Facilitating open exchange of data and information

Citation	Gallagher, J., Orcutt, J., Simpson, P., Wright, D., Pearlman, J., & Raymond, L. (2015). Facilitating open exchange of data and information. [Article in Press] Earth Science Informatics. doi:10.1007/s12145-014-0202-2
DOI	10.1007/s12145-014-0202-2
Publisher	Springer
Version	Accepted Manuscript
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

The final publication is available at link.springer.com
DOI: 10.1007/s12145-014-0202-2

Facilitating Open Exchange of Data and Information

James Gallagher · John Orcutt · Pauline Simpson · Dawn Wright · Jay Pearlman · Lisa Raymond

6 Data Publication/Data Citation

Going back to the cultural issues initially addressed in section 2, there is limited academic recognition and promotion for collection and publication of data. In the past, there was little incentive for a researcher to make data available. Publication of raw data does not carry the same weight in deciding promotion as papers that include scientific analyses of the data. Promotion criteria do not take into account the innovation and complexity of data acquisition in the ocean's challenging environments. Many times, data were needed to prepare a peer reviewed research paper, which was essential to further a research career ("publish or perish"), but there was no benefit to making data supporting the publication accessible. The effort to produce data was not highly rated, yet data are the basis of progress in science and research. Sharing data encourages multiple perspectives, helps to identify errors, discourages fraud, is useful for training new researchers, and increases efficient use of funding and population resources by avoiding duplicate data collection (Piwowar 2011).

Can data publication and data citation offer a solution to some of the human motivation issues discussed in this paper? If so, how would it best be implemented, addressing both questions of "how to publish data under the open access model and how to motivate data collectors and creators?" (Penev 2009). Within the informatics community an interesting question has been raised—should we be using the metaphor "data publication." It is argued "that there is no widely understood and accepted definition of what exactly Data Publication means." It was equally clear that "publication" carries many differing implicit assumptions that may not be true (Parsons and Fox 2013). The conclusion was that no one metaphor suits all systems or methods. The term is often used interchangeably with data sharing, but data publishing implies something more. It is a way of using best practices and standards to make sure that data really can be discovered and reused effectively, and that data owners and custodians get the recognition for making datasets public' (GBIF 2014). In this paper, we define Data Publication as making data freely accessible [or at marginal cost] and permanently available on the Internet along with information as to its trustworthiness, reliability, format and content to enable discovery and re-use. Data Publication can take a number of forms including: Standalone Data Publication; Data Publication by Proxy; Appendix Data; Journal Driven Data Archival; and Overlay Publication (Lawrence et al 2011).

6.1 From Fieldwork to Citation

There are established and/or emerging workflows for selected disciplines that enable the publishing of data and credit via citation mechanisms. However, in many disciplines, researchers are simply not aware of such workflows (WDS 2014) or have the data management support or appropriate training.

In order to motivate research scientists to engage in Open Data models, there must be a clear understanding of the benefits of participation. Such understanding should be based on the end-to-

end flow of information from fieldwork to citation. As shown in Figure 1, key elements of the flow are identified and should be primary areas for collaboration and improvement in the emerging transition to Open Data.

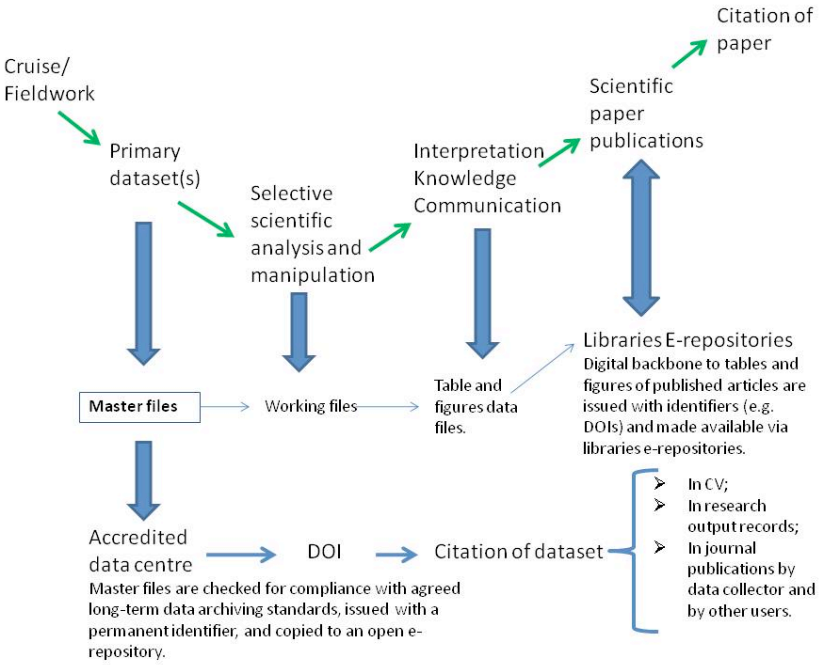

Figure 1. End-to-end flow of data and information going from collection to publishing of data. (Gwenaelle Moncoiffe, 2011).

There has been much discussion in the community on incentives for researchers to publish their data: data repositories, citation increase, DOIs, Funding Agency mandate compliance, kudos and recognition for the data creator et cetera (Costello 2009, 2013), (Piwowar, 2011, 2013) (Sayogo and Pardo 2012), (Sears 2011), (Tenopir et al 2011). Yet, "despite policies and calls for scientists to make data available, this is not happening for most environmental- and biodiversity-related data because scientists' concerns about these efforts have not been answered and initiatives to motivate scientists to comply have been inadequate" (Costello 2009). Survey results from Cragin et al. 2010 indicate the many and varied concerns of researchers including their perceptions of private sharing versus public sharing and real issues with misuse of data. Positive responses to these concerns will go some way to providing incentives to facilitate open data.

The discussion reflects that data collected by scientists and data managers, whether generated from research or operational observations, are not always deposited in national or international data repositories/archives or deposited in a format that makes them retrievable and reusable. Scientists rarely have the skills or resources needed to prepare all their data for public sharing (RIN 2008). Even when submitted, the data often lack a bare minimum of metadata. The problem is in part cultural.

As research careers are heavily dependent upon journal publications and related citations, researchers wish to hold on to "their" data as long as possible to generate more research papers. In addition, the portability of computing power (researchers can easily store years of data on their laptop) and researchers frequent lack of the most basic data management and preservation practices, makes data unavailable and constantly at risk of being lost. Added to this are the restrictions imposed by the institution or government concerning sensitive data that reduce the "open and free" exchange and access to data (see section on Data Access Models). Below we highlight early project work for data publication: within Institutional/Thematic Repositories and Data Centres, specifically implemented to assist data publication within the ocean science community.

Data publication by deposit into institutional data repositories (IRs) may ensure provenance, permanence, attribution and metadata; at present IRs do not guarantee the scientific quality of

published data which requires domain experts more likely found in domain focused data centres. Organizations such as the MBLWHOI Library (MBLWHOI Library 2014) serving the Woods Hole scientific community and supporting the Biological and Chemical Oceanography Data Management Office (BCO-DMO) (see Use Case 2); Lamont's Integrated Earth Data Applications (IEDA 2014) and Scripps' Geological Data Center for geology which includes oceanography (GDC 2014), are among 'early players' participating in World Data System/Research Data Alliance groups (WDS 2014) that are developing standards to be used across all disciplines. Repositories and services established specifically for data publication/sharing and preservation are under development including: Living Atlas of the World (Environmental Systems Research Institute 2014); Planet OS (Planet OS 2014); (UKDS Re-Share 2014). The Ocean Observatories Initiative (OOI 2014) confirms it intends to maintain access to data collected over the 25-30 year life of the observatory.

In the research community, peer review is the accepted process for evaluating the quality of scientific work. Acceptance of a community-agreed peer review procedure for data publication approaching those expected for paper publications is not currently available, yet it is essential, to support reliable and trustworthy data publication and to offer data creators the kudos for promotion.

Emerging open access data journals that publish papers on the management of data and articles on original research data (sets) are now offering peer review including data e.g. Biodiversity Data Journal and other Pensoft journals (BDJ 2014), Data Science Journal from CODATA (CODATA 2009), Earth System Science Data (Copernicus Publications 2014), F1000Research (2014), Geoscience Data Journal (Wiley 2014), Scientific Data (Nature Publishing Group 2014) and the new AGU Earth and Space Science journal (ESS 2014).

A need for a comprehensive peer review of data publication was stated in Parsons et al, (2010) and the first steps for a formal data peer review were given in 2011 by Lawrence et al (Dusterhus 2014). A new statistical scheme for quality evaluation by domain experts is also described by Dusterhus including discussion not only on the quality of data but also the quality of the metadata to provide optimal description for discovery and reuse and interestingly, the quality and availability of the reviewers. Blog comments from the 9th International Digital Curation Conference, Feb 2014 Breakout Session, evidences that the discussion on data validation (and peer review) still abounds with ideas and is ongoing (Kratz, 2014). Data peer review for Ocean Science is in the same place as other disciplines – there is 'processing' and quality control at the data centres, but not traditional external peer review.

The advent of Funding Agency mandates for Open Data, such as the requirement in the National Science Foundation Data Management Plan (NSF 2010) and the European Commission's recent recommendation for open access to scientific publications and data within Horizon 2020 (EU 2013) and Research Councils UK (2014) is expected to stimulate authors to make data available. Data repositories and data journals providing citation metrics will offer evidence of compliance and multiple venues for data publication.. In addition to standard search engines, secondary services like the Thomson Reuters Data Citation Index (Thomson Reuters 2014), will facilitate discovery, use and attribution of datasets and data studies by connecting researchers and data repositories around the world.

A successful early example of motivating data publication is the cooperative work of four organizations: The Marine Biological Laboratory/Woods Hole Oceanographic Institution (MBLWHOI) Library; the Scientific Committee on Oceanic Research (SCOR); the British Oceanographic Data Centre (BODC); and the International Oceanographic Data and Information Exchange (IODE) of the Intergovernmental Oceanographic Commission (IOC). These organizations have developed and executed a pilot project (SCOR-MBLWHOI-IODE 2014) related to two Use Cases:

1. Data held by data centres are packaged and served in formats that can be cited.
2. Data related to traditional journal articles are assigned persistent identifiers referred to in the articles and stored in institutional repositories;

The goal of the Use Cases has been to identify best practices such as Open Access Initiative (OAI) standards for web content; metadata—Dublin Core, Darwin Core; vocabularies and the ability to add other standards for tracking data provenance and clearly attributing credit to data creators/providers so that researchers will make their data accessible. The assignment of persistent

identifiers, specifically Digital Object Identifiers (DOI 2014), enables accurate data citation. The project is also investigating Uniform Resource Identifiers (W3C 2001) and NameIDs. The two project data repositories are meant to be complementary to national and international (e.g., IODE, NODCs, ICSU World Data System) and thematic data centres, rather than a replacement. A "cookbook" has been published (Leadbetter et al 2013) that provides extensive instructions and guidelines to scientists as well as the data publication process to repository managers. It identifies that some form of infrastructure and process must be created to motivate and support data publication.

6.2 Use Cases

Two uses cases were developed as exemplars in ocean science for the discussion of data publication and review. The first is the the BODC Published Data Library (PDL) (BODC 2014) and the second is the work of MBLWHOI Library Woods Hole Open Access Server (WHOAS). Both WHOAS and PDL are indexed by Thomson Reuters Data Citation index, enabling researchers to gain metrics for their data publication. For the purposes of this paper only the MBLWHOI project is described in detail. Other similar repository models for data publication include Dryad (Dryad 2014) and Pangaea.(Pangaea 2014). The sharing of repository records through harvesting also provides greater exposure for data exchange.

6.2.1 Use Case 1

The Published Data Library is implemented by the British Oceanographic Data Centre (BODC 2014). It provides snapshots of specially chosen datasets that are archived using rigorous version management. The publication process exposes a fixed copy of an object and then manages that copy in such a way that it may be referred to over an indefinite period of time. Using metadata standards adopted across NERCs Environmental Data Centres (NERC 2014), the repository assigns DOIs, obtained from the British Library/DataCite, to appropriate datasets.

6.2.2 Use Case 2

The MBLWHOI Library has successfully assigned DOIs to a number of datasets associated with published articles. In the ideal scenario, the DOI(s) should be assigned to the dataset(s) before the article is published, but within the framework of the project there is the ability to retroactively link data to articles after publication. The system has been in operation for over three years, and there is growing interest in the work . Author reaction has been very positive. "This was much easier than trying to deposit data with a publisher"; "The data will be in an open access environment, not owned by publishers"; "Great to know that if my data on my hard disks gets lost at least I have the library copy". It is interesting to note the bias against publishers, something which should be addressed as a broad spectrum solution for sustainability evolves.
Scientists are now becoming aware that DOIs offer the means to easily cite their datasets and gain important citation metrics. Librarians have been using DOIs for years and they are now becoming the de facto standard for data citation within data repositories and institutional repositories (commonly universities) and are being facilitated in such services as NASA's EOSDIS (NASA 2014), Pangaea , Dryad, et cetera. Many current data projects register their DOIs with Datacite (Datacite 2014), an organization that is working to develop standards to foster data access and reuse. The MBLWHOI Library registers DOIs with CrossRef (PILA 2013). The Library began assigning DOIs before DataCite existed and many of the major publishers use CrossRef, but there are a number of DOI Registry Agents..

Publishers are now acknowledging the importance of datasets supporting and within published articles. Nature Publishing developed a platform in 2012 and in 2013 PLoS announced a new data sharing policy (Silva 2014). Supporting data made available in a data repository provides publishers with a safe and easy means of linking the dataset to the published article without them having to publish an annex, deal with data on DVDs, or setting up their own data repository.

Many publishers have identified a specific repository for this purpose (in the medical sciences, publishers use PubMed and in fact are required to do so by such Funding Agencies as the National Institutes of Health and the Wellcome Trust. In ocean science, funding agencies like the UK Natural Environment Agency (NERC) require all data created through their grants to be deposited

in the British Oceanographic Data Centre but at present there is no one repository designated by publishers for ocean data. Many publishers do not yet have an identifiable policy dealing with supporting datasets (JoRD 2013), though this is now changing with publishers forging new partnerships to store supplemental data; for example, Taylor and Francis Journals (and others) are now using figshare (figshare 2014) who will host the supplemental data as well as provide a widget that will enable Taylor and Francis users to view data in the articles in the browser alongside the content (Research Information 2014).

Because of the assignment of DOIs, Elsevier Publishing sought a collaboration with the MBLWHOI Library. Article records in ScienceDirect (ScienceDirect 2014) now contain links to datasets deposited in the Woods Hole Open Access Server (WHOAS 2014) that are associated with Elsevier articles. This system works for DOIs assigned before and after article publication and a WHOAS statement covers copyright, "All Items in WHOAS are protected by original copyright, with all rights reserved, unless otherwise indicated." In addition some depositors request a specific Creative Commons License (Creative Commons 2014). The WHOAS system of linking data to the articles in ScienceDirect was implemented in May 2012.

Another outcome of the project includes tools and procedures developed by the MBLWHOI Library and the NSF funded Biological and Chemical Oceanography Data Management Office (BCO-DMO 2014) to automate the ingestion of data and related metadata from BCO-DMO into the WHOAS Institutional Repository (IR). WHOAS is built on the DSpace platform (DuraSpace 2014). The system also incorporates functionality for BCO-DMO to request a Digital Object Identifier (DOI) from the Library. This partnership allows the Library to work with a trusted data repository to ensure high quality data while the data repository uses library services and is assured that a permanent archived copy of the data is associated with the persistent DOI. Feedback from BCO-DMO is very positive. The Data Manager reports that the most sought after functionality is the DOI and the ability to cite the data. This use case has demonstrated that data can be successfully deposited into a library institutional repository and that the assignment of DOIs is an effective way to enable data citation.

The Library is also participating in an NSF Grant that will result in WHOAS content being published as Linked Open Data which wll expose relationships between DSpace repository content and other data sources. Linked Open Data enables knowledge discovery, sharing and integration. Exposing linked data is a concept continuing to emerge. Tim Berners-Lee's vision in 2009 (Berners-Lee 2009) was to "build a web for open, linked data that could do for numbers what the Web did for word . . . unlock our data and reframe the way we use it together."

6.3 Data Citation

Previously, researchers have not really understood how to cite data (or compile meaningful metadata) and "full citation of data is not currently a normative behaviour in scholarly writing" (Mooney and Newton 2012). However, the introduction of DOIs for data sets has been a positive encouragement welcomed by the research and informatics community. The advent of a number of Research Data Training online courses e.g., (MANTRA 2014) are welcome tools for researchers to gain RDM skills.

Citation metrics have been adopted across the sciences as a method to obtain quantitative indicators for the assessment of the quality of research and researchers, as well as the impact of research products. Systems and services such as the Science Citation Index (Thomson Reuters 2013), the h-index (or Hirsch number), or the Impact Factor of scientific journals have been developed to track and record access and citation of scientific publications. These indicators are widely used by investigators, academic departments and administration, funding agencies, and professional societies across all disciplines to assess performance of individuals or organizations within the research landscape and inform and influence the advancement of academic careers and investments of research funding. New data metrics indicating the value and impact of data publications (like those launched by the Data Citation Index in 2012) are needed to raise the value and appreciation of data and data sharing because the missing recognition for data publication in science is seen as a major cause for the reluctance of data producers to share their data (Smit 2010). Calls for data sets to be cited in a conventional manner are widespread and the growing use of persistent DOIs assigned to data sets (e.g., by MBLWHOI, DataCite and Dryad) is a major contribution leading to a call for a central registry resolving various digital identifiers (DOI, URL, URI et cetera) (Costello 2013).

The Research Data Alliance (RDA 2014) supported by the European Commission, the U.S. Government and the Australian Government is likely to have a significant impact on the research data landscape. An overall objective of the ICSU World Data System/Research Data Alliance WDS/RDA Interest Group on Data Publication - Bibliometrics is to "conceptualize data metrics and corresponding services that are suitable to overcome existing barriers and thus likely to initiate a cultural change among scientists, encouraging more and better data citations…" (WDS 2014).

Work among several groups is resulting in recommendations for data citation formats; early examples include Altman and King (2007), the UK Digital Curation Centre (Ball and Duke 2012) and the Federation of Earth Science Information Partners (ESIP 2012). Data Citation Groups have been formed; the Force 11 Data Citation Synthesis Group has released and called for endorsement of the consolidated Joint Declaration of Data Citation Principles, a collaborative effort including such groups as the CODATA-ICSTI Task Group on Data Citation (Force II 2013). Other data citation groups like that within Mendeley (Mendeley 2014) and UK Data Service (UKDS 2014) contribute to the discussion.

"Data publication and data citation is becoming increasingly important to the scientific community, as it will provide a mechanism for those who create data to receive academic credit for their work and will allow the conclusions arising from an analysis to be more readily verifiable, thus promoting transparency in the scientific process" (Lawrence et al 2011). It can create incentives for researchers to make data available with sufficient metadata, to make it discoverable and re-usable, thereby gaining citations. Of course, this is conditional upon institutional management agreeing to use data citation metrics as an element in performance assessment and career advancement decisions. The recent trend of Funding Agencies and Publishers requiring data related to publications to be accessible will accelerate data publication.

…….

Acknowledgements The authors would like to thank the other members of the NSF Research Coordination Network "OceanObsNetwork" : Milton Kampel ; Takeshi Kawano; Fred Maltz; Michael McCann ; Benoit Pirenne ; Peter Pissierssens; Iain Shepherd · Christoph Waldmann; Albert Williams III, who contributed to the report that this paper summarizes and enhances (Pearlman et al 2013). The authors acknowledge the support of the National Science Foundation through Grant Award No. OCE-1143683.

Altman M, King G (2007) A proposed standard for the scholarly citation of quantative data. D-Lib Magazine, 13(3/4). http://www.dlib.org/dlib/march07/altman/03altman.html. Accessed 19 September 2014

Australian Government (2009) Government 2.0 Task Force Report. http://www.finance.gov.au/publications/gov20taskforcereport/doc/Government20TaskforceReport.pdf. Accessed 16 September 2014.

Ball A, Duke M (2012) How to cite data sets and link to publications. Edinburgh, UK: Digital Curation Centre. http://www.dcc.ac.uk/webfm_send/525. Accessed 16 September 2014.

BDJ (2014) Biodiversity Data Journal. http://biodiversitydatajournal.com. Accessed 16 September 2014.

BCO-DMO (2014) Biological & Chemical Oceanography Data Management Office. http://www.bco-dmo.org. Accessed 16 September 2014.
Berners-Lee T (2009) The next web. TED 2009 Conference. http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html. Accessed 16 September 2014.
BODC (2014) Published Data Library. https://www.bodc.ac.uk/data/published_data_library. Accessed 16 September 2014.
Borgman CL (2012) The conundrum of sharing research data. Journal of the Association for Information Science and Technology 63(6):1059–1078, DOI:10.1002/asi.22634.
Braunschweig K, Eberius J, Thiele M, Lehner W (2012) The state of open data—limits of current open data platforms. In: Mille A, Gandon FL, Misselis J, Rabinovich M, Staab S (eds) Proceedings of the 21st World Wide Web Conference 2012, (WWW 2012), Lyon, France.
Busse S, Kutsche RD, Leser U, Weber H (1999) Federated information systems: concepts, terminology and architectures. Tech. Rep., Technical University Berlin.
Cocco M (2012) Research infrastructure and e-science for data and observatories on earthquakes, volcanoes, surface dynamics and tectonics. ICRI2012, International Conference on Research Infrastructures. http://ec.europa.eu/research/infrastructures/pdf/workshop_october_2011/16_esfri_epos_cocco.pdf. Accessed 16 September 2014.
CODATA (2009) Data Science Journal. http://www.codata.org/dsj. Accessed 16 September 2014.
Copernicus (2014) Copernicus: The European Earth Observation Programme.http://www.copernicus.eu. Accessed 16 September 2014.
Copernicus Publications (2014) Earth System Science Data: the data publishing journal. http://earth-system-science-data.net. Accessed 16 September 2014.
Costello MJ (2009) Motivating online publication of data. BioScience, 59, 418-427. doi:10.1525/bio.2009.59.5.9.
Costello M, Michelner WK, Gahegan M, Zhang Z-Q, Bourne PE (2013) Biodiversity data should be published, cited, and peer reviewed. Trends in Ecology & Evolution, 28(8), 454-461.
Costello M, Appeltrans W, Bailly N, Berendsohn W, Jong Y, Edwards M, Froese R, Huettmann F, Los W, Mess J, Segers H, Bisby F (2014) Strategies for the sustainability of online open-access biodiversity databases. Biological Conversation, vol.173, 155-165.
Cragin MH , Palmer CL, Carlson JR., Witt M (2010) Data sharing, small science and institutional repositories. Philosophical Transations of the Royal Society, A, 368, 4023-4038.
Creative Commons (2014) Creative commons license. http://creativecommons.org/licenses. Accessed 16 September 2014.
CUAHSI (2013) CUAHSI Water Data Center. http://wdc.cuahsi.org. Accessed 16 September 2014.
Datacite (2014) Datacite. https://www.datacite.org. Accessed 16 September 2014.
DataNet (2014) DataNet Federation Consortium—collaboration environments for data drivenscience. http://datafed.org. Accessed 16 September 2014.
DataONE (2014) NSF data observation network for earth (DataONE). https://www.dataone.org/about. Accessed 16 September 2014.
DOI (2014) Digital object identifier. http://www.doi.org. Accessed 16 September 2014.
Dryad (2014) Dryad Digital Repository. http://datadryad.org. Accessed 16 September 2014.
DuraSpace (2014) DSpace. http://www.dspace.org. Accessed 16 September 2014.
Dusterhus A, Hense A (2014) Automated quality evaluation for a more effective data peer review. Data Science Journal, 13, 67-78.
Earth Observations (2013) GEO BON—biodiversity observation network http://www.earthobservations.org/geobon.shtml. Accessed 16 September 2014.
Environmental Systems Research Institute (2014) Living Atlas of the World. http://doc.arcgis.com/en/living-atlas. Accessed 21 September 2014.
ESIP (2012) Federation of Earth Science Information Partners. http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/Citations. Accessed 16 September 2014.
ESS (2014) Earth and Space Science. http://agupubs.onlinelibrary.wiley.com/agu/journal/10.1002/%28ISSN%292333-5084/. Accessed 17 September 2014
EU (2006) Communication from the Commission to the Council and the European Parliament: Interoperability for pan-European government services.

http://www.epsos.eu/uploads/tx_epsosfileshare Communication-on-Interoperability_01.pdf. Accessed 16 September 2014.

EU (2013) Guidelines on open access to scientific publications and research data in Horizon 2020. Version one. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_ pilot/h2020-hi-oa-pilot-guide_en.pdf. Accessed 16 September 2014.

European Commission (2011) Digital agenda: turning government data into gold. Press release. http://europa.eu/rapid/press-release_IP-11-1524_en.htm. Accessed 16 September 2014.

F1000Research (2014). http://f1000research.com. Accessed 16 September 2014.

FGDC (2014) National spatial data infrastructure (NSDI).http://www.fgdc.gov/nsdi/nsdi.html. Accessed 16 September 2014.

figshare (2014) figshare. http://figshare.com. Accessed 16 September 2014.

Force II (2013) Force II: The future of research communication and scholarship, Joint Declaration of Data Citation Principles. http://www.force11.org/datacitation. Accessed 16 September 2014.

FSF (2014) Free software foundation. http://www.fsf.org. Accessed 16 September 2014.

Gallagher J, Potter N, Sgouros T, Hankin S, Flierl G (2007) The data access protocol—DAP 2.0. NASA ESE-RFC-004.1.1. https://earthdata.nasa.gov/our-community/esdswg/standards-process-spg/rfc/esds-rfc-004-dap-20. Accessed 16 September 2014.

GBIF (2014) Global Biodiversity Information Facility. http://www.gbif.org. Accessed 16 September 2014.

GDC (2014) Geological Data Center, Scripps Institution of Oceanography. http://gdc.ucsd.edu/. Accessed 17 September 2014

GEO/CEOS (2008) GEO/CEOS workshop on quality assurance of calibration & validation processes: Establishing an operational framework. http://qa4eo.org/workshop_washington08.html. Accessed 16 September 2014.

GeoViQua (2007) GeoViQua: QUAlity aware VIsualization for the global earth observation system of systems. http://www.geoviqua.org. Accessed 16 September 2014.

Guess A (2013) Japan embraces open data, launches multiple open projects. http://semanticweb.com/japan-embraces-open-data-launches-multiple-open-projects_b35158. Accessed 16 September 2014.

Hankin S, Blower J, Carval Th, Casey K, Donlon C, Lauret O, Loubrieu T, Srinivasan A, Trinanes J, Godoy O, Mendelssohn R, Signell R, De La Beaujardiere J, Cornillon P, Blanc F, Rew R, Harlan J (2010) NETCDF-CF-OPENDAP : standards for ocean data interoperability and object lessons for community data standards processes, Oceanobs 2009, Venice Convention Centre, 21-25 septembre 2009, Venise, publication date 2010-12-23, http://archimer.ifremer.fr/doc/00027/13832/10969.pdf. Accessed 16 September 2014.

Harley D, Kryzys Acord S, Earl-Novell S, Lawrence, S, Judson King C (2010) Assessing the future landscape of scholarly communication: An exploration of faculty values and needs in seven disciplines. Center for Studies in Higher Education, UC Berkeley. http://escholarship.org/uc/cshe_fsc. Accessed 17 September 2014.

IEDA (2014) Integrated Earth Data Applications. http://www.iedadata.org. Accessed 16 September 2014.

INSPIRE (2014) Infrastructure for spatial information in the European Community (INSPIRE). http://inspire.ec.europa.eu. Accessed 16 September 2014.

IODE (2014) International Oceanographic Data and Information Exchange. http://www.iode.org,. Accessed 16 September 2014.

JoRD (2013) JoRD: Journal research data policy bank project. http://jordproject.wordpress.com. Accessed 16 September 2014.

Klump J, Bertelmann R, Brase J, Diepenbroek M, Gross, H, Hock H, Lautenschlager M, Schingler W, Sens I, Wachter J (2006) Data publication in the Open Access Initiative, Data Science Journal, 5, 79-83.

Kozak M, Hartley J (2013) Publication fees for open access journals: Different disciplines—different methods. Journal of the American Society for Information Science and Technology 64 (12). doi:10.1002/asi.22972.

Kratz J (2014) Fifteen ideas about data validation (and peer review). Data Pub [Blog], http://datapub.cdlib.org/2014/05/08/fifteen-ideas-about-data-validation-and-peer-review. Accessed 16 September 2014.

Laakso M, Welling P, Bukvova H, Nyman L, Björk B-C, et al. (2011) The Development of Open Access Journal Publishing from 1993 to 2009. PLoS ONE 6(6): e20961. doi:10.1371/journal.pone.0020961.

Lawrence B, Jones C, Matthews B, Pepler S, Callaghan S (2011) Citation and peer review of data: moving towards formal data publication. International Journal of Digital Curation 6(12):4–37.

Leadbetter A, Raymond L, Chandler C., Pikula L, Pissierssens P, Urban E (2013) Ocean Data Publication Cookbook. Paris: UNESCO, 39pp. (Intergovernmental Oceanographic Commission Manuals and Guides 64) .http://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=10574, Accessed 2014-03-09.

Lecomte P, Stensaas G (2009) Overview of progress towards a data quality assurance strategy to facilitate interoperability. http://www.earthobservations.org/documents/committees/adc/200909_11thADC/ DA-09-01a%20QA4EO.pdf, Accessed 2014-04-22.

MANTRA (2014) Research Data MANTRA.http://datalib.edina.ac.uk/mantra. Accessed 15 September 2014.

MBL WHOI Library (2014) http://www.mblwhoilibrary.org. Accessed 15 September 2014.

Mendeley (2014) Mendeley. http://www.mendeley.com Accessed 15 September 2014.

Mooney H, Newton M.P. (2012) The anatomy of a data citation: discovery, reuse and credit. Journal of Librarianship and Scholarly Communication, 1(1), eP1035.

NASA (2014) EOSDIS: NASA's earth observing system data and information system. https://earthdata.nasa. gov. Accessed 15 September 2014.

National Research Council (2012) For attribution – Developing data attribution and citation practices and standards. National Academies Press, Washington, DC.

Nativi S, Craglia M, Pearlman J (2012) The brokering approach for multidisciplinary interoperability: A position paper. International Journal of Spatial Data Infrastructure 7:1–15, http://ijsdir.jrc.ec.europa.eu/index.php/ijsdir/ article/view/281. Accessed 15 September 2014.

Nativi S, Craglia M, Pearlman J (2013) Earth science infrastructures interoperability: The brokering approach. Journal of Selected Topics in Applied Earth Observation and Remote Sensing, 6:1118–1129, DOI 10.1109/JSTARS.2013.2243113.

Nature Publishing Group (2014) Scientific Data. http://www.nature.com/sdata. Accessed 15 September 2014.

Neilsen M (2011) Reinventing Discovery: The New Era of Networked Science. Princeton University Press.

NERC (2014) Data centres. http://www.nerc.ac.uk/research/sites/data. Accessed 15 September 2014.

NOAA IOOS (2014) Quality assurance of real time ocean data, QARTOD. http://www.ioos.noaa.gov/qartod/welcome.html. Accessed 15 September 2014.

NSB (2011) NSB 11-79 digital research data sharing and management: report of the Task Force on Data Policies. Tech. rep., National Science Board.

NSF (2010) National science foundation data management plan. http://www.nsf.gov/bfa/dias/policy/dmp. jsp. Accessed 15 September 2014.

NSF (2014) National Science Foundation Directorate for Geosciences: Earth Cube.http://www.nsf.gov/geo/ earthcube/, visited 2014-04-22.

OGC (2014) Open Geospatial Consortium. http://www.opengeospatial.org. Accessed 15 September 2014.

OneGeology (2014) OneGeology.http://www.onegeology.org, visited 2014-04-24

Onoda M (2012) GEOSS Data sharing principles and action plan. Workshop on GMES Data and Information Policy, Brussels. http://ec.europa.eu/enterprise/newsroom/cf/_getdocument.cfm?doc_id=7140. Accessed 15 September 2014.

OOI (2014) Ocean Observatories Initiative. http://oceanobservatories.org/. Accessed 18 September 2014

Open Knowledge Foundation (2012) The open data handbook. http://opendatahandbook.org/en/what-is-open-data. Accessed 15 September 2014.

Palfrey J, Gasser U (2012) Interop: The Promise and Perils of Highly Interconnected Systems. Basic Books.

Pangaea (2014) Pangaea: Data publisher for the earth & environmental sciences. http://www.pangaea.de. Accessed 15 September 2014.

Parsons MA, Fox P (2013) Is data publication the right metaphor? Data Science Journal 12:WDS32–WDS46.

Parsons MA, Duerr R, Minster J-B (2010) Data citation and peer review. Eos: Transactions American Geophysical Union, 91(34) 297-299.

Pearlman J, Shibasaki R (2008) Guest editorial: Global earth observation system of systems. IEEE Systems Journal 2(3), 302-303. DOI 10.1109/JSYST.2008.928859.

Pearlman J, Williams A, Simpson P (eds) (2013) Report of the Research Coordination Network: RCN OceanObs Network: Facilitating open exchange of data and information. NSF/Ocean Research Coordination Network Tech Rep., 46 pp.

Penev L, Erwin T, Mille, J, Chaqvan V, Motitz T, Griswold C (2009) Publication and dissemination of dataset in taxonomy. :ZooKeys working example. ZooKeys, 11, 1-8.

PILA, Inc (2013) CrossRef. http://www.crossref.org. Accessed 15 September 2014.

Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. Doi:10.1371/journal.pone.0000308

Piwowar H (2011) Who shares? Who doesn't? Factors associated with openly archiving raw research data. PLos ONE, 6(7), e18657.

Piwowar HA and Vision TJ (2013) Data reuse and the open data citation advantage. PeerJ, 1, e175. http://dx.doi.org/107717/peerj.175.

Planet OS (2014) Planet OS: big data platform... https://planetos.com/. Accessed 21 September 2014.

President Barack Obama (2013) Memorandum on Open Data Policy–Managing Information as an Asset (May 9, 2013). http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf. Accessed 17 September 2014.

Research Data Alliance (2014). Research data sharing without barriers.https://www.rd-alliance.org. Accessed 15 September 2014.

Research Councils UK (2014) http://www.rcuk.ac.uk/research/datapolicy. Accessed 15 September 2014.

Research Information (2014) Taylor & Francis partners with figshare for supplementary data.http://www.researchinformation.info/news/news_story.php?news_id=1485. Accessed 15 September 2014.

Research Information Network (2008). To share or not to share: publication and quality assurance of research data outputs. A report commissioned by the Research Information Network. http://www.rin.ac.uk/data-publication. Accessed 15 September 2014.

Sayogo DS, Pardo T (2012) Exploring the motive for data publication in open data initiative: Linking intention to action. In: Proceedings of the 45th Hawaii International Conference on System Sciences, IEEE Computer Society.

ScienceDirect (2014) http://www.sciencedirect.com. Accessed 15 September 2014.

SCOR/MBLWHOI/IODE (2014). Data publication/data citation project. http://www.iode.org/index.php?option=com_content&view=article&id=110&Itemid=129. Accessed 15 September 2014.

SeaDataNet (2014) Sea data net. http://www.seadatanet.org. Accessed 15 September 2014.Sears J (2011) Data sharing effect on article citation rate in paleoceanography. *Eos, Trans. AGU*, 92, Fall Meet. Suppl., Abstract /IN53B-1628.

Silva L (2014) PLoS new data policy: public access to data. http://blogs.plos.org/everyone/2014/02/24/plos-new-data-policy-public-access-data-2. Accessed 15 September 2014.

Smit E (2010) Preservation, access and re-use of research data. Presented at DataCite Summer Meeting 2010. https://www.datacite.org/datacite_summer_meeting_2010. Accessed 15 September 2014.

SURF (2013) Enhanced publications. Collaborative organisation for ICT in Dutch higher education and research. http://www.surf.nl/en/themes/research/research-data-management/enhanced-publications. Accessed 15 September 2014.

Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, Manoff M, Frame M (2011) Data sharing by scientists: practices and perceptions. PLos, 6(6), e21101.

Thomson Reuters (2014) The Data Citation Index. http://wokinfo.com/products_tools/multidisciplinary/dci. Accessed 11 March 2014.

Thomson Reuters (2013) Science Citation Index. http://science.thomsonreuters.com/cgi-bin/jrnlst/jloptions.cgi?PC=K. Accessed 1 March 2014.

Turnitsa C (2005) Extending the levels of conceptual interoperability model. In: Proceedings IEEE summer computer simulation conference, IEEE CS Press.

US Congress (1980) Bayh-Dole act. Public Law 96–517, also known as the Patent and Trademark Law Amendments Act; enacted by the United States Congress.

UKDS (2014) UK Data Service, Citing Data and Re-Share. http://ukdataservice.ac.uk/media/440282/publishingcitigdata.pdf. Accessed 10 March 2014.

Vision T, (2010) Open Data and the Social Contract of Scientific Publishing. BioScience 60(5): 330-1. doi:10.1525/bio.2010.60.5.2.

W3C (2001) URIs, URLs, and URNs: Clarifications and recommendations 1.0. http://www.w3.org/TR/uri-clarification. Accessed 10 March 2014.

WDS (2014) Data Publication Working Group. http://icsu-wds.org/community/working-groups/data-publication. Accessed 10 March 2014.

Whiteside A, Evans JD (2006) Web coverage service implementation specification #06-083r8, version 1.1.0.https://portal.opengeospatial.org/files/?artifact_id=18153, visited 2014-05-19

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, et al. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715.

Whitfield P (2012) Why the provenance of data matters: assessing “fitness for purpose” for environmental data. Canadian Water Resources Journal 37(1):23–36, DOI 10.4296/cwrj3701866.

WHOAS (2014) Woods Hole Open Access Server. .https://darchive. mblwhoilibrary.org. Accessed 16 September 2014.

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, et al. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715. doi:10.1371/journal.pone.0029715.

Wiley (2014) Geoscience Data Journal. http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%292049-6060. Accessed 11 March 2014.

World Meteorological Organization (2014) Information management. http://www.wmo.int/pages/themes/wis/index_en.html.. Accessed 11 March 2014.