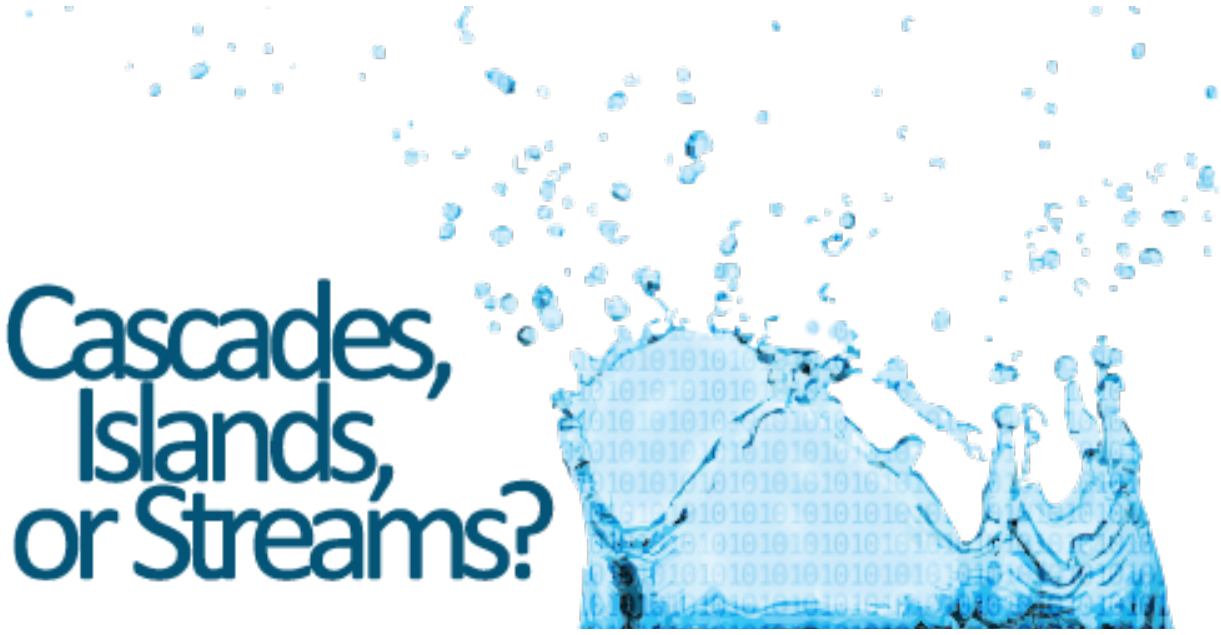

Time, topic, and scholarly activities in humanities and social science research

PRINCIPAL INVESTIGATORS:

Cassidy R. Sugimoto, Indiana University Bloomington

Vincent Larivière, Université de Montréal

Mike Thelwall, University of Wolverhampton

FUNDING AGENCIES:

National Science Foundation (US)

Social Science Humanities Research Council (Canada)

JISC (UK)

GRANTING PERIOD:

February 2012-January 2014

OVERVIEW

The objective of this project was to create and examine large-scale heterogeneous datasets to increase our understanding of the scholarly communication system, to identify and analyze various scholarly activities for creating and disseminating new knowledge, and further develop the innovative computer software developed to collect, filter and analyze data from the web and social media to discover trends in science and in scholarly communication. For years, our knowledge of the scholarly landscape, and subsequently, our understanding of innovation, productivity, and impact, has been largely informed by homogeneous and often biased corpora. However, the growth of datasets that reflect unique areas of scholarly activities (both informal and formal activities) have altered the research landscape and provide us with the opportunity to create more accurate understandings of the nature of science and how science is communicated. Also, while the growth of large-scale datasets has enabled examination *within* scientific datasets, there is a lack of research that looks *across* datasets. By looking across various dataset to investigate trends and correlations this project aimed at identifying the impact and visibility of various scholarly activities, and to identify datasets that should no longer be marginalized, but built into our understandings and measurements of scholarship. The results from the project present an argument that transformations in the scholarly communication system affect not only how scholars interact, but also the very substance of these communications. The project created many refereed journal articles, conference articles, and other publications that build on each other and help to triangulate a broad picture of the current scholarly communication system and the role of social media metrics. Some of the scientific publications written by the team members created interest beyond the scientific community and were recognized by the popular press.

PERSONNEL

The project spanned three countries, with a principal investigator (PI) at each location. Cassidy R. Sugimoto led the US portion, funded by the National Science Foundation. Vincent Larivière directed the Canadian research, funded by SSHRC. The third portion was funded by JISC in the United Kingdom and led by Mike Thelwall. The team also included a number of co-PIs, post-doctoral researchers, and doctoral students, as shown below.

Indiana University Bloomington, USA	**Université de Montréal, CAN**	**University of Wolverhampton, UK**
Cassidy R. Sugimoto, PI Ying Ding, Co-PI Stasa, Milojevic, Co-PI Timothy Bowman, PhD student Bradford Demarest, PhD student Chaoqun Ni, PhD student Scott Weingart, PhD student Grant Simpson, PhD student Andrew Tsou, PhD student Erjia Yan, PhD student Guo Zhang, PhD student	Vincent Larivière, PI Stefanie Haustein, Post-doc Isabella Peters, Visiting scholar Philippe Mongeon, PhD student	Mike Thelwall, PI Kim Holmberg, Post-doc Fereshteh Didegah, PhD student Ehsan Mohammadi, PhD student

Senior personnel. The investigators and post-doctoral researchers on this project brought a diversity of experience, expertise, and datasets to the project. They are briefly introduced below.

Ying Ding. Ying Ding is an Associate Professor at the Department of Information and Library Science, School of Informatics and Computing, Indiana University. Before she worked as a senior researcher at the University of Innsbruck, Austria and as a researcher at the Free University of Amsterdam, the Netherlands. She has been involved in various NIH and European-Union funded Semantic Web projects. She has published 150+ papers in journals, conferences and workshops. She serves as a Program Committee member for 120+ international conferences and workshops. She is the coeditor of book series called Semantic Web Synthesis by Morgan & Claypool publisher. She is co-author of the book "Intelligent Information Integration in B2B Electronic Commerce" published by Kluwer Academic Publishers. She is also co-author of book chapters in the book "Spinning the Semantic Web" published by MIT Press and "Towards the Semantic Web: Ontology-driven Knowledge Management" published by Wiley. She is the editoral board member of four ISI indexed top journals in Information Science and Semantic Web. Her current interest areas include social network analysis, Semantic Web, citation analysis, knowledge management and application of Web Technology.

Stefanie Haustein. Stefanie Haustein is a post-doctoral researcher at Université de Montréal and a research analyst at Science-Metrix in Montréal, Canada. Her research focuses on bibliometrics and altmetrics, i.e. social-media based usage data, to evaluate scholarly communication. Stefanie holds a PhD in information science from Heinrich Heine University Düsseldorf, Germany and has worked in the bibliometrics team at Forschungszentrum Jülich, Germany, where she has conducted several bibliometric analyses supporting decisions in research evaluation. Her doctoral work focused on the multidimensional evaluation of scholarly journals and was awarded the Eugene Garfield doctoral dissertation scholarship in 2011. Stefanie frequently presents her work at international conferences and has published in journals such as Scientometrics and Journal of Informetrics.

Kim Holmberg. Kim Holmberg, PhD, is a Research Associate at the Statistical Cybermetrics Research Group at University of Wolverhampton and a Research Fellow at the Department of Information Studies at Åbo Akademi University. His research interests include webometrics, scientometrics, information dissemination, social media, Web 2.0, Library 2.0 and virtual worlds in education. Of these topics he has published widely and held several talks and presentations and organized courses and workshops for companies, universities, libraries and other organizations.

Vincent Larivière. Vincent Larivière is assistant professor of information science at the Université de Montréal, where he teaches research methods and bibliometrics. He is also an associate researcher at the Observatoire des sciences et des technologies and a regular member of the Centre interuniversitaire de recherche sur la science et la technologie. His work in the area scholarly communication has been published in journals such as the Journal of the American Society for Information Science and Technology, Scientometrics and Journal of Informetrics. Vincent holds a B.A. in Science, Technology and Society (UQAM), an M.A. in history of science (UQAM) and a Ph.D. in information science (McGill), and has performed postdoctoral work at Indiana University's Department of Information and Library Science, School of Informatics and Computing.

Stasa Milojevic. Staša Milojević is an Assistant Professor at Indiana University's Department of Information and Library Science, School of Informatics and Computing. Staša received her PhD from the Department of Information Studies at the University of California, Los Angeles. Her research focuses on studying how modern scientific disciplines/fields form, organize and develop. She approaches modern scientific fields/disciplines as complex heterogeneous socio-cultural networks of people, ideas, documents and institutions. In large-scale longitudinal studies of different scientific fields she combines models, theories and methods from information science, science and technology studies, and social

network analysis. She has published in the areas of scientometrics and network science and has explored ways in which techniques from network science and scientometrics can be combined for more comprehensive studies of science.

Cassidy R. Sugimoto. Cassidy R. Sugimoto is an assistant professor at Indiana University Bloomington. She earned her bachelor's, master's and doctoral degrees from the University of North Carolina at Chapel Hill. Her work has appeared in a dozen journals, most notably in the Journal of the American Society for Information Science and Technology and Scientometrics, on whose editorial boards she serves. She is active in the American Society for Information Science & Technology, having been elected to the Board of Directors as well as chairing and serving on multiple special interest groups and chapters. Her research has been funded intra-murally as well as by professional associations (ALISE, ASIS&T) and national agencies (e.g., NSF).

Mike Thelwall. Mike Thelwall is Professor of Information Science and leader of the Statistical Cybermetrics Research Group at the University of Wolverhampton, UK and a research associate at the Oxford Internet Institute. Mike has developed tools for gathering and analysing web data, including hyperlink analysis, sentiment analysis and content analysis for Twitter, YouTube, blogs and the general web. His publications include 152 refereed journal articles, seven book chapters and two books, including Introduction to Webometrics. He is an associate editor of the Journal of the American Society for Information Science and Technology and sits on three other editorial boards.

Student emissaries. One of the jewels of this project was the student emissary program. This program was built upon the belief that the best investment we can make in the future of the scientific enterprise is in our students. Given the globalization of science, it is imperative that we train the next generation of scholars in the practice of international collaboration. To this end, we incorporated students into our research and funded their travel to international conferences and team meetings. Student authors were included on half of the journal articles and conference works. Ten students were given an opportunity to participate in full team meetings and seven students traveled to international meetings or conferences. In addition, collaborations among students were highly encouraged and many of these were initiated and continue beyond the confines of the grant.

MEETINGS

As the team involved multiple people living in five different countries, constant communication was imperative for the success of the project. This was achieved through large in-person meetings as well as smaller in-person and virtual meetings.

Full team meetings. The team was distributed across three countries: the United States, the United Kingdom and Canada. To facilitate the collaboration, each PI hosted a full team meeting at their institution. The first meeting was held March 10-12, 2012 in Wolverhampton, United Kingdom. Present at this meeting were all three PIs (Mike Thelwall (host), Vincent Larivière, and Cassidy R. Sugimoto). Also present was co-PI Stasa Milojeivc and post-doc Kim Holmberg. Co-PI Ying Ding skyped in for some of the meeting. Four students (three from US and one from the UK) were present: Erjia Yan, Chaoqun Ni, Brad Demarest, and Fereshte Didegah. This meeting was particularly important in building consensus among team members, preliminary data collection, and planning explicit action items for the project.

The second team meeting was held two months later, May 23-26, 2012, in Bloomington, United States. All three PIs (Cassidy Sugimoto (host), Vincent Larivière, and Mike Thelwall) were present, as well as

both co-PIs (Ying Ding and Stasa Milojevic). Six students from the US were involved in the meeting: Erjia Yan, Chaoqun Ni, Brad Demarest, Scott Weingart, Grant Simpson, and Guo Zhang. At this meeting, we discussed our progress on the action items described in the first team meeting, discussed problems and solutions, and spent some time doing research in small groups. We also spent a day discussing new avenues of research from those outlined in Montréal. The benefit of hosting this meeting only a couple months from the first team meeting was that many projects had begun, but were still enough in the design and data collection phases to allow for full team input. The outcome was a series of specific action items for the next year of work.

The final full team meeting was held on March 11-12, 2013 in Montréal, Canada. Two PIs were present at this meeting (Vincent Larivière (host) and Cassidy R. Sugimoto). Co-PI Ying Ding was present, as well as two post-docs (Kim Holmberg and Stefanie Haustein). Three students were present, one from Canada and two from the US (Tim Bowman, Brad Demarest, and Philippe Mongeon). A full day of this meeting was spent presenting the state of research from the grant. The second day was used to discuss what steps needed to be taken to finish grant activities and to propose post-grant activities.

Smaller meetings. There were also some smaller in-person team meetings. On June 6-9, 2013, Vincent Larivière and Stefanie Haustein came to Bloomington to meet with visiting scholar Isabella Peters. Work at this meeting focused primarily on the twitter data and involved doctoral student Tim Bowman. We were also able to have impromptu meetings at conferences where many of us were in attendance (e.g., ISSI in Vienna, ASIST in Baltimore, ASIST in Montréal, KnowEscape in Finland, and STI in Berlin).
In addition to the in-person meetings, we also had several skype and phone calls among the team PIs and small group calls between members from each of the locations working on a specific paper. At the beginning of the project, these were largely facilitated by the PIs. However, by the end of the project, the students and post-docs were skyping frequently without any PIs. We saw this as a sign of success in the project—enabling the next-generation of scholars to establish and maintain collaborations outside of the project.

DATA

As a Digging into Data initiative, it is perhaps not surprising that a large amount of data was collected and analyzed during the course of the project. We detail below the various types of data that were gathered and analyzed in our published research and research that has been gathered for research in progress.

arXiv. On theme of our research was on the relationship among various genres and the relationship and impact of novel forms of scholarly communication on the current paradigm. To this end, we utilized arXiv as an exemplary pre-print database for two studies. In the first, we downloaded the entirely of the arXiv database from 1990 to March 22, 2012 (n=744,583 pre-prints) and matched these data with the entire WoS database (Larivière, Sugimoto, Macaluso, Milojevic, Cronin, & Thelwall, in press; Larivière, Macaluso, Milojevic, Sugimoto, & Thelwall, 2012; Larivière, Macaluso, Sugimoto, Milojevic, Cronin, & Thelwall, 2013). In a subsequent study, only astrophysics preprints (n=117,913) were used for analysis (Hu, Dong, Zhang, Bowman, Ding, Yan, & Milojevic, under review; Ding, Yan, Sugimoto, & Milojevic, 2013).

Dissertations and Theses. We had access to the entire ProQuest Dissertations and Theses database, using a research license from ProQuest. In one study, we analyzed all 1,850,846 research doctorate dissertations (Ni & Sugimoto, 2012). In another study, we extracted a sample of 22,322 dissertations

with the subject category philosophy, 79,731 dissertation abstracts with the subject category physics, and 166,783 dissertation abstracts with the subject category psychology. We then used a simple undersampling approach to create similarly sized subsamples from each discipline for analysis (Demarest & Sugimoto, in press; Demarest & Sugimoto, 2013a; Demarest & Sugimoto, 2013b).

Web of Science. We consistently used Web of Science in various projects. We obtained article information for digital humanities (Bowman, Demarest, Weingart, Simpson, Larivière, Thelwall, & Sugimoto, 2013; Bowman, Demarest, Weingart, Simpson, Larivière, Thelwall, & Sugimoto, in progress), science and technology studies (STS) (Milojevic, Sugimoto, Larivière, Thelwall, & Ding, 2013; Milojevic, Sugimoto, Larivière, Thelwall, & Ding, under review), and astrophysics (Hu, Dong, Zhang, Bowman, Ding, Yan, & Milojevic, under review; Ding, Yan, Sugimoto, & Milojevic, 2013) journal articles; citation and publishing records for 27 astrophysicists (Haustein, Bowman, Holmberg, Peters, & Larivière, in press) and nearly 1000 TED talk presenters (Sugimoto, Thelwall, Larivière, Tsou, Mongeon, Macaluso, 2013); and references to TED talks (Sugimoto & Thelwall, 2013). We matched Web of Science records to PubMed (Haustein, Peters, Sugimoto, Thelwall, Larivière, 2014) arXiv (Larivière, Sugimoto, Macaluso, Milojevic, Cronin, & Thelwall, in press) and Mendeley (Mohammadi, Thelwall, Haustein, & Larivière, in press) records. We also used WoS to identify the most productive researchers in astrophysics, biochemistry, digital humanities, economics, history of science, cheminformatics, cognitive science, drug discovery, social network analysis, and sociology**.**

PubMed. We downloaded all 2010-2012 articles in PubMed for analysis (Haustein, Peters, Sugimoto, Thelwall, Larivière, 2014; Haustein, Larivière, Thelwall, Amyot, & Peters, in press). These were later matched with tweets, WoS, Altmetric scores, and Mendeley metrics.

Digital Humanities. We obtained a large amount of data relative to digital humanities, including all articles from DHQ and LLC as well as 19 other DH-related journals; full listserv history from Humanist and TEI-L; grant data from NEH and other relevant funding sources; DH syllabi; DH blogs; DH center information; and data on DH twitter communication (starting from a list of Twitter handles from digitalhumanitiesnow.org and their followers) (Bowman, Demarest, Weingart, Simpson, Larivière, Thelwall, & Sugimoto, 2013; Bowman, Demarest, Weingart, Simpson, Larivière, Thelwall, & Sugimoto, in progress).

Information science. We identified a list of all iSchool faculty members and generated 6,760 keywords from 1,168 researchers' online profiles for analysis.

Handbooks. Five handbooks were analyzed for the STS study. The title information for each of these is listed below:

- Rösing, I., & Price, D. J. d. S. (Eds.). (1977). Science, technology, and society: A cross-disciplinary perspective. London: SAGE.
- Van Raan, A. F. J. (Ed.). (1988). Handbook of quantitative studies of science and technology. Amsterdam: North-Holland.
- Jasanoff, S., Markle, G. E., Petersen, J. C., & Pinch, T. (Eds.). (1995). Handbook of science and technology studies. Thousand Oaks: SAGE.
- Moed, H. F., Glänzel, W., & Schmoch, U. (Eds.). (2005). Handbook of Quantitative Science and Technology Research. New York: Kluwer Academic Publishers.
- Hackett, E. J., Amsterdamska, O., Lynch, M., & Wajcman, J. (Eds.). (2008). The handbook of science and technology studies. Cambridge, MA: The MIT Press.

These were individually scanned, OCR'ed and manually cleaned prior to analysis.

TED Talks. We conducted a series of studies on TED talks. We downloaded a list of 1,202 videos maintained by the TED talks website. This list provided a list of the videos and associated metadata. To this, we gathered webometric metrics using Webometric Analyst (including metrics from TED, YouTube, Mendeley references, syllabi presence, etc.). We also conducted manual analyses in Google Scholar and Web of Science and added citation data to this database. The final dataset contained detailed information on each of the TED talks (Sugimoto & Thelwall, 2013). In a subsequent paper, we identified the individual presenters associated with the TED talks and manually collected demographic data on these individuals, including academic status, year of doctoral degree, and gender (Sugimoto, Thelwall, Larivière, Tsou, Mongeon, Macaluso, 2013). All these information were merged into the previous database. In the third analysis, we gathered comments from the TED website and YouTube websites. The first analysis involved six total comments for each of 405 TED talks; the second analysis involved 5880 comments associated with 196 videos (Tsou, Thewall, Mongeon, & Sugimoto, 2014). We have now gathered the transcripts for all of the original 1,202 videos as well as transcripts for videos that have been posted since our original analysis (for a total of 1,683 transcripts). These will be used in our future analyses and incorporated into our present database on TED talks.

Altmetric. Data from almetric.com was used for a number of studies (Thelwall, Haustein, Larivière, & Sugimoto, 2013; Haustein, Peters, Sugimoto, Thelwall, Larivière, 2014). The data was delivered on January 1, 2013 and includes almetric scores gathered since July 2011. These included:

- *Tweets*: Tweets from a licensed Twitter firehose are checked for citations.
- *FbWalls*: A licensed Facebook firehose is used for Wall posts to check for citations.
- *RH*: Research highlights are identified from Nature Publishing Group journals.
- *Blogs*: The blog (feed) citations are from a manually-curated list of about 2,200 science blogs, derived from the indexes at Nature.com Blogs, Research Blogging and ScienceSeeker.
- *Google+*: The Google+ Applications Programming Interface (API) is used to identify Google+ posts to check for citations.
- *MSM*: The mainstream media citation count is based on a manually curated list of about 60 newspapers and magazines using links in their science coverage.
- *Reddits*: Reddit.com posts from the Reddit API are checked for citations.
- *Forums*: Two forums are scraped for citations.
- *Q&A*: The Stack Exchange API and scraping of older Q&A using the open source version of Stack Exchange's code are used to get online questions and answers to check for citations.
- *Pinners*: Pinterest.com is scraped for citations.
- *LinkedIn*: LinkedIn.com posts from the LinkedIn API are checked for citations.

One set of studies focused exclusively on tweets to PubMed documents, provided by Altmetric.com (Haustein, Peters, Sugimoto, Thelwall, Larivière, 2014; Thelwall, Haustein, Larivière, & Sugimoto, 2013; Haustein, Thelwall, Larivière, & Sugimoto, 2013).

Twitter. We gathered other twitter data from the Twitter API, in addition to the twitter data provided by Almetric.com. For one set of studies, we used the twitter API to identify all tweets that could be downloaded (n=68,232) from 37 astrophysicists (Haustein, Bowman, Holmberg, Peters, & Larivière, in press; Peters, Bowman, Haustein, & Holmberg, 2013). A similar approach was used to gather tweets from prolific researchers in astrophysics, biochemistry, digital humanities, economics, history of science, cheminformatics, cognitive science, drug discovery, social network analysis, and sociology (Holmberg & Thelwall, 2013a; Holmberg & Thelwall, 2013b; Holmberg & Thelwall, in press). We also constructed queries to generate a sample of tweets that included links to academic articles (see table below).

Source	Twitter query	Unique tweets
Wiley digital library	"onlinelibrary.wiley.com"	39,292
Science Direct digital library	"sciencedirect.com"	33,380
SpringerLink digital library	"springerlink.com"	17,515
JSTOR digital library	"jstor.org/stable"	1,862
Nature journal	"go.nature.com"	6,784
PLOS ONE journal	"plosone.org/article"	30,657
PNAS journal	"pnas.org/content"	11,756
Science journal	"scim.ag"	12,596
DOI links	"dx.doi"	5,234

The content of these tweets was used for a number of studies (Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013) and will be used for future studies.

Mendeley. Using a set of WoS articles, we extracted 219,326 corresponding records in Mendeley and Mendeley metrics (using the Mendeley API) (Mohammadi, Thelwall, Haustein, & Larivière, in press). We also matched PubMed articles to Mendeley, using the Mendeley API (Haustein, Larivière, Thelwall, Amyot, & Peters, in press).

RESULTS

As noted in our statement of significance, our goal was to "examine how different scientific activities enable or propel scientific discovery". We wanted to "analyze the importance of various scholarly activities for creating, sustaining, and propelling new knowledge", inform "our understanding of innovation, productivity, and impact", and "identify datasets that should no longer be marginalized, but built into our understandings and measurements of scholarship." The results of our two-year funded project have met these objectives by examining the relationship and role of various genres, examining new forms of impact, analyzing novel dissemination initiatives, and developing new methods and approaches.

Role of genres. Much of our work sought to understand the growing diversity of new forms of scholarly communication, the relationship among new and established genres, and the relationship between genres and disciplinarity. For example, our work in digital humanities suggested a new scholarly communication ecosystem, where neither journal article nor monograph are king (Bowman, Demarest, Weingart, Simpson, Larivière, Thelwall, & Sugimoto, 2013; Demarest, Weingart, Simpson, Larivière, Thelwall, & Sugimoto, in progress). We studied the role of preprints in astrophysics, find that the elapsed time between arXiv submission and journal publication has shortened, that the arXiv versions are cited more promptly and decay faster, and that the arXiv versions of paper shave lower citation rates than published papers (Larivière, Sugimoto, Macaluso, Milojevic, Cronin, & Thelwall, in press; Larivière, Macaluso, Milojevic, Sugimoto, & Thelwall, 2012; Larivière, Macaluso, Sugimoto, Milojevic, Cronin, & Thelwall, 2013). Our study of STS handbooks and journal articles showed a sharp distinction between the quantitative and qualitative traditions of STS and demonstrated that handbooks do not play a leading or summative role for this area of research (Milojevic, Sugimoto, Larivière, Thelwall, & Ding, 2013; Milojevic, Sugimoto, Larivière, Thewlall, & Ding, under review). We used dissertation abstracts to show the growing level of interaction among disciplines in the social sciences and humanities (Ni & Sugimoto, 2012).

Social media metrics. One new form of communication that we studied in-depth was Twitter, a microblogging platform. Our study demonstrated that this was the altmetric with the highest degree of coverage of scholarly material (Thelwall et al., 2013b). We examined a number of disciplines (i.e., astrophysics, biochemistry, digital humanities, economics, history of science, cheminformatics, cognitive science, drug discovery, social network analysis, and sociology) and the degree to which Twitter behavior varied among these groups of scholars (Holmberg & Thelwall, in press; Holmberg & Thelwall, 2013a; Holmberg & Thelwall, 2013b). Focusing on a small group of astrophysicists, we found a negative correlation between the number of publications and tweets per day (Haustein, Bowman, Holmberg, Peters, & Larivière, in press). We analyzed the use of hashtags in scholarly twitter conversations (Peters, Bowman, Haustein, & Holmberg, 2013), finding that less experienced users make use of more unique hashtags. We studied tweets containing links to scientific articles and found that many tweets provide little more than the title of the article (Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013), suggesting that tweets provide little indication of the reaction of the tweeters to the tweeted material. We also found that correlations between tweets and citation were low, implying that impact metrics based on tweets are different from those based on citations (Haustein, Peters, Sugimoto, Thelwall, Larivière, 2014). Furthermore, we noted the significance of time in these analyses (Thelwall et al., 2013b). Similarly, we found a difference in impact in Mendeley readership vs. tweets, showing that Mendeley was primarily an academic audience, whereas Twitter represented a more general public (Haustein, Larivière, Thelwall, Amyot, & Peters, in press). This was reinforced in a subsequent study of Mendeley which showed that Mendeley users are primarily doctoral and post-doctoral students (Mohammadi, Thelwall, Haustein, & Larivière, in press).

TED talks. We conducted a series of studies on TED talks, one of the largest science dissemination initiatives in history. The first study was a webometric analysis of the impact of TED talks (Sugimoto & Thelwall, 2013). We found that TED talks primarily impact the public, rather than the academic sphere. However, we found that TED talks were often found in online syllabi, demonstrating their value as pedagogical material. Finally, we noted that videos by academics were more favorably received that those by nonacademics. The second study pushed further into the analysis of the demographics of presenters. We found that presenters were primarily male and non-academic and that male presenters were received more favorably. We found that TED talks did not result in a citation advantage for academic presenters (Sugimoto, Thelwall, Larivière, Tsou, Mongeon, Macaluso, 2013). After finding the emphasis on male and non-academic presenters, we decided to analyze the reception of the videos as shown through the commenting behavior. Our analysis of comments revealed that commenters were more likely to discuss the characteristics of a presenter on YouTube, whereas commenters tended to engage with the talk content on the TED website. Furthermore, we found that commenters made more emotional comments about the presenter when the speaker was a woman (leaving comments that were either more positive or more negative) (Tsou, Thelwall, Mongeon, & Sugimoto, 2014). The current study on this topic is an analysis of the full transcripts of the TED talks, made available on and downloaded from the TED website.

Methodological studies. Given the novelty of many of the platforms studied, we often found ourselves in the process of developing new methods and approaches to data analysis. For example, we attempted many ways of gathering tweets about science, including keywords (Holmberg, 2013), scholar names (Haustein, Bowman, Holmberg, Peters, & Larivière, in press; Peters, Bowman, Haustein, & Holmberg, 2013), and journal-associated queries (Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013). We proposed the use of a sign test for studying correlations between citations and social media metrics (Thelwall, Haustein, Larivière, & Sugimoto, 2013; Haustein, Thelwall, Larivière, & Sugimoto, 2013). We

proposed a new approach to quantitatively measure the degree and nature of differences among disciplinary discourses using the social and epistemic terms used in texts (Demarest & Sugimoto, in press; Demarest & Sugimoto, 2013a; Demarest & Sugimoto, 2013b). The approach, called discourse epistemetrics, used a support-vector model of machine learning to classify disciplines based on the relative frequencies of social and epistemic terms. To examine the relationship between genres, we proposed a mixed LDA and regression analysis approach to study the lead-lag relationship between journal articles and pre-prints in astrophysics (Hu, Dong, Zhang, Bowman, Ding, Yan, & Milojevic, under review; Ding, Yan, Sugimoto, & Milojevic, 2013). We used webometrics to gather co-words of research areas from faculty webpages, rather than the traditional use of gathering keywords from journal articles to study the landscape of iSchools (Holmberg, Tsou, & Sugimoto, 2013).

PRODUCTION AND DISSEMINATION

Publications. This was an extremely productive team and sought a diversity of dissemination channels. Fourteen journal articles were accepted for publication during the two-year grant duration. Journal articles were accepted or published in seven different venues: *JASIST*, *Information Technology*, *ASLIB Proceedings*, *PLoS ONE*, *Scientometrics*, *Information Research*, *Cybermetrics*. It was important to the team that we engage in open access publishing; therefore, we sought gold open access publications (four of our works were made freely available online by the journal) and engaged in green open access practices (making our preprints available on our personal websites as well as repositories. We also sought to make our work available at conferences, to engage in wider discussions with the academic community. Ten refereed conference papers and three refereed conference posters were accepted and presented at nine different venues (from digital humanities to scientometrics).

Presentations. We gave invited talks at for instance World Social Science Forum (Montréal, Canada), Oxford Internet Institute (Oxford, UK) and at the National Science Foundation (Arlington, VA, USA). Our work was additionally presented 18 times at conference and other workshops (in the US, UK, Canada, and Germany). Finally, we hosted two methodological webinars that were made freely available to the public. The webinar on analyzing scholarly communication was attended by 48 participants and the presentation by Kim Holmberg on conducting twitter research was attended by 113. These were supported by the special interest group for metrics for Association for Information Science & Technology (ASIST) and are hosted on the ASIST website to provide continued access. Also supported by SIG/MET was a special bulletin issue on informetrics, to which the team heavily contributed. This provided an introduction, in accessible language, to the methods used in our group.

Software development. Many of the projects involved the extraction and merging of heterogeneous datasets. These were relatively large (several gigabytes in size) and involved the development of specific algorithms and software for data retrieval. During this research project the tools built at the University of Wolverhampton to collect, filter and analyze data from the web and social media were further developed. The tools developed were mainly added to the existing program Webometric Analyst and made available free online at http://lexiurl.wlv.ac.uk. In particular, functions were created to download tweets from the Twitter API. We had to write this code twice because the Twitter API changed substantially half way through the project. In addition to the data collection code, we also added data cleaning functions and functions to summarise the results by tweeter and tweetee and to create networks of tweeter-tweetee pairs. These functions were also added to an existing program, Mozdeh, which was designed for time series analysis of blogs. Mozdeh was converted with these extra functions to be a Twitter time series analysis program, and is available free at http://mozdeh.wlv.ac.uk. In

addition, additional Twitter time series capabilities were added to Mozdeh, including sentiment analysis and various types of co-word analysis as well as integrated spam filtering methods.

Communication. To inform people of our publications and other dissemination activities, we maintained a website (http://did.ils.indiana.edu/index.php) and twitter feed (DIDCascades). Our website contained descriptions of the project and personnel, a publication list (with preprints), and connected to our twitter feed. We tweeted 20 times and gained 21 followers on our twitter account.

Continued work. Dissemination did not end with the conclusion of the granting period. A number of other manuscripts are currently in progress: a paper on STS handbooks led by Stasa Milojevic, a paper on Digital Humanities led by Tim Bowman, and a paper on lead-lag analysis led by Ying Ding will all be submitted in the next few weeks. Work is also in progress in studying the conversational and interactional aspects of twitter for astrophysicists, the use of hashtags in tweets about journal articles, and the motivations for and meanings behind scholarly use of social media. We also have another webinar scheduled for May on social media metrics, and a workshop as part of the WebSci conference on big data analysis. We anticipate that the list of relevant publications for this project will continue to grow given the number of current projects within the team analyzing the data collected during the granting period.

IMPACT

The prolific publication list and international presentation schedule of the team speaks to the high degree of impact within the scholarly community.

However, the work also received coverage in the popular press, demonstrating the importance of the work to a larger audience. For example, our work on tweeting biomedicine was covered in *Bloomberg Businessweek* (For Scientists, More Tweets Don't Mean Better Citation Numbers), *Nature* (Twitter buzz about papers does not mean citations later), *Science* (More Tweets Don't Add Up to More Citations, Study Finds), and *the Chronicle of Higher Education* (Twitter's Value as Measure of Scientific Impact Encounters New Doubt). Our paper on TED talks was covered by *US News* (TED Conferences Most Inspired Thinkers Overwhelmingly Men), *the Chronicle of Higher Education* (Giving a TED Talk Expect More Visibility But Not More Citations), and the *Vancouver Sun* (TED Talks Presenters Will Male Domination Continue in Vancouver Videos and Conference).

Social media metrics provide another indicator of potential impact. For example, our paper on TED Talks published in PLoS ONE was viewed 8,528 at the time of writing and had been shared 271 times. The Almetric score for this showed that it had been tweeted by 203, blogged by 4, and had 20 readers in Mendeley. Our PLoS ONE paper on almterics had 8,611 views, 105 shares, 10 saves, and was cited five times. The Altmetric score showed that it had been tweeted by 164, blogged by 4, and had 105 readers on Mendeley.

LIST OF PUBLICATIONS

Refereed journal articles:

1. Mohammadi, E., **Thelwall, M.,** Haustein, S., **Larivière, V.** (in press). A multi-disciplinary analysis of research impact diversity with Mendeley users' occupations. *JASIST*.
2. Haustein, S., **Larivière, V.**, **Thelwall, M.**, Amyot, D., Peters, I. (in press). Tweets vs. Mendeley

readers: How do these two social media metrics differ? *IT – Information Technology*.

3. Haustein, S., Bowman, T.D., Holmberg, K., Peters, I., **Larivière, V.** (in press). Astrophysicists on Twitter: An in-depth analysis of tweeting and scientific publication behavior. ASLIB Proceedings.
4. Demarest, B., & **Sugimoto, C.R.** (in press). Argue, observe, assess: Measuring disciplinary identities and differences through socio-epistemic discourse. *JASIST*.
5. Tsou, A., **Thelwall, M.**, Mongeon, P., & **Sugimoto, C.R.** (in press). A community of curious souls: An analysis of commenting behaviour on TED Talks videos. *PLoS ONE.*
6. **Larivière, V.**, **Sugimoto, C.R.**, Macaluso, B., **Milojevic, S.**, Cronin, B., & **Thelwall, M.** (in press). arXiv e-prints and the journal of record: An analysis of roles and relationships. *JASIST*.
7. Haustein, S., Peters, I., **Sugimoto, C.R.**, **Thelwall, M.**, & **Larivière, V.** (in press). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *JASIST*.
8. Holmberg, K. & **Thelwall, M.** (in press). Disciplinary differences in Twitter scholarly communication. *Scientometrics*.
9. Holmberg, K., Tsou, A., & **Sugimoto, C.R.** (2013). The conceptual landscape of iSchools: Examining current research interests of faculty members. *Information Research, 18*(3), paper C32. Link
10. **Thelwall, M.,** Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting links to academic articles*, Cybermetrics*. 17(1), http://cybermetrics.cindoc.csic.es/articles/v17i1p1.html
11. Holmberg, K., Tsou, A. & **Sugimoto, C.R.** (2013). The conceptual landscape of iSchools: Examining current research interests of faculty members. *Information Research,* available at http://informationr.net/ir/18-3/colis/contents.html (Colis 2013 conference issue).
12. **Thelwall, M.**, Haustein, S., **Larivière, V.**, & **Sugimoto, C.R.** (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS ONE*, 8(5), e64841. http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0064841
13. **Sugimoto, C.R., Thelwall, M., Larivière, V.,** Tsou, A., Mongeon, P, & Macaluso, B. (2013). Scientists popularizing science: Characteristics and impact of TED Talk presenters. *PLoS ONE*, 8(4): e62403. http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0062403
14. **Sugimoto, C.R. & Thelwall, M.** (2013). Scholars on soap boxes: Science communication and dissemination via TED videos. *JASIST, 64(4), 663-674.*

Refereed conference papers:

1. **Milojevic, S.** (presenter), **Sugimoto, C.R., Larivière, V., Thelwall, M., & Ding, Y. (**2013). The role of handbooks in knowledge creation and diffusion: A case of science studies. Presentation at METRICS2013: Symposium on Informetric and Scientometrics Research. Montréal, Québec, November 2, 2013. [extended abstract]
2. Peters, I., Bowman, T.D., Haustein, S., & Holmberg, K. (2013). #twinkletweet: Hashtag use of astrophysicists on Twitter. Presentation at METRICS2013: Symposium on Informetric and Scientometrics Research. Montréal, Québec, November 2, 2013. [extended abstract]
3. Haustein, S., **Thelwall, M.**, **Larivière, V.**, & **Sugimoto, C.R.** (2013). On the relation between altmetrics and citations in medicine. *STI 2013*.[work in progress paper]
4. Holmberg, K. (2013). Discovering scholarly communication on Twitter. *WIS & Collnet meeting 2013*. [full paper]
5. **Larivière, B.**, Macaluso, B., **Sugimoto, C.R., Milojevic, S.**, Cronin, B., & **Thelwall, M.** (2013). The nuanced nature of e-print use: A case study of arXiv. Proceedings of the *International Society for Scientometrics and Informetrics* conference. [full paper]
6. **Ding, Y.**, Yan, E., **Sugimoto, C.R., & Milojevic, S.** (2013). Lead-lag topic analysis: e-prints vs. journal articles in astrophysics. Proceedings of the *International Society for Scientometrics and Informetrics* conference. [short paper]

7. Holmberg, K. & **Thelwall, M**. (2013). Disciplinary Differences in Selected Scholars' Twitter Transmissions. Proceedings of the *ASIS&T European Workshop 2013*.
8. Holmberg, K. & **Thelwall, M.** (2013). Disciplinary differences in Twitter scholarly communication. Proceedings of the *International Society for Scientometrics and Informetrics* conference. [full paper]
9. Bowman, T.D., Demarest, B., Weingart, S.B., Simpson, G.L., **Larivière, V., Thelwall, M.**, **Sugimoto, C.R.** (2013). Mapping DH through heterogeneous communicative practices. Digital Humanities 2013, Lincoln, Nebraska. [short paper]
10. **Larivière, V.**, Macaluso, B., **Milojevic, S.**, **Sugimoto, C.R.**, & **Thelwall, M.** (2012). On the scientific impact of ArXiv: A case study of astrophysics. *METRICS2012: Symposium on Informetric and Scientometrics Research* (part of the ASIS&T Annual Meeting). [short paper] Pre-print

Refereed conference posters:

1. Demarest, B.F., & **Sugimoto, C.R.** (2013). Interpreting epistemic and social cultural identities of disciplines with machine learning models of metadiscourse. *Proceedings of the International Society for Scientometrics and Informetrics conference*. [short paper]
2. Demarest, B. & **Sugimoto, C. R.** (2013). Using machine learning models to interpret disciplinary styles of metadiscourse in dissertation abstracts. *iConference 2013 Proceedings* (pp. 901-904). doi:10.9776/13459
3. Ni, C., & **Sugimoto, C.R.** (2012). Using doctoral dissertations for a new understanding of disciplinarity and interdisciplinarity. *Annual Meeting of the American Society for Information Science and Technology*. Baltimore, MD. October 26-30, 2012. [short paper]

Conference presentations and panels:

1. Haustein, S., **Larivière, V.** (2014). Astrophysicists on Twitter and other social media metrics research. Harvard-Smithsonian Center for Astrophysics, Harvard University. 7 février 2014.
2. Bar-Ilan, J.,**Milojevic, S.**, Gunn, W., Haustein, S., Konkiel, S., **Larivière, V.**, Li, J. (2013). Altmetrics: Present and Future (SIG/MET). Panel at the 76th ASIS&T Annual Meeting, Montréal, Québec, November 1-5 2013. [panel]
3. Haustein, S., **Larivière, V.** (2013). Les nouveaux indicateurs de l'impact social de la recherche. World Social Science Forum, Montréal, October 14, 2013.
4. **Larivière V., Sugimoto, C., Ding, Y., Milojević, S**., Holmberg, K., **Thelwall, M.** (2013). An update on DID. Cascades, Islands, or Streams? Time, Topic and Scholarly Activities in Humanities and Social Science Research. Digging into Data, World Social Science Forum., Montréal, October 12 2013
5. Haustein, S., Bowman, T.D., Holmberg, K., **Larivière, V.**, Peters, I., **Thelwall, M.**, **Sugimoto, C.R.** (2013) Empirical analyses of scientific papers and researchers on Twitter: Results of two studies. PLOS Article-Level-Metric workshop http://article-level-metrics.plos.org/alm-workshop-2013/, San Francisco, October 10-11 2013.
6. Mohammadi, E., **Thelwall, M.**, **Larivière, V.,** Haustein, S., (2013). Mendeley Readership Altmetrics for Clinical Medicine and Engineering. Article-Level-Metric workshop http://article-level-metrics.plos.org/alm-workshop-2013/, San Francisco, October 10-11 2013.
7. Haustein, S., **Larivière, V.** (2013). Empirical Analysis of Social Media in Scholarly Communication Overview of current altmetrics research projects at University of Montréal. GESIS, Leibniz-Institut für Sozialwissenschaften, Cologne, Allemagne, September 2, 2013.

8. Joinson, A., **Thelwall, M.**, Sessions Goulet, L., Ellison, N., Boyd, D., & Contractor, N. (2013). Methodological Opportunities and Challenges in the Age of Social Media and "Big Data": Beyond the Survey. International Communication Association, London. [panel]
9. **Larivière, V.,** Macaluso, B., & **Milojevic, S.** (presenter), **Sugimoto, C.R.**, & **Thelwall, M.** (2012). Of caterpillars and butterflies: The life and afterlife of an arXiv e-print. *altmetrics12: ACM Web Science Conference 2012 Workshop*. Evanston, IL. June 21, 2012. [short paper]
10. Bobley, B. (moderator), Dempster, S., Ewing, E.T., Henry, C., Larson, R., Serventi, J., **Sugimoto, C.R.**, Thomas, C., & Tran, E. (2012).The Digging into Data Challenge: A Roundtable Discussion. *JCDL 2012 - ACM/IEEE - CS Joint Conference on Digital Libraries*. Washington, DC. June 10-14, 2012. [panel]
11. **Sugimoto, C.R.**, **Ding, Y.**, & **Thelwall, M.** [proposers] (2012). Library and Information Science in the Big Data Era: Funding, Projects, and Future. With panelists: Richard Marciano, **Vincent Larivière**, Michael Khoo, and Stephen Downie. *Annual Meeting of the American Society for Information Science and Technology*. Baltimore, MD. October 26-30, 2012. [short paper]
12. **Sugimoto, C.R.** (presenter), Cronin, B., **Ding, Y**., **Larivière, V.**, **Milojević , S.**, & **Thelwall, M.** (2012). Toward a new model of scholarly communication. *Social Science and Digital Research: Interdisciplinary Insights*. University of Oxford, UK. March 12, 2012. [abstract]

Other invited talks:

1. **Larivière V., Sugimoto, C.R., Ding, Y., Milojević, S.**, Holmberg, K., & **Thelwall, M.** (October, 2013). An update on DID. Cascades, Islands, or Streams? Time, Topic and Scholarly Activities in Humanities and Social Science Research. Digging into Data, World Social Science Forum, Montréal, Canada.
2. Haustein, S. (November, 2013). Disciplinary differences and other biases: Exploring social media metrics in scholarly context. NISO Webinar: New Perspectives on Assessment How Altmetrics Measure Scholarly Impact.
3. Haustein, S. (October, 2013). Exploring disciplinary differences in the use of social media in scholarly communication. Lightning talk at the NISO Altmetrics Project - In-person Meeting, San Francisco, CA.
4. **Larivière, V.** (May, 2013). Panel on Big Data. Conférence Canada 3.0, Toronto, Canada.
5. Haustein, S.,& **Larivière, V**. (September, 2013). Empirical Analysis of Social Media in Scholarly Communication Overview of current altmetrics research projects at University of Montréal. GESIS, Leibniz-Institut für Sozialwissenschaften, Cologne, Allemagne.
6. **Sugimoto, C.R**. (May, 2012). "Transformations in the scholarly communication system: a Big Data perspective." National Science Foundation, Arlington, VA.

Other publications:

1. **Larivière, V.**, Haustein, S. (2014) Science et médias sociaux: décoder le vrai du buzz. Découvrir. Le magazine de l'ACFAS. Février 2014. http://www.acfas.ca/publications/decouvrir/2014/02/science-medias-sociaux-decoder-vrai-buzz
2. **Larivière, V.**(2012). The decade of metrics? Examining the evolution of metrics within and outside LIS. *ASIST Bulletin, 38*(6) Aug-Sept 2012, p.12-17
3. **Milojević , S. & Sugimoto, C**.(2012). Metrics and ASIS&T: Introduction. *ASIST Bulletin, 38*(6) Aug-Sept 2012, p.9-11
4. **Sugimoto,C.** (2012). Taking the measure of metrics: Interviews with four ASIS&T members. *ASIST Bulletin, 38*(6) Aug-Sept 2012, p. 33-38.
5. **Thelwall, M.** (2012). A history of Webometrics. *ASIST Bulletin, 38*(6) Aug-Sept 2012, p. 18-23.

Webinars:

1. **Ding, Y.**, & Lin, L. (2013). Analyzing scholarly communication using topic modeling methods. August 22, 2013.
2. Holmberg, K. (2013). Conducting twitter research. December 24, 2013.