

Klump, J. (2012). Offener Zugang zu Forschungsdaten. In U. Herb (Hrsg.), *Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft* (S. 45-53). Saarbrücken: universaar.

URN: [urn:nbn:de:bsz:291-universaar-873](https://nbn-resolving.org/urn:nbn:de:bsz:291-universaar-873)

URL: <http://universaar.uni-saarland.de/monographien/volltexte/2012/87>



Dieses Werk bzw. dieser Inhalt steht unter einer Creative Commons Namensnennung 3.0 Deutschland Lizenz. <http://creativecommons.org/licenses/by/3.0/de/legalcode>

Jens Klump

Offener Zugang zu Forschungsdaten

Open Data und Open Access to Data – Die ungleichen Geschwister

Die Gründe, die für einen offenen Zugang zu Forschungsdaten sprechen, sind vielfältig: Forschungsergebnisse werden besser nachvollziehbar, was zu einer höheren Qualität der Ergebnisse beitragen kann. Daten können für neue Fragestellungen neu analysiert oder mit anderen Daten zusammengeführt werden. In Bereichen, in denen vorwiegend international vernetzte Projekte durchgeführt werden, wie z.B. in der Teilchenphysik, Geophysik, Ozeanographie oder Biodiversitätsforschung, wurden die Vorteile eines offenen Zugangs zu Forschungsdaten schon früh sichtbar. Schon lange vor der Entstehung des Internets wurden Forschungsdaten der Allgemeinheit zur Verfügung gestellt. Ein Beispiel sind die World Data Center des International Council for Science (ICSU), die aus dem International Geophysical Year (1957-1958) hervorgegangen sind. Das System der World Data Center wurde inzwischen modernisiert und in World Data System umbenannt.

Mit dem allgemeinen Zugang zum Internet entstanden sehr bald diverse Initiativen, die sich das Paradigma der „Offenheit“ im Sinne des freien Zugangs und der Nachnutzbarkeit auf die Fahnen schrieben. Auch bei Freier Software (Free and Open Source Software) waren die Barrieren in der Nachnutzbarkeit eine entscheidende Komponente der Motivation. Inspiriert von den Prinzipien der Freien Software entstand 1994 die Open Archives Initiative mit dem Ziel, wissenschaftliches Wissen, das in Fachartikeln veröffentlicht wurde, allen zugänglich zu machen – nicht nur den Abonnenten der jeweiligen Fachzeitschriften (Harnad, 1994). Die Open Archive Initiative, und ähnliche parallele Initiativen, inspirierten 2003 die „Berliner Erklärung über den Offenen Zugang zu wissenschaftlichem Wissen“ (Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities) (Berliner Erklärung, 2003). Der offene Zugang zu Forschungsdaten ist von Anfang an ein integraler Bestandteil der Berliner Erklärung.

In der Erklärung heißt es:

„Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material.“ (Berliner Erklärung, 2003)

Schon bald darauf wurde der offene Zugang zu Forschungsdaten aus öffentlich geförderter Forschung auch auf Regierungsebene thematisiert und als Richtlinie der Organisation for Economic Cooperation and Development (OECD) veröffentlicht (OECD, 2007), die inzwischen in den meisten Mitgliedsstaaten der OECD in die jeweilige nationale Gesetzgebung überführt wurde.

Damit stand der offene Zugang zu Forschungsdaten von Anfang an mit auf der Agenda der Initiative für den offenen Zugang zu wissenschaftlichem Wissen. Zudem war damit auch definiert, was mit dem Begriff „Open Access“ in Bezug auf Forschungsdaten gemeint war. Die Umsetzung dieser Ziele gestaltete sich jedoch schwierig, denn Forschungsdaten sind ein wesentlich komplexeres Feld als Texte, denn sie erscheinen in einer Vielzahl von Formaten, sie weiter zu nutzen erfordert umfangreichere Metadaten als rein bibliographische Angaben. Ungeachtet dessen haben es sich verschiedene Projekte zum Ziel gesetzt, den Zugang zu Forschungsdaten zu verbessern (Pfeiffenberger & Klump, 2006).

Um den Offenen Zugang zu Forschungsdaten mit anderen Open Data Initiativen zu vergleichen ist es wichtig, die Motivation der Initiatoren zu beleuchten. Auf der Webseite der Schwerpunktinitiative Digitale Information der Allianz der deutschen Wissenschaftsorganisationen heißt es dazu:

„Qualitätsgesicherte Forschungsdaten bilden einen Grundpfeiler wissenschaftlicher Erkenntnis und können unabhängig von ihrem ursprünglichen Erhebungszweck vielfach Grundlage weiterer Forschung sein. Dies gilt namentlich für die Aggregation von Daten aus unterschiedlichen Quellen zur gemeinsamen Nutzung. Die nachhaltige Sicherung und Bereitstellung von Forschungsdaten dient daher nicht nur der Prüfung früherer Ergebnisse, sondern in hohem Maße auch der Erzielung künftiger Ergebnisse. Sie bildet eine strategische Aufgabe, zu der Wissenschaft, Politik und andere Teile der Gesellschaft gemeinsam beitragen müssen.“ (Grundsätze zum Umgang mit Forschungsdaten, 2010)

Wenngleich offen formuliert, ist die hier vorgestellte Motivation in erster Linie durch die Sicht der Wissenschaftsorganisationen geprägt, welche die von ihnen geförderte Forschung allgemein weiter verbreitet sehen wollen, denn sie stehen der Gesellschaft gegenüber in einer Bringschuld. Ganz anders ist dagegen die Sicht der Forscher, die in erster Linie auf ihre Reputation und auf

die Wahrnehmung ihrer Arbeiten durch ihre Peers angewiesen sind. Dadurch, dass die Initiative für den offenen Zugang von Forschungsdaten von den Wissenschaftsorganisationen und Hochschulen ausging, ist Open Access to Data eine top-down Initiative und unterscheidet sich damit in seinem Ursprung von anderen Open Data Initiativen, die aus Basisinitiativen hervorgegangen sind. Es gibt auch Zwischenformen, wie z.B. Science Commons, die sowohl Basisinitiative sind, als auch institutionell unterstützt werden. Aufbauend auf den Vorarbeiten in Science Commons (Wilbanks, 2007) veröffentlichte die Open Knowledge Foundation die „Panton Principles“ (Murray-Rust, Neylon, Pollock & Wilbanks, 2010) über den offenen Zugang zu Forschungsdaten, die jedoch von den deutschen Wissenschaftsorganisationen nicht übernommen wurden. Somit unterscheiden sich die Initiatoren von Open Access to Data und Open Data und deren Motivation bereits deutlich. Auch die Definitionen von „offen“ sind uneinheitlich (s. dazu den Beitrag von Herb in diesem Band).

Noch gravierender werden die Unterschiede zwischen Open Access to Data und Open Data wenn man die angesprochenen Datenproduzenten betrachtet, die aufgefordert werden, ihre Daten zugänglich zu machen. Open Data Initiativen richten sich in erster Linie an staatliche Stellen. Diese Stellen werden von der jeweiligen Initiative aufgefordert, Daten, die sie im Rahmen ihrer Aufgaben erheben, den Bürgern frei zugänglich zur Verfügung zu stellen. Das Kernargument der Open Data Initiativen ist meist, dass die Datenerhebung durch staatliche Stellen bereits durch Mittel aus dem Steueraufkommen finanziert wurde und Bürger deshalb nicht auch noch für die Nutzung der Daten bezahlen sollten. Das Argument ist hier stark verkürzt dargestellt, da die Motivation unterschiedlicher Initiativen bereits in anderen Kapiteln dieses Buchs weit ausführlicher behandelt wird. Im Unterschied zu Open Access to Data sind die Datenproduzenten an die Weisungen des jeweiligen Dienstherrn gebunden, die „Belohnung“ erfolgt für die Ausführung der Weisung.

Völlig anders sieht die Ausgangslage für Open Access to Data aus. Die hier gesammelten oder produzierten Daten sind die Grundlage des Renommee des Forschers, es besteht ein ausgeprägter *sense of ownership*. In ihrem Wertesystem habe diese Akteure gute Gründe, ihre Daten nicht offen zugänglich zu machen, denn die Arbeit, die notwendig ist, um Daten so zu dokumentieren, dass sie für andere nachnutzbar sind, wird im Wissenschaftssystem bisher nicht honoriert (Abb. 1) (Borgman, 2010). Das heißt, dass im Wertesystem der Wissenschaftler die Aufgabe Daten zu dokumentieren immer mit anderen Aufgaben, wie z.B. neue Artikel schreiben oder Fördergelder einwerben, konkurriert. Sie hat einfach nicht den notwendigen Stellenwert und bleibt selbst bei Wissenschaftlern, die dem Offenen Zugang zu Daten wohlwollend gegenüber stehen, unerledigt.

Es gibt also kaum eine intrinsische Motivation, die Wissenschaftler dazu zu bewegen könnte, ihre Daten so zu dokumentieren, und anderen zur Verfügung zu stellen. In den heutigen Strukturen, in denen sich Verfahren zum effizienten Datenmanagement und zur Datenveröffentlichung erst herausbilden, sind Motive wie Fortschritt der Wissenschaft, Transparenz, Synergien nur schwache Anreize. Künftige Forschungsdateninfrastrukturen werden ein überzeugendes Dienstleistungsangebot für Forscher als Datenproduzenten anbieten müssen. Schon heute ist sichtbar, dass Artikel, deren Daten zugänglich gemacht wurden, deutlich häufiger zitiert werden (Botstein, 2010; Piwowar, Day & Fridsma, 2007). Datenveröffentlichung wird in Zukunft ein attraktives Angebot sein, wenn die dafür notwendigen Strukturen weiter ausgebaut sind und für mehr Wissenschaftsdisziplinen zur Verfügung stehen. Mit der Verbreitung des „Cloud“-Paradigmas wird auch die Speicherung von Daten als Dienstleistung für Forscher neue Möglichkeiten bieten, die Langzeiterhaltung von Forschungsdaten zu systematisieren (z.B. Schaeffer, 2011).

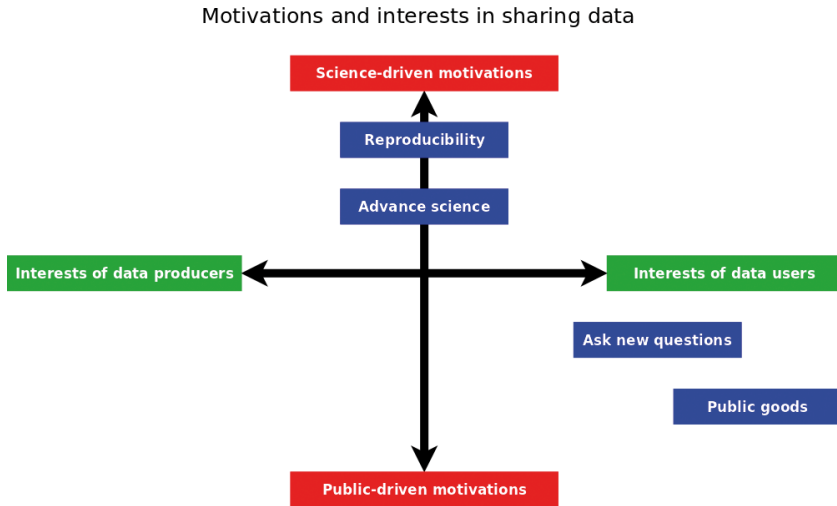


Abbildung 1: Motivation und Interessenlage beim Offenen Zugang zu Forschungsdaten. Für Datenproduzenten gibt es kaum Anreize, ihre Daten anderen zur Verfügung zu stellen (aus Borgman, 2010).

Eine Motivation durch externe Anreize lässt sich nur schwer aufbauen, denn gerade in Bezug auf den Umgang mit Forschungsdaten wird jedwede verpflichtende Regelung von den Wissenschaftlern als extrem hinderlich vehement abgelehnt (Feijen, 2011). Solche Regelungen werden als bürokratisch angesehen, die nur Mehraufwand erzeugen ohne zur Reputation des Wissenschaftlers beizutragen. Zudem bestehen Bedenken, wie geistige Eigentumsrechte an Forschungsdaten geregelt werden. Aus der Perspektive des Open Data Paradigma ergibt sich hier ein Konflikt zwischen den Interessen der Forschern, die eine Hoheit über die Daten für ihren Distinktionsgewinn brauchen, und möglichem öffentlichen Anspruch, Zugang zu Daten aus öffentlich geförderter Forschung zu erhalten. Diese Fragen werden in den Richtlinien des Dryad Datenarchivs angesprochen (Data Archiving Policy, 2010). In den Musterrichtlinien der Partnerzeitschriften des Dryad-Archivs wird darauf hingewiesen, dass die Rechte an den Daten beim Autor verbleiben und dass dieser eine Embargofrist von bis zu einem Jahr festlegen können, innerhalb derer nur die Gutachter des Artikels und sie selber Zugriff auf die Daten haben. Die Aussage über den Verbleib der Rechte an den Daten beim Autor in den Dryad Datenrichtlinien ist ein Zugeständnis an den sense of ownership der Autoren.

Die Frage des urheberrechtlichen Schutzes von Forschungsdaten ist jedoch noch nicht geklärt, die von Land zu Land stark voneinander abweichenden Rechtsprinzipien im Urheberrecht machen die Situation unübersichtlich. Ungeachtet der ungeklärten Rechtslage wurden verschiedene Vorschläge veröffentlicht, wie Forschungsdaten lizenziert werden könnten (Ball, 2011). Auch hier spiegeln sich wieder die Unterschiede zwischen den Disziplinen im Umgang mit Forschungsdaten. Vertreter des Open-Paradigmas treten für eine standardisierte, möglichst offene Lizenzierung ein (Wilbanks, 2007). Ein wichtiges Argument für standardisierte, möglichst offene Lizenzen ist, dass in der Nachnutzung von Daten in abgeleiteten Datenkompilaten Verkettungen von Lizenzen auftreten können, durch die eine Nachnutzung der Daten erschwert oder sogar unmöglich wird, ohne mit den Rechteinhabern der Daten in Verhandlungen zu treten. In jedem Fall sollten standardisierte, maschinenlesbare Lizenzen verwendet werden, um die Rechteprüfung und nachfolgende Lizenzierung des abgeleiteten Datensatzes zu erleichtern. Wilbanks Eintreten für eine gemeinfreie Lizenzierung ist jedoch umstritten, denn erst das Zitat wertet die Veröffentlichung von Forschungsdaten in dem Maße auf, dass für Forscher ein Kooperationsgewinn als Mehrwert für die zusätzliche Arbeit entsteht.

Im gegenwärtigen Wertesystem der Wissenschaft ist die Veröffentlichung das Maß der Anerkennung und des Erfolgs. Eine Einbindung der Daten in das System der wissenschaftlichen Veröffentlichungen ist, zusammen mit der Entwicklung von Werkzeugen für Datenmanagement, unerlässlich um Daten den

notwendigen Stellenwert im Wertesystem der Wissenschaft zu geben (Klump et al., 2006). In den vergangenen Jahren wurden bereits umfangreiche Vorarbeiten geleistet, um Daten den Status einer zitierbaren, qualitätsgeprüften Veröffentlichung zu verleihen (Brase & Klump, 2007; Dallmeier-Thiessen, Pfeifferberger, Puschmann & Stein, 2010; Lowry, Urban & Pissierssens, 2009). Dies hat in einzelnen Bereichen der Forschung bereits zu einem Nachdenken über neue und höhere Ansprüche an Daten geführt, die Grundlage einer Veröffentlichung sind (z.B. Frankel, Finkel & Owen, 2010; Moldwin & Rose, 2009; Reichman, Jones & Schildhauer, 2011; Yilmaz et al., 2011).

Bisher kamen Anreize, Daten zugänglich zu machen, aus den jeweiligen disziplinären Kulturen, wenn diese spezifische Anforderungen an die Dokumentation von Daten stellten. Wer diese Normen nicht erfüllte, wurde ausgeschlossen (z.B. Baggerly, 2010; Cicerone, 2010). Wenn die Einhaltung der Norm jedoch nicht kontrolliert und sanktioniert wurde, blieb sie ohne Auswirkung (Savage & Vickers, 2009), denn eine Norm, die nicht mit einer Sanktion bewehrt ist, bleibt unwirksam (Spittler, 1967). Doch selbst die „Regeln für eine gute wissenschaftliche Praxis“ der DFG (DFG, 1998) und anderer Wissenschaftsorganisationen hatten bisher keine Wirkung gezeigt, da in Bezug auf Forschungsdaten bisher keine Verstöße gegen diese Regeln verfolgt wurden. Daraus wurden inzwischen Konsequenzen gezogen in dem eine Reihe von Forschungsförderern von ihren Antragstellern einen Datenmanagementplan verlangen (z.B. DFG, 2010; Mervis, 2010).

Sind diese neuen Regeln durchsetzbar? Wissenschaft ist demokratisch verfasst. Ein offener Zugang zu Forschungsdaten setzt einen kulturellen Wandel voraus, denn er erfordert eine Veränderung des Wertesystems der Wissenschaft. Dieser kulturelle Wandel kann nur eingeleitet werden, wenn die Herausgeber der Fachzeitschriften den offenen Zugang zu Daten verlangen, die dafür notwendigen Werkzeuge bereit stehen, die Forschungsförderer Datenmanagementpläne verlangen und Datenveröffentlichung zur Reputation ihrer Autoren beiträgt. Open Access to Data und Open Data sind sich in vielem ähnlich, und dennoch sind sie ungleiche Geschwister.

Literatur

- Baggerly, K. (2010). Disclose all data in publications. *Nature*, 467(7314), 401. doi:10.1038/467401b
- Ball, A. (2011). *How to License Research Data*. JISC How-to Guides. Edinburgh, Großbritannien: Digital Curation Centre. Zugriff am 09.11.2011 unter <http://www.dcc.ac.uk/resources/how-guides/license-research-data>
- Berliner Erklärung. (2003). Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Zugriff am 09.11.2011 unter <http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/>
- Borgman, C. L. (2010). Research Data: Who will share what, with whom, when, and why? Presented at the China-North America Library Conference, Beijing, Los Angeles, CA, USA: The University of California at Los Angeles. Zugriff am 09.11.2011 unter <http://works.bepress.com/borgman/238/>
- Botstein, D. (2010). It's the Data! *Molecular Biology of the Cell*, 21(1), 4 -6. doi:10.1091/mbc.E09-07-0575
- Brase, J. & Klump, J. (2007). Zitierfähige Datensätze: Primärdaten-Management durch DOIs. In R. Ball (Hrsg.), *WissKom 2007 : Wissenschaftskommunikation der Zukunft ; 4. Konferenz der Zentralbibliothek, Forschungszentrum Jülich, 6. - 8. November 2007*, Schriften des FZ Jülich - Reihe Bibliothek (Vol. 18, S. 159-167). Jülich: Forschungszentrum Jülich. Zugriff am 09.11.2011 unter <http://edoc.gfz-potsdam.de/gfz/10493>
- Cicerone, R. J. (2010). Ensuring Integrity in Science. *Science*, 327(5966), 624. doi:10.1126/science.1187612
- Dallmeier-Thiessen, S., Pfeiffenberger, H., Puschmann, C. & Stein, D. (2010). Peer reviewed data publication in earth system sciences. *Towards Open Access Scholarship: Selected Papers from the Berlin 6 Conference* (S. 77-84). Düsseldorf: düsseldorf university press. urn:nbn:de:hbz:061-20100722-142254-7
- Data Archiving Policy. (2010, 16. November). *Dryad*. Data Centre. Zugriff am 25.08.2011 unter <http://datadryad.org/jdap>
- DFG. (1998). *Sicherung guter wissenschaftlicher Praxis*. Bonn: Deutsche Forschungsgemeinschaft. Zugriff am 09.11.2011 unter http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf

- DFG. (2010). *Merkblatt für Anträge auf Sachbeihilfen mit Leitfaden für die Antragstellung und ergänzenden Leitfäden für die Antragstellung für Projekte mit Verwertungspotenzial, für die Antragstellung für Projekte im Rahmen einer Kooperation mit Entwicklungsländern* (No. 10.2). Bonn: Deutsche Forschungsgemeinschaft (DFG). Zugriff am 26.08.2011 unter http://www.dfg.de/download/formulare/1_02/1_02.pdf
- Feijen, M. (2011). *What researchers want*. Utrecht, The Netherlands: SURF-foundation. Zugriff am 10.03.2011 unter <http://www.surffoundation.nl/en/publicaties/Pages/Whatresearcherswant.aspx>
- Frankel, K. L., Finkel, R. C. & Owen, L. A. (2010). Terrestrial Cosmogenic Nuclide Geochronology Data Reporting Standards Needed. *EOS, Transactions, American Geophysical Union*, 91(4), 31-32.
- Grundsätze zum Umgang mit Forschungsdaten*. (2010). Bonn: Allianz der deutschen Wissenschaftsorganisationen. Zugriff am 26.08.2011 unter <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze/>
- Harnad, S. (1994). I. Overture: The Subversive Proposal. In A. S. Okerson & J. J. O'Donnell (Hrsg.), *Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing*. Washington, D.C.: Association of Research Libraries. Zugriff am 10.05.2011 unter <http://www.arl.org/sc/subversive/i-overture-the-subversive-proposal.shtml>
- Klump, J., Bertelmann, R., Brase, J., Diepenbroek, M., Grobe, H., Höck, H., Lautenschlager, M., et al. (2006). Data publication in the Open Access Initiative. *Data Science Journal*, 5, 79-83. doi:10.2481/dsj.5.79
- Lowry, R., Urban, E. & Pissierssens, P. (2009). A New Approach to Data Publication in Ocean Sciences. *EOS, Transactions, American Geophysical Union*, 90(50), 483-485.
- Mervis, J. (2010, 5. Mai). NSF to Ask Every Grant Applicant for Data Management Plan - ScienceInsider. *Science Insider*. Zugriff am 10.05.2010 unter <http://news.sciencemag.org/scienceinsider/2010/05/nsf-to-ask-every-grant-applicant.html>
- Moldwin, M. B. & Rose, S. (2009). Documenting Precision and Accuracy in the Open Data Policy Era. *EOS, Transactions, American Geophysical Union*, 90(32), 276.
- Murray-Rust, P., Neylon, C., Pollock, R. & Wilbanks, J. T. (2010, February 19). Panton Principles - Principles for Open Data in Science. *Panton Principles*. Zugriff am 26.08.2011 unter <http://pantonprinciples.org/>

- OECD. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Paris, France: Organisation for Economic Co-operation and Development. Zugriff am 25.08.2011 unter <http://www.oecd.org/dataoecd/9/61/38500813.pdf>
- Pfeiffenberger, H. & Klump, J. (2006). Offener Zugang zu Daten - Quantensprung in der Kooperation. *Wissenschaftsmanagement*, (Special 1), 12-13.
- Piwowar, H. A., Day, R. S. & Fridsma, D. B. (2007). Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE*, 2(3), e308. doi:10.1371/journal.pone.0000308
- Reichman, O. J., Jones, M. B. & Schildhauer, M. P. (2011). Challenges and Opportunities of Open Data in Ecology. *Science*, 331(6018), 703 -705. doi:10.1126/science.1197962
- Savage, C. J. & Vickers, A. J. (2009). Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. *PLoS ONE*, 4(9), e7078. doi:10.1371/journal.pone.0007078
- Schaeffer, P. (2011, 22. Februar). Archiving legacy data, or “Why is Dryad better than a floppy disk?” *Dryad Blog*. Zugriff am 25.08.2011 unter <http://blog.datadryad.org/2011/02/22/archiving-legacy-data-or-why-is-dryad-better-than-a-floppy-disk/>
- Spittler, G. (1967). *Norm und Sanktion. Untersuchungen zum Sanktionsmechanismus*. Olten, Schweiz: Walter Verlag.
- Wilbanks, J. T. (2007). Protocol for Implementing Open Access Data. *Science Commons*. Zugriff am 25.08.2011 unter <http://sciencecommons.org/projects/publishing/open-access-data-protocol/>
- Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., Gilbert, J. A., et al. (2011). Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIXS) specifications. *Nature Biotechnology*, 29(5), 415-420. doi:10.1038/nbt.1823