:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655927/?tool=pubmed>

- Nuijten, M. B., Borghuis, J., Veldkamp, C. L. S., Alvarez, L. D., van Assen, M. A. L. M., & Wicherts, J. M. (2017, July 13). Journal Data Sharing Policies and Statistical Reporting Inconsistencies in Psychology. Retrieved from [psyarxiv.com/sgbta](http://psyarxiv.com/sgbta)

Overige bronnen:

- [presentatie] “Replication and Economics Journal Policies”; Jan H. Hoeffler, University of Goettingen, 7 januari 2017 op Annual Meeting van American Economic Association , thema on “Replication and Ethics in Economics: Thirty Years After Dewald, Thursby and Anderson, retrieved from <https://www.aeaweb.org/conference/2017/preliminary/1530?page=5&per-page=50>, 10 februari 2017.

We investigate the impact of the introduction of replication policies for leading journals in economics on citations. As has previously been shown for other social sciences, there is an indication that the introduction of a replication policy increases the number of citations for a journal, presumably because readers use the data for their own investigation, possibly also because of a reliability effect. We see our results as an incentive for journals to introduce and enforce replication policies. Lamentably, only a minority of journals so far enforce their policies in a way that ensures replicability of most of the empirical work. With several examples we show how replication becomes difficult if policies are not enforced, and we suggest a pool of replicability editors as a solution: Since it would be too much to expect from journals to have experts for every single topic and software package, a joint effort of journals for such a pool of experts could help to ensure each empirical study is published with data, code, and instructions how to use them together such that all published results can easily be replicated. Reviewers can join the effort for replicability by following the principles of the Agenda for Open Research and refuse to comprehensively review empirical work that does not guarantee fully replicable empirical results. Further study is needed to investigate the citation impact on single articles, and we suggest a design for such research.

- [blogpost] “Journal policies that encourage data sharing prove extremely effective”, LSE Impact blog, retrieved from <http://blogs.lse.ac.uk/impactofsocialsciences/2017/09/05/journal-policies-that-encourage-data-sharing-prove-extremely-effective/>, October 6, 2017

There is currently little incentive for researchers to share their data. But what if it was enough for journals to simply ask authors to make their data available? Michèle B. Nuijten reports on a recent study that found journal policies that encourage data sharing to be extremely effective, with a steep increase in the percentage of articles with open data from the moment these policies took effect. Even something as seemingly frivolous as offering a badge to display on your paper as a reward for sharing data can have a transformative effect, not only on the overall availability of data but also on its relevance, usability and completeness, as well as on the rigour and quality of science as a whole.

- [Guide] Research data for journal editors.

This ANDS Guide provides a starting point for Journal Editors considering developing or improving research data availability policies for their journals.

<http://www.andis.org.au/working-with-data/publishing-and-reusing-data/data-journals/data-policies-and-journals>

### **c. Onderzoeken naar (niet) beschikbaarheid data van journals met een DAP:**

- Roche, D. G., Kruuk, L. E. B., Lanfear, R., & Binning, S. A. (2015). Public Data Archiving in Ecology and Evolution: How Well Are We Doing? PLOS Biology, 13(11), e1002295.

<http://doi.org/10.1371/journal.pbio.1002295>

“Policies that mandate public data archiving (PDA) successfully increase accessibility to data underlying scientific publications. However, is the data quality sufficient to allow reuse and reanalysis? We surveyed 100 datasets associated with nonmolecular studies in journals that commonly publish ecological and evolutionary research and have a strong PDA policy. Out of these datasets, 56% were incomplete, and 64% were archived in a way that partially or entirely prevented reuse. We suggest that cultural shifts facilitating clearer benefits



to authors are necessary to achieve high-quality PDA and highlight key guidelines to help authors increase their data's reuse potential and compliance with journal data policies.”

- Vanpaemel, W., Vermorgen, M., Deriemaeker, L., & Storms, G. (2015). Are We Wasting a Good Crisis? The Availability of Psychological Research Data after the Storm. *Collabra*, 1(1), Art. 3. DOI: <http://doi.org/10.1525/collabra.13>

“To study the availability of psychological research data, we requested data from 394 papers, published in all issues of four APA journals in 2012. We found that 38% of the researchers sent their data immediately or after reminders. These findings are in line with estimates of the willingness to share data in psychology from the recent or remote past. Although the recent crisis of confidence that shook psychology has highlighted the importance of open research practices, and technical developments have greatly facilitated data sharing, our findings make clear that psychology is nowhere close to being an open science.”

#### d. data policies en data archiveren

- Vlaeminck, Sven; Herrmann, Lisa-Kristin (2015) : Data Policies and Data

Archives: A New Paradigm for Academic Publishing in Economic Sciences?, In: Schmidt, Birgit Dobрева, Milena (Ed.): *New Avenues for Electronic Publishing in the Age of Infinite Collections and Citizen Science*. Proceedings of the 19th International Conference on Electronic Publishing, IOS Press, Amsterdam, pp. 145-155; <http://dx.doi.org/10.3233/978-1-61499-562-3-145>

In our paper we summarise the findings of an empirical study in which a sample of 346 journals in economics and business studies were examined. We regard both the extent and the quality of journals' data policies, which should facilitate replications of published empirical research. The paper presents some characteristics of journals equipped with data policies and gives some recommendations for suitable data policies in economics and business sciences journals. In addition, we also evaluate the journals' data archives to roughly estimate whether these journals really enforce data availability. Our key finding is that we are currently not able to determine a new publishing paradigm for journals in economic sciences.

#### e. data policies en de rol van bibliotheken / instellingen

- Vlaeminck, Sven (2013) : Data Management in Scholarly Journals and Possible Roles for Libraries - Some Insights from EDaWaX, *LIBER Quarterly*, ISSN 2213-056X, Vol. 23, Iss. 1, pp. 49-79, <http://nbn-resolving.de/urn:NBN:NL:UI:10-1-114595>

In this paper we summarize the findings of an empirical study conducted by the EDaWaX-Project. 141 economics journals were examined regarding the quality and extent of data availability policies that should support replications of published empirical results in economics. This paper suggests criteria for such policies that aim to facilitate replications. These criteria were also used for analysing the data availability policies we found in our sample and to identify best practices for data policies of scholarly journals in economics. In addition, we also evaluated the journals' data archives and checked the percentage of articles associated with research data. To conclude, an appraisal as to how scientific libraries might support the linkage of publications to underlying research data in cooperation with researchers, editors, publishers and data centres is presented.

- (Poster) Local assessment of science RDM practices as examined through journal policies, Dylanne Dearborn, Steve Marks, University of Toronto Libraries; IDCC conference 2015

[http://www.dcc.ac.uk/sites/default/files/documents/IDCC15/210\\_rdmpractices.pdf](http://www.dcc.ac.uk/sites/default/files/documents/IDCC15/210_rdmpractices.pdf), retrieved 10 februari 2017; NB: het onderzoek is herhaald en resultaten worden op IDCC17 gepresenteerd.

Research data management practices vary greatly across areas within the physical and applied sciences. Challenges for librarians and other data service providers include understanding the differences in disciplinary practices and identifying researchers receptive to engaging with RDM services. Prominent journals requiring that supplementary data be made publicly available could be one possible driver of increased interest in the

development of institutional support for RDM practices in the sciences. Using this driver as an instrument to identify potential “clients”, we propose a replicable methodology for local assessment by examining high impact journal policies and harvesting institutional publishing data to determine current local RDM practice across disciplines. The results of the data gathered in this analysis can supplement traditional researcher interviews and identify clusters to engage in further RDM discussion or projects, as well as allow targeted services, training and/or advocacy efforts.

#### **f. Over (on)mogelijkheid data policy registry:**

- Naughton, L. & Kernohan, D., (2016). Making sense of journal research data policies. Insights. 29(1), pp.84–89. DOI: <http://doi.org/10.1629/uksg.284>

“The project undertook an analysis of 250 journal research data policies to assess the feasibility of developing a policy registry to assist researchers and support staff to comply with research data publication requirements. The evidence shows that the current research data policy ecosystem is in critical need of standardization and harmonization if such services are to be built and implemented.”

- Projectblog: JoRD, Journal Research Data Policy Bank, <https://jordproject.wordpress.com/reports-and-article/journals-and-their-policies-on-research-data-sharing/>

#### **g. overzichten van uitgeversbeleid t.a.v. onderzoeksdata**

- Universiteit Wageningen: <http://www.wageningenur.nl/en/Expertise-Services/Data-Management-Support-Hub/Browse-by-Subject/Journal-requirements.htm>

- Radboud universiteit: <http://www.ru.nl/library/@1020388/journal-requirements/>

- Sociale Wetenschappen: <https://jordproject.wordpress.com/project-data/social-science-journals-that-have-a-research-data-policy/> (>2014)

- Economie (mei 2014): [http://www.edawax.de/wp-content/uploads/2016/02/Data-Policies\\_mit-Links.pdf](http://www.edawax.de/wp-content/uploads/2016/02/Data-Policies_mit-Links.pdf)

- Fairsharing: <https://fairsharing.org/policies/>

## Bijlage I : vragen bij inventarisatie data availability policies & data repositories

<b>Policy</b>	<b>Possible answers</b>
Data sharing	Unknown / no mention Required as condition of publication, barring exceptions Required but may not affect editorial decisions Explicitly encouraged/addressed, may be reviewed and/or hosted Implied
Code sharing	Unknown / no mention Required as condition of publication, barring exceptions Required but may not effect editorial decisions Explicitly encouraged/addressed, may be reviewed and/or hosted Implied
Supplemental materials	Unknown / no mention Required as condition of publication, barring exceptions Required but may not effect editorial decisions Explicitly encouraged/addressed, may be reviewed and/or hosted Implied
Repositories	Unknown / no mention List of required repositories Explicitly encouraged/addressed, may be reviewed and/or hosted
Names	--- open question
Link to policy	--- open question
Data citation	Unknown / no mention Mentioned with citation style Mentioned
Policy Status	Unknown / no mention Same for all journals Some journals may have other policies
Policy status journals	Unknown / no mention Specific to this journal Publisher policy
Open Access	Yes No Unknown / no mention
<b>Repository</b>	
Commercieel	Unknown/No mention Academic Business
Citaties	Refers to citing publications Refers to original publication

	No references Unknown	
Type data	Field specific General	
Set size	Small: underlying publication (supplementary) Only larger sets Supplementary and broader data	
Data review	Unknown/no mention Full review Review of metadata and file format Review of file format	
Copyrights	Unknown/no mention CC-0 CC-BY  Publisher Author	Free to use Free to use with attribution Publisher/repository gets copyright Copyright remains with author
Supplementary	Allowed Not allowed (only broader) Unknown/No mention	Underlying publication