

Inventarisatie

UKB Werkgroep Research Data Management

Sub-werkgroep Datapublicatie en -citatie

Hans Fransen

April 2014

Rapportage

Doelstelling

De sub-werkgroep heeft zich ten doel gesteld om:

- a. Een overzicht maken van manieren waarop data gepubliceerd worden met typische voorbeelden van de verschillende verschijningsvormen.
- b. Een overzicht te maken van wijzen waarop data geciteerd worden met voorbeelden van citaties bij de belangrijkste wetenschappelijke disciplines en aangeven welke ontwikkeling het citeren van datasets de afgelopen jaren heeft doorgemaakt totaal en bij de belangrijkste disciplines. Tevens wil de groep aangeven welke indicatoren een rol kunnen spelen bij de bevordering van het citeren van datasets.

Tijdens de inventarisatie bleek al snel dat het citeren van datasets nog niet die brede omvang heeft om een goed overzicht te kunnen maken voor de verschillende disciplines. Daarom is het werk ingeperkt tot een overzicht bieden van de belangrijkste ontwikkelingen op dit gebied.

Samenstelling van de sub-werkgroep

- Maarten van Bentum, Servicecentrum Bibliotheek & Archief, Universiteit Twente
- Jan de Boer, Universiteitsbibliotheek, Universiteit Utrecht
- Hans Fransen, Universiteitsbibliotheek, Universiteit Leiden
- Harrie Knippenberg, Universiteitsbibliotheek, Radboud Universiteit Nijmegen
- Jaap de Lange, Bibliotheek TU Delft

Werkwijze

Literatuur en internetadressen die in 2013 zijn verzameld in het kader van het project Front Office Datamanagement van UB Leiden vormde een goede basis voor verder onderzoek. Ook is gebruik gemaakt van informatie die verkregen is tijdens bezoeken aan congressen en workshops zoals het 8th International Digital Curation Conference in Amsterdam januari 2013 en de Knowledge Exchange workshop: Making Data Count in Berlijn april 2013.

Hoewel er een duidelijke samenhang is tussen het publiceren van data en het citeren van gepubliceerde data zijn deze onderwerpen in dit rapport apart behandeld.

Voor beide onderwerpen is dankbaar gebruik gemaakt van publicaties en informatie van organisaties die zeer actief zijn met betrekking tot het toegankelijk maken van de wetenschappelijke data en het zoeken naar een goede inbedding van data in de wetenschappelijke communicatie. Met name zijn dat:

- ICSU's Committee on Data for Science and Technology (CODATA)
 - Data Publication Working Group (<http://www.icsu-wds.org/community/working-groups/data-publication>)
 - Data Citation Standards and Practices Task Group (<http://www.codata.org/taskgroups/TGdatacitation/index.html>)
- Force11 (<https://www.force11.org/>)
 - The data citation synthesis group (<https://www.force11.org/node/4432>)
- De Research Data Alliance (RDA) Research Data Alliance (<https://rd-alliance.org/>)
 - RDA Data Citation WG (<https://rd-alliance.org/working-groups/data-citation-wg.html>)
- DataCite (<https://www.datacite.org/>),
- Data Observation Network for Earth (DataONE) (<http://www.dataone.org/>)
- SURF (<http://www.surf.nl/themas/onderzoek/management-van-onderzoeksdata/index.html>)
 - SURF Special Interest Group (SIG) Research Data (<http://www.surf.nl/themas/onderzoek/management-van-onderzoeksdata/special-interest-group-sig-research-data/index.html>)

Publiceren

Er is een inventarisatie gemaakt van plaatsen waar data en datasets worden gepubliceerd. Het gaat dan om de volgende te onderscheiden dragers, data journals, institutionale repositories, disciplinaire repositories en algemene repositories. Daarnaast is gekeken naar de koppeling van datasets aan artikelen in tijdschriften en naar websites met verzamelingen van repositories. Dit alles is in een Excel bestand opgenomen (zie bijlage 1)
Tevens is een overzicht gemaakt van de diverse aspecten die bij het publiceren van data aan de orde komen. Bij elk van die aspecten is een korte toelichting gegeven en een aantal referenties voor meer informatie.

Citeren

Voorbeelden van citing guidelines zijn verzameld van met name een aantal organisaties die nationaal en internationaal actief zijn met betrekking tot data citatie. (zie bijlage 2)

Daarnaast een overzicht is gemaakt van enkele belangrijke zaken die bij het citeren van data spelen.

Resultaten

In dit hoofdstuk komen zaken aan de orde die spelen bij datapublicatie en datacitatie. Gezien de overvloed aan informatie over deze onderwerpen zullen er steeds slechts enkele voorbeelden worden genoemd.

Publiceren:

Waarom publiceren?

Er is een groot aantal redenen aan te wijzen om tot publicatie van onderzoeksdata over te gaan. Controleerbaarheid van het onderzoek wordt groter, hergebruik van data wordt mogelijk, onderzoek wordt zichtbaarder hetgeen kan leiden tot extra citaties. Ook moet worden voldaan aan de toenemende eisen van onderzoeksfinancierders om data toegankelijk te maken. Onderzoeksinstituten komen met beleid ten aanzien van het publiceren van data en ook de wetenschappelijke tijdschriften willen steeds vaker openheid van data die aan artikelen ten grondslag liggen.

Op veel plaatsen zijn opsommingen te vinden van argumenten om over te gaan op datapublicatie.

Enkele voorbeelden zijn te vinden op websites van universiteiten:

Leiden University Libraries: [Toegang tot onderzoeksdata: waarom?](#)

Oxford University Research Data: [Benefits, Oxford policy, Funder requirements](#)

MIT Libraries - [Reasons to Manage and Publish Your Data](#)

Beleid universiteiten: In het Standard Evaluation Protocol 2015 – 2021 van VSNU en KNAW staan datasets als indicatoren van onderzoekskwaliteit bij "Demonstrable products" en hergebruik van datasets bij "Demonstrable use of products" genoemd. Onderzoek gefinancierd door de EU in het kader van het nieuwe [Horizon 2020](#) programma zal binnen afzienbare tijd ook openbaarmaking van de onderzoeksdata gaan vragen, er loopt nu een pilot.

Een voorbeeld van eisen die een tijdschrift ten aanzien van data stelt is Nature: ".An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. Therefore, a condition of publication in a Nature journal is that **authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications**. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission...."

PLOS heeft net [een nieuw beleid](#) bekendgemaakt ten aanzien van data voor alle journals inclusief het belangrijke open access journal PLOS ONE: "Authors must make all data publicly available, without restriction, immediately upon publication of the article."

Waar publiceren?

Data(sets) worden in repositories of archives opgeslagen. Dit verzekert de terugvindbaarheid en conservering. Er bestaan veel verschillende repositories. Verschillend qua inhoud, omvang en kwaliteit.

Een onderscheid kan gemaakt worden tussen institutionele, nationale, internationale, disciplinaire, en multidisciplinaire repositories of archives. Bijlage 1 geeft een overzicht van de belangrijkste vertegenwoordigers van die groepen.

Een andere plaats waar datasets gepubliceerd kunnen worden zijn de datajournals. Van Nature is onlangs het eerste nummer van [Scientific Data](#) verschenen. Datajournals zijn gevuld met artikelen die databases of datasets beschrijven en koppelingen naar die data bevatten. Voorbeelden van datajournals zijn ook in bijlage 1 te zien.

Reguliere tijdschriften die artikelen koppelen aan onderliggende datasets geven vaak aan in welke repositories die data opgeslagen kunnen worden. Elsevier geeft bijvoorbeeld [een uitgebreide uitleg](#) over het linken van de data en repositories die per discipline gekozen kunnen worden.

Veel repositories zijn opgenomen in grote collecties zoals [Databib](#), [DataCite](#), [BioSharing](#) en [re3data](#).

Deze collecties zoeken ook weer samenwerking: "DataCite, re3data.org, and Databib Announce Collaboration"

<http://www.re3data.org/>

De betrouwbaarheid en kwaliteit van repositories is natuurlijk van belang. DCC heeft een overzicht gemaakt van [repository audit en assessment](#). Een daarvan is de Trusted Repositories Audit & Certification (TRAC). Zij maakten een [Checklist for Audit and Certification](#).

In Nederland heeft DANS sinds 2005 gewerkt aan een datakeurmerk, het [Data Seal of Approval](#). Dit zijn minimeisen voor (wetenschappelijke) data-archieven.

Er bestaan ook rankings, Dryad bijvoorbeeld geeft [een lijst van meest populaire repositories](#) in haar bestand. Evenals [DataCite](#).

The "[The Ranking Web of World repositories](#)" is een initiatief van het Cybermetrics Lab, een Spaanse onderzoeksgroep.

Open Access publiceren

OpenAIRE, Open Access Infrastructure for Research in Europe is de Europese organisatie die in het kader van de onderzoeksfinancieringsprogramma's (FP7, H2020) informatie geeft over open access.

<https://www.openaire.eu/>

Onlangs verscheen met betrekking tot data: "Policy RECommendations for Open Access to Research Data in Europe (RECODE) Project website: www.recodeproject.eu

The first deliverable of the RECODE project on Open Access to research data is now available. The report, [Stakeholder values and ecosystems](#), undertakes a case study analysis of stakeholder values and inter-relationships in five different disciplines; physics, earth sciences, archaeology, health and bioengineering."

Eerder werd met de steun van LIBER en COAR in het kader van Horizon 2020 een Pilot for Open Research Data gestart.

https://www.coar-repositories.org/files/Horizon_2020_Open_Data_Pilot_20130703_final.pdf

Ook werden er guidelines voor data-archieven opgesteld:

https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_Data_Archives

Met CERN, the European Organization for Nuclear Research samen heeft OpenAIRE het research data repository Zenodo gelanceerd:

[http://home.web.cern.ch/about/updates/2013/05/cern-and-openaireplus-launch-european-repository](http://home.web.cern.ch/about/updates/2013/05/cern-and-openaireplus-launch-european-research-repository)

Hoe publiceren?

Er bestaan wel werkprocedures voor het publiceren van data maar die zijn vaak specifiek voor een bepaalde omgeving en veel onderzoekers zijn er niet mee bekend of het voldoet niet aan hun behoefte.

Een voorbeeld van een specifieke instructie is het "IOC Manuals and guides No. 64 - [Ocean Data Publication Cookbook](#)"

"Het zegt: This cookbook provides a step-by-step guide to the data publication process and showcases some best practices for data publication. This cookbook is an outcome of the 5th session of the SCOR/IODE/MBLWHOI Library Workshop on Data Publication."

Een tweede voorbeeld is er op het terrein van de biodiversiteit, de [Integrated Publishing Toolkit \(IPT\)](#) van De Global Biodiversity Information Facility (GBIF) met de GBIF [Integrated Publishing Toolkit User Manual](#)

Een werkgroep van RDA, onderdeel van de RDA-WDS Publishing Data Interest Group gaat bestaande werkprocedures inventariseren en zoeken naar een generieke oplossing. Dit in het kader van het programma [RDA Data Publishing 2020: Proposal for a Coordinated Approach](#) .

Belangrijke aspecten bij het publiceren zijn:

Standards/Metadata

Van wetenschappers wordt steeds meer verwacht dat data gedeeld en hergebruikt kunnen worden en dan wel op een standard conforme manier. Dat is niet eenvoudig want:

- "Different groups and individuals have different definitions for what constitutes a 'data standard'.
- Even within one domain, no one standard is the 'right' standard across all cases; rather, one must select a standard (or even specific pieces of a standard) based on one's particular needs.
- Integrated resources and registries are needed to help researchers navigate the fluid standards landscape and to choose and implement the right standard for their respective project."

(J Am Med Inform Assoc 2014;21:2 193 doi:10.1136/amiajnl-2014-002666)

Toch probeert men bij de diverse disciplines naar een standard toe te werken. [BioSharing](#) bijvoorbeeld werkt aan een data standard voor de life sciences

Het [Data Documentation Initiative \(DDI\)](#) is een poging om internationale standaard te maken voor het beschrijven van data in de sociale, economische en gedragswetenschappen.

Goede metadata zijn onontbeerlijk voor toegang tot en hergebruik van data. Om te bepalen welke metadata gebruikt kunnen worden is een lastige opgave. Veel wetenschappelijke disciplines hebben zich al ingezet voor initiatieven om metadata specificaties te formaliseren. Het Britse Digital Curation Centre geeft [een overzicht van metadata voor de verschillende disciplines](#).

DataCite publiceert het [Metadata Schema](#) "a list of core metadata properties chosen for the accurate and consistent identification of data for citation and retrieval purposes, along with recommended use instructions".

ANDS Best practice guidelines for citation metadata

<http://ands.org.au/guides/cpguide/cpgcitation-bestpractice.html>

ANDS over Metadata <http://www.ands.org.au/metadata/index.html>

Persistent identifier PID

Belangrijk bij de metadatering is het toekennen van een persistent identifier aan een dataset. Hierdoor wordt het citeren van de data gemakkelijk gemaakt, het gebruik van de data kan getraceerd worden en het voegt citaties toe aan die welke de onderzoeker al krijgt voor de gewone publicaties. Het kunnen de 'Digital Object Identifier' (DOI) zijn of andere identifiers zoals Handle of ARC. Zie ook dit [overzicht van de University of California](#). Bij INSPIRE, het informatiesysteem van High Energy Physics, krijgen de datasets een DOI en gebruikers krijgen [een uitleg](#) over hoe deze sets te koppelen aan artikelen. ANDS over Identifiers <http://www.ands.org.au/metadata/identifiers.html>

Overige aspecten van publiceren

Hoewel de meeste data in open access worden gepubliceerd zijn er toch ook voorbeelden van organisaties die "data processing charges" in rekening brengen. [Dryad bijvoorbeeld](#). Als een auteur een artikel plaatst in een open access journal en daarvoor APC betaalt dan kan het voorkomen dat daarbovenop nog een DPC (**data processing charges**) voor Dryad moet worden betaald.

Hoewel **peer review van datasets** zeker niet algemeen gebruikelijk is, wordt op veel plaatsen onderzocht of dit mogelijk en haalbaar is zoals bij de earth sciences die [PREPARDE](#) kennen: the Peer REview for Publication & Accreditation of Research data in the Earth sciences.

Een RDA subwerkgroepje is bezig met de **workflow for publishing data**:

"A number of workflows for publishing data exist, all of them individually adapted to their specific environments, but many researchers are unaware of them or they may not suit their needs. This WG will classify a representative range of existing workflows and identify generic components, which might be reused or repurposed elsewhere. These might include the challenges of technical and research QA/QC and peer-review as well as the role of researchers, their institutions, data centres, academic publishers and funders in the data publication process. The WG will then identify use case(s) in which one or more generic workflow component may be tested so that we can evaluate the benefits for all stakeholders involved and enable improved data discovery and reuse."

De Europese Commissie betaalde mee aan het CHAIN-REDS project (<http://www.chain-project.eu/>)

Die "are working on proposing a Data Accessibility, Reproducibility and Trustworthiness challenge workflow, which aims to cover the whole cycle of a research based on previous data: find a specific dataset (or publication with associated data); retrieve it; use these data to either reproduce some experiments or perform new calculations on a computational platform; obtain new results, publish them; and, store, if any, the new produced data. This cycle is being made possible by the use of data and metadata standards (OAI-PMH, Dublin Core, SPARQL...) and well accepted formats such as Persistent Identifiers. More information can be found in the public document 'D4.2 Analysis of Data Infrastructures and Data repositories' ((scroll down at <http://www.chain-project.eu/deliverables> to find Deliverable D4.2)."

Een andere workgroup van RDA werkt aan **Services for publishing data**: "During the last century, services for publishing data have concentrated on registration of data entities. Services to cross-reference research data and literature or to publish data have only started recently and are limited in scope and functionality. The WG will investigate content and interoperability requirements for data centres and academic publishers. Building on existing components, the WG will concentrate on the conception and implementation of a one-for-all **cross-referencing** service.

Tenslotte word er nog een ander aspect van het publiceren van data onderzocht door een RDA workgroup n.l.: The **costs for publishing data** : "General services to publish data are not available and the necessary editorial process leading to quality assured and efficiently usable data requires resources to be quantified. At present, there is an imbalance between the capacities and functionality of existing data centres and data repositories and the global production of scientific data. Budgets of data centres generally cover a precise scope, mostly data production of the host institution. The WG will supply cost estimates and elaborate a business model to compensate for additional costs of publishing data in an open access environment.

OpenAIRE deed onderzoek naar **Licensing, usage rights, and Privacy** met betrekking tot gepubliceerde data: "Any e-infrastructure which primarily relies on harvesting external data sources (e.g. repositories) needs to be fully aware of any legal implications for re-use of this knowledge, and further application by 3rd parties. OpenAIRE's legal study will put forward recommendations as to applicable licenses that appropriately address scientific data in the context of OpenAIRE."

https://www.openaire.eu/en/about-openaire/publications-presentations/public-project-documents/doc_download/621-openairelicensingstudysummary

Citeren:

Waarom citeren en hoe?

In het vorige hoofdstuk werd aangegeven dat de publicatie van data steeds belangrijker wordt en dat deze vaker worden gezien als een gelijkwaardig product aan een artikel of een boek wat betreft wetenschappelijke output. Het citeren van die data is dan ook een logisch vervolg op die ontwikkeling. Daarbij vereenvoudigt citatie het hergebruik en de verificatie van die data en het geeft krediet aan de auteur. In het The Amsterdam Manifesto on Data Citation Principles werd gesteld: "1. Data should be considered citable products of research." ('Beyond the PDF' conference in Amsterdam, 20 March 2013)

In de aanhef bij de draft Declaration of Data Citation Principles geeft de [Data Citation Synthesis Group](#) aan dat: "Sound, reproducible scholarship rests upon a foundation of robust, accessible data. For this to be so in practice as well as theory, data must be accorded due importance in the practice of scholarship and in the enduring scholarly record. In other words, data should be considered legitimate, citable products of research. Data citation, like the citation of other evidence and sources, is good research practice."

Veel organisaties, repositories, data centres, vakdisciplines en uitgevers gaven guidelines uit over het citeren van data. In bijlage 1 staat een overzicht van de belangrijkste voorbeelden.

Ook in de traditionele styleguides voor journals zoals APA, MLA, Chicago kwamen voorschriften voor het citeren van datasets. Een goed overzicht geeft:

Ball, A. & Duke, M. (2012). 'How to Cite Datasets and Link to Publications'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online:

http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How_to_Cite_Link.pdf

Hierin komen de belangrijkste zaken met betrekking tot het citeren aan de orde zoals de motivatie om data te citeren en te linken aan publicaties, zaken die nu spelen en uitdagingen, waaraan data citaties moeten voldoen, citatie infrastructuur en het implementeren ervan.

Er ontstond op verschillende plaatsen een behoefte aan regulering van al deze initiatieven met betrekking tot het citeren van data en te komen tot aan aantal standaards.

Het eerder genoemde Amsterdam Manifesto kwam met een lijst van acht data citatie principes met als uitgangspunt: "We wish to promote best practice in data citation, to facilitate access to data sets, and to enable attribution and reward for those who publish data."

In de Verenigde Staten bestond de CODATA-ICSTI Task Group on Data Citation Standards and Practices (onder voorzitterschap van Christine Borgman, Jan Brase and Sara Callaghan) al sinds 2010.

Hun werk resulteerde in de publicatie: "Out of cite, out of mind – the current state of practice, policy, and technology for the citation of data". Redactie Yvonne M. Socha.

<http://openscholar.mit.edu/sites/default/files/dept/files/outofcite.pdf>

Gebaseerd op een review van opkomende praktijken en literatuuronderzoek naar de citatiepraktijk werd daarin de volgende lijst van "first principles" voor data citatie opgenomen:

1. Status of Data: Data citations should be accorded the same importance in the scholarly record as the citation of other objects.
2. Attribution: Citations should facilitate giving scholarly credit and legal attribution to all parties responsible for those data.
3. Persistence: Citations should be as durable as the cited objects.
4. Access: Citations should facilitate access both to the data themselves and to such associated metadata and documentation as are necessary for both humans and machines to make informed use of the referenced data.
5. Discovery: Citations should support the discovery of data and their documentation.
6. Provenance: Citations should facilitate the establishment of provenance of data.
7. Granularity: Citations should support the finest-grained description necessary to identify the data.
8. Verifiability: Citations should contain information sufficient to identify the data unambiguously.
9. Metadata Standards: Citations should employ widely accepted metadata standards.
10. Flexibility: Citation methods should be sufficiently flexible to accommodate the variant practices among communities but should not differ so much that they compromise interoperability of data across communities.

Intussen werd door de Force11 Group een vervolg gegeven op het werk van het eerder genoemde Amsterdam Manifesto. Uit verschillende hoeken werd toen contact opgenomen met Force11 om activiteiten en initiatieven te coördineren. CODATA-ICSTI, RDA, ICSU, W3C, DataCite, OpenAire en vele anderen. Zo ontstond in juli 2013 de "Data Citation Synthesis Group".

De groep aanvaardde het volgende "mission statement":

"The data citation synthesis group is a cross-team committee leveraging the perspectives from the various existing initiatives working on data citation to produce a consolidated set of data citation principles (based on the Amsterdam Manifesto, the

CODATA and other sets of principles provided by others) in order to encourage broad adoption of a consistent policy for data citation across disciplines and venues. The synthesis group will review existing efforts and make a set of recommendations that will be put up for endorsement by the organizations represented by this synthesis group. The synthesis group will produce a set of principles, illustrated with working examples, and a plan for dissemination and distribution. This group will not be producing detailed specifications for implementation, nor focus on technologies or tools."

De Synthesis Group publiceerde:

Force11 Manifesto: https://www.force11.org/white_paper

Synthesis group wiki: <https://www.force11.org/node/4432>

Een lijst van deelnemers:

Joint declaration <https://www.force11.org/datacitation>

De laatste staat nu open voor "endorsement" door individuen, maar vooral ook door organisaties.

De declaratie kent de volgende principes:

1. Importance

Data should be considered legitimate, citable products of research. Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications.

2. Credit and Attribution

Data citations should facilitate giving scholarly credit and normative and legal attribution to all contributors to the data, recognizing that a single style or mechanism of attribution may not be applicable to all data.

3. Evidence

In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited.

4. Unique Identification

A data citation should include a persistent method for identification that is machine actionable, globally unique, and widely used by a community.

5. Access

Data citations should facilitate access to the data themselves and to such associated metadata, documentation, code, and other materials, as are necessary for both humans and machines to make informed use of the referenced data.

6. Persistence

Unique identifiers, and metadata describing the data, and its disposition, should persist -- even beyond the lifespan of the data they describe.

7. Specificity and Verifiability

Data citations should facilitate identification of, access to, and verification of the specific data that support a claim. Citations or citation metadata should include information about provenance and fixity sufficient to facilitate verifying that the specific timeslice, version and/or granular portion of data retrieved subsequently is the same as was originally cited.

8. Interoperability and flexibility

Data citation methods should be sufficiently flexible to accommodate the variant practices among communities, but should not differ so much that they compromise interoperability of data citation practices across communities.

Complexiteit van het citeren van data

Datasets verschillen van tijdschriftartikelen omdat ze in een aantal opzichten moeilijker zijn te citeren. Twee problemen die zich voordoen zijn "deep citation" of "granularity" en "dynamic data".

De volgende publicaties geven hier meer informatie over:

REPORT ON BEST PRACTICES FOR CITABILITY OF DATA AND ON EVOLVING ROLES IN SCHOLARLY COMMUNICATION, 21 June 2012, Rachael Kotarski, Susan Reilly, Sabine Schrimpf, Eefke Smit, Karen Walshe. http://www.stm-assoc.org/2012_07_10_STM_Research_Data_Group_Data_Citation_and_Evolving_Roles_ODE_Report.pdf

Citation Developments, Posted on October 11, 2013 by John Kratz in Blog: Data Pub, California Digital Library, Conversations about data.

<http://datapub.cdlib.org/category/data-citation-2/>

Blogpost in: Citing Bytes - Adventures in Data Citation, Tuesday, 26 November 2013, Citing dynamic data, Sarah Callaghan

<http://citingbytes.blogspot.nl/2013/11/citing-dynamic-data.html>

'What to cite: Versioning and granularity of research data for effective citation' <http://www.dcc.ac.uk/blog/data-citations-versions-and-granularity>

Scalable Dynamic Data Citation - RDA-WG-DC Position Paper

<https://rd-alliance.org/groups/data-citation-wg/wiki/scalable-dynamic-data-citation-rda-wg-dc-position-paper.html>

Data metrics

Bibliometrie voor gepubliceerde data bestaat nog nauwelijks. Als het gebeurt dan is het inconsistent en er wordt voornamelijk nog gekeken naar het aantal downloads van datasets.

De volgende publicaties geven meer informatie:

Costas, R., Meijer, I., Zahedi, Z. and Wouters, P. (2013). The Value of Research Data - Metrics for datasets from a cultural and technical point of view. A Knowledge Exchange Report, available from www.knowledge-exchange.info/datametrics

This report was presented at the Workshop: 'Making Data Count: Research data availability and research assessment, 11-12 April 2013, 11:30-13:30, Berlin

<http://www.knowledge-exchange.info/Default.aspx?ID=576>

Thomson-Reuters geeft aan dat haar **The Data Citation Index on the Web of Science™ platform**

"Expose the total impact of research output at your institution, beyond published content"

<http://thomsonreuters.com/data-citation-index/>

Data verbinden en workflow

Het ODIN project:

"ODIN will build on the ORCID and DataCite initiatives to uniquely identify scientists and data sets and connect this information across multiple services and infrastructures for scholarly communication. It will address some of the critical open questions in the area:

- Referencing a data object
- Tracking of use and re-use
- Links between a data object, subsets, articles, rights statements and every person involved in its life-cycle.

<http://odin-project.eu/>

Het VIVO-ISF project van de Coalition for Networked Information (CNI) .

"VIVO is an effort to provide linked open data about researchers in the context of their research activities."

<http://www.cni.org/>

Bijlage 1a

Institutional repositories

SMARTech, Georgia Institute of Technology

<https://smartech.gatech.edu/browse?value=Dataset&type=type>

ICPRSE, interuniversity consortium

<http://icpsr.blogspot.nl/2013/08/new-releases-through-2013-08-11.html>

Edinburgh DataShare

<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository>

Bijlage 1b

Disciplinary repositories

ADS - Astrophysical Data System

<http://adsabs.harvard.edu/abs/2003ApJ...594....1T>

Pangaea (Earth Sciences)

<http://www.pangaea.de/>

BioSharing (Life Sciences, werkt samen met Oxford en re3data)

<http://www.biosharing.org/biodbcore>

ICPSR (social and behavioural sciences)

<http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp>

Woods Hole Scientific Community (WHOAS, Marine Biology))

<http://darchive.mblwhoilibrary.org/>

HepData (physics)

<http://hepdata.cedar.ac.uk/>

The Language Archive, Max Planck Institute for Psycholinguistics, Nijmegen

<http://tla.mpi.nl/resources/data-archive/>

Bijlage 1c

Other repositories

Zenodo	http://zenodo.org/collection/datasets
Figshare	http://figshare.com/
Dryad	http://datadryad.org/
Canadensys	http://www.canadensys.net/publication/data-publication-guide
ECDS	http://www.smhi.se/ecds/Support/Data-management/data-publication-plan-1.19537
EUDAT	https://b2share.eudat.eu/
Force11	https://www.force11.org/catalog#cat:database
Dataverse	http://thedata.org/book/dataverse-networks-around-world
DANS - Easy	https://easy.dans.knaw.nl/
3TU Datacentrum	http://data.3tu.nl/repository/

Bijlage 1d

Repository Collections

re3data.org <http://service.re3data.org/search/results?term=>

Databib	http://databib.org/index.php
DataCite	http://www.datacite.org/repolist
Open Access Directory	http://oad.simmons.edu/oadwiki/Data_repositories
BioMedCentral OpenData	http://www.biomedcentral.com/about/supportingdata
DataOne - OneMercury	https://cn.dataone.org/onemercury/

Bijlage 1e

Data journals:

Overzicht	Sarah Callaghan			http://proj.badc.rl.ac.uk/preparde/blog/DataJournalsList
Discipline	Titel	ISSN	Uitgever	URL
Biology	Database: The Journal of Biological Databases and Curation	1758-0463	Oxford UP	http://database.oxfordjournals.org/content/
Archaeologie	Journal of Open Archaeology Data (JOAD)	2049-1565	Ubiquity Press	http://openarchaeologydata.metajnl.com/
Psychology	Journal of Open Psychology Data (JOPD)	2050-9863	Ubiquity Press	http://openpsychologydata.metajnl.com/
Medical	Open Public Health Data (JOPHD)	2053-2407	Ubiquity Press	http://openpublichealthdata.metajnl.com/
General	Dataset Papers in Science	2314-8497	Hindawi	http://www.hindawi.com/journals/dpis/
Geo Sciences	GeoSciences Data Journal			http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%292049-6060
Biology	Biodiversity Data Journal			http://biodiversitydatajournal.com/
General	Nature - Scientific data	2052-4463	Nature	http://www.nature.com/scientificdata/

Bijlage 2.

Data citing guidelines:	
Naam	URL
DANS	http://www.dans.knaw.nl/content/data-citeren
3TU Datacentre	http://dataintelligence.3tu.nl/iii-technische-vaardigheden/data-citeren/
ANDS Australië	http://ands.org.au/cite-data/
DCC UK	http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/data-citation-and-linking
ICPSR	http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/citations.jsp
DataOne	http://www.dataone.org/citing-dataone
Dryad data citation guidelines	http://wiki.datadryad.org/Citing_Data
Dataverse	http://thedata.org/book/standard
ESRC	http://www.esrc.ac.uk/funding-and-guidance/grant-holders/data-citation.aspx
USGS	http://www.usgs.gov/datamanagement/describe/citation.php
Roper Center	http://www.ropercenter.uconn.edu/data_access/data/how_to_cite.html