

# BULLETIN

of the ASSOCIATION for  
Information Science and Technology

*asis&t* APRIL/MAY 2013

Volume 39, Number 4  
ISSN: 1550-8366

The New  
ASIS&T  
Unveiled!

*asis&t*  
ASSOCIATION for INFORMATION  
SCIENCE and TECHNOLOGY

page 5

## SPECIAL SECTION

### Altmetrics: What, Why and Where?

#### 8] Introduction

by Heather Piwowar, Special Section Guest Editor

#### 10] The Power of Altmetrics on a CV

by Heather Piwowar and Jason Priem

#### 14] Open Access and Altmetrics: Distinct but Complementary

by Ross Mounce

#### 18] Ask Not What Altmetrics Can Do for You, But What Altmetrics Can Do for Developing Countries

by Juan Pablo Alperin

#### 22] New Opportunities for Repositories in the Age of Altmetrics

by Stacy Konkiel and Dave Scherer

#### 27] The Many Faces of Article-Level Metrics

by Jennifer Lin and Martin Fenner

#### 31] Five Challenges in Altmetrics: A Toolmaker's Perspective

by Jean Liu and Euan Adie

#### 35] Are Alternative Metrics Still Alternative?

by Mike Buschman and Andrea Michalek

## FEATURE

#### 40] The Reception of Suzanne Briet in the United States

by Michael K. Buckland

## DEPARTMENTS

[2]

Editor's Desktop

[3]

President's Page

[5]

Inside ASIS&T

## COLUMNS

[42]

IA

Enabling Action:  
Digging Deeper into  
Strategies for Learning  
by Thom Haller

[44]

RDAP

Collaborative Annotation  
for Scientific Data  
Discovery and Reuse  
by Kirk Borne

NEXT PAGE >

by Irene L. Travis

**IRENE L. TRAVIS**

Editor

*Bulletin of the Association  
for Information Science and  
Technology*

[Bulletin<at>asisis.org](mailto:Bulletin@asisis.org)

Welcome to the inaugural issue of the *Bulletin of the Association for Information Science and Technology*. ASIS&T has just changed its name from *American Society for Information Science and Technology* to the *Association for Information Science and Technology*, while keeping the ASIS&T acronym. This move reinforces the Association's focus on becoming a truly international organization. The *Bulletin* will continue to play its part in covering international developments and publishing contributions from our worldwide ASIS&T community. For additional information, please see the [President's Page](#), where ASIS&T President Andrew Dillon discusses the name change as well as the ongoing review of our online presence and our efforts to enhance the visibility and understanding of the information professions.

The international theme is also primary as [Michael Buckland](#) discusses the reputation of French information scientist Suzanne Briet (1894-1989) in the United States. Because only a small part of her work had been translated into English, she was not well known among U.S. information scientists and librarians when she visited in 1951-52, but historians of information science, particularly ASIS&T members, have worked successfully in recent years to expose many in the field to the writings and ideas of "the antelope lady."

Altmetrics is the subject of this issue's [special section](#) assembled by guest editor Heather Piwowar. Altmetrics are

alternative measures that can supplement citation counts and journal impact factors as measures of the impact of scholarly communications. Such measures are generally derived from online activity such as mentions, downloads, tweets, blog posts, Facebook "likes," bookmarking and other similar evidence of attention. Some of the altmetrics services also allow readers to follow links back to actual events to determine the context and tenor of comments or actions. Altmetrics are especially important as article-level measures for material posted to open access journals or journals in developing countries or emerging disciplines, for scholarly communication generating public discussion and for measuring the impact of alternative types of scholarly activity such as datasets, software or performances. The section covers the topic from many different perspectives, including those of repositories, open access publishers, third world scholars and leading developers of altmetric services.

In our other regular columns, guest [RDAP columnist](#), Kirk Borne, professor of astrophysics and computational science at George Mason University, advocates *human computation*, utilizing humans in the ways they can be uniquely effective in partnership with computing machines. The focus of his attention is on the use of human computation in collaborative annotation of large datasets. Thom Haller, associate editor for information architecture, examines models for the search process in the [IA Column](#). ■

by Andrew Dillon

**ANDREW DILLON**

Dean and professor  
School of Information  
University of Texas at Austin  
[adillon@school.utexas.edu](mailto:adillon@school.utexas.edu)

**EDITOR'S SUMMARY**

As the president of the newly renamed Association for Information Science and Technology, Dillon recognizes critical behind-the-scenes work being done to improve ASIS&T from the inside out. The Task Force on Web Presence will develop a plan to enhance internal communications and to keep information current and easy to find. Developing a moderated discussion forum on information will help solidify the organization's position as a thought leader. An information resource on what information professionals do will help define the field for public viewers and solicit thoughts on careers and opportunities. Recognizing the value of volunteer involvement, input and assistance, Dillon invites all ASIS&T members to participate actively at this turning point for the global association.

**KEYWORDS**

information associations  
organizational communication  
Internet information resources  
information professionals  
international aspects

Welcome to the first column by a president of the Association for Information Science and Technology. Since the last issue of the *Bulletin*, the name change has become official and while it will take a few more steps to complete the changes on every listing and publication, I suspect the greater challenge will be for all of us to get used to uttering the new name without hesitation. We are still ASIS&T, of course, but it will take time for us to unpack the acronym automatically. Regardless, this is a landmark event in the history of our society – now *association* – and it represents true recognition of our international nature.

While the name change has occupied much of our collective, public discussion over the last few months, several other important projects are afoot within ASIS&T. Primary among these is the work of Diane Neal's Task Force on Web Presence, which is charged with examining the ASIS&T website, as well as our use of social media and online communication tools. The goal is to recommend to the Board a plan to improve the delivery and use of online resources for all members. ASIS&T has successfully launched a webinar program, but the view of many members is that our website is not always current, that information needed for the effective organization of committees, chapters and SIGs can be difficult to find and that as a leading information association, we do not make the best use of available tools and resources to reach and serve our members. Obviously, no one plan can solve everything, and it is clear that there will be non-trivial costs involved – but we must address these issues. If you have views on these issues, be sure to let me know.

**For ASIS&T, this is a moment of opportunity to ensure our status as a defining voice in the world of information. Get involved.**

One area where I believe ASIS&T needs to take the lead is in the provision of an appropriately moderated discussion forum for information ideas. If, like me, you subscribe to other lists in this field, you are probably tired of the endless puff pieces or announcements of events and talks that you could not possibly attend, which seemingly make up the majority of traffic on such lists. Yet when I attend a conference or meet up with colleagues from other programs, the conversations we fall into tend to be far more compelling, detailed and relevant than what appears online. It has always been a desire of mine to participate in a thoughtful, informative discussion group for our field, and I believe ASIS&T should take the lead in providing one. The best lists I know are invitation-only, heavily moderated and somewhat private. I understand why these work as they do, but I would much rather have ours be open and give us the ability to conduct year-round conversations and idea-sharing discussions that characterize the best conference experiences. After all, it's hard to be a thought leader if you don't share your thoughts, and within ASIS&T we should provide the means to do so.

Another initiative worth attention is the work started by past president Diane Sonnenwald to create a new resource for information professionals that would serve as point of reference outlining the range of careers and opportunities for

people interested our field. ASIS&T is taking the lead in helping to shape broader understanding of the information profession and has reached out to other relevant groups and professional associations for potential partnership. Again, this proposition is possibly expensive, but one has to wonder why, in 2013, our profession is so poorly understood. Do a search on *information professional* and take a moment to read the results. Is this you? One website will not provide all the answers, but if we can address this matter constructively, in partnership or on our own, it would be progress.

All this activity tends to go on behind the scenes, but it speaks to the essence of making our society more interactive, more publicly engaged and more communicative. We are a volunteer organization, and if you want to be part of the improvements, you need to participate actively. I have been impressed so far in my presidency by the willingness of so many to step up, but we are always seeking input and assistance, so please don't wait to be asked. Not sure where to start? Try a local chapter or a SIG or send me a line. The future of all professional associations is unclear at this time as members seek increased value and return for their dues. For ASIS&T, this is a moment of opportunity to ensure our status as a defining voice in the world of information. Get involved. ■

## The New ASIS&T Unveiled

The wait is over – ASIS&T is now the Association for Information Science and Technology! Following the membership balloting in 2012 in which nearly 90% of all ballots cast approved the name change, the various legal

requirements involved in such action have now been met.

While the ASIS&T acronym stays the same, the name change recognizes the growing influence of ASIS&T in the international arena. The

opportunities and challenges with respect to the science and technology of information are increasingly international in focus and

scope. ASIS&T supports members around the globe in addressing these opportunities and challenges.

In 2012, then-ASIS&T president Diane H. Sonnenwald, working with her immediate past president Linda C. Smith and then-president-elect Andrew Dillon, called for a membership vote on the name change issue, explaining, “The word *American* in our name often makes it difficult for individuals outside the United States to receive recognition for belonging to and participating in ASIS&T. It also fails to recognize the important contributions members outside the United States make to our association and to our discipline.”

In addition, Sonnenwald noted that increasing international participation in ASIS&T will provide additional opportunities for all members to learn from and share expertise and knowledge with colleagues who have different expertise and knowledge.

Coincidentally, but reflective of the growing international nature of ASIS&T and its members and interests, the 2013 ASIS&T Annual Meeting will be held in Montreal, Canada, in November.

The new name reflects the commitment of our members to international cooperation and global efforts to increase the influence of information science in education, research and applications to ensure the best access, management and use of information in an increasingly interconnected world. ■



## ASIS&T History Fund Seeks Entries for Research and Publication Awards

In 2000, the ASIS&T Board of Directors established the history fund to support and encourage research and publication in the history of information science and technology. Supported by donations from ASIS&T members and others, the fund’s advisory board has established two competitive awards to achieve the board’s goal of recognizing the historical study of information science and technology.

### ASIS&T History Fund Research Award

All topics relevant to the history of information science and technology are eligible for this award. Proposals should

include the central topic or question to be researched and an extended abstract, qualifications of the researcher (brief vita should be included), a budget and how the funds will be expended. All funds must be expended by June 30, 2014. The award for the best research proposal considered will be for a maximum of \$1000. Proposals are due May 15, 2013.

### ASIS&T History Fund Best Paper Award

All topics relevant to the history of information science and technology are eligible for this award. Submitted papers must not have been previously published

or submitted to a journal. Papers should not exceed 30 pages double-spaced, including notes and references, using the *APA Style Manual*. The award for the best paper will be for a maximum of \$500. Papers are due by May 15, 2013.

### History Fund Advisory Board

Members of the ASIS&T History Fund Advisory Board are **Sarah Buchanan**, chair; **Lai Ma**, chair-elect; **Michael Buckland**, **Samantha Hastings** and **Trudi Bellardo Hahn**.

For more information on this year’s history fund competitions, visit [www.asis.org/awards/history\\_fund\\_awards.html](http://www.asis.org/awards/history_fund_awards.html) or contact Lai Ma at [malai0008 <at>gmail.com](mailto:malai0008@gmail.com). ■

The International Calendar of Information Science Conferences ([icisc.neasist.org/](http://icisc.neasist.org/)) is a nonprofit collaboration between the Special Interest Group / International Information Issues (SIG/III) and the European (ASIST/EC) and New England (NEASIST) chapters of the Association for Information Science and Technology, with the additional support of Haworth Press.

News about ASIS&T Members

**Don Kraft**, former ASIS&T president, longtime editor of the *Journal of the American Society for Information Science and Technology*, winner of the prestigious ASIS&T Award of Merit and retired member of the Louisiana State University computer science department, has been elected as a 2013 Fuzzy Fellow of the International Fuzzy Systems Association (IFSA). The IFSA Fuzzy Fellow award bestows honor and recognition on people who have made outstanding contributions to the field of fuzzy sets and the related disciplines. Criteria for selection are based upon outstanding technical contributions, pioneering applications and support and development of the infrastructure of the fuzzy community.

**Diane Kelly**, associate professor at the University of North Carolina at Chapel Hill’s School of Information and Library Science (SILS), earned the prestigious 2012 Karen Spärck-Jones Award from the information retrieval specialist group of the British Computer Society. The annual award honors and commemorates Karen Spärck-Jones, a champion of bringing women into the world of computing and the 2002 recipient of the ASIS&T Award of Merit. In honoring Diane Kelly for this award, the panel recognized Diane’s important contributions to the analysis of information-seeking behaviors and to the development of new experimental methods and systems to support information seeking and analysis.

**Blaise Cronin**, Rudy Professor of Information Science at Indiana University, will receive the 2013 Derek de Solla Price Award and medal. This award, conferred by the journal *Scientometrics*, honors scholars who have made outstanding contributions to the fields of quantitative studies of science. Cronin is the editor of the *Journal of the American Society for Information Science and Technology* and former editor of the *Annual Review of Information Science and Technology* and a recipient of the ASIS&T Award of Merit.

Goodyear Professor of Knowledge Management **Denise Bedford** and professor **Marcia Zeng**, both at Kent State University, were among the co-organizers of a NKOS/CENDI Workshop, *Magnet for the Needle in a Search Haystack*. Bedford also delivered a presentation titled “The 11 Views of Semantic Search.”

Kent State professor **Marcia Zeng**, presented a webinar on linked open data for libraries, archives and museums (LAM), one of six webinars organized and offered by the Food and Agriculture Organization (FAO) of the United Nations through the division of Agricultural Information Management Standards (AIMS). The series of free webinars introduces the concept of linked open data to the agricultural information management community and are offered in the seven UN languages. Zeng delivered her webinar in Chinese language to more than 50 participants from the United States and China.

News about ASIS&T Institutional Members

The **School of Information and Library Science (SILS)** at the **University of North Carolina at Chapel Hill** is one of the most recent organizations to join the Open Planets Foundation (OPF), an organization established to provide practical solutions and expertise in digital preservation, building on the research and development outputs of the Planets project (see [www.planets-project.eu/](http://www.planets-project.eu/)).

OPF addresses core digital preservation challenges by engaging with its members and the community to support the development of practical and sustainable tools and services to ensure long-term access to digital content. Its mission is to ensure that its members around the world are able to meet their digital preservation challenges with a solution that is widely adopted and actively being practiced by national heritage organizations and beyond. The international foundation includes members from around the world and hosts a variety of conferences and events including webinars and hackathons.

SILS is the first iSchool to join OPF.

News about ASIS&T Chapters

The **New England Chapter of the Association for Information Science & Technology (NEASIS&T)** offered a timely workshop in March entitled *So You Want to Publish an Ebook? Tools, Standards and Metadata for Creation and Innovation!* The half-day event featured presentations on an exciting project on open-source

online textbook creation, a discussion about the latest EPUB standard which supports wildly innovative e-publishing capabilities and a demonstration of open-source publishing tools you can use to publish your own creative content.

News from ASIS&T SIGs

**SIG/III Active on Two Award Fronts**  
**13th International Paper Competition**

Special Interest Group/International Information Issues (SIG/III) is accepting submissions for its 13th International Paper Competition, which will culminate with winning authors presenting their papers at the 2013 Annual Meeting, November 1-6, 2013, in Montreal.

Building on the overall conference theme, the theme for this year's paper contest is **Beyond the Cloud: Rethinking Information Boundaries**. Papers could discuss issues, policies and case studies on specific aspects of the theme from a global and/or international perspective. Up to three winners will be selected by a panel of judges: **Maqsood Shaheen**, IRC, US Embassy Islamabad; **Alma Rivera**, Universidad Iberoamericana Ciudad de México; and **Fatih Oguz**, University of North Carolina at Greensboro.

Papers are judged on their originality; relevance to the contest theme; and quality of argument, presentation and organization.

**Eligibility:** Only papers by a principal author who is a citizen of and resides in a developing country are eligible. Winners of the last four competitions (2009-2012) are not eligible. The papers should be

original, unpublished and submitted in English.

**Award:** Each winner will be awarded a two-year individual membership in ASIS&T. In the case of multiple authors, the principal author will be awarded the ASIS&T membership. In addition, depending on SIG/III fundraising for this competition, the first place winner will be awarded a minimum of \$1,000 toward travel, conference registration and accommodations while attending the ASIS&T Annual Meeting in Montreal, Quebec, November 1-6, 2013.

**Deadline:** Authors must submit manuscripts, not to exceed 5,000 words, by May 31, 2013, to Maqsood Shaheen at [ShaheenMA@gmail.com](mailto:ShaheenMA@gmail.com), preferably as Microsoft Word or PDF attachments.

Full contest information is available at [www.asis.org/SIG/SIGIII/](http://www.asis.org/SIG/SIGIII/).

**2013 InfoShare Awards**

Through the SIG/III InfoShare program, information professionals from developing countries, where the cost of ASIS&T membership would be a burden, are eligible to receive one year of membership. SIG/III officers vote on a roster of candidates nominated by ASIS&T colleagues or others.

The 2013 memberships are funded by monies raised at the International Reception during the 2012 ASIS&T Annual Meeting in Baltimore. InfoShare awards go to **Yared Mammo**, Ethiopia; **Teklemichael T. Wordofa**, Ethiopia; **Janakiraman**, India; **Mahmood Khosowjerdi**, Iran; **Maryam Zakerhamidi**, Iran; **Ibrahim Ramjaun**, Mauritius; **Fatima Zahra**, Pakistan. ■

In Memoriam

**Charles David Batty**

On February 11, 2013, at home in Silver Spring, Maryland, after an eight-year long battle with cancer, **Charles David Batty** – known to ASIS&T members as David – passed away in his sleep. He is survived by his wife of 38 years, Gayle Batty; children Philip (Lina), Miles and Sarah; and three grandchildren.

Memorial contributions may be made to Good Shepherd Episcopal Church, 818 University Blvd. West, Silver Spring, MD, in his name.

**Glynn Harmon**

**Glynn Harmon**, long-serving professor at the University of Texas at Austin School of Information, died quietly at his home in Austin in mid-February. He was 79 years of age.

Following his first appointment and four years teaching as assistant professor of library and information science at the University of Denver, Glynn enjoyed a 43-year career as a professor at the University of Texas. Beginning with his 1970 appointment as associate professor and continuing with promotion to professor five years later, he served as acting dean in 1990, as well as interim dean for the school from 1997 to 1999. Glynn was much loved by generations of students, and received a variety of teaching honors over his career.

Glynn's research centered on fundamental questions of the nature of information and human reasoning, with specific applications of information science to medical informatics, information economics, intelligent systems and education.

"Glynn was a pioneer," said iSchool Dean Andrew Dillon. "He envisioned a scientific discipline of information before the first iSchool was ever imagined and deserves to be recognized as a founding father of the field. He was also a true gentleman and friend who will be missed by generations of graduates and colleagues, many of whom he continued to assist, long after they had left Texas."

He is survived by his wife, Kitty. ■

# Introduction

## Altmetrics: What, Why and Where?

by Heather Piwowar, Guest Editor

### EDITOR'S SUMMARY

The ASIS&T *Bulletin* special section on altmetrics presents alternative metrics as a new and critically needed approach to measuring the impact of scholarly research. With long-established citation-based metrics unable to capture the increasing variety of online references to a scholar's work, alternative indicators offer a different view of the influence of that work. Contributed papers demonstrate how altmetrics can work on a personal level to enhance a scholar's CV and on a broad, even global level, to transform scholarly communication through its interaction with open access, digital repositories and research in emerging countries. One article suggests altmetrics should soon be included among mainstream metrics, and other contributions describe specific indicators and altmetric software considerations. The need for innovative measurement and the advantages of altmetrics in particular bode well for their wide acceptance and continuing development.

### KEYWORDS

altmetrics  
measurement  
indicators  
impact of scholarly output  
citation impact  
scholarly publishing

Heather Piwowar is a postdoc at Duke University, studying the adoption and use of open research data. She is also a co-founder of ImpactStory (<http://impactstory.org/>), an open-source web tool that helps scholars track and report the broader impacts of their research. @researchremix.

## Altmetrics: What, Why and Where?

**A**ltmetrics is a hot buzzword. What does it mean? What's behind the buzz? What are the risks and benefits of using alternative metrics of research impact – altmetrics – in our discovery and evaluation systems? How are altmetrics being used now, and where is the field going?

This special section of the *Bulletin of the Association for Information Science and Technology* focuses on these questions. Essays from seven perspectives highlight the role of altmetrics in a wide variety of settings.

The collection begins with its most general article, one I authored with my ImpactStory co-founder Jason Priem, motivating the role of altmetrics for individual scholars through “The Power of Altmetrics on a CV.”

The next few papers highlight ways that altmetrics may transform scholarly communication itself. Ross Mounce, a doctoral student and Panton Fellow of the Open Knowledge Foundation, explores the relationship between open access and altmetrics in “OA and Altmetrics: Distinct but Complementary.” Juan Pablo Alperin, doctoral student and developer with the Public Knowledge Project, encourages us to “Ask Not What Altmetrics Can Do for You, but What Altmetrics Can Do for Developing Countries.” respectively, discuss how altmetrics can empower institutional repositories in “New Opportunities for Repositories in the Age of Altmetrics.”

Completing the collection are three more perspectives from the builders of hot altmetrics tools. Jennifer Lin and Martin Fenner, both of PLOS, explore patterns in altmetrics data in “The Many Faces of Article-level Metrics.” Jean Liu, blogger, and Euan Adie, founder of Altmetric.com, consider “Five Challenges in Altmetrics: A Toolmaker's Perspective.” Finally, Mike Buschman and Andrea Michalek, founders of Plum Analytics, wrap up the collection asking, “Are Alternative Metrics Still Alternative?”



Before you dive in, if you are new to altmetrics, let me give you a quick informal introduction. For decades, the most common metric for evaluating research impact has been the number of times a research article is cited by other articles. This metric is sometimes represented by the raw count of citations received by the specific article in question or sometimes through an impact-by-association proxy – the number of citations received by the journal that published the article, summarized using a formula called the *journal impact factor*.

Citations are not the only way to represent the impact of a research article. A few alternative indicators have been the subjects of webometrics and bibliometrics research for years, including download counts and mentions in patents. However, as scholarly communication moves increasingly online, more indicators have become available: how many times an article has been bookmarked, blogged about, cited in Wikipedia and so on. These metrics can be considered altmetrics – **alternative metrics** of impact. (Appropriately enough, the term *altmetrics* was first proposed in a tweet [<https://twitter.com/asnpriem/status/25844968813>].)

We might even consider nontraditional applications of citation metrics to be altmetrics – citations to datasets as first-class research objects, for example. Other examples include citation counts filtered by type of citation, like citations by editorials or citations only from review articles or citations made only in the context of experimental replication. All of these are alternative indicators of impact.

Altmetrics offer four potential advantages:

- A more nuanced understanding of impact, showing us which scholarly products are read, discussed, saved and recommended as well as cited.
- Often more timely data, showing evidence of impact in days instead of years.
- A window on the impact of web-native scholarly products like datasets, software, blog posts, videos and more.
- Indications of impacts on diverse audiences including scholars but also practitioners, clinicians, educators and the general public.

Of course, these indicators may not be “alternative” for long. At that point, hopefully we’ll all just call them *metrics*.

Dive in, read all about it and let us know what you think. Continued conversation, background information and crowdsourced lists of new research and resources can be found on twitter using the hashtag #altmetrics (<https://twitter.com/search/realtime?q=%23altmetrics>), in the altmetrics Mendeley group ([www.mendeley.com/groups/586171/altmetrics/papers/](http://www.mendeley.com/groups/586171/altmetrics/papers/)) and probably at a conference near you.

Thanks very much to all authors in this collection for voluntarily making their articles openly available for reuse under a Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/3.0/>).

Happy reading! ■



# The Power of Altmetrics on a CV

by Heather Piwowar and Jason Priem

## EDITOR'S SUMMARY

Alternative metrics demonstrate the value and influence of scholars' work apart from traditional citation counts and can enhance the impact of a CV. Altmetrics provide additional, supplementary information and can balance misleading metrics tied to particular journals. More timely than traditional metrics, altmetrics quickly reveal the impact of recent work and add authority to different types of scholarly products not captured as articles. Altmetrics can capture social media references that escape traditional metrics and reflect public engagement prompted by scholarly writing. The availability of altmetrics expands publishing opportunities to include new venues and stimulates innovative strategies for evaluating research. When included in a CV altmetrics must be accurate, clear and meaningful.

## KEYWORDS

altmetrics  
career development  
impact of scholarly output  
social web

Heather Piwowar is a postdoc at Duke University, studying the adoption and use of open research data. She is also a co-founder of ImpactStory (<http://impactstory.org/>), an open-source web tool that helps scholars track and report the broader impacts of their research. @researchremix

Jason Priem is a PhD student and Royster Fellow, studying information science at the University of North Carolina at Chapel Hill. He is also a co-founder of ImpactStory (<http://impactstory.org/>). @jasonpriem

## Altmetrics: What, Why and Where?

**A**ltmetrics, tools measuring scholarly impact in an online environment, are displayed in a wide variety of places: journal article webpages, university press officer dashboards, data repository grant applications and many others.

In this article we focus on one particular application: including altmetrics on a scholar's CV. Ambitious scholars have been including altmetrics on their CVs for years, for example indicating that a paper was recommended by Faculty of 1000, received a "Highly Accessed" download badge on BMC or was widely discussed in the media. As tools improve, we can anticipate these early-adopters will begin to incorporate a much wider range of altmetrics on a much wider range of products.

However, if we expect these early adopters to be joined by their more cautious peers, scholars will need a clearly articulated case for value. What are the benefits that will stand the test of time and that should motivate early and ongoing action? Librarians can help in this process.

We discuss 10 benefits to scholars and scholarship when altmetrics are embedded in a CV. Altmetrics as a class of measures

1. provide additional information;
2. de-emphasize inappropriate metrics;
3. uncover the impact of just-published work;
4. legitimize all types of scholarly products;
5. recognize diverse impact flavors;
6. reward effective efforts to facilitate reuse;
7. encourage a focus on public engagement;
8. facilitate qualitative exploration;
9. empower publication choice; and
10. spur innovation in research evaluation.

There are also risks to including altmetrics in a CV when it is done without care. We give several suggestions on how one should – and shouldn't – include altmetrics in a CV. Finally, we close with a few ways that academic libraries can empower scholars to include altmetrics in their CVs, starting today.

### Provide additional information

The primary benefit of including altmetrics on a CV is the inclusion of additional information. Readers of a CV can still assess the CV items just as they have always done: based on title, journal and author list, perhaps augmented by accessing the full research product for a custom qualitative assessment. In addition to exploring this data, with altmetrics, if the readers so choose, they can dig into post-publication impact of the work.

### De-emphasize inappropriate metrics

Evaluating an article based on its journal title or journal impact factor is generally regarded as poor form: high journal impact factors vary across fields, an article often receives more or less attention than its journal container suggests, the authors may have selected a “low ranking” journal for the speed of its peer review or its open access status rather than its journal rank, and so on. For further details see [www.zotero.org/groups/impact\\_factor\\_problems/items](http://www.zotero.org/groups/impact_factor_problems/items).

Yet what else are readers of a CV to do? Most of us don't have sufficient domain expertise to dig into each item and assess its merits based on a careful reading, even if we did have time. We need help, but traditional CVs don't provide enough information to assess the work on anything *but* journal title.

Providing article-level citations and altmetrics in a CV gives readers more information, thereby de-emphasizing evaluation based on journal rank.

### Uncover the impact of just-published work

Why not suggest that we include citation counts in CVs, and leave it at that? Why go so far as altmetrics? The reason is that altmetrics have benefits that complement the weaknesses of a citation-based solution, as we'll cover in the next few points.

One of the most obvious benefits of altmetrics is timeliness. Citations take years to accrue. This delay is a big problem for graduate students who

are applying for jobs soon after publishing their first papers and for those promotion candidates whose most profound work is published only shortly before review.

Multiple research studies have found that counts of downloads, bookmarks and tweets correlate with citations, yet accrue much more quickly, often in weeks or months rather than years. Using timely metrics allows researchers to showcase the impact of their most recent work.

### Legitimize all types of scholarly products

Funders and institutions are beginning to explicitly welcome inclusion of datasets, software and other scholarly products in biosketches and CVs. This greater flexibility is great news for recognizing all worthwhile forms of research output, but how can readers of a CV know if the included dataset or software project is any good? What is the size and type of its contribution? Should they be impressed?

We often assess the quality and impact of a traditional research article based on the reputation of the journal that published it. This approach isn't possible with alternative products. Data and software can't be evaluated with a journal impact factor or journal ranking since repositories seldom select entries based on anticipated impact; they don't have an impact factor; and, even if such a metric were possible, we surely don't want to propagate the poor practice of judging the impact of an item by the impact of its container.

How, then, can alternative scholarly products be more than just space-filler on a CV, something that an evaluator counts but can't appreciate?

Product-level metrics (like article-level metrics, but for more than just articles) provide the needed evidence to convince evaluators that a product has made a difference. Furthermore, because alternative products often make impacts in ways that aren't fully captured by established attribution mechanisms, altmetrics will be key in communicating the full picture of how research products have influenced conversation, thought and behavior.

### Recognize diverse impact flavors

The impact of a research paper has a flavor. Let's imagine it as an ice cream flavor. The impact flavor might be champagne: a titillating discussion

piece of the week. Or maybe it is a dark chocolate mainstay of the field. Strawberry: a great methods contribution. Tiger-stripe black licorice: controversial. Bubblegum: a hit in the classrooms. Low-fat vanilla: not very creamy, but it fills a need.

There probably aren't 31 clear flavors of research impact. How many are there? Maybe five or seven or 12? We don't know. But it would be a safe bet that, just like ice cream, our scholarship and society need them all. It depends whether we have a cone or a piece of apple pie. The goal isn't to compare flavors: one flavor isn't objectively better than another. They each have to be appreciated on their own merits for the needs they meet.

To appreciate the impact flavor of items on a CV we need to be able to tell the flavors apart. Imagine that for ice cream all you had to go by was a sweetness metric. Not happening, right? So too, citations alone can't fully inform what kind of difference a research paper has made on the world. Important, but not enough.

We need more dimensions to distinguish the flavor clusters from each other. This is where altmetrics come in. By analyzing patterns in what people are reading, bookmarking, sharing, discussing and citing online we can start to figure out what kind – what flavor – of impact a research output is making.

It is worth noting that flavors are important for research products other than just papers. For example, some publicly available research datasets are used all the time in education but rarely in research; others are used once or twice by really impactful projects; others across a field for calibration; and so on. Understanding and recognizing these usage scenarios will be key in recognizing and rewarding the contributions of dataset creators.

More research is needed to understand the flavor palette, how to classify impact flavor and what it means (<http://researchremix.wordpress.com/2012/01/31/31-flavours/>). In the meantime, exposing raw information about downloads, shares, bookmarks and the like starts to give a peak into impact flavor beyond just citations.

### Reward efforts to facilitate reuse

Reusing research – for replication, follow-up studies and entirely new purposes – reduces waste and spurs innovation. Unfortunately, our common

method of disseminating research through subscription-based static flat pdf articles makes research difficult to reuse.

If they choose, authors can make their research easier to reuse. For example, authors can make article text available for free or available for free with broad reuse rights (open access). They can choose to publish in places with liberal text-mining policies and investment in disseminating machine-friendly versions of articles and figures.

Authors can write detailed descriptions of their methods, materials, datasets and software. They can make their associated datasets and software openly available for reuse. Authors can go further, experimenting with executable papers, versioned papers, open peer review, semantic markup and so on.

When these additional steps are useful – when they do indeed result in additional reuse – the increased use will likely be reflected in downloads, bookmarks, discussions and possibly citations. Including altmetrics in CVs will reward investigators who have made these investments and encourage others to do so in the future.

### Encourage a focus on public engagement

The research community, as well as society as a whole, benefits when research results are discussed outside the Ivory Tower: engaging the public is essential for future funding, recruitment and accountability. Today, however, researchers have little incentive to spend time engaging in public outreach or making their research accessible to the public. By highlighting evidence of public engagement like tweets, blog posts and mainstream media coverage, including altmetrics in a traditional CV will reward researchers who invest in public engagement activities.

### Facilitate qualitative exploration

Including altmetrics in a CV isn't all about the numbers! Just as we hope many people who skim our CVs will stop to read our papers and explore our software packages, so too we can hope that interested parties will click through to explore the composition and details of altmetrics engagement for themselves.

Who is discussing an article? What are they saying? Who has bookmarked

a dataset? What are they using it for? As we discuss at the end of this article, including provenance information is crucial for trustworthy altmetrics; it also provides great information to move beyond the numbers and jump into qualitative exploration of impact.

### Empower publication choice

Publishing in a new or innovative journal is risky: many authors are hesitant to publish their best work somewhere unusual, somewhere without a sky-high impact factor. Altmetrics will help to change this situation by highlighting work based on its post-publication impact rather than its journal title. Authors will be empowered to choose publication venues they feel are most appropriate, leveling the playing field for what might otherwise be considered risky choices.

Successful publishing innovators will also benefit. New journals won't have to wait two years to get an impact factor before they can compete. Publishing venues that increase access and reuse will be particularly attractive. This change will spur new innovation and support the many new publishing options that have recently debuted, such as *eLife*, *PeerJ*, *F1000 Research*, *Digital Humanities Quarterly* and others.

### Spur innovation in research evaluation

Finally, including altmetrics on CVs will engage researchers directly in research evaluation. Researchers are evaluated all the time, but often behind closed doors, using data and tools they don't have access to (and frankly wouldn't want to take the time to learn). Encouraging researchers to tell their own impact stories on their CVs, using broad sources of data, will help spur a much-needed conversation about how research evaluation is done and should be done in the future.

There are also risks to including altmetrics data on a CV, particularly if the data is presented or interpreted without due care or common sense.

Altmetrics data should be presented in a way that is accurate, auditable and meaningful. Accurate data is up-to-date, well-described and has been filtered to remove attempts at deceitful gaming. Auditable data implies completely open and transparent calculation formulas for aggregation,

navigable links to original sources and access by anyone without a subscription. Meaningful data needs context and reference. Categorizing online activity into an engagement framework (<http://blog.impactstory.org/2012/09/14/31524247207/>) helps readers understand the metrics without becoming overwhelmed. Reference is also crucial. How many tweets is a lot? What percentage of papers are cited in Wikipedia? Representing raw counts as statistically rigorous percentiles, ideally localized to domain or type of product, makes it easy to interpret the data responsibly.

Assuming these presentation requirements are met, how should the data be interpreted? We strongly recommend that altmetrics be considered not as a replacement for careful expert evaluation but as a supplement. Particularly in these early days, we should view altmetrics as way to ground subjective assessment in real data – to start conversations, not end them.

Given this approach, at least three varieties of interpretation are appropriate: signaling, highlighting and discovery. A CV with altmetrics clearly signals that a scholar is abreast of innovations in scholarly communication and serious about supporting and communicating the impact of scholarship in meaningful ways. Altmetrics can also be used to highlight research products that might otherwise go unnoticed: a recent paper with a lot of tweets or a highly downloaded dataset or a track record of F1000-reviewed papers suggests work worthy of a second look. Finally, as we described in the exploration section above, auditable altmetrics data can be used by evaluators as a jumping off point for discovery about who is interested in the research, what they are doing with it and how they are using it.

### How to Get Started

How can you add altmetrics to your own CVs or, if you are a librarian, also empower scholars to add altmetrics to theirs? Definitely start by experimenting with altmetrics for yourself. Play with the tools, explore and suggest improvements. Librarians can also spread the word on their campuses and beyond through talking, writing, teaching and outreach. Explicitly welcome evidence of impact when you solicit CVs for new positions, awards and grants. Last but not least, try it out for yourself! ■

# Open Access and Altmetrics: Distinct but Complementary

by Ross Mounce

## EDITOR'S SUMMARY

Open access to publications through electronic journals has dramatically expanded downloading and use of this literature and spurred the rise of alternative metrics to assess article impact. Open access publications have been shown to gain more citations than articles with restricted access, and seven of the 10 most popular articles in 2012 were free public access with the most response, as documented through altmetrics, coming from non-scientists. Altmetrics also enables post-publication filtering and peer review in a nearly immediate timeframe for very large open access journals. Online activity measured through altmetrics highlights attention to the article on its own, less dependent on the validation of a journal name. The field of altmetrics is young, still limited to certain open websites, but with potential for considerable expansion, development and application. Open access and altmetrics can be expected to grow in a complementary and mutually supportive manner.

## KEYWORDS

altmetrics  
open access publications  
social web  
impact of scholarly output  
electronic publishing  
electronic journals

Ross Mounce is currently a Ph.D. candidate at the University of Bath, where he is studying the impact of fossils in phylogeny using informatics-based approaches. He is also a Panton Fellow of the Open Knowledge Foundation, working to develop and promote a culture of open science and open data sharing to enable and encourage re-use. He can be reached at [ross.mounce@okfn.org](mailto:ross.mounce@okfn.org).

## Altmetrics: What, Why and Where?

**E**lectronic publishing with dissemination via the Internet has hugely changed the landscape of academic publishing in the 21st century. Now, few journals are print-only. Many are available in print and online, while perhaps a slight majority in science, technology and medicine (STM) are published online only. This distribution is a reflection of the new reality that, for most disciplines, electronic journals have become the preferred method for discovering and accessing journal literature. Publishing content electronically, with dissemination online, is obviously less costly than publishing and disseminating print copies around the world, so there is also certainly an economic incentive for this trend, not just a social and functional preference. Alongside this growth and preference for online journals, there has been a notable rise in the growth and popularity of a particular type of online journal – open access (OA) journals, which expressly allow anyone on the Internet to read them for free without paying. Such journals make it even easier for people to discover, access and re-use journal literature.

With this change in the consumption pattern of journal content to online, new ideas such as altmetrics have arisen to help us better assess the influence and impact of online journal articles. This article considers the complementary relationship between OA journal publishing and altmetrics, scholarly impact measures derived from online activity, as a means of capturing and measuring some of the influence of online journal articles.

## Open Access

Open access was first formally defined as follows by the Budapest Open Access Initiative (BOAI), as published on February 14, 2002, in a version that anyone could endorse with a signature:

“...free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.” [1]

A brief background:

- Not all journals that label themselves as “open access” strictly adhere to or fully comply with this definition but they are all at least free to read. The *Directory of Open Access Journals (DOAJ)* lists over 8000 such publications.
- Over 300 research funders and institutions now have some form of OA mandate. Notable among these are the National Institutes of Health and the Howard Hughes Medical Institute (United States), Research Councils UK (RCUK), the Medical Research Council (UK) and the Wellcome Trust (UK), the Australian Research Council and the National Health and Research Council (Australia).
- Some estimates put the percentage of articles published as OA at greater than 16% per year [2] and significantly more if one includes self-archived manuscripts in this count.
- Many commentators, both publishers and researchers alike, have stated that it is inevitable that in the future the vast majority of research will be published as OA.

A major driver behind the OA movement is the observation that the research behind the majority of academic publications is public-sector funded (by taxpayers). The logical rationale therefore is that if the public sector funded the research, then the public sector should have the right of OA to publications arising from this research. Thus OA represents a significant improvement in access to research for scholars and non-scholars alike. With traditional subscription-access journals very few people in the world have easy access to article content, and those with this privileged, paid-for, subscription access are likely to be highly educated people affiliated with higher education institutions. By contrast OA allows access to articles by anyone on the

Internet, regardless of affiliation, education, wealth, age, gender or ethnicity. Well-controlled studies have shown that this heightened online accessibility is significantly associated with doubling the number of full-text downloads of research articles [3]. Open access articles are therefore particularly interesting to measure with altmetrics.

### Complementarity Between Altmetrics and Open Access

Altmetrics help both expand and broaden our view of the impact of academic research outputs. One can track the impact of code and data with altmetrics, not just publications, but for this article I will focus just on publications. In the new reality of online availability of research more and more people are trying to access it. JSTOR, for instance, registers 150 million failed attempts every year to gain access to articles they keep behind the paywall [4]. Articles made available via such traditional pay-to-read business models may not achieve the impact they could have simply because all potential readers may have neither institutionally provided access to the resource nor the money to buy access to it themselves. Many papers have found that OA has a citation advantage relative to subscription access articles. This effect may also be true in terms of altmetrics. For example, of the 10 most popular articles in 2012 as measured with altmetrics by Altmetric.com (Table 1), 7 out of 10 were freely accessible articles [5]. Even

**TABLE 1.** The 10 most popular papers in 2012, as measured by Altmetric.com altmetrics

1. *The biological impacts of the Fukushima nuclear accident on the pale grass blue butterfly* (77% of tweets sent by members of the public)
2. *Association of coffee drinking with total and cause-specific mortality* (64% of tweets sent by members of the public)
3. *Rape-related pregnancy: Estimates and descriptive characteristics from a national sample of women* (82% of tweets sent by members of the public)
4. *Food for thought. What you eat depends on your sex and eating companions* (98% of tweets sent by members of the public)
5. *Bright minds and dark attitudes: Lower cognitive ability predicts greater prejudice through right-wing ideology and low inter-group contact* (79% of tweets sent by members of the public)
6. *Unilateral dermatoheliosis* (79% of tweets sent by members of the public)
7. *Higher social class predicts increased unethical behavior* (74% of tweets sent by members of the public)
8. *Science faculty's subtle gender biases favor male students* (59% of tweets sent by members of the public)
9. *Measuring the evolution of contemporary western popular music* (83% of tweets sent by members of the public)
10. *Classic Nintendo games are (NP-)hard* (78% of tweets sent by members of the public)

more remarkably, none of these 10 articles were from either of the two most widely read academic journals, *Nature* and *Science*, which both predominately publish articles behind a paywall. All of the top 10 articles clearly captured the public imagination and engagement, with the majority of activity on Twitter coming from accounts that were not identifiably scientists, science communicators or practitioners. Many of these papers may show rather unremarkable citation counts – a more traditional measure of academic impact. Their significant public impact is only revealed in a standardized way by altmetrics – services like Altmetric and ImpactStory [6] even attempt to normalize altmetrics to provide even greater context and meaning to the numbers, as well as providing open data to ensure the numbers are independently verifiable.

Altmetrics may also be of particular use for demonstrating the impact of articles published in OA megajournals. These megajournals (for example, *PLOS ONE*, *PeerJ*, *SAGE Open*, and *Scientific Reports*) do not reject articles on the basis of the perceived impact that they may have and accept article submissions as long as they are well-reported and technically sound contributions to the academic literature. Thus article-level altmetrics may be particularly key to these megajournals as a means of post-publication filtering and peer review to differentiate among the many thousands of articles that pass through them. The immediacy of altmetrics relative to more traditional measures, such as citations, also helps this filtering process. While citations take many years to accrue, tweets, facebook shares, blog posts and reference management bookmarks tend to occur much more quickly after publication. If we trace and read online conversations across the social networks about research articles, the conversations can in some cases indicate whether other researchers think the paper is particularly good or bad. Indeed, it cannot be stressed highly enough that altmetrics are about *more than just the numbers*: the greater context and content of web activity is also hugely meaningful. For many, publishing a paper in an OA journal is a truer test of their personal brand and the quality of their work than relying on the prestige and high journal impact factor of a traditional subscription-access journal. A work published in an OA journal is more clearly regarded on its own merits, not that of the journal it appears in. In this way, altmetrics

can help good OA articles shine and get the attention and respect they deserve by accurately and verifiably capturing the online activity around them.

### Narcissism or Optimization of Reach and Impact?

Some have criticized altmetrics as being a “technology of narcissism” and “gameable” [7]. Yet from a personal perspective I think their depth and variety of information is highly useful for self-assessment and improvement, while the transparency is a disincentive to attempts to game it. Knowing the geographic reach of one's work, the social networks it is shared on, the people that share it, bookmark it, tweet and favor it is useful and interesting. It is not just the positive data, either. Absence of impact in certain flavors of altmetrics could indicate room for improvement and spaces in which a researcher could profit from being more visible and active. One could even envisage a situation in which an altmetric service provides near real-time alerts about one's research, which would allow the researcher, if interested, to join in and interact with the social web activity going on around the paper. I think this interaction would be a good mechanism to increase the dynamism around research publications. Even with hundreds of thousands of articles currently allowing for public comment at the end, vanishingly few create productive conversations post-publication.

### A Bright and Intertwined Future for OA and Altmetrics

While OA is relatively established, altmetrics are still fairly new, relatively unexplored and underdeveloped. A great many published studies will certainly evaluate the usage of altmetrics for various purposes, stakeholders, subjects and disciplines. It is important to keep in mind that we can only use altmetrics in certain social spheres on the web that support them through their openness and standardized APIs. As more of the web opens up access to its data and links we will see both a wider variety of altmetrics and perhaps a greater richness and depth to these measures in future. The standardization and openness fits well with the OA publishing model and less well with subscription-access publishing. Research evaluation has for a long time utterly depended upon commercially provided citation databases such as Elsevier's *Scopus* or Thomson Reuters'



*Web of Knowledge*, but a host of new organizations, both for-profit and not-for-profit, have recently sprung up to market altmetrics-based information services. This development can only be a good thing to increase competition and diversity. Not all web activity around articles can be captured by

altmetrics – there will always be “dark social” [8] sharing and paywall-protected research, but this model of transparency for both publishing and assessing research will undoubtedly have an interesting and intertwined future ahead of it. ■

### Resources Mentioned in the Article

- [1] Chan, L., Cuplinskas, D., Eisen, M., Friend, F., Genova, Y., Guedon, J-C., Hagemann, M., et al. (February 14, 2002) *Budapest Open Access Initiative*. Retrieved February 24, 2013, from [www.opensocietyfoundations.org/openaccess/read](http://www.opensocietyfoundations.org/openaccess/read).
- [2] Laakso, M., & Bjork, B. C. (October 22, 2012) Anatomy of open access publishing: A study of longitudinal development and internal structure. *BMC Medicine*, 10:124. Retrieved February 24, 2013, from <http://dx.doi.org/10.1186/1741-7015-10-124>.
- [3] Davis, P. M., Lewenstein, B. V., Simon, D. H., Booth, J. G., & Connolly, M. J. L. (2008). Open access publishing, article downloads, and citations: Randomised controlled trial. *BMJ* 337:a568. Retrieved February 24, 2013, from <http://dx.doi.org/10.1136/bmj.a568>.
- [4] Howard, J. (January 13, 2012) JSTOR tests free, read-only access to some articles [blog post]. *The Chronicle of Higher Education: Wired Campus*. Retrieved February 25, 2013, from <http://chronicle.com/blogs/wiredcampus/jstor-tests-free-read-only-access-to-some-articles/34908>.
- [5] Van Noorden, R. (December 21, 2012). What were the top papers of 2012 on social media [blog post]? *Nature News*. Retrieved February 25, 2013, from <http://blogs.nature.com/news/2012/12/what-were-the-top-papers-of-2012-on-social-media.html>.
- [6] *ImpactStory*. <http://impactstory.org/>.
- [7] Wouters, P., & Costas, R. (February 2012) *Users, narcissism and control – tracking the impact of scholarly publications in the 21st century*. Utrecht: SURFfoundation. Retrieved February 25, 2013, from [www.surf.nl/en/publicaties/Pages/Users\\_narcissism\\_control.aspx](http://www.surf.nl/en/publicaties/Pages/Users_narcissism_control.aspx).
- [8] Madrigal, A.C. (October 12, 2012) Dark social: We have the whole history of the web wrong. *The Atlantic*. Retrieved February 25, 2013, from [www.theatlantic.com/technology/archive/2012/10/dark-social-we-have-the-whole-history-of-the-web-wrong/263523/](http://www.theatlantic.com/technology/archive/2012/10/dark-social-we-have-the-whole-history-of-the-web-wrong/263523/).

# Ask Not What Altmetrics Can Do for You, But What Altmetrics Can Do for Developing Countries

by Juan Pablo Alperin

## Altmetrics: What, Why and Where?

### EDITOR'S SUMMARY

Traditional citation counting for evaluating scholarly impact unfairly benefits those in North America and Europe and shortchanges the alternative scholars of the developing world. Alternative metrics more accurately measure the impact of scholarly writings, better serve all scholars and can foster a research culture that supports national development goals. The current system favors dominant journals and topics of interest to the prevailing scientific community, captured by the leading bibliographic databases. Yet publishing on platforms more open to underrepresented journals and scholars in developing nations would promote a greater range of ideas and scholarly exchange. With facilitating international development in mind, scholarly communication should encourage research on topics of local and national relevance and be presented through globally accessible channels, disseminated by social media. Publishing technology barriers to participation must be lowered. The value of altmetrics will be evident, providing advantages to alternative scholars, serving public needs and revealing scientific contributions long underrepresented in the standard literature.

### KEYWORDS

altmetrics	cultural values
impact of scholarly output	scholarly publishing
developing countries	citation indexes
international aspects	social web

Juan Pablo Alperin is a doctoral student in the Stanford School of Education and a researcher and systems developer with the Public Knowledge Project. He can be reached at [juan<at>alperin.ca](mailto:juan@alperin.ca).

The future of scholarly communication needs to be inclusive of diverse local contexts of the developing world and tailored to address national development goals, not simply address the needs of scholars working from privileged positions in the global North, including North America and Europe. However, a recent and unprecedented number of visionary minds have been attempting to reshape scholarly communications through the introduction of new journals, new technologies, new services and, most recently, new metrics, specifically *altmetrics*, which focus on measuring scholarly impact in the online environment and away from citation counting. These efforts are borne from a desire to break away from the limitations of the current system of communicating and carrying out research. Yet, most scholars, who have a tremendous effect on global scholarly communication practices, have simply ignored the fact that the legacies of the current system are the most pervasive, or most abhorrent, in the developing world. In this essay, I argue for the potential for altmetrics to serve what I am calling “alternative scholars” and, by doing so, challenge the current global order of scholarly communications that heavily favors the North.

The legacies of the current system work against developing countries in countless ways, but the most emblematic example is the role of international bibliographic databases. These repositories purport to be objective but, in reality, provide journal rankings that highly handicap underrepresented journals from developing nations. Unfortunately, these rankings continue to be used by universities, funding agencies and governments (including those from developing countries themselves) to determine the incentives that guide authors’ decisions as to where to publish. These incentives lead to an outpouring of the best scholarship from developing countries to journals of the North. This, in turn, has two perverse effects. First, it encourages authors

in developing countries to research and publish on topics of interest to the North American and European scientific communities. Second, it discourages dialogue among scholars in developing countries. These effects result in diminished local capacity – or research that is less likely to speak to local needs – and a less cohesive, less national collaborative network of scientists.

Yet, despite such examples, the future of scholarly communication around the world may look markedly different from the closed and self-serving system of the past. As economic, political and technological forces reshape the scholarly communication landscape, it is now possible to imagine forms of scholarly exchange that can help foster a platform for those who have traditionally been excluded from it. The alternative metrics movement – altmetrics – is one such force.

There are many reasons to be excited about the increasing interest and development of altmetrics, many of which are outlined in this special issue of the *Bulletin* and in the growing body of research around altmetrics. For example, altmetrics is lauded by those who believe the current models and tools for scholarly communication are stifling innovation and hindering discovery [1] [2]. While I agree and share in this excitement, I will argue in this essay that the future of scholarly communication should be about more than bringing better knowledge discovery, increased productivity and new ways of measuring impact to the scientific elite. Instead, I argue that scholarly communication needs to serve all scholars, including those working from the scientific periphery. Fortunately, the two goals need not be incompatible.

### The Alternative Scholar

Altmetrics supporters can help us move beyond the journal article as the only source of scholarly output that is valued and beyond citations as the only way of validating what we read. That is, the “alt-“ part simultaneously means alternative types of research products [3] and alternative types of metrics (that is, alternatives to citation counting) [4]. However, I am proposing yet another meaning of “alt-“: an alternative scholar. By alternative scholar, I am suggesting that altmetrics can serve those at the margins of the current system: those working with scarce resources or in areas of scholarship that

are not in vogue or without the latest in publishing technologies and, most importantly, those for whom research and publishing are carried out with different objectives than in the North. While alternative scholars can be found all over the world, they are disproportionately found in the developing world. And while the remainder of this essay focuses there, much is applicable to alternative scholars everywhere. By addressing the needs of alternative scholars, altmetrics can have a very different (alternative!) type of impact. Most importantly, altmetrics can foster a research culture that supports national development goals.

To understand what I mean, we must first look more broadly at the relationship between research and national development. Many view the link between research and development primarily in economic terms: the argument is premised on the idea that promoting research leads to knowledge creation, innovation, patents, links to industry and so forth – all of which boost the economy with increased efficiencies or by capturing the revenue of new products. While true in many respects, this mechanism is predicated on a mature and consolidated research system. It does not explain how changes in scholarly communication can support this consolidation. Furthermore, proponents of this view of development undervalue research that does not have direct economic benefit, regardless of its potential relevance to the society it is derived from. What, then, does a system of scholarly communications that supports national development goals look like? And what is the role of altmetrics in the requisite transformation?

First, scholarly communications for development must facilitate and encourage the formation of research communities at the local, national and regional levels. Altmetrics help cultivate such networks by bringing an otherwise invisible community of researchers onto the global stage with the use of social media. While an article may not receive many citations, data on who is downloading, bookmarking, tweeting or Facebook “liking” an article can allow academics to see who is utilizing their research and provide the opportunity to engage with their community of users. Even if researchers focused solely on increasing their altmetric scores, the effect, regardless of the motivation, is the same in the end: attention through social media use promotes ongoing conversations about otherwise unseen research.

Second, scholarly communication for development must encourage research that is socially relevant and that is inclusive of local contexts. And altmetrics can play an important role in this regard. As elsewhere, altmetrics may change the way that research is evaluated by promotion and tenure committees, by research funders and by governments in the developing world. The successful adoption of altmetrics in researcher evaluations would allow scholars to count (and thereby measure and be rewarded for) non-scholarly use of their research (that is, public use), as well as for the use of non-traditional research products. Such a shift would foster a system that is attuned to the public interest and that addresses local, public needs. By contrast the current system of citation measures rewards researchers for producing research articles and publishing them in databases where developing country journals are underrepresented. The result is an incentive system that rewards researching problems of North American and European interest. Altmetrics may change this focus and offer the opportunity to direct funding to researchers, journals, projects or institutions that have proven themselves relevant to interests of developing nations.

Finally, scholarly communication for development must allow scholars from the developing world to assert their worldview on a global stage. Altmetrics can provide the necessary measures for garnering attention from an audience that might otherwise consider the work irrelevant or of low quality. It can also facilitate discovery through social media channels that are more inclusive and democratic than publishers and citation databases. Metrics and tools that can do both of these things have the potential to reverse decades of marginalization in the current system. However, the reversal is neither guaranteed nor automatic. Social media usage is more prominent in the North than in the South, which could lead to skewed results if care is not taken at the time of displaying and interpreting the metrics. While the implications of this imbalance for altmetrics are unclear, no similar metrics have existed until now. And any set of metrics that gives scholars in the developing world the possibility of being on equal footing with scholars in other parts of the world have promise.

Many of altmetrics' benefits can be realized through the standard altmetrics tools available today. However, there are certain limitations that

must still be addressed for altmetrics to serve the alternative scholar. First and foremost, the available metrics must be relevant for the developing country scholar communities. The sources of metrics queried by altmetric tools must, for example, include non-English databases. Mentions in Wikipedia must query all available languages, not just English. Citations must come from regional sources such as *SciELO*, not just the standard *Web of Science* and *Scopus*. Mentions in blogs cannot be harvested only from blog collections that are predominantly from the North. What sources are included in altmetrics data aggregators is not a neutral choice. By tailoring altmetrics to the scientific community of the North, altmetrics would inadvertently replicate the exclusion of developing countries that has plagued the traditional measures of impact – that is, a system biased to give higher scores to journals of the North.

Furthermore, in order to be relevant in the developing world, altmetric tools must also be compatible with the less sophisticated publishing technologies employed by small publishers. The current over-reliance on Crossref's digital object identifier (DOI) by altmetric providers is problematic. DOIs provide a permanent identifier to articles so that if the URL of an article changes over time, or if there are multiple copies of an article, they can all be uniquely and permanently identified. While the service that Crossref provides is invaluable, DOIs are also disproportionately allocated to journals from the North (and even there, they are disproportionately used by STEM [science, technology, engineering and mathematics] journals over those from the social sciences and humanities). While Crossref has made concerted efforts to lower the barriers for participation, the financial, administrative, technological and language barriers continue to hamper their adoption in the rest of the world. Yet, there are no technical reasons why altmetrics need to rely on the presence of DOIs. DOIs simply provide a convenience that allows altmetrics to be implemented more easily, at the expense of the countless researchers who publish in journals that have not yet adopted the system.

I do not purport to be qualified to provide a complete list of implementation shortfalls. This task is one that I put to altmetric providers and that I ask everyone working to improve scholarly communication to consider. In this essay, I only intend to broaden the possibilities and chart

new directions as the altmetrics movement moves forward. These possibilities are far greater than increased efficiency and broader dissemination, but they can only be brought about if we consider the needs of all scholars, including the alternative scholar.

Other parts of the developing world, Latin America being the most notable example, are already experimenting in a big way with different models of research communication. There is a realization that scholarly communication can serve the public interest and that the modes of scholarly communication from the North are not the appropriate models to follow. The wide adoption of open access (nearly 100% of all journals based in Latin America) is indicative of the region's desire to make the shift for itself. We are seeing an adoption of open source tools and a general amity towards

openness in the developing world, just as the scholarly communication networks and channels in these regions consolidate. All these developments suggest that altmetrics would likely be well received and quickly adopted.

Once adopted, the impact of altmetrics will be more meaningful and far-reaching than in the North. To start, bringing altmetrics to work for the alternative scholar would be a step towards improving systems of higher education, towards steering research towards public needs and towards helping a large swath of the world population have a voice in a global order of science that currently undervalues their contributions. After that, the possibilities are endless. By bringing altmetrics to the developing world, and to alternative scholars everywhere, we can begin to generate a change in scholarly communication that reaches far beyond the walls of academia. ■

### Resources Mentioned in the Article

- [1] Bourne, P.E., Clark, T., Dale, R., Waard, A. De, Herman, I., Hovy, E. H., et al. (Eds.) (February 19, 2012). *Force 11 Whitepaper: Improving the Future of Research Communications and e-Scholarship*. Retrieved March 4, 2012 from [www.force11.org/sites/default/files/book\\_attachments/Force11Manifesto20120219.pdf](http://www.force11.org/sites/default/files/book_attachments/Force11Manifesto20120219.pdf).
- [2] Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). *Altmetrics: A manifesto*. Retrieved March 4, 2013, from <http://altmetrics.org/manifesto/>.
- [3] Piwowar, H. A. (January 10, 2013). Value all research products. *Nature*, 493, 159. Retrieved March 4, 2013, from [www.nature.com/nature/journal/v493/n7431/full/493159a.html](http://www.nature.com/nature/journal/v493/n7431/full/493159a.html).
- [4] Priem, J., Piwowar, H. A., & Hemminger, B. H. (n.d.). *Altmetrics in the wild: An exploratory study of impact metrics based on social media*. Presentation at Metrics 2011: Symposium on Informetric and Scientometric Research. New Orleans, Louisiana. Retrieved March 4, 2013, from <http://jasonpriem.org/self-archived/PLoS-altmetrics-sigmetrics11-abstract.pdf>.

# New Opportunities for Repositories in the Age of Altmetrics

by Stacy Konkiel and Dave Scherer

## EDITOR'S SUMMARY

For institutional repositories, alternative metrics reflecting online activity present valuable indicators of interest in their holdings that can supplement traditional usage statistics. A variable mix of built-in metrics is available through popular repository platforms: Digital Commons, DSpace and EPrints. These may include download counts at the collection and/or item level, search terms, total and unique visitors, page views and social media and bookmarking metrics; additional data may be available with special plug-ins. Data provide different types of information valuable for repository managers, university administrators and authors. They can reflect both scholarly and popular impact, show readership, reflect an institution's output, justify tenure and promotion and indicate direction for collection management. Practical considerations for implementing altmetrics include service costs, technical support, platform integration and user interest. Altmetrics should not be used for author ranking or comparison, and altmetrics sources should be regularly reevaluated for relevance.

## KEYWORDS

altmetrics  
digital repositories  
impact of scholarly output  
statistics  
collection management  
social web

Stacy Konkiel is an eScience librarian at Indiana University. She can be reached at [skonkiel@indiana.edu](mailto:skonkiel@indiana.edu).

Dave Scherer is a scholarly repository specialist at the Purdue ePubs Repository. He can be reached at [dscherer@purdue.edu](mailto:dscherer@purdue.edu).

## Altmetrics: What, Why and Where?

University administrators are increasingly trying to find new ways to measure the impact of the scholarly output of their faculty, students and researchers through quantitative means. By reporting altmetrics (alternative metrics based on online activity) for their content, institutional repositories can add value to existing metrics – and prove their relevance and importance in an age of growing cutbacks to library services. This article will discuss the metrics that repositories currently deliver and how altmetrics can supplement existing usage statistics to provide a broader interpretation of research-output impact for the benefit of authors, library-based publishers and repository managers, and university administrators alike.

### Metrics Repositories Currently Deliver

Many repository platforms measure usage statistics such as download counts and page views. Less often, repositories report citation counts and altmetrics culled from the social web for their holdings. Here, we will look at usage statistics that are commonly reported on the three most popular repository platforms in use today: Digital Commons, DSpace and EPrints.

**Digital Commons.** Digital Commons is a proprietary institutional repository and journal-publishing platform run by Bepress. Relying on proprietary, COUNTER-compliant download counts [1] and Google Analytics as a source for metrics on access, the platform records download counts, search terms and referral links for all content held in each repository. These metrics are communicated to repository managers, series administrators and authors via email. The platform provides metrics on publications available to date in each repository, downloads to date, and downloads during the lifetime of the repository. Authors also receive statistics on their deposits through a private Author Dashboard interface.

The platform also operates a federated search and discovery mechanism,

the Digital Commons Discipline Browser, that provides repository managers, authors and users with usage metrics within the network of three-tiered, taxonomy-based disciplines contained in Digital Commons repositories. Google Analytics results provide metrics on access, including number of visitors, number of unique visitors, pageviews and bounce rate. Currently, there is no option to display these metrics to repository visitors or journal readers.

Bepress recently partnered with Altmetric.com ([www.altmetric.com](http://www.altmetric.com)) to display an Altmetric “badge” for items in selected series and collections within Digital Commons and Bepress journals. Altmetric.com can display article level metrics related to social bookmarking (Mendeley, Citeulike, Connotea) and social media (Facebook, Twitter, Google+, Reddit and LinkedIn).

**DSpace.** DSpace is an open-source, configurable platform that delivers only download counts to researchers as part of its base install (Versions 1.6+). Metrics can either be displayed openly (where enabled by repository managers) or to administrators only. Repository, community, collection and item-level download counts are displayed via an HTML table. Citation metrics are available if a DSpace plug-in is installed (Versions 1.6+), where the institution has subscription access to the SciVerse Scopus API ([www.developers.elsevier.com/devcms/cited-by-count-api](http://www.developers.elsevier.com/devcms/cited-by-count-api)).

Repositories running DSpace with the help of the BMC-backed Open Repositories service can display altmetrics at the item level. These repositories openly report metrics related to social bookmarking (Del.icio.us, Citeulike, Connotea) and social media (Facebook, Stumbleupon, Digg and LinkedIn) in addition to download counts.

**EPrints.** Similarly, the open source repository platform EPrints tracks only downloads as an out-of-the-box feature. However, they have fairly robust reporting tools available: line, bar and pie graphs; HTML tables and CSV exports. Download counts can be measured at the repository, collection and item levels. Statistics can be hidden from public view, accessible only to repository administrators. Citation metrics are available as a repository plug-in (Versions 3.2+) (<http://files.eprints.org/641/>); like the DSpace plug-in, subscription access to the SciVerse Scopus API is required. The few EPrints repositories to offer altmetrics have implemented them through “homegrown” means.

## How Metrics Can Be Used

Usage and citation statistics can reveal many things to both authors and repository administrators, including the demographics of those accessing their scholarly outputs and what types of content are most popular. Authors can use these numbers to gain basic insight into the reach of their scholarship and can supplement their tenure and promotion dossiers with numbers more granular (and some say more meaningful) than journal impact factors. Repository administrators can use usage statistics to help promote similar content within their institutional repositories (IR), supplement their collection development policies and provide evidence to university administration as to the impact of their university’s intellectual output [2]. Using altmetrics, some repositories have been successfully able to showcase the social importance of repository content to the general public in non-academic settings [3].

What usage statistics do not always reveal is the nature of use or the context for how scholarship is consumed [2]. Altmetrics can help to fill in some of the knowledge gaps that usage statistics alone cannot address. In the following we provide some possible use cases for altmetrics as a supplementary type of measure that is of use to three different user groups: authors, repository managers and university administrators.

## Value of Altmetrics to Authors

1. *Altmetrics can help authors better understand the readership of their open access (OA) content.* Many altmetrics tracking services, including Altmetric.com and ImpactStory, not only document basic usage statistics, but also capture information about readers and how they use content. For example, ImpactStory’s inclusion of Topsy (a Twitter feed archiving platform) metrics’ links to the individual tweets that mention specific articles showcases not only who is reading and sharing scholarship, but also what they are saying about it. Altmetric.com’s content dashboard also showcases sophisticated demographic reports for readers. Giving authors insight into their readership can help them better understand how their OA content stored in IRs is making an impact.

2. *As supplements to the journal impact factor, altmetrics can help authors document the impact of their research when compiling tenure and promotion dossiers.* The journal impact factor (JIF) is the de facto standard in many academic fields for determining the quality of articles. Many researchers include the JIFs for journals in which they have published on their vitas when going up for tenure or promotion, as a means of documenting the impact of their work. By also including supplemental measures of impact (usage counts and altmetrics) for traditional publications, as well as grey literature and other outputs deposited in IRs, faculty can more fully document the impact of their scholarship.

### Value of Altmetrics to Repositories

3. *Repository managers can use altmetrics to persuade potential depositors that there is value in making their content openly accessible.* As Harnad contends, “The prospect of increasing their usage and citation metrics (and their attendant rewards) is an incentive to researchers to provide Open Access to their findings.” [4, p. 6] The possibility of increasing altmetrics counts would arguably have a similar effect on deposit rates.

4. *Gathering numbers beyond general usage statistics can better communicate to repository funders – most often, university administrators – the value of the repository as a platform for hosting OA content.* While general usage statistics might not tell a very informative story about the impact of a particular repository deposit, seeing how content is used and shared (on which websites, by which demographics and for what purposes) can. Similarly, by tracking non-academic use of content, repository managers can build a case for community engagement.

5. *Altmetrics can supplement existing usage statistics to help plan collection development, resource allocation and marketing/outreach.* Altmetrics such as F1000 scores and scholarly social bookmarking sites, in particular, can provide insight into what scholarship is making an impact within specific user groups. By tracking which collections and subjects are popular within a repository, IR administrators can better plan outreach activities. Such altmetrics can also be used to strengthen departmental engagement, which in turn could help build collections.

### Value of Altmetrics to University Administrators

6. *Administrators can use altmetrics as supplementary indicators of impact when showcasing university scholarship to both internal and external stakeholders.* In particular, tracking altmetrics alongside traditional metrics can shed light on impact for university trustees and state legislatures when requesting budget increases, recruiting faculty, etc. [5].

7. *Altmetrics can be used by faculty review committees (such as awards boards or promotion and tenure review systems) to better understand how a particular researcher’s work has been received by scholarly and lay communities* [5]. Altmetrics as supplemental measures of impact not only help authors and IR managers better understand the reach of scholarship, but can also help faculty review committees do so.

Repositories will likely decide if they will implement altmetrics based on a number of factors, including possible service costs, technical support needs, platform integration restrictions (open source or proprietary) and user interest. In addition to the most popular out-of-the-box altmetrics services (Altmeter, ImpactStory and Plum Analytics), there are many ad hoc possibilities for mining and displaying altmetrics for repository content by way of web service APIs and open source tools like PLOS’s Article Level Metrics package.

Library adoption will likely be customized to meet the demands not only of authors and repository managers, but also of university administrators. There are two “flavors” of impact (to borrow a term) – scholarly and popular – that repository managers should keep in mind when considering implementing altmetrics. Metrics that fall within those two categories (which are by no means mutually exclusive) should be judged in tandem with the authority and relevance of the web services that provide them and the possible value those metrics would provide to stakeholders.

There are a number of traditional and alternative metrics that track scholarly impact, including citations (sourced from *Scopus* and *PubMed Central*), Bookmarks (*Mendeley*, *CiteULike*), *Faculty of 1000* reviews, and blog mentions on research blog networks. These metrics are sourced from websites and services that track usage of scholarship at various points in the research life cycle, from reading to writing to post-publication peer review.



Studies have shown that for OA content, traditional measures of scholarly impact (citations) are often closely related to altmetrics measures (social reference management bookmarks) and that for a variety of reasons some scholarly altmetrics can be a better indicator of impact than traditional usage statistics [6], [7], [8]. As supplementary metrics, scholarly altmetrics can prove value for OA content, including content held by repositories.

Popular impact metrics generally rely upon measuring the social web, including Wikipedia citations, Bit.ly clicks and shares, Facebook likes and shares, Del.icio.us bookmarks, Reddit mentions, Twitter mentions and influential tweets, general interest blog mentions and news media mentions. They can be useful when determining the reach of scholarship within a lay audience, though it is worth noting that many researchers use social media for scholarly pursuits, and so at least a portion of popular metrics are accounted for there. Occasionally, popular impact metrics can predict later citations [9].

Page views and download counts fall within a gray area of possible impact, as usage statistics generally reveal little about the end users and what they will do with that which they download. Studies have shown that page views and download counts for OA content are correlated with

scholarly citations and Facebook shares, alike [10] [8]. These metrics can provide general insights and should be considered carefully alongside other metrics when reporting the possible impact of research.

Altmetrics excel over current impact measures such as citations and usage metrics in the area of sentiment analysis. Though in its infancy, some researchers have shown that by combining text mining with altmetrics you can begin to understand how users regard the content they are sharing, liking and bookmarking [6], [11].

Existing barriers to participation are cost, IR technical support resources, inability to incorporate tools into proprietary platforms, limited DOI (digital object identifier) implementation in most repositories, author disambiguation issues and the political implications of displaying nonexistent metrics for relatively unpopular IR materials. Areas for caution are using altmetrics (or any other metrics) to rank or compare researchers and conflating the types of impact with one another – scholarly metrics usually cannot replace popular metrics and vice versa. As certain web services lose their relevancy, their inclusion in altmetrics reports should be reconsidered (for instance, Digg). ■

### Resources Mentioned in the Article

- [1] Counting Online Usage of Networked Electronic Resources (COUNTER). (April 2012). *The COUNTER Code of Practice for e-Resources: Release 4*. Retrieved February 27, 2013, from [www.projectcounter.org/r4/COPR4.pdf](http://www.projectcounter.org/r4/COPR4.pdf).
- [2] Organ, M. K. (2006). Download statistics - What do they tell us? The example of Research Online, the open access institutional repository at the University of Wollongong, Australia. *D-Lib Magazine*, 12(11). Retrieved February 27, 2013, from [www.dlib.org/dlib/november06/organ/11organ.html](http://www.dlib.org/dlib/november06/organ/11organ.html).
- [3] Digital Commons. (December 11, 2012). Purdue Libraries creates buzz at the Indiana State Fair [blog post]. *DC Telegraph*. Retrieved February 27, 2013, from <http://blog.digitalcommons.bepress.com/2012/12/11/purdue-libraries-create-buzz-at-the-indiana-state-fair/>.
- [4] Harnad, S. (May 30, 2008). Validating research performance metrics against peer rankings. *Ethics and Science in Environmental Politics*, 8(11), 103-107. Retrieved February 27, 2013, from <http://eprints.ecs.soton.ac.uk/15619/>.
- [5] Carr, L., Weal, M., & White, W. (April 16, 2010). Research assessment and a diverse role for repositories [slides]. Talk presented at the 5th International Conference on Open Repositories, Madrid, Spain, July 6-9, 2010 (OR2010). Retrieved February 27, 2013, from <http://eprints.soton.ac.uk/270842/7/or2010Diverse.pdf>.
- [6] Haustein, S., & Siebenlist, T. (2011). Applying social bookmarking data to evaluate journal usage. *Journal of Informetrics*, 5(3), 457-446.

*Resources continued on next page*

### Resources Mentioned in the Article (cont.)

- [7] Kraker, P., Körner, C., Jack, K., & Granitzer, M. (2012). Harnessing user library statistics for research evaluation and knowledge domain visualization. *Proceedings of the 21st International Conference Companion on World Wide Web - WWW '12 Companion* (pp. 1017–1024). New York: ACM Press.
- [8] Priem, J., Piwowar, H. A., & Hemminger, B. M. (March 2012). Altmetrics in the wild: Using social media to explore scholarly impact. *arXiv.org*. Retrieved February 27, 2013, from <http://arxiv.org/html/1203.4745v1>.
- [9] Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research*, 13(4). Retrieved February 27, 2013, from [www.jmir.org/2011/4/e123](http://www.jmir.org/2011/4/e123).
- [10] Brody, T., Harnad, S., & Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact. *Journal of the American Society for Information Science and Technology*, 57(8), 1060–1072.
- [11] Parra, C., Birukou, A., Casati, F., Saint-Paul, R., Wakeling, J. R., & Chlamtac, I. (June 15, 2011). UCount: A community-driven approach for measuring scientific reputation. [Version 0]. Abstract accepted for presentation at altmetrics11, Koblenz, Germany, June 14-15, 2011. Retrieved from <http://altmetrics.org/workshop2011/parra-v0/>.

# The Many Faces of Article-Level Metrics

by Jennifer Lin and Martin Fenner

## EDITOR'S SUMMARY

The Public Library of Science (PLOS) collects a range of alternative metrics about the articles they publish to provide different, more meaningful and granular insights into reader response. PLOS captures usage statistics, social shares, academic bookmarks and both scholarly and non-scholarly citations, all offering distinct types of information. Early interest in an article is more apparent through HTML views and mentions on social sharing sites than usage statistics, and Mendeley bookmarking reflects interest but does not correlate with citation count. An article's appearance in citation databases commonly takes at least two years. Mentions in blogs often stimulate commentary and critique. Instead of presenting only a simplistic citation number, PLOS offers article level metrics (ALM) signposts reflecting ALM that capture the variety of response, audience, timing, purpose and impact of a scientific article.

## KEYWORDS

altmetrics	scholarly publishing
impact of scholarly output	social web
citation impact	collaborative filtering

Martin Fenner is the technical lead for the PLOS Article-Level Metrics project. Previously he worked as a medical doctor and cancer researcher. He can be reached at [mfenner@plos.org](mailto:mfenner@plos.org) or @mfenner.

Jennifer Lin is passionate about open access and its political and social impacts. As a former business consultant with Accenture, she worked with Fortune 500 companies as well as governments to develop and deploy new products and services. Jennifer received her Ph.D. in political philosophy and has served as an instructor at Johns Hopkins University. She can be reached at [jlين@plos.org](mailto:jlين@plos.org) (<http://plos.org>).

## Altmetrics: What, Why and Where?

The Public Library of Science (PLOS) is collecting and displaying a large variety of metrics about the articles they publish (Table 1). They include traditional citations, usage stats and altmetrics.

The sheer variety of metrics can be overwhelming, and it would be much easier if we simply substitute a single number for them all. Such economy – though perhaps convenient – is not feasible if we are to preserve the breadth of information that these metrics offer as well as maintain fidelity to their different natures. We will discuss how individual article-level metrics (ALMs) in the PLOS suite measure different things:

- Different audiences: public vs. scholarly interest
- Different dimensions: attention, self-promotion, or impact
- Different timepoints after publication: days, months, or years.

The nature of these measurements is quite dissimilar. As such, they offer different information that is of value in disparate ways. We conclude that it is necessary to look at these metrics altogether as a group, shifting focus to the most relevant ones based on the questions that need to be addressed.

In our early analysis of PLOS ALM data, we have observed some typical

**TABLE 1. Metrics collected and displayed as part of the PLOS Article-Level Metrics**

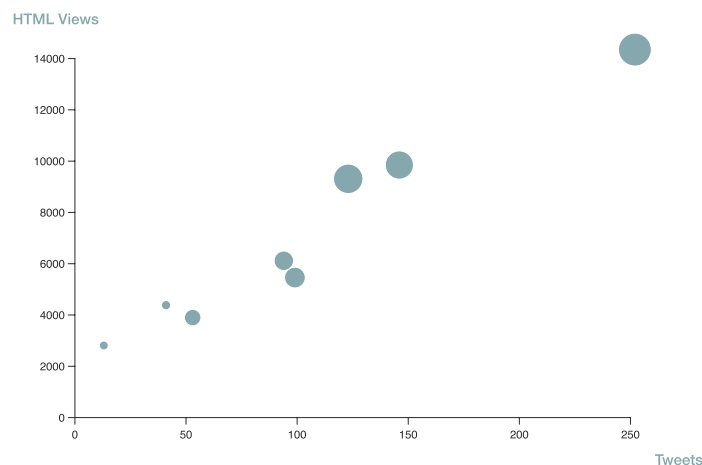
1. **Usage Stats**
  - a. PLOS website
  - b. PubMed Central
2. **Social Shares**
  - a. Facebook
  - b. Twitter
3. **Academic Bookmarks**
  - a. Mendeley
  - b. CiteULike
4. **Scholarly Citations**
  - a. CrossRef
  - b. Scopus
  - c. Web of Science
  - d. Pubmed Central
5. **Non-scholarly Citations**
  - a. ResearchBlogging
  - b. ScienceSeeker
  - c. Nature Blogs
  - d. Wikipedia

patterns that have emerged around certain areas – early attention, scholarly attention, citation impact and non-scholarly citations – which go far to illustrate the diversity of researcher engagement and the subsequent need for different measurements adequate to the task.

### Early Attention

To illustrate the first distinction – different audiences – let us take a quick view of early ALM activity across the PLOS corpus. Early attention for a newly published paper is best described using HTML views and social shares via Twitter and Facebook. There is a strong interaction between social shares and HTML views, and attention via social shares seems to amplify usage (Figure 1).

**FIGURE 1.** Tweets vs. HTML views at the PLOS website for articles in the PLOS Medicine Big Food Collection. Circle size correlates with Facebook activity (likes, shares, comments). Data collected August 19, 2012.



Even though a strong correlation emerges, social shares add an important dimension missing from usage stats: they include information about who is sharing a link to the article and what they are saying.

Since PLOS is an open-access publisher that makes all content openly available in full text, a substantial amount of early activity could be attributed to public attention. For scholarly activity in the first days after publication it is better to look at PDF downloads or at the usage stats from PubMed Central,

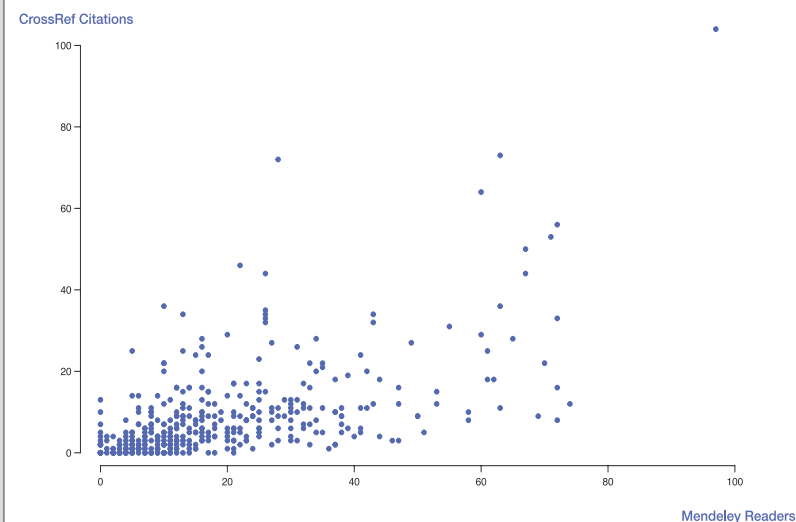
as these numbers do not seem to be much influenced by social shares. In our data, about 90% of PLOS articles show a ratio of HTML views to PDF downloads that is very close to 4:1. Articles with a much higher ratio (e.g. 10:1) regularly show strong activity in Twitter and Facebook. This observation suggests that the HTML/PDF ratio, together with the numbers of social shares, is a better indicator of public interest in an article than absolute numbers of usage stats.

### Scholarly Interest

PLOS collects metrics from Mendeley and CiteULike, both scholarly bookmarking services. Although we often see some activity in the days after publication of a paper, they typically take a few weeks or months to accumulate and, in contrast to usage stats, don't taper off as quickly. Both bookmarking services are used by scholars, so the numbers don't necessarily correlate strongly with the social shares from Twitter and Facebook, which also see a lot of public activity.

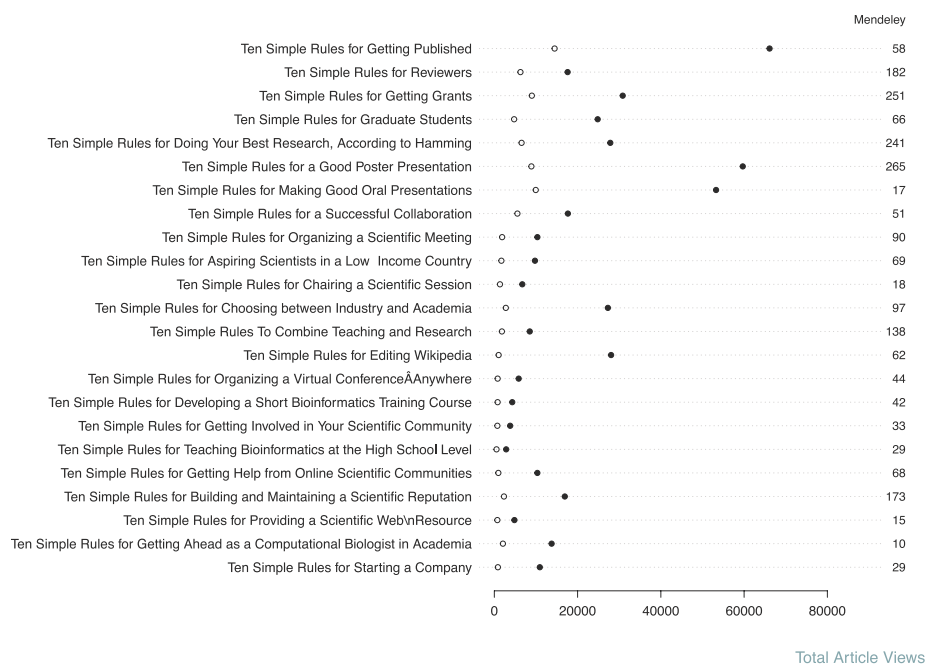
Mendeley bookmark numbers show no real correlation with citation counts, suggesting that storing a paper in a reference manager and citing the paper may not be strongly correlated (Figure 2).

**FIGURE 2.** CrossRef Citations vs. Mendeley Readers for all PLOS Medicine articles published 2010-2011. Data collected January 24, 2013.



Content popular with a relatively large number of scientists – for example, the *PLOS Computational Biology* 10 Simple Rules Collection – often has both high usage stats and high numbers of scholarly bookmarks (Figure 3).

**FIGURE 3.** PDF downloads (open circles), HTML pageviews (closed circles) and Mendeley readers (numbers on the right) for 23 editorials in *PLOS Computational Biology*. Articles are sorted by descending age.



For the most part, however, the correlation between usage stats (PDF downloads and usage stats from PubMed Central included) and Mendeley bookmarks is not especially strong. The number of scholarly bookmarks likely at least in part indicates something beyond citations and usage stats. Such low correlation between these ALM groupings is understandable if we view them as measuring fundamentally different modes or dimensions of researcher engagement.

## Non-scholarly Citations

We have seen further evidence of fundamental differences in the dimensions of research activity captured with the non-scholarly citations ALMs. PLOS tracks three science blog aggregators (ResearchBlogging, Nature Blogs and ScienceSeeker). Blog posts represent a fundamentally different medium than social sharing platforms. The absolute number of times that an original article is cited in blog posts is not the most indicative of its influence and reach as less than 5% of articles are discussed in science blogs. Instead, the enduring value of this measurement lies in the richness of the content that it provides. Authors of blog posts commonly use the open format of a blog posting to engage the cited articles in great detail, providing deep commentary and critique that neither a bookmark or social share can by its very nature. PLOS also began collecting non-scholarly citations on Wikipedia as of September 2012. (Similarly, Wikipedia links to 6% of our articles. An analysis of the kind of PLOS content used in Wikipedia is forthcoming.)

## Citation Impact

Scholarly citations have traditionally been used to look at the impact of a paper. PLOS is collecting metrics from four different citation databases (CrossRef, Web of Science, Scopus, PubMed Central). Although the numbers differ somewhat, they generally show a very strong correlation to each other. Within this group, all the indices capture the same activity relative to each other and as such, reflect the same event time horizon. But they do so in stark contrast to the other ALMs. Citations are much slower to accumulate than the other metrics. We generally see a two- to five-year lag before meaningful numbers emerge. These measurements not only capture a different dimension of engagement but also a dramatically contrasting timescale after publication than the other elements in the suite.

## Conclusions

Article-level metrics describe many different aspects to the broad spectrum of research engagement and can never be expressed in a single number. From the considerations enumerated above, one recommended approach is to focus on the metrics that correspond and are most relevant to a particular use case

(for example, immediate attention after publication). Alternatively, we can deploy a set of metrics to describe an article, be it composed of the entire suite or a select view. PLOS has introduced ALM signposts, an aggregated view of usage stats, social shares, academic bookmarks and scholarly citations. They appear at the top of each article as well in key navigational places on the site. Although the signposts only represent a subset of ALMs, they are meant to provide an easy at-a-glance view of the article's activity across the different groups of measurements (audiences, dimensions and timescales).

By providing a broader spectrum of metrics rolled-up into a summary view, we offer researchers a more manageable set of numbers that hopefully has not replaced deeper explanation for simplification. We have evidence that these components each offer a different view, depending on the user and need. In short, not all “faces of ALM” may be useful or even relevant to any one circumstance. But each, whether a single “face” or set, offers a more adequate outlook than one that is applied without distinction to every circumstance. ■

# Five Challenges in Altmetrics: A Toolmaker's Perspective

by Jean Liu and Euan Adie

## EDITOR'S SUMMARY

The experience of Altmetric LLP, an altmetric tool developer, reveals common issues that demand attention when designing alternative metrics for response to scholarly writings. Identifying what can and should be measured for different user groups is fundamental. A default is to count all relevant mentions in a set of online sources, permitting drill down for more qualitative information. Data source selection varies by need, ranging from government documents to social media comment sites. Since the topic of discussion can be elusive, a tracking method must point backward to original articles or data. Text mining helps for text documents, but audio and video are less workable. Multiple versions of a single article and subsections of books and datasets add ambiguity and redundancy. Valid interpretation depends on context and the relevance and timeliness of data and sources, requiring continual reassessment.

## KEYWORDS

altmetrics  
selection  
relevance  
redundancy

Jean Liu is data curator and blog editor at Altmetric LLP, and Euan Adie is the founder. They can be reached at <http://altmetric.com>.

## Altmetrics: What, Why and Where?

**D**riven by the development of new tools for measuring scholarly attention, altmetrics constitute a burgeoning new area of information science. It is an exciting time to be involved in the field since there are so many opportunities to contribute in innovative ways.

We develop altmetrics tools and related services at Altmetric LLP, a small London-based start-up founded in 2011 [1]. Like all developers of new altmetrics tools, we frequently encounter challenges in defining what should be measured, accurately collecting attention from disparate sources and making sense of the huge amount of compiled data. We outline five of these challenges in this piece, illustrating them with examples from our experience.

It is worth noting that the altmetrics community as a whole comes together regularly to discuss these and other issues, with two open workshops held in 2012 and more planned for the future.

### 1. What can and should we measure?

The term *altmetrics* is often used loosely to refer to all non-traditional measures of re-use, engagement and impact, though emphasis is usually placed on the latter. However, impact is a multi-faceted concept, [2] and different audiences have their own views of what kind of impact matters and the context in which it should be presented: researchers may care about whether they are influencing their peers, funders may care about re-use or public engagement and universities may wish to compare their performance with competing institutions. Accordingly, altmetrics data and methodologies are inevitably used in a variety of different ways to suit a variety of different purposes. This situation is arguably how it should be, with each interested community deciding for itself the kinds of impact or engagement it wants to track and then cherry-picking from the available tools and data.

At Altmetric, we currently only offer one new off-the-shelf metric: we try to sum up the online attention surrounding a journal article by automatically counting all the relevant mentions from a set of online sources (covering mainstream news outlets, social media and more). We then use these counts, along with the relative influence of each source, to create an aggregate metric, called the Altmetric score [3].

The Altmetric score is only one possible, subjective measure of online attention, and ultimately any such measure is only as good as the data upon which it is based. We encourage users to drill down into the underlying data wherever possible, and to this end, we keep a clear audit trail for any activity that has contributed to the score. All of the relevant tweets, posts and other types of mention may be viewed directly. Users are free to perform their own quantitative analyses of the data or even create new metrics and tools that are suitable to measure the specific kinds of impact that they are interested in.

Having *metrics* in the name suggests that altmetrics is a purely quantitative affair, but this perception isn't necessarily the case. Arguably, the current crop of tools is best used during qualitative assessment. By looking at the underlying data, one may take relevant material into account when assessing a piece of work.

## 2. What sources of data should be used?

Where the underlying data of altmetrics should come from is another key challenge. Typically, different data sources are required to measure different types of impact. For example, to measure impact on policy, you may need to look at government documents. Or to look at how work has influenced practitioners, you may need to monitor the online communities in which they congregate. To see how successful public outreach has been, you may want to look at Twitter and Facebook.

Each of the currently available altmetrics tools (discussed elsewhere in this issue) measures a different, though overlapping, set of sources. This diversity is partly attributable to practical considerations, as each data source has different licensing terms, collection issues and risks associated with it. It is also partly because deciding the usefulness of any one data

source remains a fairly subjective process at this point.

Further complexity is added by the fact that online attention from one data source can often be measured in many different ways. For example, quantification of the mentions of scholarly articles on Facebook could take into account either all or just public wall posts, and these posts might be further parsed into the number of wall posts with an article mention or the number of “likes” and comments on that wall post. Each number emphasizes something different and thus paints a slightly different picture of engagement with an article on Facebook.

To make it easy to mix and match data from different altmetrics tools, common standards are required; however, so far, developing these standards has taken a back seat to developing the actual tools themselves.

## 3. How can we identify what research outputs are being discussed?

Once data sources have been identified, an altmetrics tool must be able to map the constituent attention to specific research outputs. Current tools, ours included, typically track attention through links to articles or artifacts such as datasets and presentation slides, resolving these links to unique identifiers like a DOI, PubMed ID or Handle.

A pressing day-to-day issue stems from this reliance on links. Although most tweeters, science bloggers and digitally native media outlets diligently include direct links to the journal articles they discuss, traditional news outlets have no such standard practice. As a result, a large number of science, health and technology news reports fail to include links to the research that they mention. At Altmetric, we have circumvented this particular issue by developing a text-mining mechanism that analyzes the content of news articles. This “news tracker” retrieves relevant keywords like journal titles and author names from the text, performs a search in literature databases and then matches journal articles probabilistically with their associated news coverage.

Text-mining technology might have solved some of our own product's issues around accurately tracking the news, but identifying research output mentions within online multimedia sources has proven to be more challenging.



In podcasts and videos, direct links to research outputs are only very occasionally included in an item's metadata. References to research tend to be made verbally, and altmetrics tools lack the capabilities and resources for analyzing audio and video content to determine what has been mentioned.

There are also understandable concerns that altmetrics may be gamed or artificially increased, either by authors engaging in excessive self-promotion or inadvertently by spammers. Right now such gaming of the system is rare, but simple to spot both algorithmically; in the case of Twitter spam, where hundreds of fake accounts will suddenly engage in meaningless, random retweets, all of the accounts are quite new, follow each other and have never mentioned a scholarly article before.

In the future, more sophisticated methods of detection will certainly be required. Here, advice from experienced groups like SSRN (Social Science Research Network) and COUNTER (Counting ONline Usage of NeTworked Electronic Resources), both of whom regularly deal with these issues as they relate to download statistics, may be invaluable.

#### 4. You say tomato, I say tomahto

Along with the issue of missing links to papers, an opposing problem exists: sometimes different versions of the same article will appear online on multiple sites and with different identifiers. For example, the PubMed Central version of an article may have only a PubMed Central ID and the original article on the publisher's website only a DOI, with no simple way of reconciling the two.

This scatter dilutes the altmetrics for the article, as it is split among different versions, but end-users rarely care for the distinction. It is therefore necessary for altmetrics tools to maintain mappings between different sets of identifiers or to try to automatically match bibliographic metadata to known articles in literature databases. We do both of these things at Altmetric, although items sometimes still slip through the cracks.

A slightly more complicated case is that of datasets, book chapters or other items that are related to a parent book or article. Should attention paid to a dataset be reflected in the altmetrics of the journal article describing it? What if the article is cited extensively, but was written by somebody who

was not the dataset's creator? This kind of scenario already occurs, with research data deposited in sites like figshare.com and Dryad getting their own DOIs. As such, this complication calls for flexibility from altmetrics tools.

#### 5. How do we interpret the data?

The number of scholars who regularly discuss research using social media and/or blogs has been increasing [4], which in turn means that the number of article mentions seen by Altmetric has also been on the rise. Since launching in July 2011, we've collected attention and Twitter demographic information for well over one million unique articles. Already, we have an abundance of data, which will be invaluable for determining trends in the use of particular communication channels over time. As technologies progress and the landscapes of scholarly communication and publishing change, developers of altmetrics tools need to be mindful of how relevant the collected metrics are. It is potentially dangerous to create and encourage adoption of metrics based on sources that might unpredictably cease to be relevant in the future. In other words, what is considered "significant" attention according to a specific measure today (for example, number of times pinned on Pinterest) may become much less meaningful in a few years. How should we account for this?

One approach is to always put any such metric into context. There are many potential ways to do this: we benchmark the Altmetric score (see above) based on other articles within the same journal and from the same time period, as well as across the whole database. As an example, the most popular article in the Altmetric database received an incredible amount of online attention relative to other items appearing in the same journal (*Canadian Medical Association Journal*). Accordingly, the article-level metrics page included a context statement, indicating that the article's Altmetric score "is one of the highest ever scores in this journal (ranked #1 of 940)." *ImpactStory*, too, benchmarks the numbers it presents, by displaying percentiles calculated from large representative samples of articles in *Web of Science*.

With no gold standard of attention to refer to, optimizing thresholds (what's a "good" level of engagement?) and benchmarks is a big challenge for making sense of altmetrics. Establishing context by comparing article-

level attention within journals or against a set of other articles is a good way to start tackling the issue of potentially changing metrics, but context could be further enriched by making comparisons across articles by the same research group or even across articles of the entire discipline. Arguably, the latter comparison would be most useful. A particle physics article that is popular among a small audience of specialists could have lower Altmetric score than an average molecular genetics article that is being discussed by geneticists and members of the general public (a broad audience), and so putting the attention in perspective would be valuable.

Academics from some disciplines may prefer to use certain communication channels over others; for instance, we see more chemists than expected actively participating in academically and professionally oriented discussions on LinkedIn. Moreover, certain disciplines, notably medicine, receive a disproportionately high volume of attention in the mainstream media, and thus, online discussions of these subject areas might include numerous non-specialist participants. Ideally, various discipline-specific norms or trends could be compiled into indicators of the typical level of attention for a particular field. Readers can then interpret the quantitative altmetrics scores in light of this typical level of attention.

However context ends up being defined, detailed records of an article's performance (in relation to others within a similar grouping) will remain informative, even if certain metrics disappear in the future. The challenge,

therefore, is to create robust, informative standards of context that can withstand minor changes in technology and online scholarly communication. Much more research on the usage of particular publishing platforms and social media networks is needed in order to construct and refine typical threshold levels of attention according to specific groupings.

### Concluding Thoughts

When developing altmetrics tools, a number of important considerations must be made with regards to defining metrics, improving measurement capabilities and providing contextual details for present and future data interpretation. Altmetrics toolmakers need to be flexible enough to accommodate the needs of different communities, while still guiding people towards best practice. The use of altmetrics, however they might end up being defined and measured, gives scholars the power to showcase new and unconventional forms of research impact that have previously gone unrecognized.

Certainly an important hurdle not mentioned above has been that some have felt highly skeptical towards the utility of altmetrics. As data research further validates altmetrics as useful measurements of impact, the availability of off-the-shelf tools will also drive wider adoption of altmetrics. Consequently, increased community participation will help to inform new top-down solutions for key tool-development problems. ■

### Resources Mentioned in the Article

- [1] Adie, E. & Roe, W. (2012) Altmetric: Enriching scholarly content with article-level discussion and metrics. *Learned Publishing*, 26(1), 11-17. doi:10.1087/20130103
- [2] Sutherland, W. J, Goulson, D., Potts, S. G., & Dicks, L. V. (2011) Quantifying the impact and relevance of scientific research. *PLoS ONE*, 6(11): e27537. doi:10.1371/journal.pone.0027537. Retrieved March 6, 2013 from [www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0027537](http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0027537).
- [3] *How is the Altmetric score calculated?* Retrieved March 6, 2013, from <https://altmetric.uservoice.com/knowledgebase/articles/83337-how-is-the-altmetric-score-calculated->.
- [4] Priem, J., Piwowar, H. A., & Hemminger, B. M. (March 2012). Altmetrics in the wild: Using social media to explore scholarly impact. *arXiv.org*. Retrieved February 27, 2013, from <http://arxiv.org/html/1203.4745v1>.

# Are Alternative Metrics Still Alternative?

by Mike Buschman and Andrea Michalek

## EDITOR'S SUMMARY

Since the 1960s citation counts have been the standard for judging scholarly contributions and status, but growing awareness of the strategy's limitations should lead to acceptance of alternative metrics. Citation analysis drawbacks include lack of timeliness, self citation and citations that are superfluous, negative and incomplete, and traditional counts reflect only a small fraction of actual usage. A better categorization of scholarly impact would cover usage, captures, mentions and social media in addition to citations. Metrics should include mentions in blogs and other nontraditional formats, open review forums, electronic book downloads, library circulation counts, bookmarks, tweets and more. Such alternative metrics provide a more complete view of peer response to scholarly writings and better demonstrate the relative position of a research grant applicant and potential for influential work. Altmetrics are readily available, and their value for evaluating scholarly work should be recognized.

## KEYWORDS

altmetrics	scholarly publishing
citation analysis	social web
impact of scholarly output	evaluation

Mike Buschman and Andrea Michalek are the co-founders of Plum Analytics. Mike has worked at Microsoft as a librarian and program manager for Microsoft Academic and Book Search and most recently was the director of product management for ProQuest's Summon Discovery Service. Mike lives in Seattle and can be reached at [mike@plumanalytics.com](mailto:mike@plumanalytics.com).

Andrea is a serial entrepreneur with a focus on search and information retrieval products. She ran a technology consulting firm and, prior to founding Plum Analytics, was the director of technology for ProQuest's Summon Discovery Service. Andrea lives in Philadelphia and can be reached at [andrea@plumanalytics.com](mailto:andrea@plumanalytics.com).

## Altmetrics: What, Why and Where?

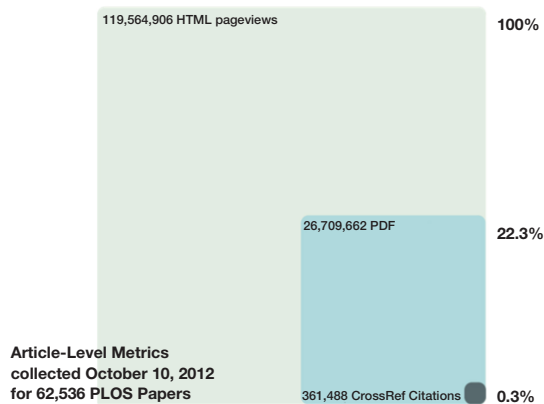
Citation counts have long been the tried and true measure of academic research usage and impact. Specifically, published articles in prominent journals citing other published articles in other prominent journals equate to prestige and tenure. This scheme for determining impact was developed in the 1960s, and while so much else about collecting and disseminating information has changed since that time, the citation count mechanism continues to dominate the way research is evaluated. Yet, there are many well-known problems with this system.

The most obvious problem is that, as the pace of scholarly communication and science advancement has increased, citation analysis is a lagging indication of prestige. Brody and Harnad [1] found that it takes five years for a paper in physics to receive half of the cited-by references that the article will ever acquire. Another issue is self-citation. While there are often very good reasons authors cite themselves in an article, it is also a practice that has been criticized as a tool to increase citation counts and thus potentially artificially inflate prestige and influence. Another is the known practice of publications pressuring academics to pad their papers with superfluous citations. This pressure is applied for a variety of reasons, the most nefarious being that publishers can elevate the status of their own journals with increased citations. Then there is the problem of negative citations. Just because a paper is cited does not mean that it is cited positively; yet, there is no distinction between positive and negative references when evaluating citations counts.

A further problem with citation analysis is the acknowledgement that not all influences are cited in an article, thus leaving the whole measure incomplete. In fact, MacRoberts and MacRoberts reported in their 2010 study that only 30% of influences are typically cited [2]. There are several reasons for these omissions, including authors not citing informal influence

**FIGURE 1.** Chart showing relative reuse metrics for PLOS papers [3]. Used with permission.

**Citations are only a small fraction of how a paper is reused**



or citing a review paper and hence not citing the original work. Citations are a victim of the Matthew effect where the rich get richer. That is, for a variety of reasons, authors tend to cite well-cited material from well-cited journals and ignore other valid work.

In an analysis by PLOS, citation counts only represent a small fraction of how a paper

is used; in fact, citation counts represent less than 1% of usage for an article. Therefore, an article reaches many people but citation counts do not begin to capture the extent of that reach.

**Five Categories of Impact**

It is not surprising that a metric created in the pre-digital world of the 1960s misses a lot of impact and usage. That failure does not make citation analysis inherently bad; it is still a useful tool. But, it does make it inadequate for a complete picture of the usage and impact both of research articles and other research artifacts. To create that complete picture, Plum Analytics studied all of the ways that research artifacts, from articles to videos and everything in between, are made available and used. That research led to the following categorization of impact (Table 1).

These are example sources only; the full list of metrics supported by Plum Analytics can be found at [www.plum-analytics.com/metrics.html](http://www.plum-analytics.com/metrics.html).

By capturing valuable metrics in all of these categories and creating a more complete representation of research and researchers, Plum is able to

provide a more holistic picture than traditional citation analysis. While many will claim that these newer metrics are “alternative,” it is our position that all these metrics are anything but alternative. They are readily available, abundant and essential.

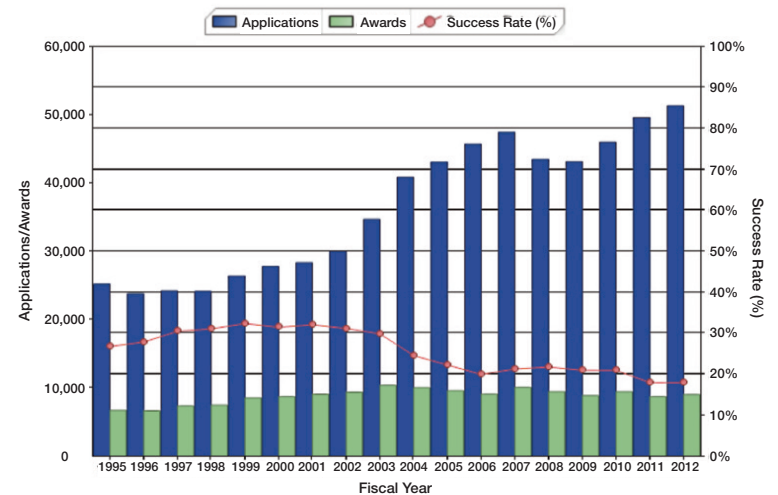
**Era of Increased Competition**

The world of scholarly research is getting more and more competitive. Research budgets are tightening, and funding sources are not meeting the increased demand. Figure 2 shows how applications for

**TABLE 1.** Categorization of impact of scholarly research

Category	Metrics	Example Sources*
Usage	Downloads Views Book Holdings ILL Document Delivery	PLOS WorldCat ePrints Vimeo dSpace
Captures	Favorites Bookmarks Saves Readers Groups Watchers	CiteULike Slideshare Github Mendeley YouTube
Mentions	Blog Posts News Stories Wikipedia Articles Comments Reviews	Wikipedia Facebook SourceForge Reddit
Social Media	Tweets +1s Likes Shares Ratings	Facebook Twitter Google Plus
Citations	Citation Count	Pubmed Scopus

**FIGURE 2.** NIH Research Grants: Applications, Awards and Success Rates [4]



NIH grants have been steadily increasing and success rates have dropped by over 43% in the last 10 years.

When applying for grants, researchers need to show reasons why their use of the award will provide the greatest impact possible. Currently researchers rely on classic citation analysis for relating impact – generally for work that is several years old in order to collect the maximum number of cited-by references. Their latest work, however, is often most relevant to the grant application at hand and may not have had the time to acquire the requisite citation counts. Also there is the possibility that highly cited work that is several years old may have already spurred the most interesting new related research, hurting the chances for award success (for good reason). This paradox exposes another problem of relying only on traditional impact metrics for this purpose.

If researchers can show that their recent research is generating a lot of interaction in the scholarly community, that information can provide an advantage in this tight funding environment. A large number of downloads, views, plays and so forth can show not only early interaction with research, but also how open and accessible the scientists are making their research – a more and more important indicator for funding bodies. In addition if peers are following, saving and bookmarking a researcher’s output, it may portend future citations. Early adopters of these newer impact metrics can reap a noticeable advantage in standing out before the full range of impact metrics becomes universal.

### Metrics for Funders

The funding bodies themselves can gain new understanding and better measure their own success with more timely and holistic metrics about the research they fund. As Bill Gates wrote in his 2013 annual letter, “Given how tight budgets are around the world, governments are rightfully demanding effectiveness in the programs they pay for. To address these demands, we need better measurement tools to determine which approaches work and which do not.” [5] As the success rates for grant funding go down, funding bodies will need to make sure they are making the right choices and are able to defend their decisions over time.

### Negative Results and Other Forms of Research Output

Most researchers agree that both positive and negative results help advance science. Not sharing negative results can lead to unnecessary duplication and incomplete understanding of positive results. The current promotion system, however, discourages publishing research with negative results. Fanelli [6] finds that articles showing negative results have declined in the literature, and positive articles have grown 22% across all disciplines and geographies in the last 20 years. As a reaction, new journals have cropped up such as *Journal of Negative Results in Biomedicine* (<http://jnrbm.com>), *All Results Journal* ([www.arjournals.com/ojs/](http://www.arjournals.com/ojs/)) and *Journal of Pharmaceutical Negative Results* (<http://pnrjournal.com>). These journals, however, by their very nature, make it difficult to discover these findings being grouped by negativity rather than by the scientific niche in general. Many scientists are using blogs to show more details of their research including negative results. Being able to measure the impact of this output in non-traditional venues and formats will encourage scientists to share more details of their research.

Besides negative results, blogs are being used in other ways to communicate research output. Descriptions of methods and settings are increasingly being posted. A good example can be found on Nicholas Pyenson’s blog post ([http://nmnh.typepad.com/pyenson\\_lab/nature-rorquals-organ.html](http://nmnh.typepad.com/pyenson_lab/nature-rorquals-organ.html)) about a recent article published in *Nature*. It contains additional content that makes the study more accessible to lay people, as well as discussing what was not included in the study. Maps and photographs of archeological and paleontological sites as well as other visual artifacts from field study can be widely found these days. Humanities researchers are creating open-review manuscripts using WordPress with the CommentPress Core Plug-in so that the community can comment as the work is coming together and, in the process, create something altogether new, where the comments become preserved peer-review and can themselves lead to new avenues of research. *Writing History in the Digital World* (<http://writinghistory.trincoll.edu>) and *Subjecting History* (<http://subjectinghistory.org>) are two examples of this in action.

With so much research relying on large data coming from sources such

as sensor networks, telescopes, instruments, surveys and simulations, the datasets themselves are often the new “output.” Being able to measure interaction and future science based on a researcher’s data set incentivizes scientists to share their datasets. Sites such as figshare (<http://fishare.com>), Dryad (<http://datadryad.org>) and DataCite (<http://datacite.org>) now allow for datasets to be better hosted, shared and found.

As impact measurements become more accepted and researchers receive credit for these other forms of research output, researchers will be freer to utilize the right scholarly communications format for their output without having to conform to a publication model with the limitations of a print journal.

### Challenges of Disciplines Where Journal Articles Do Not Apply

Classic citation analysis has been applied most readily to those disciplines where journal articles have been the dominant format of research output. In disciplines where books and book chapters prevail, however, it is more difficult to impose the cited-by reference model. And while Elsevier’s *Scopus* and *Web of Science* from Thomson Reuters have recently added some book and data citation sets, citation analysis does not properly offer defensible impact metrics for these disciplines.

Metrics that take into account usage, such as library holdings, library circulation, course readings and eBook downloads, add a layer of impact that is more meaningful for these disciplines. Another category of impact comes with reviews – published and informal – as well as comments and other mentions.

### Web Scale Is a Must

It is easy to see that a system that can support collecting, analyzing and calculating the plethora of metrics for the world’s scholarly research output requires web-scale architecture. There are millions of researchers with hundreds of millions of pieces of research output. A core technology challenge in this space is combining metrics for the same research artifact when it appears in many separate digital locations. For example, the same article can exist in a preprint repository, on the final publisher’s website or in open

access repositories as well as being directly downloadable from a researcher’s homepage. A full representation of the use of this article should capture and algorithmically combine metrics from each of these locations. The problem gets even more complicated since in order to capture the sharing of links to the article, it is necessary to determine all of the URLs that might take a user to that article. There can be multiple URLs that are valid on each website that hosts the article. The process of identifying all of these disparate sources of the same article is called identity Resolution. Although linking articles together by well-known identifiers such as DOI will get partial coverage, this method is insufficient for a full identity resolution solution.

### People Not Papers

The greatest opportunity for applying these new metrics is when we move beyond just tracking article-level metrics for a particular artifact and on to associating all research outputs with the person that created them. We can then underlay the metrics with the social graph of who is influencing whom and in what ways. We can further examine and compare sets of cohorts, whether such a set is a particular lab, institution or set of researchers. Of course this data is valuable to administrators seeking to get a picture about how groups compare to one another and to their peer groups in other institutions. It is also extremely useful for giving researchers the context of their impact with groups meaningful to them. For example, some researchers might be more interested in how they stack up against others in their discipline, their institution as a whole or other researchers at a similar career stage.

### Conclusion

In 1903, with only 150 miles of paved roads in the entire country, Horatio Nelson Jackson was the first to drive a car across the United States. He did not wait for highways to be built. In the same way the popularity of the car created the demand for better highways, the availability of more complete impact metrics for research will surely change the current system. However, even before the system as a whole changes, new metrics are already available to those conducting, supporting or funding research today.

There is a temptation to see this new paradigm for measuring impact as a passing fad: interesting, but too early, or simply not serious with regard to scientific research. The question arises: Does the process for granting tenure need to be changed in order for these measures to be accepted? A better question is why a demonstrably sub-standard process whose faults and

drawbacks are so well known has persisted for so long. The easy answer is that it is all we have had for five decades, but the truth is that decision-makers want quantifiable data for making decisions. Promotion, hiring and grant funding processes will continue to evolve, but those changes will not be prerequisites for including more holistic measurements. ■

### Resources Mentioned in the Article

- [1] Brody, T. & Harnad, S. (2005). *Earlier web usage statistics as predictors of later citation impact*. Retrieved March 5, 2013, from <http://arxiv.org/ftp/cs/papers/0503/0503020.pdf>.
- [2] MacRoberts, M. H., & MacRoberts, B. R. (2010). Problems of citation analysis: A study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology*, 61(1), 1-12.
- [3] Public Library of Science. (November 2012). Citations are only a small fraction of how a paper is reused [slide]. *Article Level Metrics: Analyzing Value in the Scholarly Content*. Presentation by Richard Cave at the Charleston Conference 2012, Charleston, SC, November 7-10, 2012.
- [4] National Institutes of Health. (2012). Research project grants: Applications, awards, and success rates [chart]. *Research Portfolio Online Reporting Tools (RePORTS)*. Retrieved March 5, 2013, from <http://report.nih.gov/NIHDatabook/Charts/Default.aspx?showm=Y&chartId=20&catId=2>.
- [5] Gates, B. (2013). *2013 Annual Letter from Bill Gates*. Gates Foundation. Retrieved March 6, 2013, from <http://annualletter.gatesfoundation.org/>.
- [6] Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90, 891–904. Retrieved March 6, 2013, from <http://eloquentscience.com/wp-content/uploads/2012/02/Fanelli12-NegativeResults.pdf>.

# The Reception of Suzanne Briet in the United States

by Michael K. Buckland

**Editor's Note:** A shortened version of this article in French appeared as a panel in Fayet-Scribe, S. (2012). *Connaissez-vous Suzanne Briet?* *Bulletin des Bibliothèques de France* 57(1), p. 40-44. <http://bbf.enssib.fr/consulter/bbf-2012-01-0040-007>.

## EDITOR'S SUMMARY

A library pioneer of the mid-20th century, Suzanne Briet was a driving force in expanding and modernizing library service and early information science in France. She is most well known for insightful writings on the concepts of bibliography and of documents and documentation. Despite international connections, Briet's writings, almost exclusively in French, received little recognition in the United States during her lifetime. One influential commentator's review suggested misunderstanding of her key ideas. But historical interest in information science, largely from within ASIS&T, and translation and online distribution of her seminal writings have renewed scholarly interest in Briet's work and established a place for her among notable information scientists.

## KEYWORDS

information science history  
 scholars  
 librarianship  
 information scientists  
 France

Michael Buckland is professor pmeritus at the University of California, Berkeley. He can be reached at [buckland@at.ischool.berkeley.edu](mailto:buckland@at.ischool.berkeley.edu).

Suzanne Briet (1894-1989) was one of a small group of women who did much to transform library service in France in the mid-20th century [1] [2]. Her principal professional achievements were to establish reference service at the French National Library and to be director of studies for an early iSchool, the National Institute for Documentation Techniques in Paris. She also contributed substantially through multiple professional organizations before she retired in 1960 and turned to historical and literary studies [3, pp. 1-7, 65-69] [4] [5].

When she died 30 years later, she and her contributions to library and information science had been thoroughly forgotten. Yet now, after another 30 years, she is widely known for her forceful manifesto *Qu'est-ce que la documentation?* (What is documentation?), with its argument that bibliography is really about access to *evidence* and that what constitutes evidence (*documents*) can take many forms, even an antelope [6] [3].

Briet was known personally to individual Americans active in international organizations (FID, IFLA, UNESCO) or to those she met during her tour of the United States in 1951-52. She wrote extensively in French, but apart from English translations of UNESCO documents, only one, about bibliographies in the Bibliothèque Nationale, appeared in English in the United States [7], although when Briet visited the Library of Congress, a sympathetic review of her manifesto *Qu'est ce que la documentation* appeared in the *LC Information Bulletin* [8].

Documents, for Briet, are things that document (verb!), regardless of their material form. Unfortunately, when the influential Jesse H. Shera summarized Briet's ideas he wrote that her theory was "materialistic rather than functional" [9, p. 194]. This statement and his other dismissive comments indicate he



had misunderstood Briet's argument. Briet's extensive writings on professional education and other topics were presumably known to specialists capable of reading French, but no evidence of any impact by Briet is evident.

Later Briet became known in the United States, primarily through Mary Maack [1] [10] and also Buckland [4], as a leader in the modernization of librarianship in France and as a rare woman pioneer of information science. The editors of a special issue on women pioneers requested an article on her [2].

A new interest in the history and theory of information science developed during the 1990s, mainly through the activities of the Special Interest Group/History and Foundations of Information Science of the American Society for Information Science and Technology. Two related articles of mine from the 1990s that included Briet's ideas, one more analytical

("Information as Thing") [11] and one more historical ("What Is a 'Document'?") [12], were widely read by students in schools of library and information science. The idea that an antelope placed in a zoo (and in a taxonomy) could become a document stimulated students' imaginations. T-shirts showing Briet's antelope even became popular at three universities.

The extreme scarcity of copies of *Qu'est ce que la documentation* and the lack of an English translation impeded recognition of Briet. Fortunately, a careful English translation by Ron Day and Laurent Martinet now exists with commentary and bio-bibliographical material [3]. Both the French original and the English translation are now available online. The inclusion, now, of Briet's ideas in the writings of Ron Day (for example, [13]) and others suggest that she is now established in the English-language discourse on information science. ■

### Resources Mentioned in the Article

- [1] Maack, M. N. (1983). Women librarians in France: The first generation. *Journal of Library History*, 18, 407-449.
- [2] Fayet-Scribe, S. (2009). Women professionals in France during the 1930s. *Libraries and the Cultural Record*, 44(2) (2009): 201-219.
- [3] Briet, S. (2006). *What is documentation?* Transl. & ed. by R. E. Day & L. Martinet. Lanham, MD: Scarecrow Press. With bio-bibliography material and commentary. See <http://ella.slis.indiana.edu/~roday/briet.htm>.
- [4] Buckland, M. (1995). The centenary of "Madame Documentation": Suzanne Briet, 1894-1989. *Journal of the American Society for Information Science*, 46(3), 235-237.
- [5] Buckland, M. (N.d.) *Suzanne Briet, 1894-1989: "Madame Documentation."* [Website] Retrieved February 27, 2013, from <http://people.ischool.berkeley.edu/~buckland/briet.html>.
- [6] Briet, S. (1951). *Qu'est ce que la documentation?* Paris: EDIT. Also at <http://martinetl.free.fr/suzannebriet/questcequeladocumentation>.
- [7] Briet, S. (1950). Bibliography in the basement. *Special Libraries*, 41, 52-55.
- [8] Clapp, V. W. (January 28, 1952). [Review of *Qu'est ce que la documentation?*] *Library of Congress Information Bulletin*, 11(5), 1-3.
- [9] Shera, J. H. (October 1952). Special librarianship and documentation. *Library Trends*, 1(2), 189-199. Retrieved February 27, 2013 from [www.ideals.uiuc.edu/bitstream/2142/5464/1/librarytrendsv1i2D\\_opt.pdf](http://www.ideals.uiuc.edu/bitstream/2142/5464/1/librarytrendsv1i2D_opt.pdf).
- [10] Maack, M. N. (2004). The lady and the antelope: Suzanne Briet's contribution to the French Documentation movement. *Library Trends*, 53, 719-747. Pre-press version retrieved February 27, 2013 from [www.gseis.ucla.edu/faculty/maack/BrietPrePress.htm](http://www.gseis.ucla.edu/faculty/maack/BrietPrePress.htm).
- [11] Buckland, M. (1991). Information as thing. *Journal of the American Society for Information Science*, 42(5), 351-360. Preprint retrieved February 27, 2013, from <http://people.ischool.berkeley.edu/~buckland/thing.html>.
- [12] Buckland, M. (1997). What is a "document"? *Journal of the American Society for Information Science*, 48, 804-809. Preprint retrieved February 27, 2013, from <http://people.ischool.berkeley.edu/~buckland/whatdoc.html>.
- [13] Day, R. (2001). *The Modern Invention of Information: Discourse, History, and Power*. Carbondale, IL: Southern Illinois University Press.

## EDITOR'S SUMMARY

A central goal for information architects is to understand how people use information, make choices as they navigate a website and accomplish their objectives. If the goal is learning, we often assume it relates to an end point, a question to answer, a problem to which one applies new understanding. Benjamin Bloom's 1956 taxonomy of learning breaks down the cognitive process, starting from understanding needs and progressing to action and final evaluation. Carol Kuhlthau's 1991 outline of the information search process similarly starts with awareness of a need, progresses through exploring options, refining requirements and collecting solutions, and ends with decision making and action. Recognizing the stages of information browsing, learning and action can help information architects build sites that better meet searchers' needs.

## KEYWORDS

information architecture  
learning  
information needs  
information seeking  
browsing  
information use

## Enabling Action: Digging Deeper into Strategies for Learning

by Thom Haller, associate editor for information architecture

When I teach, my students have an opportunity to explore how humans use information. “Let’s help the humans,” we cry. If we know more about how people interact with the words and images we create, the odds are better that we can develop products that people can use.

To help us understand humans, we explore how people respond visually, make connections, navigate space and act – how they make choices as they work to get their jobs done.

“How?” we wonder, “can we help people accomplish what they want to accomplish?” Attend to tasks, we decide – we can look for these mini-goals – looking specifically at questions formed in a user’s mind and what we expect as a visible result, an observable end.

But what if there is no “observable end?” What if a human’s knowledge is limited and he or she leans more toward browsing. Maybe, they would refer to their interest as “learning”?

Certainly, I regard myself as a fan of learning. Questions first drove me to the WorldWideWeb (no spaces in 1994). I was curious about the Māori, indigenous Polynesian people

Thom Haller – teacher, speaker, writer and user advocate – teaches principles of performance-based information architecture and usability. Thom volunteers time supporting advocates of plain language. Thom can be reached via email at [thom@thomhaller.com](mailto:thom@thomhaller.com).

of New Zealand. Go to the Mosaic browser, type in Maori, out comes information. How novel. How exciting.

Students and clients often tell me they go online because they “want information.”

“Click anywhere, there it is: information.” I respond. “We drown in information.” But as students of information structure, we begin to dig into what “wanting information” means. I often explain how our reading, especially workplace reading, focuses on “reading to do” or reading that specifies action. I contrast this with “reading to learn” or reading that centers on a topic – as in a classroom.

During a recent class, one of my students – a department manager – challenged the notion that her audiences were reading to do, telling me “our audiences read to learn.” “Sure they want to learn,” I responded. “But you can find a reason for this learning – they want to understand, comprehend and apply.”

I referred to an organizing structure on human cognition first introduced by Benjamin Bloom in 1956. His taxonomy explored how understanding (identifying what specifics or theories the person wants to know) leads to comprehension (organizing, comparing, translating) which leads to application (solving problems based on what you’ve learned), analysis (breaking information into parts to enable closer examination), synthesis (assembling ideas and reshaping them to meet needs) and evaluation (developing and presenting judgments).

Did I win her over and help her begin to emphasize doing? No.

The manager stared at me.

I began to question – or continue to question – my reliance on Bloom. Yes, it's a great taxonomy for helping to understand what people do with information – but it may not be so useful for web development specialists crafting information to help people learn.

So I left class with my own homework assignment: learn more about phases we follow when we look at text “to learn.”

I found (in our information science literature) a six-stage process of information seeking, specifically directed at learning: Carol Kuhlthau's information search process. Introduced in 1991, Kuhlthau identified six stages of learning: initiation/awareness, selection, exploration, formulation, collection and action.

How can we apply these stages in architecting usable structure and content? Fortunately, my exploration directed me to Jared Pomranky's explanation on his site Net Profit Marketing ([http://bit.ly/stages\\_learning](http://bit.ly/stages_learning)). Pomranky states that “by understanding what people are looking for at each stage of their learning process, we can design websites that [effectively] guide them.” He identifies each step and applies them to our work in user experience:

- **Initiation (Awareness)** – Users becomes aware that they need information. Generally, it's assumed that

visitors to your website already have this awareness, but there are circumstances in which you can generate awareness as well.

- **Selection** – Users weigh topics against criteria such as personal interest, project requirements, available information and time – predicting the outcome of possible choices and choosing a topic with the potential for success.
- **Exploration** – Users see the available options to choose among. Quite often, especially online, “analysis paralysis” can set in and make learners quit at this stage because they can't decide which of the options are worth further pursuit.
- **Formulation** – Users see they're going to have to create further requirements before they're able to make a final selection, and they make decisions to narrow the field. Confidence returns.
- **Collection** – Users have clearly articulated their precise needs and are able to evaluate potential options. They gather all available solutions and begin to weigh them based on relevant criteria.
- **Action** – Users make their final decisions and act on them based on their understanding.

So let's help the humans. Start by exploring Carol Kuhlthau's research. We have a lot to learn. At least I do.

**Editor's note:** For further reading, please see the following two articles by Carol Kuhlthau that have appeared in the *Bulletin*:

Kuhlthau, C. C. (June/July 2007). Reflections on the development of the model of the information search process (ISP): Excerpts from the Lazerow Lecture, University of Kentucky, April 2, 2007. *Bulletin of the American Society for Information Science and Technology*, 33 (5), 32-37. Retrieved March 15, 2013, from <http://onlinelibrary.wiley.com/doi/10.1002/bult.2007.1720330511/full>.

Kuhlthau, C. C. (February/March 2005). Accommodating the user's information search process: Challenges for information retrieval system designers. *Bulletin of the American Society for Information Science and Technology*, 25(3), 12-16. Retrieved March 15, 2013, from <http://onlinelibrary.wiley.com/doi/10.1002/bult.115/full>.

**EDITOR'S SUMMARY**

Human classification alone, unable to handle the enormous quantity of project data, requires the support of automated machine-based strategies. In collaborative annotation, humans and machines work together, merging editorial strengths in semantics and pattern recognition with the machine strengths of scale and algorithmic power. Discovery informatics can be used to generate common data models, taxonomies and ontologies. A proposed project of massive scale, the Large Synoptic Survey Telescope (LSST) project, will systematically observe the southern sky over 10 years, collecting petabytes of data for analysis. The combined work of professional and citizen scientists will be needed to tag the discovered astronomical objects. The tag set will be generated through informatics and the collaborative annotation efforts of humans and machines. The LSST project will demonstrate the development and application of a classification scheme that supports search, curation and reuse of a digital repository.

**KEYWORDS**

knowledge discovery  
classification schemes  
automatic taxonomy generation  
machine aided indexing  
digital repositories  
information reuse  
astronomy

# Collaborative Annotation for Scientific Data Discovery and Reuse

by Kirk Borne

The enormous growth in scientific data repositories requires more meaningful indexing, classification and descriptive metadata in order to facilitate data discovery, reuse and understanding. Meaningful classification labels and metadata can be derived autonomously through machine intelligence or manually through human computation. Human computation is the application of human intelligence to solving problems that are either too complex or impossible for computers. For enormous data collections, a combination of machine and human computation approaches is required. Specifically, the assignment of meaningful tags (annotations) to each unique data granule is best achieved through collaborative participation of data providers, curators and end users to augment and validate the results derived from machine learning (data mining) classification algorithms. We see very successful implementations of this joint machine-human collaborative approach in citizen science projects such as Galaxy Zoo and the Zooniverse (<http://zooniverse.org/>).

In the current era of scientific information explosion, the big data avalanche is creating enormous challenges

Kirk Borne is professor of astrophysics and computational science at George Mason University in Fairfax, Virginia. He can be reached at [kborne@gmu.edu](mailto:kborne@gmu.edu).

for the long-term curation of scientific data. In particular, the classic librarian activities of classification and indexing become insurmountable. Automated machine-based approaches (such as data mining) can help, but these methods only work well when the classification and indexing algorithms have good training sets. What happens when the data includes anomalous patterns or features that are not represented in the training collection? In such cases, human-supported classification and labeling become essential – humans are very good at pattern discovery, detection and recognition. When the data volumes reach astronomical levels, it becomes particularly useful, productive and educational to crowdsource the labeling (annotation) effort. The new data objects (and their associated tags) then become new training examples, added to the data mining training sets, thereby improving the accuracy and completeness of the machine-based algorithms.

Humans and machines working together to produce the best possible classification label(s) is *collaborative annotation*. Collaborative annotation is a form of human computation [1]. Humans can see patterns and semantics (context, content and relationships) more quickly, accurately and meaningfully than machines. Human computation therefore applies to the problem of annotating, labeling and classifying voluminous data streams.

### Resources Mentioned in the Article

- [1] von Ahn, L. (2009). Human computation. *ACM/IEEE Design Automation Conference*, 46, 418-419
- [2] Borne, K. D. (2010). Astroinformatics: Data-oriented astronomy research and education. *Journal of Earth Science Informatics*, 3, 5-17.
- [3] Feigelson, E. D., Ivezić, Z., Hilbe, J., & Borne, K. D. (2013). New organizations to support astroinformatics and astrostatistics. arXiv:1301.3069v1. Retrieved February 25, 2013 from <http://arxiv.org/pdf/1301.3069.pdf>.

The best annotation service in the world is useless if the tags (markup) are not scientifically meaningful (that is, if the tags do not enable data discovery, reuse and understanding). Therefore, it is incumbent upon science disciplines and research communities to develop common data models, taxonomies and ontologies. Since these concepts do not appear spontaneously out of large data collections, they require research and study. The data science of discovery informatics is focused on these research problems: how to enable discovery, access, interoperability, integration, reuse and mining of large distributed data. The disciplines of bioinformatics, geoinformatics and medical informatics are examples of well-established discovery informatics sub-disciplines within their larger scientific disciplines. A similar emerging research domain in the field of astronomy is astroinformatics, which targets the big data flood in astronomy [2]. New professional organizations within astronomy are now established in this area [3].

In astronomy, the proposed LSST (Large Synoptic Survey Telescope, [www.lsst.org](http://www.lsst.org)) project would carry out a systematic 10-year observation program to image the entire southern sky repeatedly throughout the night, every night for 10 years. The resulting data repository would include over 100 petabytes in the final image archive and over 20 petabytes in the final scientific database of extracted science measurements, parameters and metadata. The discovery potential of this data collection would be enormous, and its long-term value (through careful data management and curation) would thus require (for maximum scientific return) the participation of scientists and citizen scientists as well as science educators and their students in a collaborative knowledge mark-up

(annotation and tagging) data environment. To meet this need, we envision a collaborative tagging system called AstroDAS (Astronomy Distributed Annotation System). AstroDAS is similar to existing science knowledge bases, such as BioDAS (Biology Distributed Annotation System, [www.biodas.org](http://www.biodas.org)). AstroDAS is distributed in the sense that the source data and metadata are distributed, and the users are distributed. “Annotation” includes tagging both individual data granules and subsets of the data. It is a “system” in the sense that it is based on a formal, explicit, unified schema for the annotation database, applicable to all astronomy data collections, not only LSST. The DAS provides a distributed system for scientists (professional or citizen) anywhere to annotate individual astronomical objects with labels (known classes), attributes (known features) and new characterizations (newly discovered patterns and behaviors). These annotations can be applied to other astronomical data/metadata within distributed digital data collections. The annotations provide curation, provenance and semantic (scientifically meaningful) metadata about the data source and the data object being studied.

The design and specification of a unique, meaningful, searchable and scientifically impactful set of tags can be achieved through collaborative (human-plus-machine) annotation efforts and through discovery informatics research. These steps will produce a searchable classification and indexing scheme for the curation, classification, discovery, reuse, interoperability, integration and understanding of digital repositories. These efforts will assist scientific data librarians in reaching the holy grail of semantic annotation of data, information and knowledge. ■