Metrics for Openness

David M. Nichols
Department of Computer Science, University of Waikato, Hamilton 3240, New Zealand.
E-mail: dmn@cs.waikato.ac.nz

Michael B. Twidale
Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, Champaign, IL 61820-6211, USA.
E-mail: twidale@illinois.edu

The characterisation of scholarly communication is dominated by citation-based measures. In this paper we propose several metrics to describe different facets of open access and open research. We discuss measures to represent the public availability of articles along with their archival location, licences, access costs and supporting information. Calculations illustrating these new metrics are presented using the authors' publications. We argue that explicit measurement of openness is necessary for a holistic description of research outputs.

Introduction

Metrics are developed in order to serve a variety of purposes, including to help us understand and manage a phenomenon. Scholarly communication is one such phenomenon - of interest to information scientists, managers of universities, funding bodies, governments and wider society. Citation-based measures, as a proxy for quality and impact, have come to dominate the characterisation of research publications. In particular, the h-index has become a popular summary measure for the publishing impact of an author.

In this paper we propose that a further facet of scholarly communication, that of openness, is sufficiently important to merit the development of dedicated metrics. These new measures are intended both to inform understanding of current practice and to encourage increased openness of the research literature. The open access movement (Willinsky, 2006) has highlighted the issue of access to information in the scholarly record. The openness of research is a concern both for researchers without institutional access (Bradley and Soldo, 2011) and for the general public (Zuccala, 2009).

All the aspects of research articles (the text, graphics, its citations, the data it contains, the data and code it uses to substantiate its claims) can become input data for other researchers. These additional uses of the scholarly record (beyond reading the text) are not well-represented by current metrics. In particular, researchers' efforts to make these diverse research outputs open are not captured; consequently, their "invisible work" (Star & Strauss, 1999) is not recognised.

The contribution of this paper is to gather existing measures of openness together and to propose new metrics to inform discussion on aspects of 'openness' within scholarly communication. In particular, we propose an individual Openness Index that characterises the accessibility of an author's publications.

The next section reviews previous work on metrics, including citations and the more recent altmetrics proposals. We summarise existing studies on openness and then present several metrics for characterising openness in scholarly communication together with examples of their use. We discuss the design, value, feasibility and robustness of the proposed metrics.

Background

Information science has frequently used citations to provide a quantitative basis for investigating the nature of scholarly communication. The use of citation data in university rankings systems (Van Raan, 2005; Shin, Toutkoushian & Teichler 2011) has helped in cementing the importance of citations in academia. Other data sources, such as web metrics or altmetrics, have played a secondary role to citations. In this section we briefly review citation-based metrics and other metrics.

Citation-based Metrics

The journal impact factor (JIF) counts citations made

This is a preprint of an article accepted for publication in Journal of the Association for Information Science and Technology © 2016 (American Society for Information Science and Technology).
Nichols, D.M. and Twidale, M.B. (2016) Metrics for openness. *Journal of the Association for Information Science and Technology*.
https://doi.org/10.1002/asi.23741

in the current year in the scholarly literature to items published over the two previous years. The JIF was originally intended to help libraries decide which journals to purchase (Garfield, 2006) but has been widely used as a proxy for the quality of journals, articles and authors. The JIF has been criticised for both its formulation and for its extended (albeit unintended) use in evaluating articles and authors (Seglen, 1997; Ewing 2006).

Hirsch (2005) proposed an "easily computable index, h, which gives an estimate of the importance, significance, and broad impact of a scientist's cumulative research contributions". This h-index, and its many variants, has proved popular and several systems, including Scopus and Google Scholar, now automatically calculate this measure for authors. Finch (2010) reviews several citation-based variants of the h-index and concludes: "perhaps trying to find a single metric by which to evaluate authors is a quixotic endeavour at best".

The dominance of citations in the characterisation and evaluation of scholarly communication has led to calls for alternative measures to be devised (Adler and Harzing, 2009; Boleslaw, Szymanski, de la Rosa & Krishnamoorthy, 2012). Lane (2010) suggests that a new metrics infrastructure should aspire to capture "the essence of what it means to be a good scientist".

Non-citation-based metrics

Citations are often used as a useful and easy-to-measure proxy for the impact and influence of a document. However, citations are restricted to the scholarly literature. If an article has influence on someone but they do not author a document of the right kind, then a citation-based metric will not capture that relationship.

References to academic work can appear in a variety of document genres facilitated by the low-cost publishing environment of the web, including blog posts, Twitter and Wikipedia. Priem and Hemminger (2010) note that these sources can be mined for references to academic articles and suggest seven specific sources: "bookmarking, reference managers, recommendation services, comments on articles, microblogging, *Wikipedia*, and blogging." Measures derived from these alternative sources are typically referred to as altmetrics. Piwowar (2013) suggests that altmetrics derived from social media capture properties not reflected in citations: contributing to a more holistic view of scholarly use.

The scholarly record itself is increasingly present on the web. This digital representation allows the capture of richer usage data than is possible in a paper-based environment. Kurtz and Bollen (2010) extend the citation-based bibliometrics of scholarly articles with usage data such as views and downloads. A variety of other metrics have been proposed to capture other facets of scholarly communication, including: Author Affiliation Index (Gorman and Kanet, 2005), Journal Internationality Index (Buela-Casal, Perakakis, Taylor & Checa, 2006) and International Publication Ratio (Leite, 2011).

Metrics based on citations, web presence, downloads, views, affiliation and international collaboration do not address one of the major themes of scholarly communication in recent years: open access.

Existing measures of openness

Table 1 summarises several studies that have measured the open availability of published papers, noting the approach used and the percentage of articles assessed as open. The results of the studies listed in Table 1 can only be approximately compared as there are several variations in their methods.

One measure of openness that has been evaluated on an individual basis is choice of publication venue. Hughes (2008) tracked (by journal policy rather than individual article status) the publishing behaviour of the signers of the PLOS (Public Library of Science) "Open Letter to Scientific Publishers", which called for greater access to the scholarly record. Overall, she found that the signers had largely chosen to publish in open venues.

Measures of openness are also starting to be used as components in other indicators. Aguillo, Ortega, Fernández & Utrilla (2010) use the accessibility of full-text files in institutional repositories as one indicator in compiling a repository ranking. Prost & Schöpfel (2014) likewise survey institutional repositories but instead compute openness rates across different document genres, e.g. working papers (96%), PhD theses (76%) and book chapters (17%). Willmott, Dunn & Duranceau (2012) define an *Accessibility Quotient* which combines information on price, citations and shareability to characterise a publishing environment for a group or individual. This measure is based on publishers' agreements and so reflects potential accessibility rather than actual availability to information seekers. Shotton (2012) includes open access in a five criteria framework for article-level evaluation; distinguishing between articles with different re-use rights and different financing models.

There are two main reasons for wishing to extend these current approaches to measuring openness: to better characterise the landscape of scholarly communication and as an aid for open access advocacy. The studies outlined above provide a good overview of openness for some cases, but they are neither comprehensive nor detailed enough to reflect the variety of observed publication behaviour. In the following sections we present metrics that aim to capture a richer description of openness in scholarly communication.

TABLE 1. Summary of several studies of open access articles.

Paper	Domain / Selection	Search Method	Open
Björk, Laakso, Welling and Paetau (2010)	1837 articles published in 2008 from Scopus	title search in Google	20%
Mercer (2011)	3,873 articles published in 2008 from Library and Information Science Abstracts	web searching for individual articles	41%
Way (2011)	922 articles from 20 Library and Information Science journals with highest impact factor in 2007 *Journal Citation Reports Social Science Edition*	title search in Google Scholar.	27%
Matsubayashi et al. (2009)	4,667 biomedical journal papers from 2005 via PubMed	searching using top 20 results from PubMed Central, Google Scholar, Google and OAIster	27%
Kurata et al. (2010)	1,955 biomedical articles from PubMed	searching using PubMed, PubMed Central and Google	51%
Lyons and Booth (2011)	452 articles from 60 business and management journals from 2009	title search in Google	36%
Xia, Wilhoite and Myers (2011)	812 articles from 20 Library and Information Science journals with highest impact factor in 2006 *Journal Citation Reports*	Google Scholar	51%
Hedlund (2010)	1216 articles in English in 2008 from Scopus, following approach of Björk et al. (2010)	replicating Björk et al. (2010)	22%
Koskinen et al. (2010)	407 journal articles from 2007-08 from the University of Helsinki publication database	title search in Google, Google Scholar, OpenDOAR, Scientific Commons and HELDA (University of Helsinki institutional repository)	49%
Norris, Oppenheim & Rowland (2008a)	journal articles from Journal Citation Reports: 628 in ecology from 2003 966 in economics from 2003 925 in sociology from 2004	OAIster, OpenDOAR, Google and Google Scholar	 35% 54% 24%
Wren (2005)	48,516 articles from 13 journals between 1994 and 2004 from Medline	custom program using the Google API	0-48%
Archambault et al (2013)	320,000 articles from Scopus	custom harvester	43%

Designing Metrics for Openness

The studies in Table 1 have certain recurrent features, including:

- title searching in Google as a method to determine open availability on the internet
- the main metric calculated is the percentage of open items found from the sample
- article sample sets are mainly limited to journals
- no explicit consideration of pre-print or other aspects of document status
- no consideration of rights beyond 'access for reading'
- limited analysis of the location where the open articles were found
- no consideration of the costs to access non-open documents
- no consideration of the availability of supplementary material such as data and code
- analysis is conducted at the level of a discipline, institution, or a group of journals; rather than at the level of an author.

This list reveals a tension in any metric design process: the desirability of having an easy-to-compute measure versus one that addresses more of the subtleties of the scholarly publication process. Each of these observations suggests extensions that add complexity for any proposed metric. For example, the restriction to homogeneous collections of publications, usually journal articles, simplifies analysis but does not accurately reflect the variety of publication types: journal articles, conference papers, book chapters, books, technical reports, white papers, working papers etc.

Similarly, a work (in say an institutional repository) can exist in several inter-related forms that can be easily confused by a potential reader (or a computer program), including: submitted version, post-review author's version, proofed version and final published version. It is not clear that any of the existing studies have attempted to distinguish these different versions, but if we value the roles of peer review and editing then measures of openness should attempt to reflect these differences.

Measuring the availability of a document on the internet via search engines has been used as a proxy for open access. Carroll (2011) argues that full open access, with rights to remix and re-use, is important and authors should not be satisfied with just 'open to read' status for their work. Computational access to the entire scholarly literature has the potential to create new applications and research opportunities (Lynch, 2006; Shadbolt, Brody, Carr & Harnad, 2006). These extended use cases highlight the difference between a license to read an article and a licence that allows anyone to process, reuse and remix an article.

From the perspective of measuring open access there are a variety of copyright licences and access restrictions. Matsubayashi et al. (2009) measure a category of "restricted OA (e.g., user must register to gain [free] access)" but only at 0.4% of all articles; Kurata et al. (2010) find 0.6% restricted in this way. SPARC (2013) provide a six-faceted approach to the evaluation of a journal's policies: reader rights, reuse rights, copyrights, author posting rights, automatic posting and machine readability. Shotton (2012) provides a five point scale which includes a distinction between access to read and wider rights of reuse.

For the purposes of metric design we suggest a simplification distinguishing three broad groups:

- Closed documents: behind a paywall or other access restriction
- Freely readable documents: articles deposited under conditions that allow free access for readers
- Freely processable and remixable: articles that, at most, require attribution to allow applications such as text-mining

The dominant use case is 'access for reading' but there is also value in quantifying the distribution of less restrictive licenses. A further use case of 'access for reproducibility' stresses the availability of supplementary information such as data and code. We return to these issues later in the paper.

We now present several metrics that illustrate alternative methods for characterising the openness of the scholarly literature. We follow Shotton (2012) in using our own work to demonstrate the ideas.

A Practical Openness Index

The measures reviewed earlier, with the exception of Hughes (2008), do not address openness on an individual level. However, there is a growing interest in developing metrics to characterise openness; with the discussion mainly occurring outside the scholarly literature (Nichols, 2012; Lewis, 2012; Eaves, 2013; Eisen, 2013;

McCormick, 2013; Piowar & Priem, 2014). Our first suggestion is that, just as with an individual h-index, each author should have an *Openness Index* that reflects the accessibility of their outputs. Although, as with the h-index, this index can be defined for different facets (e.g. departments, institutions, journals, funders, countries, career level, disciplines etc.) we believe it should start where the documents are created—with the authors.

We also believe that a simple metric is a good first step. One option is to follow the pattern of the studies outlined above and just consider published journal articles. However that seems to ignore a substantial component of scholarly practice; in computer science peer-reviewed conferences play a substantial role (Wainer, Goldenstein & Billa, 2011). We acknowledge a bias towards the research traditions we are most familiar with, but we think this is a useful place to begin. It allows us to bring deeper understanding of the complexities that metrics will need to address when applied in our own domain, and allows us to begin by experimenting on ourselves, computing our own metrics and assessing the degree to which they seem a reasonable proxy for what we care about. Naturally this localised approach should not determine applicability for the whole of scientific publishing, but it is a useful analytic starting point.

As well as journals and conferences, other common formats are book chapters in edited volumes and monographs. These can be even harder to make available openly, and so we choose to make an arbitrary choice to begin with just journals and conferences.

A further issue is whether there should be a difference between the 'value' of pre-review and post-review versions. As an initial approximation, we suggest that a useful openness metric can be constructed based on full-text read access to a submitted, accepted or published version of a journal or conference paper. This threshold focuses on the types of items that constitute the bulk of published research, addresses the dominant use-case of reading and also partially recognises the value of peer review. Although we count submitted pre-review versions we are only including papers that eventually went on to be published. Troll Covey (2009) doesn't distinguish between preprints and postprints, on practical grounds, when measuring openness across the organisational units of a university.

One option is to develop a thorough but impractical and costly to produce version of the openness index – one we might term the *Idealistic Openness Index*. However for the reasons outlined above we settle for a more practical openness index. Therefore, we propose to estimate an individual's *Practical Openness Index* (POI) as:

$$OI = (\text{number of open } C + J) / (\text{number of } C + J)$$

(where C is conference papers and J is journal papers)

An author with all of these types of works freely available online would have an POI of 1; an author with all these works behind a paywall would have an POI of 0. Following the approach of earlier studies we performed a title search for each conference and journal publication in our CVs in Google and Google Scholar: restricting ourselves to items returned on the first page of results. When testing openness with public search engines there is the possibility of institutional, geographical and personal customisation of result listings. For example, access to research documents is likely to be greater from a university campus location than from home. Similarly, personalisation of search results can occur if the testing user (or program) is simultaneously logged into a service (such as via a Google Account). These effects can influence both absolute openness (is this document accessible?) and practical openness (is this version near the top of result listings?). The tests in this paper were performed at the authors' homes without using university services, though when using opaque external services (such as Google) there is no way to remove all forms of possible result customization (Micarelli, Gasparetti, Sciarrone & Gauch, 2007).

This approach means that we have excluded the following types of our work: book chapters, books, reports, occasional papers, technical reports and working papers. For other scholars it would also exclude items such as compositions, art works, performances, exhibitions etc.

Calculating a Practical Openness Index

As an example, the results of manually calculating a Practical Openness Index for ourselves across our research careers are Twidale = 0.82 and Nichols = 0.91: these calculations proved to be a salutary experience. We had aimed to avoid a lot of complexities in calculation by having a broad definition of openness and ignoring whole categories of scholarly output (such as books and book chapters) that can be particularly difficult to make open. As ever, performing a process by hand before delving into algorithm development is a valuable exercise. We found a surprising number of complexities even in computing just our own data points—where of course we enjoyed considerable insider knowledge. For example, there were often times when we knew that an open copy existed online, but it did not appear on the first page of results. The detail of the calculations and selectivity criteria used are described in the Supplementary Material.

Our experiences raises interesting issues about access versus discovery. If a paper is freely available online, but hard to find, is it truly open? Advocates for open access, ourselves included, note the desirability of access for both scholars and members of the general public. How hard must someone search to find a free copy of your paper?

A specific example of the limitations of these methods occurs when a free version can only be accessed via an author's home page (such as via the ACM *Author-Izer* service (Delman, 2011)), but that version may not be discovered by a simple Google search of the title. Is it reasonable to assume that those interested in accessing the paper for free would know that it can be worth trying a researcher's web page?

We are pleased to report one desirable result. Computing one's POI immediately made us want to increase it; particularly for those papers we are especially proud of that were not freely available. This creates a testable hypothesis: to what extent will others make the effort to improve their score as soon as they become aware of an openness metric?

OI-Broad

A broader version of the openness index would include book chapters.

OI-Broad = (number of open C + J + BC)
/ (number of C + J + BC)
(where BC is book chapters)

Book chapters are typically peer reviewed (but often not as rigorously as certain journals and conferences) and are part of the scholarly literature. However book copyright agreements often make it harder to produce open access versions of chapters. Consequently, we expect that OI-Broad would often lead to a lower score than a basic POI. As a result we prefer the basic POI as an initial metric to encourage greater openness. Other more exacting metrics might be introduced or popularised later, rather as energy efficiency metrics (such as star ratings) or product safety standards are progressively ratcheted up as people become more used to them. An even broader version of an openness index would include books and all other academic outputs.

Effective Openness Index

The ability of an author to make a post-review work open is often constrained by the copyright policies of particular publication venues. Many academic authors have signed copyright agreements that prevent archiving of post-review open access versions of their papers. As a result the simpler metrics above can appear to be unfair: particularly for older researchers where it may be difficult to locate both relevant document versions and copyright agreements.

To reflect this legal environment we can additionally characterise authors' behaviour by drawing a distinction between papers where a version can be made openly available and those where copyright restrictions limit authors' options. Such a restriction would have more effect on items included in a Broader-OI (e.g. books) than it would for narrower measures (e.g. conferences and journal articles) where pre-review versions can usually be made open. We suggest the following measure to answer the question: of the items an author *could* have made open, how many *are* actually open?

EOI = number of open publications
/(number of total publications - number of copyright restricted publications)

Troll Covey (2009) refers to the non-open fraction (1-EOI) as "the gap between opportunity and practice". Mercer (2011) estimates this gap for library and information science articles using the only current tool available - the SHERPA/RoMEO database of journal policies (Jenkins, Probets, Oppenheim & Hubbard, 2007). A more fine-grained index could distinguish between document versions (e.g. submitted, accepted, final etc.) reflecting whether the closest-to-final-version is available. This refined index would reward authors who make an effort to distribute the best version given the constraints of their prior agreements. In attempting to embrace the complexity of copyright in scholarly communication these indices are much harder to calculate than the basic OI.

Assessing the copyright policies of journals can often be more difficult than might be expected; provision of clear information on copyright transfer agreements is far from universal; even in LIS journals (Coleman, 2007). These issues are exacerbated for articles published many years ago, where current policies may not apply retroactively. Further, this general policy approach will fail to detect individual agreements and exceptions (such as for a special issue of journal). Although the Effective Openness Index is more difficult to accurately calculate than a simple Openness Index, it does specifically address the archiving behaviour of authors within the boundaries set by copyright policies. Examining the items which are restricted can produce a 'Blocking List' of publishers who are, in effect, limiting the Openness Index of an author (or other unit of analysis).

Preservation-Friendly Openness Index

As noted earlier, existing studies of openness have only partially addressed the *location* of open documents. Way (2011) found 29% of open access papers were identified as being on a "personal web site". This fits a definition of openness in terms of freely accessible from a search engine. But there is no guarantee that the document will remain open. What if the website is no longer maintained? What if it is a faculty member's web page and the faculty member moves? Or retires? Or dies? Openness based on using links from a faculty member's web page (e.g. the ACM *Author-Izer* service (Delman, 2011)) will end when that page disappears. Troll Covey (2009) reports that some faculty are aware of these issues and have "expressed concern about the preservation of their "legacy.""

Nothing is guaranteed to be permanent, but certain resources such as institutional repositories or libraries are likely to last longer than others. Given the desirability of open access as a lasting value, we may want to count what proportion of a person's work is archived in some 'long-term' location. Institutional and subject repositories and online journals all count as 'long-term' locations whereas personal web spaces would not. Research group websites and ftp servers are somewhere in between in terms of durability, so, as ever, a decision must be made. We choose to determine that these do not make it into our *Preservation-Friendly Openness Index* (PFOI). A PFOI only counts open items in these long-term locations, treating other items as if they were closed. Continuing our earlier example, our PFOI values are Twidale = 0.45 and Nichols = 0.84. These are counts based on the earlier accessibility criteria via search engines; alternative approaches could locate items using repository-specific searches or personal knowledge.

As ever, edge cases crop up: an example from our evaluation is Citeseer. This is an interesting test case. It has a degree of longevity, having been around for at least a decade, a long time in terms of the web. Given its history of development we can hope that if it were not to continue, its maintainers would find a way of preserving it, perhaps by submitting it to the Internet Archive.

Björk et al. (2014) report a "trend towards increased use of subject and institutional repositories [for open access documents] in comparison to home and departmental web pages." Submitting a document to an institutional service is often more work than putting it on a personal web page. Terras (2012) directly addresses the work of self-archiving: "Is it worth me digging out the full text, running the gamut with the UCL [University College London] repository, and trying to spend the time putting my previous research online?" A preservation-aware metric captures this invisible work of making a resource open for the long term.

Although authors may view online journals as a secure venue for the continued open availability of their work Laakso et al. (2011) report that 28% of their sample of 'born open access' journals from 2000 were no longer active in 2009. This result suggests that it might be useful to highlight the unexpected absence of open articles through a separate *Loss Index* which would measure the proportion of such open items that were no longer available.

Acce$$ Index

The defining characteristic of the non-open fraction of a researcher's output is the financial cost to access the items. Consequently we can have metrics that examine all these different non-zero costs. The simplest metric is just to sum the costs to produce an *Acce$$ Index*:

Acce$$ Index = sum of price of all non-open items

As with the other indices this can be calculated over different kinds of items. Our example Acce$$ Index values using the POI criteria are: Twidale = $1,484 (over 20 items), Nichols = $183 (over 4 items). We counted the total cost to buy each paper independently but more complex and subtle versions of the Acce$$ Index are possible. We might allow for joining a society to reduce costs where many articles are from the same publisher. For example, a single paper published by the ACM may cost $15, but an annual subscription to the ACM digital library costs $199 (including ACM membership), so if the author has many papers published in the ACM, we might set an upper limit of $199 for that set.

Book chapters can be a problematic case as they may not be available to be purchased individually. This entails purchasing a whole book in order to access one chapter. This is a prime example of Ruscio's (2012) note that extreme values may disrupt the calculation of a metric, for example purchasing an old book may be very difficult - or it may be possible to buy it second hand for a low price. The Acce$$Index is by design a polemical metric, as such potential extreme costs support its intention of inciting change. Indeed, even its name is a component of its design.

Other versions might acknowledge the non-monetary costs involved in accessing material; for example, the costs of travel to a physical location, such as a library, that offers access. Even obtaining price metadata for some items required registration with a publisher's website, incurring time costs and requiring the provision of personal information. We note that automating the calculation of an Acce$$ Index would require machine-readable price metadata to be more widely available. Access cost metrics may be currently impractical to implement as their creation costs most likely outweigh their benefits. But versions of them may prove useful as one-off illustrations to make the case for increasing open access. For example they could show the hidden costs of not digitizing a book collection.

The Acce$$ Index can be calculated across any set of papers; we think one potentially useful set is that of all the publications produced by an institution in a year. Imagine one hypothetical reader who sits down to read the entire annual output of a university. An *Everyperson Index* (in North America this could be referred to as a *John Q. Public Index*) measures what it would cost our reader to access all the non-open outputs of an institution. We are not aware of any estimates of this figure but suggest it might be an interesting piece of future work. Many institutions have mission statements, charters and visions that emphasise an ambition to support the public good of advancing knowledge. Quantifying the cost of accessing the research outputs of an institution would cast new light on the relationship between the institution and wider society. It is possible that universities might compete to see who has the lowest *Everyperson Index*: or indeed on any of the other openness metrics. In particular, if a University ranking system included an openness metric it would create a powerful organizational incentive to increase open access.

Actual Individual Purchase Index

The previous access cost metrics were hypothetical, and indeed polemical in that it is highly unlikely anyone would pay these exorbitant costs. Nevertheless, some people actually do pay money to access information but authors do not know how much is paid to access most of their works. Book purchases are the main exception but most scholarly communication is not in book form.

This metric simply sums the actual payments made across a set of works over a defined time period. The design intent of this metric is to encourage authors to consider: firstly, how their work freely handed over to the publisher causes individuals (rather than institutions via journal subscriptions) to pay, and secondly, whether they might do something to let those people and, others, access their work for free. The complexities around measuring access costs as part of institutional subscriptions lead us to propose a simplified index based on individual purchase decisions.

Bergstrom (2014**)** provides insights into the costs of library subscriptions—which have often been confidential. However, even our simplified index relies on information that is internal to commercial publishers. The airing of the metric challenges the difference between an author actually receiving information on book sales (even if the royalty income is just pennies per copy), and failing to receive the same information on article sales with a royalty of $0. A further consequence of conceptualising use and payment in this way suggests that data on access charges could become a bargaining chip in negotiations between libraries and publishers. Is it reasonable for a university library to demand data on payments to access the work of scholars from their institution?

Although this metric, as with others presented, can be applied across different sets of works it appears to have considerable force at the level of individual authors. The *Actual Individual Purchase Index* represents actual costs paid by readers to access their work. The figure may be rather revealing - but we have no idea what this index would be for our publications. Just because a publisher asks for, say, $37 to let someone read our paper does not mean that anyone has paid for access. On the other hand, maybe hundreds of people have. We would like to know.

Openness Cost Index

The Acce$$ Index measures the cost of consumption of scholarship. A complementary metric, the Openness Cost Index, measures how much it costs to make a work open. This includes the sum of any access fees/page fees etc. It also includes the effort taken.

In trying to understand the nature of resistance by faculty to participating in making scholarly output open, perceived costs and benefits are likely to play a role. Substantial work has been done on trying to measure benefits. This could be complemented by relatively modest attempts at measuring costs (Research Consulting, 2014). How much work is it for a faculty member at a particular institution to make her papers open? Is it a matter of just consenting and someone else does all the work? Does she have to supply the correct legal version of her paper and its bibliometric details in an email? Or does she have to manually enter it into the Institutional Repository? If so, how much effort it that? What is the average time or number of clicks necessary to upload each paper?

We believe it is important for open access advocates to be clear that open access is not free. Money and effort has to be spent to create, maintain and populate institutional repositories and other OA sources. Those costs should be made explicit so that they can be put against the benefits of OA.

Open Reference Index

In the same way we can ask whether a paper is open, we can also ask the same of *all* the items that the paper cites. The *Open Reference Index* (ORI) is the proportion of all the cited works of a paper that are themselves open access. The ORI for this paper at the time of writing is 0.92 (99 items of which 91 are openly available); details in the Supplementary Material. This measure gives a broad idea of how open the referenced literature is, reflecting the ease with which someone could access the cited support for the arguments in the paper.

An Open Access paper with an ORI of one is an *Open Paper*: all of the immediate supporting papers are themselves Open Access. In principle, we can also recursively search through all the citations of the supporting papers until we identify the set of all linked papers: all the items in the citation graph. This set of all linked papers can be characterised in its openness, e.g. as a straightforward fraction of open papers (as above). Pushing the idea of the openness of the supporting literature to its natural extreme is the concept of a *Fully Open Paper*:

- which itself is Open Access, and
- where all of the references it relies on are *Fully Open Papers*

In other words, the entire set of supporting literature for a Fully Open Paper is openly available. Clearly, Fully Open Papers are going to be rare: we wonder whether there are any significant papers that meet the criteria to be Fully Open.

The ambition of open access advocates is that all academic research should be free to access via the internet. However, we suggest there is value in the analysis of the openness of important documents beyond academic papers. Interesting targets include government policy documents, legislation, legal judgements, reports from policy institutes, white papers and reports from non-governmental organisations. In particular, when a government implements a policy based on research findings, what are the barriers to members of public who wish to understand the research support of the policy? For these types of documents then several openness metrics may be usefully applied including: the ORI, the full supporting citation graph, supporting data and source code. We consider specific metrics for code and data below.

A corollary of the ORI is that there is a cost associated with accessing the non-open references of a document. When an Acce$$ Index is calculated over this set of closed references it is an *Acce$$ Support Index*. As an example, the Acce$$ Support Index for the 8 non-open items cited in this paper is $3,662 (details in the Supplementary Material): this value is mainly from the book *Encyclopedia of Library and Information Sciences* (Bates & Maack, 2009). This result is a good illustration of the influence of extreme values as noted by Ruscio et al. (2012): the encyclopedia costs $3,282.

Illegality Index

Openness studies have typically used the presence of documents on the internet as a measure of access. We are not aware of studies that have attempted to quantify how many open research papers are available on the internet in violation of signed publishers' copyright agreements. Antelman (2006) notes that many authors are unaware of the detail of their copyright agreements and that "authors who know their rights or publisher requirements do not necessarily abide by them." Troll Covey (2009), in a study of personal and departmental websites, reports a significant number of deposits that run counter to publisher policies and that "the most frequent problem is self-archiving the publisher PDF when publisher policy prohibits it." It seems likely that substantial numbers of open access research documents on the internet are available illegally.

We propose as a thought experiment the *Illegality Index* to measure the proportion of a group of works that is available in contravention of existing copyright agreements. This is another polemically named index intended to draw attention to a major problem of making documents open access. Our assumption is that the majority of contraventions are inadvertent. The author simply does not know or does not remember the details of the copyright agreement signed with her publisher. Concerns over such accidental violations and fears about their repercussions can lead to a reluctance to make certain papers open, even when those fears are unfounded. For example, a researcher may be reluctant to upload a version of an older paper to a repository because of fear of copyright violations. They may have no idea if it can be uploaded, perhaps because the journal or conference it appeared in is no longer in existence. This is a variant of the 'orphan works' problem in copyright law.

As anyone can distribute a document on the internet without regard to copyright then an Illegality Index does not necessarily reflect the actions of an author or institution. Illegal copies of movies and music have largely not been distributed by the artists or their publishers. Consequently, directly associating this metric with an individual can be misleading and we do not report this measure for ourselves. Nevertheless as a thought experiment it usefully focusses an author's mind on the copyright agreements they have signed. In calculating the ORI for this paper we found several instances of papers and books that appeared to be available outside the scope of copyright. Furthermore, it can serve as an opportunity to address unfounded fears about the copyright terms associated with specific versions of documents.

Data and Code Archiving Indices

The metrics introduced so far refer to access to the written article, however definitions of open access recognise the importance of access to elements such as supporting data and code. There is widespread agreement on the increasing importance of viewing the scholarly record as wider than simply the main text. This holistic view of research has been expressed through proposals to encapsulate research outputs as "scientific publication packages" (Hunter, 2006) and "research objects" (Bechhofer et al., 2013). These aggregations aim to capture data, code, and other items necessary for reproducibility. As with the archiving of papers in institutional repositories, the work needed to produce such enhanced research outputs should be explicitly recognised through the metrics used to describe scholarly communication.

Data Archiving Index

The importance of supporting data for published results is widely recognised (Borgman, 2012; Stodden & Miguez, 2014; Piwowar, 2013). Several studies focussing on reproducibility have highlighted the inaccessibility of underlying data across several domains (Anderson, Greene, McCullough & Vinod, 2008; Wicherts, Borsboom, Kats, & Molenaar, 2006; Evangelou, Trikalinos & Ioannidis, 2005; Alsheikh-Ali, Qureshi, Al-Mallah, & Ioannidis, 2011). Boulton et al. (2012) urge that curation of datasets is explicitly rewarded and Fecher, Friesike and Hebing (2015) suggest better incentives for data sharing.

Nichols, Twidale and Cunningham (2012) proposed the concept of a *Data Archiving Index* to characterise the data archiving behaviour of authors. This index would be derived by examining the data-using papers that had their data archived and expressed as a simple percentage. Although this index is simple to state at a high level it is difficult to perform precise calculations: how many separate datasets are used or generated in a specific paper? Should there be separate calculations for the "input" and "output" datasets?

Shotton (2012) suggests five categories for supplementary information which include a distinction between data and data in an actionable format. A further refinement of availability can incorporate Berners-Lee's five star open data model (Berners-Lee, 2009) which addresses formats and licensing but does not consider the location of the data (i.e. curated repository or personal website). Vines et al. (2014) note that the availability of research data in their sample fell by 17% per year. This finding highlights that the location of datasets is crucial for their long term access and should be reflected in data archiving metrics. Following the Preservation-Friendly OI we suggest a preservation-friendly variant of a Data

Archiving Index where data in curated repositories is counted, whereas data found on less formal locations (e.g. personal websites) is excluded. Goodman et al. (2014) provide ten "simple rules" for sharing scientific data. Two of the rules encourage repository use to make data open and available: this behaviour is captured in a Preservation-Friendly Data Archiving Index. One rule suggests rewarding researchers who share their data: calculation of a Data Archiving Index achieves this goal of public recognition.

Tenopir et al. (2011) conducted a survey of 1300 scientists and found that the desire to use other researchers' data sets exceeded the desire to share their own data. Pham-Kanter, Zinner & Campbell (2014) surveyed data sharing in the life sciences and found that there were "very few formal or informal sanctions for data sharing noncompliance; if a scientist fails to share as required or expected, she or he faces few penalties from other scientists." They also note the "public goods feature" of data sharing and suggest that funding agencies and journals could play a role as an "independent third party norm-setter and enforcer" to level the playing field (Pham-Kanter et al., 2014). Vines et al. (2013) report that explicit data archiving mandates are associated with higher rates of data deposit and are one mechanism to encourage archiving. A complementary approach would be the provision of a public researcher Data Archiving Index which could establish sharing norms without explicit sanctions. The "public goods feature" of data sharing suggests that *public* indicators are needed to reflect researcher behaviour and would be a counter-balance to the private informal data sharing culture reported by Wallis, Rolando and Borgman (2013).

Code Archiving Index

There is growing agreement on the importance of providing supporting source code for research results: "anything less than the release of source programs is intolerable for results that depend on computation" (Ince, Hatton & Graham-Cumming, 2012).

Much of the previous discussion on data archiving equally applies to the archiving of source code (Peng, 2011). Although code is typically smaller than datasets it is has complexities that can make it at least as difficult to successfully share, e.g. use of code libraries, specifics of the execution environment, version dependency etc. However, the basic principle of a *Code Archiving Index* (Nichols et al., 2012) mirrors that for data, it simply calculates a proportion of those papers with code that have open code. Similarly, the location and the licensing of the code can be assessed to create index variants that consider preservation and re-use.

Summary

Stodden and Miguez (2014) list best practices for managing data and code throughout the research process. Morin et al. (2012) characterise research with an absence of open code as a 'black box' and recommend policy changes to require that source code be made available at publication time. Publicising a Code Archiving Index for researchers may well be an effective tool for informing a debate on code archiving policies and research practices.

The journal *Biostatistics* has a specific reproducibility policy that publicises both the availability of data and code (Peng, 2009); papers that provide enough material to enable reproduction are labelled with a 'D' (for data availability), a 'C' (for code) or an 'R' (for 'reproducible'). For example, Chang, Peng and Dominici (2011) is marked with a 'C' on the first page of the PDF of the paper. The *Biostatistics* policy is a leading exemplar for open data and code; it serves a similar 'public indicator' purpose as the openness metrics proposed in this paper.

On examining our publications it is chastening to observe that little of our own code and data is publicly available (let alone stored in a curated preservation-friendly location). However, we are not alone; there appears to be very little supporting information attached to papers in locations such as the ACM Digital Library or the archives of JASIST.

Discussion

A set of metrics allows researchers to view authors through different lenses; the metrics introduced here allow additional complementary views of scholarly communication. In particular, openness metrics reward and recognise the invisible work performed by researchers to make their outputs more widely available. The altmetrics movement stresses wider measures of impact beyond citation. Similarly, openness metrics stress wider contributions than simple publication.

Shaw and Vaughan (2008) characterise a "typical professor" in LIS through citations and publication venues. Sugimoto and Cronin (2012) add notions of collaboration patterns and productivity at various career stages. We propose that a further part of a holistic researcher profile should be the openness of their works.

Harnad (2011) suggests that to increase research availability universities should mandate their researchers to deposit their final manuscripts in institutional repositories. He notes that metrics, such as citations and downloads, should reflect the greater visibility and impact of open access works. We suggest that *directly* measuring authors' behaviour is an important additional metric for open access and that the broader studies reviewed earlier should be complemented with the individual metrics that we have proposed. It should be as natural to enquire about a researcher's Openness Index as it is currently about their h-index. Although generally applicable to researchers, Openness Indices are particularly relevant in LIS—where access to information is a core element of the discipline. In addition to characterising *other* researchers, these indices can also be used as a self-assessment tool by authors to reflect on the openness of their publication record.

Although we have focussed on applying metrics at an individual level, most of the openness measures can be applied to any set of works. At the institutional level this allows for interesting new comparisons. We can imagine institutions comparing themselves on openness criteria and even openness scores being included in university ranking systems. Internally, measuring openness allows universities to evaluate themselves against the visions in their charters, mission statements and university mottos espousing knowledge for all. Under this framing, costs for 'gold' open access can be interpreted as a positive measure: the more an institution spends on making works open, the more committed it is to sharing knowledge to all of society. These ideas might be particularly relevant for taxpayer-funded public universities rather than for private institutions.

The extension from citizen access to research papers (Zuccala, 2009) to a recognition of broad access to supporting data is explicitly recognised by Kowalczyk and Shankar (2011): "making data broadly available can promote public understanding of science, evidence-based advocacy, educational uses, and citizen-science initiatives." This observation emphasises that the metrics that we design and apply should reflect this wider context of scholarly communication. An example of the broadening importance of access to data beyond the scientific community is Elsevier's recent initiative to provide free access to Science Direct for journalists (van Gijlswijk & Boucherie, 2014). This laudable extension solves one problem (giving journalists better/easier access to the scientific literature, encouraging better informed articles) but makes another problem even more glaring - the reader of the journalists' articles cannot read the articles the journalists cite. As scientific data and research findings are increasingly used to inform public decision-making and the framing of legislation, there is another reason for measurement of the ease of access to these resources. In particular, the Open Reference, Data and Code Archiving Indices can be viewed as foundations (or even pre-requisites) of a data-literate citizenry (Twidale, Blake & Gant, 2013).

As the various metrics outlined above have shown, there are several challenges in actually computing a metric. Often the problem is that the metadata needed is missing and this is particularly challenging for older documents. The metrics therefore also function as a call for richer metadata for the scholarly record. For example, distinguishing between articles that are free-to-read and those that grant further use rights is limited by the availability of machine-readable rights metadata. Although the SHERPA-RoMEO service (Jenkins et al., 2007) is often used as a reference for journal policies there is not an easy method to take a document identifier, such as a DOI, and determine the copyright status of various versions of that specific article. Kiley, interviewed in Hodgson (2014), observes that "license metadata isn't always included at the article level or done in a consistent way" and that "a standardized method and taxonomy is needed to express licensing at the article level in a machine-readable way." Recent PLOS papers (e.g. Goodman et al., 2014) include a <license> element in the XML version (which leads to a machine-readable RDF license) however this level of detail is rare across the whole scholarly record. The recent NISO proposal to include <free_to_read> and <license_ref> metadata elements provides a potential route for richer analysis of the scholarly record (NISO, 2015): including facilitating calculation of some of the metrics proposed in this paper. The proposed cost metrics require that pricing metadata is also available and that it is clear precisely which rights would be acquired through a purchase.

Metadata describing supplementary information is also largely absent; inhibiting the calculation of the Data and Code Archiving metrics. The code availability information assessed for a paper such as Chang, Peng and Dominici (2011) is not currently represented in the metadata provided at the journal or, consequently, in aggregation sources such as PubMed, Scopus and the Web of Science. The work done in making the judgement is not recorded in a structured manner and is thus not available for automated calculation. Recent papers at PLOS (e.g. Wicherts, Bakker & Molenaar, 2011) include `<supplementary-material>` elements in the XML source but only include the format (e.g. `application/msword`) rather than the type (e.g. source code). The NISO recommended practice for metadata describing supplementary information (NISO, 2013) would significantly improve this situation. However, the experience of the slow adoption of the ONIX-PL licensing framework (Pesch & Lamoureux, 2013) suggests it may take many years before such metadata is in widespread use. Nichols, Twidale & Cunningham (2012) noted that metadata describing supplementary information is likely to be controlled by publishers rather than necessarily being open data itself; possibly presenting barriers to the open calculation of some metrics. Where licence metadata does exist then more detailed metrics can be computed; for example, as suggested by a reviewer, a Reusability Index could reflect the precise license terms of both paper and supplementary material. However, Google Scholar is well-placed to publicise a basic Openness Index from its current database.

The proposed estimate of a researcher's Openness Index uses conference papers, journal articles and (in a revised version of the index) book chapters. Books are clearly a significant output type that is missing. However, there are others: music compositions, exhibitions, art works, performances etc. These diverse output types have been recognised as important in national research evaluation activities (e.g. the Research Excellence Framework in the UK, the Performance-Based Research Fund in New Zealand). These outputs may be harder to make open access, and may require different ways to open up access (Crossick, 2015). Investigating whether, and if so how, these heterogeneous works should be part of an openness measure is an area for future work. Furthermore, disciplines vary in the way they value various publication formats. Separate indices for conferences, journals, book chapters and books might be of interest to some. However, for our main purpose of sensitizing people to open access, we initially suggest a simpler metric.

Many critics of metrics argue that existing measures don't recognise the diversity of academic work and can also create inappropriate incentives. The challenge of metric design is that different metrics measure different attributes so the choice of target is central. However, just as with altmetrics or citation-based measures of 'impact', there is no single concept of openness; as is shown by the various indices we have suggested. We believe the best approach is to use a number of metrics measuring different aspects of openness alongside metrics of other dimensions of scholarly communication.

For any proposed measure there will be initial uncertainty (over averages, ranges, disciplinary differences etc.) which can lead to problems in making comparisons between people (Ruscio et al., 2012). Wider adoption of the measures may lead to greater understanding and refinements (e.g. discipline normalisation)—as has occurred with the h-index.

Conclusion

Adler and Harzing (2009) invite researchers to "innovate and design more reliable and valid ways to assess scholarly contributions that … best fulfil the university's fundamental purpose"

Many of the ideas behind the metrics described above have been implicitly understood by information researchers, and some have been discussed on blogs and at conferences. The contribution of this paper is to formalise and extend these ideas into quantitative measures that can be used systematically to further characterise the landscape of scholarly communication. We anticipate that these measures can, and will, be used to encourage greater access to information. The simple act of measuring current practice can be a powerful incentive to alter that practice**:** we suggest authors could start with calculating their own Practical Openness Index. Where that measurement is impeded by a lack of metadata an explicit statement of a potential benefits can support moves to enhance metadata provision.

A further benefit to quantifying concepts relating to the openness of published research is to provide a basis for management and policy decision-making. The frequently repeated maxim; that to control something you must first measure it, applies here. We might add that measurement also has a publicity component: one way to raise the profile of an issue is simply to measure it: what gets measured gets *noticed*. Indeed, it may well be that what gets measured gets to frame the argument. From an open access advocacy perspective, we suggest that it should be just as common for authors to publicise their Openness Indices as it is to publicise their h-index. We have presented our indices (and found creating them a valuable learning process); we invite readers to calculate their own Openness Indices.

References

Adler, N. & Harzing, A.W. (2009). When Knowledge Wins: Transcending the sense and nonsense of academic rankings, *The Academy of Management Learning & Education*, 8(1), 72-95.

Aguillo, I.F., Ortega, J.L., Fernández, M. & Utrilla, A.M. (2010). Indicators for a webometric ranking of repositories. *Scientometrics,* 82(3), 477-486.

Alsheikh-Ali, A.A., Qureshi, W., Al-Mallah, M.H. & Ioannidis, J.P.A. (2011). Public Availability of Published Research Data in High-Impact Journals. *PLOS ONE,* 6(9): e24357. doi:10.1371/journal.pone.0024357

Anderson, R.G., Greene, W.H., McCullough, B.D. & Vinod, H.D. (2008). The Role of Data/Code Archives in the Future of Economic Research. *Journal of Economic Methodology*, 15(1), 99-119.

Antelman, K. (2006). Self-archiving practice and the influence of publisher policies in the social sciences. *Learned Publishing*, 19(2), 85-95.

Archambault, E., Amyot, D., Deschamps, P., Nicol, A., Rebout, L. & Roberge, G. (2013) *Proportion of Open Access Peer-Reviewed Papers at the European and World Levels—2004-2011*, Science-Metrix Inc., Montreal, Canada. Retrieved February 3, 2016, from http://www.science-metrix.com/pdf/SM_EC_OA_Availability_2004-2011.pdf

Bates, M.J. & Maack, M.N. (Ed.) (2009). *Encyclopedia of Library and Information Sciences* (3rd ed.). CRC Press, Boca Raton, FL, USA.

Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., Couch, P., Cruickshank, D., Delderfield, M., Dunlop, I., Gamble, M., Michaelides, D., Owen, S., Newman, D., Sufi, S. & Goble, C. (2013). Why linked data is not enough for scientists. *Future Generation Computer Systems*, 29(2), 599–611.

Bergstrom, T.C. (2014). Evaluating big deal journal bundles, *Proceedings of the National Academy of Sciences,* 111(26), 9425-9439.

Berners-Lee, T. (2009). *Design Issues*. Retrieved February 3, 2016, from http://www.w3.org/DesignIssues/LinkedData.html

Björk B-C., Welling P., Laakso M., Majlender P., Hedlund, T. & Guðnason, G. (2010). Open Access to the Scientific Journal Literature: Situation 2009. *PLOS ONE,* 5(6), e11273. doi:10.1371/journal.pone.0011273

Björk, B.-C., Laakso, M., Welling, P. & Paetau, P. (2014). Anatomy of green open access. *Journal of the Association for Information Science and Technology*, 65(2), 237–250. doi: 10.1002/asi.22963

Boleslaw K., Szymanski, B.K., de la Rosa, J.L. & Krishnamoorthy, M. (2012). An Internet measure of the value of citations, *Information Sciences*, 185(1), 18-31.

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059–1078.

Boulton, G., Campbell, P., Collins, B., Elias, P., Hall, W., Laurie, G., O'Neill, O., Rawlins, M., Thornton, J., Vallance, P. & Walport, M. (2012). *Science as an Open Enterprise*. The Royal Society Science Policy Centre Report 02/12, The Royal Society, UK.

Bradley, L. & Soldo, B. (2011). The New Information Poor: How Limited Access to Digital Scholarly Resources Impacts Higher Education, *The Serials Librarian*, 61(3-4), 366-376.

Buela-Casal, G., Perakakis, P., Taylor, M. & Checa, P. (2006). Measuring internationality: reflections and perspectives on academic journals, *Scientometrics*, 67(1), 45–65.

Carroll M.W. (2011). Why Full Open Access Matters. *PLOS Biololgy,* 9(11), e1001210. doi:10.1371/journal. pbio.1001210

Chang, H.H., Peng, R.D. & Dominici, F. (2011). Estimating the acute health effects of coarse particulate matter accounting for exposure measurement error, *Biostatistics,* 12(4), 637-652. doi:10.1093/biostatistics/kxr002

Coleman, A. (2007). Self-archiving and the Copyright Transfer Agreements of ISI-ranked library and information science journals. *Journal of the American Society for Information Science and Technology*, 58(2), 286–296.

Crossick, G. (2015) *Monographs and Open Access: A report to HEFCE*. Higher Education Funding Council for England, UK. Retrieved February 3, 2016, from http://www.hefce.ac.uk/media/hefce/content/pubs/indirreports/2015/Monographs,and,open,access/2014_monographs.pdf

Delman, S.E. (2011). ACM offers a new approach to self-archiving. *Communications of the ACM*, 54(11), 14.

Eaves, D. (2013). Beyond Property Rights: Thinking About Moral Definitions of Openness, Retrieved February 3, 2016, from http://techpresident.com/news/wegov/24244/beyond-property-rights-thinking-about-moral-definitions-openness

Eisen, J. (2013). How Open Are You? Part 1: Metrics to Measure Openness and Free Availability of Publications, *The Tree of Life*, Retrieved February 3, 2016, from http://phylogenomics.blogspot.ca/2013/06/how-open-are-you-part-1-metrics-to.html

Evangelou, E., Trikalinos, T. A., & Ioannidis, J. P. (2005). Unavailability of online supplementary scientific information from articles published in major journals. *The FASEB Journal*, *19*(14), 1943-1944.

Ewing, J. (2006). Measuring journals, *Notices of the American Mathematical Society*, 53(9), 1049-1053.

Fecher B., Friesike S. & Hebing, M. (2015). What Drives Academic Data Sharing? *PLoS ONE* 10(2): e0118053. doi: 10.1371/journal.pone.0118053

Finch, A. (2010). Can we do better than existing author citation metrics? *Bioessays*, 32(9), 744–747. doi: 10.1002/bies.201000053

Garfield, E. (2006). The History and Meaning of the Journal Impact Factor, *Journal of the American Medical Association*, 295(1), 90-93.

Goodman, A., Pepe, A., Blocker, A.W., Borgman, C.L., Cranmer, K., Crosas, M., Di Stefano, R., Gil, Y., Groth, P., Hedstrom, M., Hogg, D.W., Kashyap, V., Mahabal, A., Siemiginowska, A. & Slavkovic, A. (2014). Ten Simple Rules for the Care and Feeding of Scientific Data. *PLOS Computational Biology,* 10(4), e1003542. doi: 10.1371/journal.pcbi.1003542

Gorman, M.F. & Kanet, J.J. (2005). Evaluating operations management-related journals via the author affiliation index. *Manufacturing & Service Operations Management*, 7(1), 3–19.

Harnad, S. (2011). Open Access to Research: Changing Researcher Behaviour Through University and Funder Mandates. *eJournal of eDemocracy and Open Government*, 3(1), 33-41.

Hedlund, T. (2010). Open Access availability of articles by Nordic authors, *ScieCom Info*, 7(4). Retrieved February 3, 2016, from http://www.sciecom.org/ojs/index.php/sciecominfo/article/view/4760

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output, *Proceedings of the National Academy of Sciences*, 102(46), 16569-16572.

Hodgson, C. (2014). Open Access Infrastructure: Where We Are and Where We Need to Go, *Information Standards Quarterly*, 26(2), 4-14.

Hughes, A.M. (2008). *Publication Behaviors of the Signers of the Public Library of Science (PLoS) "Open Letter to Scientific Publishers"*, Masters Thesis, School of Information and Library Science, University of North Carolina at Chapel Hill, USA. Retrieved February 3, 2016, from https://cdr.lib.unc.edu/record?id=uuid%3aaa695765-1f66-4637-848c-dea704fe79a4

Hunter, J. (2006). Scientific publication packages—a selective approach to the communication and archival of scientific output, *International Journal of Digital Curation,* 1(1), 33-52.

Ince, D.C., Hatton, L. & Graham-Cumming, J. (2012). The case for open computer programs. *Nature,* 482, 485–488. doi:10.1038/nature10836

Jenkins, C., Probets, S., Oppenheim, C. & Hubbard, B. (2007). RoMEO Studies 8: self-archiving: The logic behind the colour-coding used in the Copyright Knowledge Bank, *Program: electronic library and information systems*, 41(2), 124-133.

Koskinen, K., Lappalainen, A., Liimatainen, T., Nevalainen, E., Niskala, A. & Salminen, P.J. (2010). The current state of Open Access to research articles from the University of Helsinki, *ScieCom Info*, 6(4), 1-7.

Kowalczyk, S. & Shankar, K. (2011). Data sharing in the sciences. *Annual Review of Information Science and Technology*, 45, 247–294.

Kurata, K., Matsubayashi, M., Mine, S., Yokoi, K. & Morioka, T. (2010). Enhancing open access in the biomedical field. *Proceedings of the American Society for Information Science and Technology*, 47(1) 1–2. doi:10.1002/meet.14504701383

Kurtz, M.J. & Bollen, J. (2010). Usage bibliometrics. *Annual Review of Information Science and Technology*, 44, 3-64.

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

Lane, J.I. (2010). Let's make science metrics more scientific. *Nature*, 464, 488–489.

Lee, M. & Park, H.W. (2012). Exploring the web visibility of world-class universities, *Scientometrics,* 90(1), 201–218.

Leite, P., Mugnaini, R. & Leta, J. (2011). A new indicator for international visibility: exploring Brazilian scientific community. *Scientometrics,* 88(1), 311-319.

Lewis, S. (2012). *Open Access Index*, Retrieved February 3, 2016, from https://www.youtube.com/watch?v=deHD2XFNEh0

Lynch, C. (2006). Open computation: beyond human reader-centric views of scholarly literatures, in Jacobs, N. (ed.) *Open Access: Key strategic, technical and economic aspects*, Chandos Publishing. 185-193.

Lyons, C. & Booth, H.A. (2011). An Overview of Open Access in the Fields of Business and Management, *Journal of Business & Finance Librarianship*, 16(2), 108–124.

McCormick, T. (2013). Proposal: the Open Scholar Rating, Retrieved February 3, 2016, from http://tjm.org/2013/08/07/proposal-the-open-scholar-rating/

Matsubayashi M., Kurata K., Sakai Y., Morioka T., Kato S., Mine S. & Ueda, S. (2009). Status of open access in the biomedical field in 2005. *Journal of the Medical Library Association* 97(1) 4–11.

Mercer, H. (2011). Almost Halfway There: An Analysis of the Open Access Behaviors of Academic Librarians, *College & Research Libraries*, 72(5), 443-453.

Micarelli, A., Gasparetti, F., Sciarrone, F. & Gauch, S. (2007). Personalized search on the world wide web. In Brusilovsky, P., Kobsa, A. and Nejdl, W. (Eds.) *The Adaptive Web: Methods and Strategies of Web Personalization*, LNCS4321. Springer-Verlag, Berlin/Heidelberg. 195-230.

Morin, A., Urban, J., Adams, P. D., Foster, I., Sali, A., Baker, D. & Sliz, P. (2012). Shining light into black boxes. *Science*, 336, 159–160.

Nederhof, A.J. (2008). Policy impact of bibliometric rankings of research performance of departments and individuals in economics. *Scientometrics,* 74(1), 163-174 .

Nichols, D.M. (2012). *Metrics for Openness*, presented at the Center for Informatics Research in Science and Scholarship, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. 4th May 2012. Retrieved February 3, 2016, from http://cirssweb.lis.illinois.edu/Events/eventDetails.php?id=45

Nichols, D.M., Twidale, M.B. & Cunningham, S.J. (2012). Metadatapedia: A proposal for aggregating metadata on data archiving, *Proceedings of the 2012 iConference (iConference '12)*. ACM. 370-376.

NISO (2013) *Recommended Practices for Online Supplemental Journal Article Materials*, NISO RP-15-2013, National Information Standards Organization (NISO)/National Federation of Advanced Information Services (NFAIS), Baltimore, MD, USA. http://www.niso.org/publications/rp/rp-15-2013

NISO (2015). *Access and License Indicators,* NISO RP-22-2015, National Information Standards Organization (NISO), Baltimore, MD, USA. http://www.niso.org/publications/rp/rp-22-2015

Norris, M., Oppenheim, C. & Rowland, F. (2008). The citation advantage of open-access articles. *Journal of the American Society for Information Science and Technology*, 59(12), 1963-1972.

Peng, R.D. (2009). Reproducible research and biostatistics, *Biostatistics,* 10(3), 405-8.

Peng, R.D. (2011). Reproducible research in computational science, *Science,* 334, 1226-1227.

Pesch, O. & Lamoureux, S.D. (2013). ONIX-PL License Encoding Project, *The Serials Librari*an, 65(3-4), 241-251.

Pham-Kanter G., Zinner D.E. & Campbell, E.G. (2014). Codifying Collegiality: Recent Developments in Data Sharing Policy in the Life Sciences. *PLOS ONE,* 9(9), e108451. doi: 10.1371/journal.pone.0108451

Piwowar, H. (2013). Altmetrics: Value all research products, *Nature,* 493, 159. doi:10.1038/493159a

Piowar, H. & Priem, J. (2014). What level of Open Access scholar are you?, *Impactstory blog*, http://blog.impactstory.org/what-level-of-open-access-scholar-are-you/

Priem, J. & Hemminger, B.M. (2010). Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web, *First Monday*, 15(7). doi:10.5210/fm.v15i7.2874

Prost, H. & Schöpfel, J. (2014). Degrees of Openness: Access Restrictions in Institutional Repositories. *D-Lib Magazine*, 20(7/8). doi:10.1045/july2014-prost

Research Consulting (2014). *Counting the Costs of Open Access*. Retrieved February 3, 2016, from http://www.researchconsulting.co.uk/wp-content/uploads/2014/11/Research-Consulting- Counting-the-Costs-of-OA-Final.pdf

Ruscio, J., Seaman, F., D'Oriano, C., Stremlo, E. & Mahalchik, K. (2012). Measuring Scholarly Impact Using Modern Citation-Based Indices, *Measurement: Interdisciplinary Research and Perspectives*, 10(3), 123-146.

Seglen, P.O. (1997). Why the impact factor of journals should not be used for evaluating research, *British Medical Journal*, 314, 498-502.

Shadbolt, N., Brody, T., Carr, L. & Harnad, S. (2006). The open research web, in Jacobs, N. (ed.) *Open Access: Key strategic, technical and economic aspects*, Chandos Publishing. 195-208.

Shaw, D. & Vaughan, L. (2008). Publication and citation patterns among LIS faculty: Profiling a "typical professor", *Library and Information Science Research*, 30(1), 47-55.

Shin, J.C., Toutkoushian, R.K. & Teichler, U. (Ed.) (2011). *University Rankings*, Springer.

Shotton, D. (2012). The Five Stars of Online Journal Articles — a Framework for Article Evaluation, *D-Lib Magazine*, 18(1/2). January/February 2012. doi:10.1045/january2012-shotton

SPARC (2013). *HowOpenIsIt? Open Access Spectrum*. Scholarly Publishing and Academic Resources Coalition, Washington, DC. http://www.arl.org/sparc/bm~doc/howopenisit-guide.pdf

Star, S.L. & Strauss, A. (1999). Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work*, 8(1) 9-30.

Stodden, V. & Miguez, S. (2014). Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research, *Journal of Open Research Software*, 2(1), e21, July 2014. doi:10.5334/jors.ay

Sugimoto C.R. & Cronin B. (2012). Biobibliometric profiling: An examination of multifaceted approaches to scholarship, *Journal of the Association for Information Science and Technology*, 63(3) 450-468.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Maribeth Manoff, M. & Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLOS ONE* 6(6): e21101. doi:10.1371/journal.pone.0021101

Terras, M. (2012). The impact of social media on the dissemination of research: results of an experiment, *Journal of Digital Humanities*, 1(3) Summer 2012. http://journalofdigitalhumanities.org/1-3/the-impact-of-social-media-on-the-dissemination-of-research-by-melissa-terras/

Troll Covey, D. (2009). Self-archiving journal articles: A case study of faculty practice and missed opportunity. *portal: Libraries and the Academy*, 9(2) 223-251.

Twidale, M.B., Blake, C. & Gant, J.P. (2013). Towards a data literate citizenry, *Proceedings of the 2013 iConference (iConference'13)*, 247-257. doi:10.9776/13189

Van Gijlswijk, E. & Boucherie, S. (2014). Free access to

ScienceDirect for journalists, *Elsevier Connect*. http://www.elsevier.com/connect/free-access-to-sciencedirect-for-journalists

Van Raan, A.F.J. (2005). Fatal Attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1), 133-143.

Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., Moore, J-S., Moyers, B.T., Renaut, S., Rennison, D.J., Veen, T. & Yeaman, S. (2013). Mandated data archiving greatly improves access to research data. *The FASEB Journal*, *27*(4), 1304-1308.

Vines, T. H., Albert, A. Y., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., Gilbert, K.J., Moore, J-S., Renaut, S. & Rennison, D.J. (2014). The availability of research data declines rapidly with article age. *Current Biology*, 24(1), 94-97.

Wainer, J., Goldenstein, S. & Billa, C. (2011). Invisible work in standard bibliometric evaluation of computer science. *Communications of the ACM,* 54(5), 141-146.

Wallis J.C., Rolando E. & Borgman, C.L. (2013). If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLOS ONE,* 8(7), e67332. doi: 10.1371/journal.pone.0067332

Way, D. (2010). The Open Access Availability of Library and Information Science Literature, *College & Research Libraries*, 71(4), 302-309.

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLOS ONE*, 6(11), e26828.

Wicherts, J. M., Borsboom, D., Kats, J. & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61(7), 726-728.

Willinsky, J. (2005). *The Access Principle: The case for open access to research and scholarship*. Cambridge, MA, USA. MIT Press.

Willmott, M.A., Dunn, K.H. & Duranceau, E.F. (2012). The Accessibility Quotient: A New Measure of Open Access. *Journal of Librarianship and Scholarly Communication,* 1(1), eP1025. http://jlsc-pub.org/jlsc/vol1/iss1/7

Wren, J.D. (2005). Open access and openly accessible: a study of scientific publications shared via the internet. *BMJ*, 330(7500), 1128.

Xia, J., Wilhoite, S.K. & Myers, R.L. (2011). A “librarian-LIS faculty” divide in open access practices, *Journal of Documentation*, 67(5), 791-805.

Zuccala, A. (2009). The layperson and open access. *Annual Review of Information Science and Technology*, 43, 359-396.