What counts in research? Dysfunction in knowledge creation & moving beyond

by Heather Morrison

Bio statement: Heather Morrison is an Associate Professor in the School of Information Studies, and cross-appointed to the Department of Communication, University of Ottawa. Her research focus is the economics of transition of scholarly publishing from purchase and subscriptions to support for production to facilitate open access. Heather is the Principle Investigator of the SSHRC Insight Grant (2016-2021) supported *Sustaining the Knowledge Commons* research program.

Abstract

This chapter begins with a brief history of scholarly journals and the origins of bibliometrics and an overview of how metrics feed into university rankings. Journal impact factor (IF), a measure of average citations to articles in a particular journal, was the sole universal standard for assessing quality of journals and articles until quite recently. IF has been widely critiqued; even Clarivate Analytics, the publisher of the Journal Citation Reports / IF, cautions against use of IF for research assessment. In the past few years there have been several major calls for change in research assessment: the 2012 San Francisco *Declaration on Research Assessment* (DORA), the 2015 *Leiden Manifesto* (translated into 18 languages) and the 2017 Science Europe *New vision for meaningful research assessment*. Meanwhile, due to rapid change in the underlying technology, practice is changing far more rapidly than most of us realize. IF has already largely been replaced by item-level citation data from Elsevier's *Scopus* in university rankings. Altmetrics illustrating a wide range of uses including but moving beyond citation data, such as downloads and social media use are prominently displayed on publishers' websites. The purpose of this chapter is to provide an overview of how these metrics work at present, to move beyond technical critique (reliability and validity of metrics) to introduce major flaws in the logic behind metrics-based assessment of research, and to call for even more radical thought and change towards a more qualitative approach to assessment. The collective agreement of the University of Ottawa is presented as one model for change.

A brief history of journals, bibliometrics & rankings

In 1665, two scholarly entrepreneurs independently seized the potential of the printing press and the postal system, and invented the modern scholarly journal. Guédon's (2001) *In Oldenburg's long shadow* presents an overview of the history of the scholarly peer-reviewed journal from its inception in 1665 with Oldenburg's *Philosophical Transactions* and de Sallo's *Journal des Sçavans,* to the end of the 20th century. The idea of peer review has evolved over time, but the format of journals has remained largely the same. Odlyzko (1994) predicted the impending demise of scholarly journals. Print and mail are in the process of becoming obsolete as the standard for production and dissemination of scholarly work, as this becomes electronic and web-based. The continuity of the print-based format, with online journals closely resembling print ones, reflects acceptance of the scholarly journal article as the gold standard for publication in many academic disciplines.

The growth of scholarly journals and articles since 1665 has been remarkably constant. This was first documented by De Solla Price (1963, p. 17) in *Little Science, Big Science*, and updated by Mabe & Amin (2001) and Mabe (2003). There is an average annual scholarly

journal and article growth rate of about 3 - 3.5% per year from the 1600s to the present day. If there were still just two scholarly journals producing a small volume of articles on an annual basis, it would be feasible for every scholar to read every scholarly article. However, as the volume of production grew, journals began to specialize in particular disciplines and sub-disciplines, at rates varying with the growth of the disciplines.

As production continued to increase, specialization was not enough. Guédon (2001) argues that the growing numbers of journals was the inspiration for a tendency to want to define "core journals". The purpose of the "core journal" concept was to address two problems that arose as the number of journals grew. One problem was the "serials crisis" documented by the Association of Research Libraries (1989), a combination of increasing numbers of journals and average price rises for journals beyond inflation, year after year, leading libraries to cancel subscriptions. A second problem was the increasing difficulty scholars had in keeping up with the growing literature. In that sense, identifying "core journals" would help busy scholars prioritize their readings and publication venues.

Garfield (1955) proposed "a bibliographic system for science literature that can eliminate the uncritical citation of fraudulent, incomplete, or obsolete data by making it possible for the conscientious scholar to be aware of criticisms of earlier papers". Another proposed purpose of this system was to facilitate communication among scientists. It was in this article that Garfield first coined the term "impact factor" (IF), a then-hypothetical measure of the influence of a highly cited article.

Garfield (2006) describes the history of the development of citation indexing and IF. With support from the U.S. National Institutes of Health, IF became the basis for the development of first the *Genetics Citation Index* and later the *Science Citation Index*. IF is a metric applied to journals rather than articles. It is based on 2 elements, a numerator consisting of the number of citations in the current year to items published in a particular journal in the previous 2 years, and the denominator, the number of substantive reviews and articles published in the same 2 years. In other words, IF is the average number of citations to an article in a particular journal for the previous two years. IF varies considerably by discipline and sub-discipline as well as by journal, and is often evaluated on the basis of the status by quartile within a discipline.

The *Science Citation Index* developed by Garfield and colleagues in 1961 has morphed and grown into *Web of Science*, including the Science Citation Index, the Social Science Citation Index, the Arts & Humanities Citation Index, the Emerging Sources Citation Index, the Book Citation Index, the Conference Proceedings Citation Index, in addition to optional specialized collections. *Web of Science* is one of a suite of inter-related products produced and sold by Clarivate Analytics (until recently published by Thomson Reuters), and the basis for their *Journal Citation Reports* (JCR) which provides reports of journal IF.

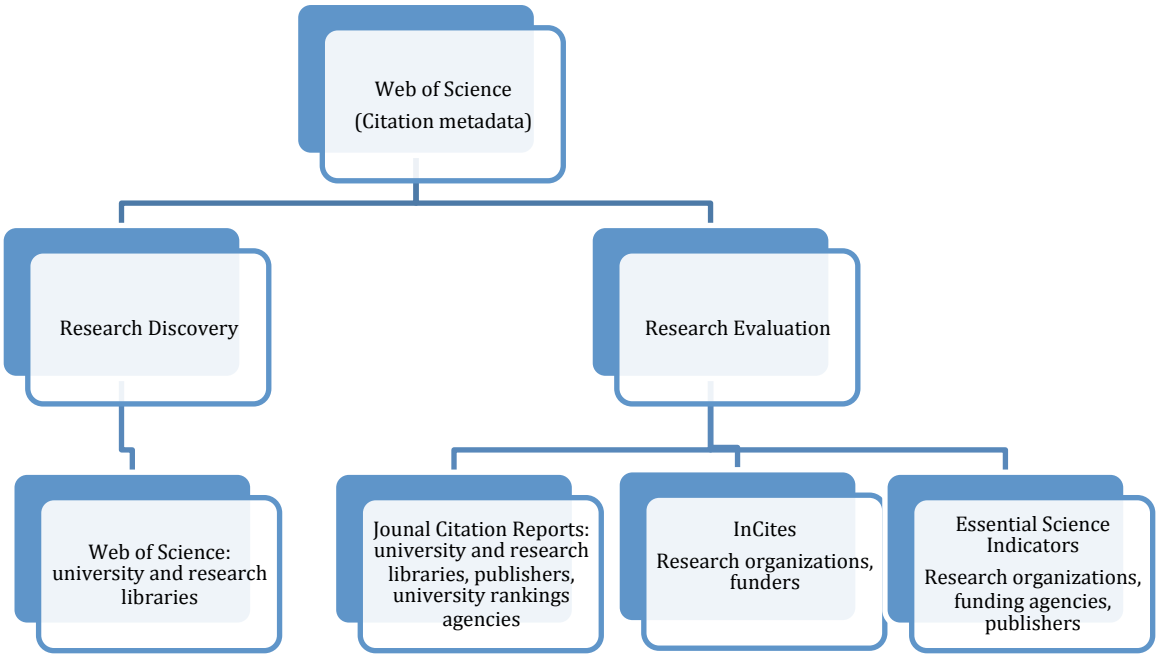

Table 1: core Clarivate products

Table 1 illustrates the relationship and evolution of the core products related to research and research metrics offered by Clarivate analytics. The initial core underlying product is a massive database of citations to journal articles, Web of Science. The research discovery tool called Web of Science is extensively used by researchers at university and research libraries for research discovery. The same underlying metadata is used for the traditional Journal Citation Reports (JCR). This is the tool originally envisioned by Garfield as a means of identifying a set of "core" or most highly cited journals so that researchers could prioritize these for reading and libraries for purchase. JCR is important as a branding tool for publishers. JCR is also used by university rankings agencies. On the right, more recent tools InCites and Essential Science Indicators focus exclusively on metrics for evaluation. These are tools for measuring researchers, not assisting researchers in their work.

At a surface level, information on different pages on the Clarivate website might appear confusing and contradictory. This is because Clarivate offers services to a diverse group of stakeholders that have different goals that are not always compatible. On the JCR website, Clarivate claims that JCR "gives you a systematic, objective means to evaluate the world's leading scientific and scholarly journals. By analysing citation references… JCR measures research influence and impact at the journal and category levels, and shows the relationship between citing and cited journals" (Clarivate Analytics, n.d.). JCR includes 11,000 journals from over 230 disciplines. This is about a third of the total active peer-reviewed journals reported by Ware & Mabe (2015) in a recent state of the industry overview report produced for the International Association of Scientific, Technical and Medical Publishers (STM) (28,100 in the English-language, 6,450 in other languages than English).

The target markets for JCR, according to the Clarivate website as of October 16, 2018 are librarians, to inform purchase and cancellation decisions; publishers and editors, to assess the effectiveness of journals in the marketplace; researchers, to identify the most influential journals in which to publish, and research managers and information analysts to "track

publication and citation patterns to aid your strategy and policy decisions".

Clarivate Analytic's (n.d.) web advice on the suitability of the use of JCR and other data derived from *Web of Science* in research assessment is dependent on the target audience of their different products. There is a marked contrast between advice on the JCR website (librarians as the primary target audience) and InCites (research organizations and funding and policy organizations as the two primary target audiences).

On the JCR website, Clarivate warns against the use of IF in assessing journal quality, stating that "Clarivate Analytics does not depend on the impact factor alone in assessing the usefulness of a journal, and neither should anyone else… The impact factor should be used with informed peer review. In the case of academic evaluation for tenure it is sometimes inappropriate to use the impact of the source journal to estimate the expected frequency of a recently published article. Again, the impact factor should be used with informed peer review. Citation frequencies for individual articles are quite varied."

In contrast, the title of the Clarivate website for InCites (like JCR, based on Web of Science data) states that it is "an **objective** analysis of **people**, programs and peers [emphasis added]". Recommended use of InCites data for assessing individual researchers as a primary use is implied for each target audience in their "Who's it for". For research organizations, InCites is presented as a means to "identify and manage research activities and their impact" as well as to "identify experts". For funding organizations, uses suggested include "identify emerging…researchers and experts" and "manage funding activity from submissions to progress reports through outcomes". This assumes that emerging researchers and experts can be objectively identified through Web of Science data, i.e. researchers who publish in high IF journals are presumably "emerging" and "expert". It is also assumed that publication in high IF journals and high citation counts are objective measures of the quality of research. Publishers are told that InCites provides a means to "identify the best authors and reviewers".

Ware and Mabe (2015) discuss increasing industry criticism of the use of citation data, particularly IF, to judge the quality of individual researchers and departments. Will this industry recognition lead to change, and if so what form will this change take? Based on major university ranking agencies' descriptions of their methods and Elsevier's description of *Scopus* (n.d.), it appears that an evolution from journal-based metrics (IF) to bibliometrics based on individual works (articles, books and book chapters, conference proceedings) has already taken place in a large sector of the market. This shift addresses a major technical critique of IF, using journal IF as a surrogate for article impact. Elsevier's *Scopus* is the major source of data for the Times Higher Education (THE) and QS *World University Rankings*, among others.

Assessment of research and researchers is often conducted by researchers themselves, for example in promotion and tenure decisions and in review of grant applications. While university rankings systems are moving towards article-level metrics, researchers' own practices are deeply ingrained in academic culture and continue to rely primarily on IF. Stephan, Veugelers & Wang (2017) discuss what they call "back-door bibliometrics", in which researchers and reviewers report and/or use journal IF in assessment even when this is not required. They also discuss the formal use of bibliometric indicators, such as the use of rankings derived from journal IF in Spain in promotion and salary increase decisions, and payment of bonuses in China according to the prestige of the journal in which a

researcher is published. In some regions such as Flanders and Brazil journal IF is used in allocating resources to universities.

When researchers focus on their own areas of specialization, one might assume that they have the background knowledge to understand commonly used metrics. However, reliance on measures and surrogate measures of journal and article influence is common, although research metrics per se is not a common research speciality. To fully understand university rankings, we need to know who produces the data that feeds into the rankings, and how they produce it.

Elsevier's *Scopus* data is the basis for 38.5% of the ranking for THE's *World University Rankings* (WUR) (THE 2018, p. 82-83). Citations or research influence account for 30% of the THE WUR. These are based on "almost 62 million citations to more than 12.4 million journal articles, article reviews, conference proceedings and book and book chapters published over 5 years". According to THE, these data "help to show us how much each university is contributing to the sum of human knowledge…whose research has stood out…been picked up and built on by other scholars…" (p. 83). THE's "Research Productivity" is a count of the "number of papers published in the academic journals indexed by Elsevier's *Scopus* per scholar, scaled for institutional size and normalised for subject". Under "Institutional Outlook", "International Collaboration" is a measure of portion of the university's total research journal publications that have at least one international co-author. The title paper of the THE report states that the work is "in partnership with Elsevier".

QS' (2018) *World University Rankings* uses Citations per Faculty, a straightforward count of citations to the works of scholars at the university being evaluated, using the *Scopus* database. Maclean's (Dwyer, 2017) added bibliometrics indicators in 2015, publications per faculty and a field-weighted citation impact factor, drawn from *Scopus*.

New or alternative metrics-based approaches

Article metrics is at present in a process of rapid evolution. Several basic trends are observable: a shift from journal to article level citation metrics, discussed above; new types of metrics or altmetrics that illustrate different types of indicators of usage, such as views, downloads, and social media usage; and inclusion of metrics and links to downstream citing articles, social media, etc., on publisher websites. In this section, I will present a brief overview illustrating the rapid implementation of diverse approaches, explain in plain terms what researchers and publishers in this area are aiming to accomplish, and argue that while some aspects of these developments are useful for research, there is a problematic lack of critical reflection on the impact of these developments.

In 2012, Haustein published a comprehensive book on the technical details and flaws of scholarly bibliometrics as of that point in time, concluding with a recommendation for a multidimensional approach to metrics to overcome the flaws evident in any one method. Many other authors, such as Khodiyar, Rowlett, & Lawrence (2014), have similarly discussed the changing nature of assessment of scholarly work.

At present, the state of practice has far outpaced scholarly conceptions of new approaches. As discussed in the previous section, while researchers continue to assume that the journal IF is state of the art in metrics-based evaluation, major university rankings and the world's

largest commercial scholarly journal publisher, Elsevier, have already moved to article-level citation metrics using data from their product Scopus.

Meanwhile, publishing practice already reflects heavy use of new or altmetrics that include beyond citations. The state of practice can be easily observed by browsing the websites of scholarly journals.

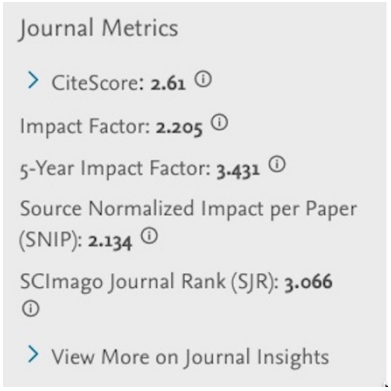

Figure 1. Elsevier's Journal of Development Economics. Source: https://www.journals.elsevier.com/journal-of-development-economics

On the right-hand side of the home page of Elsevier's *Journal of Economic Development* there is a means for readers to filter articles that includes the options "most downloaded" and "most cited". On the left-hand side of the page, as illustrated in Figure 1, there is a list of 5 journal metrics and a link to "view more on journal insights"; this page includes even more metrics. Each metric has an icon "I" for more information; hovering over the icon brings forward the technical explanation for each metric.

Figure 2: Trending. Source: *Nature Scientific Reports*. https://www.nature.com/srep/

If you scroll down the home page of Nature's open access journal *Scientific Reports* you will see a section called "Trending" with the word "Altmetric" prominently displayed at the top left hand corner as illustrated in Figure 2.

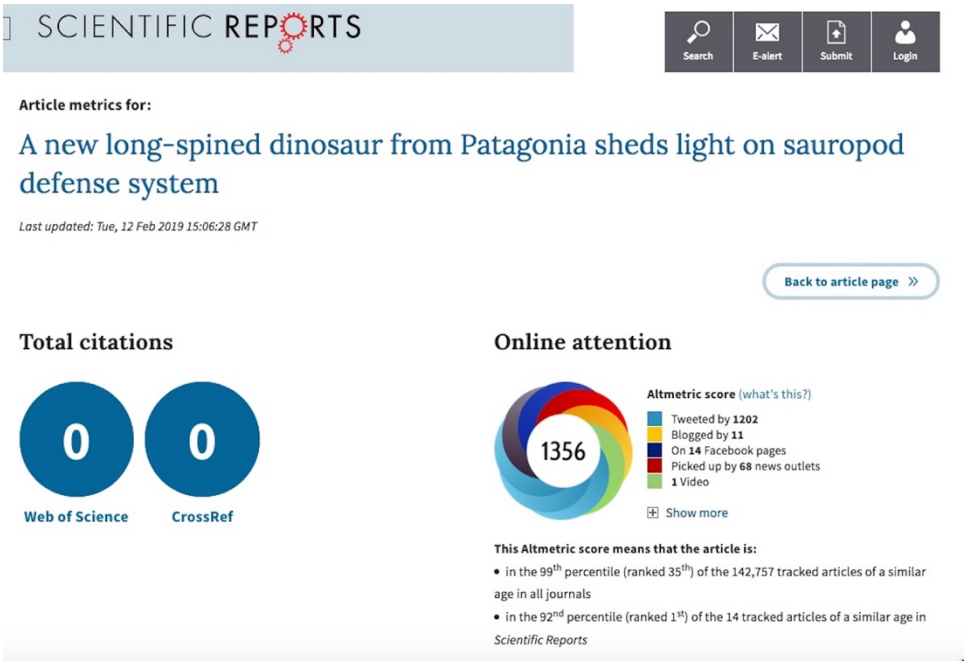

Figure 3. Article metrics for dinosaur article. Source: https://www.nature.com/articles/s41598-018-37943-3/metrics

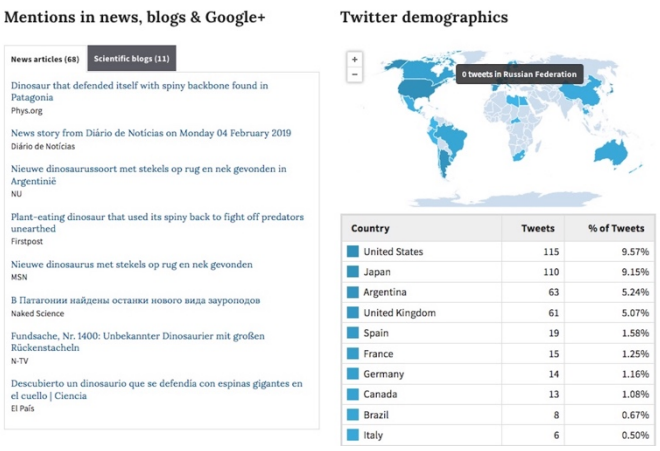

Figure 4. Mentions and Twitter demographics for dinosaur article. Source: https://www.nature.com/articles/s41598-018-37943-3/metrics

Clicking on the top trending article by altmetric (dinosaur article) reveals more detail about the metrics involved. As of February 12, 2019, there are 0 citations from Web of Science. Online Attention indicates tweeting, Facebook, and media attention as illustrated in the following Figures. Figure 3 shows total citations, online attention and the altmetrics score. Figure 4 illustrates further detail that can be found by scrolling down from Figure 3, a clickable list of media references and an option to switch to scientific blogs, as well as a map illustrating Twitter references and a list of tweeting countries in descending order by number of tweets. Further down on the page (not illustrated here) are explanations of terms and sources.

These Elsevier and Nature journals each report new or altmetrics, but not the same ones. Why? Elsevier uses citation data from its own product, *Scopus*. Nature Publishing Group uses citation data from Clarivate's Web of Science and CrossRef for *Scientific Reports* rather than *Scopus* data. Could this be because the two companies are rivals? Nature Publishing Group owns Springer, the world's second-largest commercial scholarly publisher and hence a major competitor for Elsevier. Perhaps Nature prefers not to display Elsevier's *Scopus* data, or it might be that Elsevier prefers not to provide Nature with a reasonable price for use of *Scopus* data to a competitor.

Public Library of Science (PLOS) provides a detailed explanation of their article-level metrics (ALMs) on the PLOS website: https://www.plos.org/article-level-metrics. PLOS defines ALMS as: "quantifiable measures that document the many ways in which both scientists and the general public engage with published research". Suggested uses of ALMs for researchers are to communicate impact in general and to funders, to raise a researcher's career profile, and to find collaborators.

Calls for change in research assessment

DORA

The first major call to action is the 2012 San Francisco *Declaration on Research Assessment* (DORA) initiated by the American Society for Cell Biology. DORA's recommendations state "the need to **eliminate** [emphasis added] the use of journal-based metrics such as Journal Impact Factors in funding, appointment, and promotion considerations". As of October 2018, DORA has been endorsed by over 600 organizations and 13,000 individuals, myself included.

DORA does not question the concept of measurement per se, stating that: "funding agencies, institutions that employ scientists, and scientists themselves, all have a desire, and need, to assess the quality and impact of scientific outputs. It is thus imperative that scientific output is measured accurately and evaluated wisely". Implicit in DORA is an assumption that the peer-reviewed journal article will continue to be the most frequent means of dissemination of new knowledge in the foreseeable future. Development of metrics to include new forms of research outputs such as datasets and software is encouraged.

DORA includes general recommendations and specific recommendations for funding agencies, institutions, publishers, organizations that supply metrics, and researchers. The latter, when involved in committees making decisions such as hiring, promotion, and tenure, are encouraged to make assessment based on scientific content rather than publication metrics. One of the deficiencies of all citation-based metrics noted in DORA is a skew towards review articles as authors tend to cite review articles rather than the original works that are reviewed. Publishers and researchers are called upon to encourage authors to cite original research. It is recommended that funding agencies and institutions "consider a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice". However, publishers are encouraged to "make available a range of article-level metrics" and researchers to "use a range of article metrics and indicators", which suggests a deeper quantitative rather than a qualitative approach. Publishers are asked to "encourage responsible authorship practices and the provision of information about the specific contributions of each author." Qualitative information about

author contributions would appear to advance the potential for enhanced qualitative assessment.

Leiden Manifesto

A group of self-described scientometricians, social scientists and research administrators highlight some of the major issues with journal IF and other metrics and list 10 principles to guide research evaluation, principles that were crystallized at the 19th Annual Conference on Science and Technology Conference Indicators held in Leiden in 2014; hence this list is referred to as the Leiden Manifesto (Hicks et al., 2015). The Leiden Manifesto http://www.leidenmanifesto.org/ has been translated into 18 languages.

The first principle states: "quantitative evaluation should support qualitative, expert assessment". The authors cite an "impact factor obsession", stating "soaring interest in one crude measure – the average citation counts of items published in a journal in the past two years – illustrates the crisis in research evaluation". Critique of metrics is not limited to IF. Simple publication counts can be problematic as well. An example is provided of a relatively low rating of a group of European historians in a national peer-review exercise simply because historians tend to write books rather than journal articles. H-factor increases with the researcher's age, even if the researcher does not produce new papers. The *h*-factor is also database-dependent. Computer scientists can have an *h*-factor of 10 in Web of Science but 20 – 30 in Google Scholar. Precision matters; "a single highly cited publication slightly improves the position of a university in a ranking that is based on percentile indicators, but may propel the university from the middle to the top of a ranking built on citation averages" The authors suggest that relying on a single measure will "invite gaming and goal dis-placement (in which the measurement becomes the goal)", and one of their proposed solutions is multiple metrics.

Leiden Principle 2, is to "Measure performance against the research missions of the institution, group or researcher". There is no single metric that makes sense in every research context. For example, consider an action research project designed to help a community group address an issue of concern to them. Ideally, design of the project's goals and evaluation measures should be undertaken in consultation with, or led by, the community group. The optimal measures of success would probably be real-world indicators of change. Homeless people want homes, not citations to articles about homelessness. Publication of results, written in plain language, in venues that are physically and intellectually accessible to the community, such as the community's own newsletter or blog or a local workshop may be more effective in meeting the goals of the research than publication in scholarly journals using academic jargon that the group may not have access to or understand, or presentations at scholarly conferences that group members cannot afford to attend. The same principles would apply to academic / industry and academic / government collaborations. This is not to say that traditional academic ideas of excellence do not apply, rather that measuring excellence by number of publications and citations in prestigious journals is not the optimal way to evaluate every type of research project.

Leiden Principle 3 states: "protect excellence in locally relevant research". One example of the problem is that of Spanish law which states the desirability of Spanish scholars publishing in high-impact journals. In sociology, the highest impact factor journals are published in English in the United States (US); likely as a result of this, highly cited Spanish sociologists are those who focus on either abstract models or US social problems.

Science Europe's *New vision for more meaningful research assessment*

In July 2017, Science Europe (2017a), a non-profit organization based in Brussels representing major research organizations across Europe, issued a position statement "*on a new vision for more meaningful research assessment".* The preamble contrasts the broad impact of research on society, often gained through a gradual development of new knowledge, with metrics designed to measure the impact of a specific study, and points out that it is not always possible to connect societal impact with a particular research study. The preamble goes on to discuss the concept of the value of research, a broader notion of research impact that includes societal values. Societal progress draws from both research outputs and other sources; too narrow an emphasis on concrete impact may generate unintended, and not necessarily beneficial, effects on research activity.

The conclusion states "there is great diversity in the ways in which research brings its immense value to society. Some of these ways are indirect or intangible and cannot easily be measured by strictly defined impact assessment criteria. Others are long-term or unpredictable and may not yet be visible at the time that the research is evaluated...ultimately, the best way to maximise the value of research to society is by ensuring that the research produced meets the highest standards of quality and excellence". Science Europe's statement emphasizes the importance of trust between researchers and society. This is reflected in Priority 2 of Science Europe's (2017b) *Taillinn Call for Action*, which calls for research organizations and funders to "recognise a broad notion of impact that acknowledges the societal value of research for policy and practice" and calls on policymakers, research funders, and academic to "foster the necessary cultural change to embrace the broad notion of impact".

Implementation of change in approaches to research evaluation is in an early stage. In September 2018 the European Commission endorsed Plan S and Coalition S, an ambitious plan to accelerate the transition to open access publishing; "the way we evaluate research outputs" is identified as one of the barriers to change (European Commission, 2018). Science Europe's President Marc Schiltz (2018) in a statement called *Why Plan S*, states that "We commit to fundamentally revise the incentive and reward system of science, using the San Francisco Declaration on Research Assessment (DORA) as a starting point". The basic idea is for all funding agencies, particularly Europe but the aspirations are to inspire change globally, to commit to changing how research is evaluated, from traditional to new or altmetrics. It will be interesting to observe progress towards implementation of this ambitious plan over the next few years.

Discussion

Metrics *versus* quality in research

Can metric systems capture the essentially qualitative nature of the concept of quality in research? The main goal driving development of citation indexing as described by Garfield in 1955 was so that researchers could track forwards from published work to citing works that might point out critique of the original. I argue that the whole idea of using metrics to assess the quality of research and researchers is relatively new and has not received the critical attention that it deserves. This section aims to begin a process of applying scholarly

critique to this area.

As DORA notes, there is a skew in citations towards review articles rather than original research. This raises a question: if researchers are citing review articles rather than original research, are they even reading the original research, never mind tracking downstream citations? If we assume that review articles contain all the important information from every reviewed article, capture it accurately, and that the original articles are never critiqued, retracted or refuted by subsequent research, then reading review articles is not problematic, but then would be no need for citation linking to uncover critique.

Current approaches to research assessment assume that when it comes to scholarly publishing, more is better. Given the constant increase in the volume of production of scholarly works and the availability of citation indexing to permit more careful checking, shouldn't researchers today be spending relatively more time reading rather than publishing? If they were, wouldn't they be publishing less rather than more? Assessment systems based on the premise that more is better seem likely to risk increasing errors such as invalid results. Research into current practice would be helpful. One might survey researchers on whether they actually read all of the works that they cite, whether they rely on secondary sources such as reviews or go to the originals, and whether they use citation indexes to check downstream citing sources. Or, one might analyse written publications to see whether there are errors that might have been caught with more in-depth reading.

Advancing our knowledge requires questioning underlying assumptions in addition to building on existing work. Two assumptions in the area of research assessment that should be challenged are the assumption that "impact" itself is necessarily positively correlated with good quality work and that impact is inherently desirable. The second most highly cited retracted paper according to the *Retraction Watch* (n.d.) blog is the infamous 1998 paper by Wakefield et al. published in the highly prestigious journal *Lancet*, purporting to make a connection between vaccination and autism. This article has been cited over 1,000 times in the list of journals included in *Web of Science*, with 640 citations before retraction and 468 citations after. Any of the existing or emerging metrics-based approaches to research assessments would find that this study has had a lot of impact. The article was published in a high IF journal. It is a highly cited article, which would result in high article-level rankings and boost the h-factor of all of the authors. If we consider real-world impact, the influence of this article in the anti-vaccination movement and the subsequent return of diseases such as measles demonstrate an exceptional real-world impact for a single article. This illustrates the danger of assuming that impact is necessarily good. Like almost all qualities of things in the real world, impact is neither good nor bad in and of itself, but rather must be interpreted in context.

If universities and research funders are relying on university rankings, this provides an additional incentive to focus on traditional forms of publication. For example, the bibliometrics partner of Times Higher Education (2018) in producing the *World University Rankings* is Elsevier. To measure research productivity, Elsevier counts "the number of papers published in the academic journals indexed by Elsevier's Scopus database per scholar". To calculate citations or research influence, Elsevier "examined…citations to journal articles, article reviews, conference proceedings and books and book chapters". This count of citations to particular types of works is a de facto endorsement of these types of works. It is probably not a coincidence that Elsevier is a highly profitable publisher of exactly these types of works.

My own experience as a scholar confirms this focus on a limited range of formats. Our university's online CV system is likely typical in categorizing types of publications – books and book chapters, peer-reviewed journal articles, non-peer-reviewed articles and so forth. This categorization is understandable for historical reasons, however the end result is that the majority of my works, and almost all that I consider my most important and leading-edge works, such as the open data discussed above and my scholarly blogs, *Sustaining the Knowledge Commons* and *The Imaginary Journal of Poetic Economics* are labelled as "other" and would count for little or nothing under existing metrics-based assessment approaches.

Retraction Watch http://retractionwatch.com/ bloggers track and report on published articles that were retracted after publication. Unlike JCR, InCites, or Scopus, Retraction Watch addresses the original main goal of citation indexes as proposed by Garfield in 1955, "a bibliographic system for science literature that can eliminate the uncritical citation of fraudulent, incomplete, or obsolete data by making it possible for the conscientious scholar to be aware of criticisms of earlier papers". Most retractions reflect errors discovered after publication but also occasionally fraud is uncovered and reported. Retraction Watch's (n.d.) top 10 most highly cited retracted papers lists 10 papers that have been cited more than 550 times in journals indexed in Web of Science (as of January 2018), and this list includes 2 papers that have been cited more often after retraction than before. The blog also tracks evidence that citation of retracted papers is an ongoing problem.

Retraction Watch demonstrates a fundamental flaw with current approaches to research assessment, which focuses on the *impact* of scholarly work, whether measured indirectly through metrics such as journal IF or directly through article-level metrics, and for the most part neglects the more important question of the *accuracy* of scholarly work.

Bibliometrics and the economic sustainability of scholarly communication

Increasing market concentration was the subject of investigation by the United Kingdom (UK) Office of Fair Trading (2002). As described by Morrison (2012), industry concentration and growing profits of a few large commercial scholarly journal publishers was accompanied by a significant decrease in the average number of copies of scholarly monographs produced and sold. Recently, LaRivière, Haustein & Mongeon (2015) report an increase in concentration in the scholarly publishing market, with the top 5 publishers accounting for more than 50% of the articles indexed in Web of Science. The Scholarly Publishing and Academic Resources Coalition (SPARC) (n.d.) maintains a list of journal big deal cancellations by university libraries, library groups and state library systems, and national coalitions such as the Consortium on Core Electronic Resources in Taiwan (CONCERT). This is not a healthy system; the high prices and profits of a few commercial scholarly publishers cannot be sustained by academic libraries, and the economic clout behind the big package deals result in little funding left over for publishing scholarly monographs and supporting the journals of smaller publishers, particularly in the humanities and social sciences. My research suggests the same trend of commercial concentration, involving the same companies, is emerging in open access publishing. As of 2017, the largest open access journal publisher by number of journal titles was Springer Nature (including BioMedCentral), followed by Elsevier (Morrison, 2017).

This development is not an anomaly. Identifying a sub-set of journals as "core" and therefore more desirable to publish in and more essential to purchase increases their

market value. “Core” is in quotes to emphasize that this is an essentialization of the concept for market purposes. The priorities of for-profit publishers are returning profit to shareholders or private owners, not the health of the scholarly publishing ecosystem. Thus, it is logical that journal IF exacerbates the problem of affordability of scholarly publishing and similarly logical to hypothesize that new bibliometrics-based approaches will have a similar effect.

Qualitative focused assessment: how, why, and the University of Ottawa as model

There are models for assessing research at the level of evaluation of individual researchers, programs, and institutions that exemplify an understanding of the broader value of research to society and address the complexity of the diversity of research as expressed in the Science Europe vision and the Leiden Manifesto. At my own university, the University of Ottawa, the criteria for evaluating faculty members for promotion and tenure is collaboratively developed by faculty and administration and governed by the Association of University Professors of the University of Ottawa (APUO) *Collective Agreement* (2016-2018). The full text of all sections of the collective agreement directly relevant to research assessment can be found in the Appendix.

The language of the APUO agreement addresses the question of diverse and evolving forms of scholarly works. Section 23.3.1 (h) regarding the types of material that members may submit for assessment, states “it is understood that since methods of dissemination may vary among disciplines and individuals, dissemination shall not be limited to publication in refereed journals or any particular form or methods”; one example of other forms is listed in section (a), “in the case of literary or artistic creation, original works and forms of expression”. Canada’s Social Sciences and Humanities Research Council recognizes “research-creation” as a valid form of dissemination.

Typically, a new faculty member at the University of Ottawa is hired at the rank of Assistant Professor and, after six years, applies for the senior rank of Associate Professor, a promotion that automatically invokes tenure. The criteria for research assessment at this stage are covered under the APUO Collective Agreement Section 25.3.2.2 (c). To achieve tenure, a new faculty member must demonstrate production of good quality scientific, literary, artistic, or professional works, that go beyond work done in the completion of the doctorate and that show continuous progress. Evaluation is conducted by three outside evaluators and reviewed by committees at the faculty and university-wide level. A similar process is followed when a faculty member applies for promotion to the rank of full or titular Professor. This is a holistic career-level peer review process.

Ironically, and somewhat mysteriously, in spite of this qualitative and inclusive approach to assessment of research and researchers rather than simplistic metrics, the University of Ottawa does very well in metrics-based rankings. According to the University of Ottawa & Government of Ontario Ministry of Advanced Education and Skills Development 2017-2020 *Strategic Mandate Agreement* (p. 17), “Independent national and International rankings (such as Research Infosource, QS World University and the Times Higher Education) consistently place uOttawa among the top three Ontario universities, among Canada’s top 10 research universities and among the top two per cent of the world’s universities”.

Why is this? The answer to this question is not easy to ascertain by reviewing ranking agencies’ descriptions of methodology. Times Higher Education claims for the 2019

rankings to have the first audited university rankings and appears to be the most transparent. The overall ranking for the University of Ottawa from 2012 to 2019 varied from a high of 171 (2013) to a low of 251-300 in 2017, rising to 201 – 250 in 2018 and 176 in 2019. The areas where the University of Ottawa appears to score relatively well are citations, industry income, and international outlook. Citations count for 30% of the overall weight and so are likely the major factor. There is no obvious reason from the detailed method description of THE as to why the average citations to research published by University of Ottawa faculty would have changed during this time frame. Over this time frame, the full-faculty faculty complement has decreased slightly; over time, one would expect this to decrease the number of citations with less research being produced. A slight decrease in the faculty complement would account for a slightly higher average citation rate if the faculty complement were factored in, however the THE methodology does not state that this is the case. Industry income may reflect local economic conditions. The University of Ottawa's central location in the nation's capital (close to downtown, City Hall, Parliament, surrounded by embassies, close to national corporate head offices) may be a factor in the university's strengths in industry income and international outlook. However, the wide variation in overall standing over recent years, given a stable university, suggest that changes in how rankings are calculated are a large factor in current standings.

At the University of Ottawa, regular comprehensive assessment processes that include peer assessment of faculty and student research are already in place at the program level. At minimum, each program undergoes a provincially mandated cyclical review every 7 years. This is far more in-depth assessment than counting publications or citations. For example, external reviewers conduct in-person interviews separately with faculty, students, and administrators and have the opportunity to ask questions not just about research outputs, but also about institutional support for research in terms of time, facilities, assistance with grant applications and so forth. In addition, some programs, particularly professional programs, undergo professional accreditation processes, that also review research undertaken at the departmental level. For example, the School of Information Studies undergoes a rigorous accreditation process coordinated by the American Library Association.

For a university with these in-depth, holistic research assessment practices already in place, a turn to greater reliance on simplistic metrics based on a limited and backwards-looking understanding of formats and what constitutes good quality scholarly work would be a step backwards.

The primary mission of Elsevier is returning profit to shareholders of its parent company RELX; the primary mission of Clarivate is returning profit to its private owners. The mission of the University of Ottawa (2017-2020), with respect to its role as a research-intensive university: "we provide our students with an outstanding education and enrich the intellectual, economic and cultural life of Canada, helping our country play an important and valued role among the nations of the world". The mission of each university will differ slightly, but will tend to revolve around the central functions of teaching, research as an activity designed to further our collective knowledge, and service to the academy and to society as a whole. I argue that we should trust scholars and the academy to design and implement assessment mechanisms that reflect and prioritize our goals (missions, vision), not those of outside parties whose primary interests are inherently different from our own.

Conclusion: the irrational rationality of metrics-based assessment of research

It is logical for people to want to measure progress towards the goals that we desire. Many measures are valid and logical. However, when we focus on the measures per se rather than the goals, we can end up with results that do not achieve our goals. This is what I call a superficially rational (mathematical, calculating) approach that is actually irrational in terms of what we are actually attempting to achieve, or irrational rationality. Current and emerging forms of metrics-based assessment of research and researchers display major problems with irrational rationality that create incentives that are not compatible with a goal of producing and disseminating quality research. These problems merit urgent attention before our current fixation with metrics further entrenches existing problems and new or altmetrics introduces new ones.

There are valid, logical reasons for use of some metrics in assessing research, researchers, research institutions, and publishers. A university should be able to point to a substantial corps of faculty with a research mandate and a body of research works produced by its faculty to call itself a research university. An individual researcher should be able to point to a collection of published works and/or substantive work-in-progress in order to be considered a productive researcher. Individual researchers and research teams may find it helpful to develop specific measurable goals that make sense for their own projects. Journals and other publishers can use metrics to assess marketing efforts. Bibliometrics is a useful research method for generating new knowledge. However, just because some metrics are helpful, it does not follow that ubiquitous metrics are helpful. A bit of salt adds flavor to food; excess sodium causes high blood pressure, increasing the risk of heart attack or stroke.

Evaluation based on metrics *looks* scientific, doesn't it? Numbers are objective. Metrics-based evaluation is rational and calculating; scientists often use lots of data. However, the resemblance is superficial. Logic is a powerful tool; but the validity of logical arguments depends on the validity of the underlying assumptions. In order to assess whether we are making progress in science, we need to understand how science works. As discussed above, our current approaches to science are compatible with the production of dangerously erroneous "facts" such as the equation of vaccines with autism and false belief in the safety of drugs; irreproducible research; and, a tendency to cite literature reviews that makes me wonder how often the original studies are actually read. If this is the situation in science, what about other branches of knowledge? As Camic, Gross & Lamont (2011) discuss, while there has been some reflection on practice in the sciences since publication of Kuhn's (1962) *The Structure of Scientific Revolutions* in the 1960's and subsequent development of the field of science and technology studies (STS), parallel study of processes in the area of the social sciences is just beginning. If we do not even know what scholars in the social sciences do, how can we claim to know how to measure whether they are doing it well?

The study of philosophy and practice of science, while more advanced than the study of social sciences, raises more questions than answers about metrics-based approaches. For example, it is logical to assume that the paradigm shifts described by Kuhn (1962) will lead to situations where whether works are cited, and by whom, depends on the phase of development of new ideas. One might hypothesize that works that fit a current paradigm will be cited more than pioneering works, in which case using citations to assess the value of research will tend to incentivize conservatism over innovation.

I argue that Kuhn's work itself illustrates the problem. Thomas Kuhn was a young, white, highly educated male based in the United States who wrote about scientific revolutions in the 1960's. This work was widely read, studied, and cited, within a short time after publication. An earlier work, Fleck's 1930's ground-breaking *Genesis of a scientific fact* (Fleck, 1979), did not enjoy this immediate acclaim. In the foreword to the 1979 edition, Kuhn describes finding this work, written in German, by happenstance while browsing in the stacks of a library. Fleck was a Jewish intellectual lacking formal credentials whose work was published in Germany in the 1930's. Unlike the popular reception of Kuhn's work, only 600 copies were printed of Fleck's work, only 200 were sold, and only 6 were delivered to the United States. Kuhn's philosophy was inspired by this work, but this does not render the work obsolete as Kuhn's ideas complement rather than supersede those of Fleck.
For example, where Kuhn emphasized sudden ruptures in scientific thinking, Fleck emphasized continuity of basic premises in apparently revolutionary advances in knowledge.

What does this have to do with university rankings? I have described a deeply flawed system, with an illusory appearance of scientific basis, that incentivizes quantity of production of research works over quality, novelty over rigorous critique and replication. The metrics behind this system feed into university rankings, and the rankings reinforce this trend toward irrational rationality.

The current trend towards new or altmetrics will create even more irrational rationality. It is logical to expect that these new metrics, particularly metrics that do not depend on academic citations, will amplify existing problems with metrics-based evaluation and/or create new ones. I predict that such metrics will reflect pre-existing social biases. The extent to which individual works are cited, downloaded, and/or shared via social media are likely to correlate with gender and ethnic biases as well as the popularity of topics studied. In addition, metrics that do not depend on academic citations (downloads, tweets, etc.) are far more vulnerable to deliberate manipulation. The fossil fuel industry can afford to hire people to download and tweet evidence of climate change denial, for example. Another factor that should be considered before using such data as a surrogate for quality of research is the impact that usage of such metrics could have on the research itself. For example, if cancer researchers find it helpful to use social media, they can and should do so. But if metrics based on non-academic use were to form the basis of assessment and research in future, this could result in a redirection of efforts from cancer research to social media sharing.

As for the impact of popularity of topics in new metrics, on a cynical note I take comfort in the possibility that someday I may have reason to move forward with a study along the lines of "correlates of perceived attractiveness of juvenile felines on YouTube" [academese for how come those kitty cats on YouTube are so darn cute] to prove my worthiness as a researcher. On a serious note, it is my experience as a long time practitioner of open research that the popularity of my works does not correlate with the importance of its contribution. My ground-breaking book chapter "The implications of usage statistics as an economic factor in scholarly communications" begins some of the discussion that continues with this chapter and introduces important but counter-intuitive ideas (Morrison, 2005). This work does not enjoy even a small fraction of the social media popularity of my *Dramatic Growth of Open Access* blog series, designed to support the advocacy efforts of a global movement.

The evolution of research metrics described in this chapter captures the irrational rationality of metrics-based research assessment. In the 1950's and 1960's we developed tools to help researchers and libraries cope with the ever-expanding volume of scholarly literature by connecting citing and cited works and identifying highly cited journals to help libraries with decisions about purchases and cancellations, and researchers with decisions about reading and publishing. The resulting metric, IF, became a yardstick for evaluating the worthiness of research and researchers even when it was acknowledge by experts and the producing company that this metric was not at all suitable for this purpose. Attempts to address the technical flaws of IF (connected to journals rather than articles) are adding a new layer of metrics based on citations to individual works that ranking companies are already incorporating into assessment of universities and that are being marketed as a means of assessing researchers. The profit goals of metrics based companies (scholarly publishers, citation metrics and ranking services) are overtaking the research missions of universities. We develop tools to help us achieve our goals, then we become slaves to the tools. That is irrational rationality. In the future, if we continue on the current trajectory, we should expect an additional layer of much more illogical metrics-based control of research and researchers in the form of altmetrics based on usage beyond the academy.

If excess reliance on metrics is the problem, what is the remedy? Let's develop and use metrics where they make sense, based on the goals of individual research projects and institutions. But let's do so with a grain of salt and not rely on metrics where such reliance is not scientific and may be counter-productive by creating perverse incentives for quantity and novelty over quality and favour particular formats, even as they are become obsolete. How can we implement this remedy? We can use approaches that appropriately weight quality and that recognize the diverse forms of research. Instead of translating a research dossier into un-scientific metrics, read and review the works. This isn't new, and shouldn't be hard. This is what we do now when we assess a thesis or peer-review the works of other researchers. At the University of Ottawa, we have a collective agreement that acknowledges the diversity of research and its products; details are in the Appendix if anyone would like to consider this as one potential model for change.

References

APUO *Collective Agreement* (2016-2018). Retrieved January 25, 2018 from http://www.apuo.ca/collective-agreement/collective-agreement/

Association of Research Libraries (ARL). (1989). *Report of the ARL serials prices project: A compilation of reports examining the serials prices problem*. Washington, DC: The Association of Research Libraries. Retrieved January 15, 2018 from http://catalog.hathitrust.org/Record/001527850

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454 (26 May 2016) doi:10.1038/533452a

Bruckner, Till, & Ellis, Beth. (2017). Clinical trial transparency: a key to better and safer medicines. Retrieved August 1, 2017 from https://media.wix.com/ugd/01f35d_0f2955eb88e34c02b82d886c528efeb4.pdf

Camic, C.; Gross, N., Lamont, M. (2011). *Social knowledge in the making*. Chicago: The University of Chicago Press.

Casadevall, A. & Fang, F. (2010). Reproducible science. *Infection and Immunity* 78:12. doi: 10.1128/IAI.00908-10. Retrieved May 2, 2017 from http://iai.asm.org/content/78/12/4972.full

Clarivate Analytics (n.d.). *Products: Journal Citation Reports*. Retrieved January 15, 2018 from https://clarivate.com/

Dwyer, M. (2017). Measuring excellence: how we rank Canada's universities. *Maclean's*. Retrieved January 23, 2018 from http://www.macleans.ca/education/measuring-excellence-how-we-rank-canadas-universities/

Elsevier *Scopus* (n.d.). Retrieved January 23, 2018 from https://www.elsevier.com/solutions/scopus

European Commission (2018). Plan S and cOAlition S – accelerating the transition to full and immediate Open Access to scientific publications. (Statement). Retrieved Feb. 18, 2019 from https://ec.europa.eu/commission/commissioners/2014-2019/moedas/announcements/plan-s-and-coalition-s-accelerating-transition-full-and-immediate-open-access-scientific_en

Fleck, L. (1979.). *Genesis and development of a scientific fact.* (T. Trenn & R.K. Merton, Eds; T. Trenn & F. Bradley, Trans.) Foreword by Kuhn, T.S. Chicago : University of Chicago Press.

Garfield (1955). Citation indexes to science: a new dimension in documentation through association of ideas. *Science*. 1955: 122:108-111. Retrieved January 15, 2018 from http://garfield.library.upenn.edu/essays/v6p468y1983.pdf.

Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of the American Medical Association* 295:1.

Guédon, J. C. (2001). *In Oldenburg's long shadow*. Washington, D.C.: Association of Research Libraries.

Haustein, S. (2012). *Multidimensional journal evaluation: analyzing scientific periodicals beyond the impact factor*. Berlin: De Gruyter / Saur.

Hicks, D.; Wouters, P.; Waltman, L.; De Rijcke, S. & Rafois, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature* 520, 429–431 (23 April 2015) doi:10.1038/520429a.

Khodiyar, V.K., Rowlett, K.A. and Lawrence, R.N. (2014). Altmetrics as a means of assessing scholarly output. *Learned Publishing*, 27: S25–S32. doi:10.1087/20140505

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

LaRivière, V.; Haustein, S. & Mongeon, P. (2015). The oligopoly of academic publishers

in the digital era. *PLOS ONE* June 15, 2015. Retrieved January 15, 2018 from https://doi.org/10.1371/journal.pone.0127502

Mabe, M. (2003). The growth and number of journals. *Serials*, 16(2), 191-197.

Mabe, M., & Amin, M. (2001). Growth dynamics of scholarly and scientific journals. *Scientometrics*, 51(1), 147-162.

Mahood, S. (2015). Privatized knowledge and the pharmaceutical industry. In: Elliott, P. & Hepting, D. *Free Knowledge: confronting the commodification of human discovery*, pp. 26 - 38. Regina: University of Regina Press.

Mignolo, Walter D. (2002). The Geopolitics of Knowledge and the Colonial Difference. *South Atlantic Quarterly, 101*(1), 57-96.

Morrison, H. (2005). The implications of usage statistics as an economic factor in scholarly communications. In: *Usage statistics of e-resources*, D. Fowler (ed.). Haworth Press.

Morrison, H. (2012). The commercialization and rationalization of scholarly publication. In: Morrison, H. *Freedom for scholarship in the internet age*. Doctoral dissertation, Simon Fraser University. Retrieved Jan. 15, 2018 from http://summit.sfu.ca/item/12537

Morrison, H.; Brutus, W.; Dumais-Desrosier, M.; Laprade, K.; Merhi, S.; Ouerghi, A.; Salhab, J.; Volkanova, V.; Wheatley, S. (2017), *Open access article processing charges 2016*, doi:10.5683/SP/KC2NBV, Scholars Portal Dataverse.

Morrison, H. (2017). Elsevier as an open access publisher. *The Charleston Advisor* 18:3, pp. 53 – 59. Retrieved January 15, 2018 from http://www.ingentaconnect.com/content/charleston/chadv/2017/00000018/00000003/art00014

Morrison, H.; Brutus, W.; Dumais-Desrosiers, M.; Kakou, T.L.; Laprade, K.; Merhi, S.; Ouerghi, A.; Salhab, J.; Volkanova, V.; Wheatley, S. (2017). Open Access Article Processing Charges (OA APC) Longitudinal Study 2016 Dataset. *Data 2*:13.

Odlyzko, A. M. (1994). Tragic loss or good riddance? the impending demise of scholarly journals. *Journal of Universal Computer Science*, 0(0), 1-52.

Price, D. J. d. S. (1963). *Little science, big science*. New York: Columbia University Press.
QS World University Rankings (2018). *Methodology*. Retrieved January 22, 2018 from https://www.topuniversities.com/qs-world-university-rankings/methodology

Retraction Watch (n.d.). *Top 10 most highly cited retracted papers*. Website. Retrieved January 22, 2018 from http://retractionwatch.com/the-retraction-watch-leaderboard/top-10-most-highly-cited-retracted-papers/

San Francisco *Declaration on Research Assessment* (2012). Retrieved January 15, 2018 from http://www.ascb.org/dora/

Schafer, A. (2015). Pseudo-evidence-based medicine: when biomedical research becomes an adjunct of pharmaceutical marketing. In: Elliott, P. & Hepting, D. *Free Knowledge: confronting the commodification of human discovery*, pp. 39 - 55. Regina: University of Regina Press.

Schiltz, M. (2018). *Why Plan S?* Retrieved Feb. 18, 2019 from https://www.coalition-s.org/why-plan-s/

Scholarly Publishing and Academic Resources Coalition (SPARC) (n.d.). *Big deal cancellation tracking*. Website. Retrieved January 15, 2018 from https://sparcopen.org/our-work/big-deal-cancellation-tracking/

Science Europe (2017a). *Science Europe position statement: on a new vision for more meaningful research assessment*. Retrieved January 15, 2018 from http://www.scienceeurope.org/tallinn-call-for-action-takes-up-science-europe-recommendation-on-research-impact/

Science Europe (2017g). *Taillinn Call for Action*. Retrieved January 17, 2018 from http://www.scienceeurope.org/tallinn-call-for-action-takes-up-science-europe-recommendation-on-research-impact/

Smith, R. (2006). *The trouble with medical journals*. London: Royal Society of Medicine Press.

Stephan, P.; Veugelers, R. & Wang, J. (2017). Reviewers are blinkered by bibliometrics. *Nature* 544, p. 411 - 412 (27 April 2017) doi:10.1038/544411a.

Times Higher Education (2018). *World University Rankings 2018*. In partnership with Elsevier. Retrieved Jan. 22, 2018 from https://www.timeshighereducation.com/world-university-rankings/2018/world-ranking

U.K. Office of Fair Trading. (2002). *The market for scientific, medical and technical journals.* No. OFT 396 U.K. Office of Fair Trade. Retrieved September 13, 2011 from http://www.oft.gov.uk/advice_and_resources/publications/reports/media/

University of Ottawa & Government of Ontario Ministry of Advanced Education and Skills Development (2017). *Strategic Mandate Agreement* 2017-2020.

Ware, M. & Mabe, M. (2015). *The STM report: an overview of scientific and scholarly journal publishing*. 4th ed. The International Association of Medical, Technical and Scientific Publishers (STM). Retrieved May 2, 2017 from http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf

Appendix

The Association of University Professors of the University of Ottawa (APUO) Collective Agreement (2016-2018) section 23.3 (assessment of scholarly activities) states:

> "23.3.1 (a) The member may submit for assessment articles, books or contributions to books, the text of presentations at conferences, reports, portions of works in progress, and, in the case of literary or artistic creation, original works and forms of expression.
> (h) It is understood that since methods of dissemination may vary among disciplines and individuals, dissemination shall not be limited to publication in refereed journals or any particular form or methods".

Typically, a new faculty member is hired at the rank of Assistant Professor and, after six years, applies for the senior rank of Associate Professor, a promotion that automatically invokes tenure. The criteria for research assessment at this stage is covered under the APUO Collective Agreement Section 25.3.2.2 (c) and reads as follows:

> "(c) The Member has produced scientific, literary, artistic, or professional works -- or a combination thereof -- which are, in accordance with the criteria set forth in 23.3.3.2, deemed of good quality. This assessment shall be made following an overall evaluation of the Member's scholarly works, carried out in accordance with the provisions of section 23.3, during which the opinion of three (3) outside evaluators will have been obtained, in accordance with 23.3.2".

23.3.3 Level of performance of scholarly activities states:

> *23.3.3.1 Whenever this agreement refers to satisfactory performance of scholarly activities by a Faculty Member, it refers to a situation where the Member is regularly engaged in scholarly activities the results of which indicate that her performance, in comparison to a relevant group of peers of comparable rank and experience, is satisfactory.
>
> *23.3.3.2 The Member's scholarly works shall be considered good if they represent a contribution in addition to that contained in the Member's doctoral thesis or to the work that has been taken to be the equivalent of a doctorate, and if, subsequent to that work:
>
> (a) in the case of research, they demonstrate continuous progress in the development of the Member's research activities and contribute to the advancement of knowledge in the Member's field of specialization;
>
> (b) in the case of literary or artistic works, they attest to continuous creative activity, well-reputed in the literary or artistic community outside the University of Ottawa;
>
> (c) in the case of professional works, they attest to the practice of a profession above and beyond that which is generally expected of a non-teaching, practicing professional, or they can be considered as a valuable contribution to the advancement of the profession itself.