

Institution: University of Wolverhampton		
Unit of Assessment: 34 Communication, Cultural and Media Studies, Library and Information Management		
Title of case study: Software and Consultancy to Support the Information Industry		
Period when the underpinning research was undertaken: 2001 - 2020		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Professor Michael Thelwall	Professor of Data Science	1989 - Present
Dr Kayvan Kousha	Postdoctoral Researcher	2012 - Present
Dr Amalia Mas-Bleda	Postdoctoral Researcher	2015 - Present
Tamara Nevill	Senior Lecturer	2010 - Present
Period when the claimed impact occurred: 2014 - 2020		
Is this case study continued from a case study submitted in 2014? N		
1. Summary of the impact <p>Information industry actors have developed improved commercial tools and services, supporting the UK to become a world leader in one digital publishing sector with software, consultancy and indicators from the Statistical Cybermetrics Research Group (SCRG). Specifically, our methods and findings helped London-based Digital Science create research impact indicators and alternative impact indicators (altmetrics) for scholarly publications for their world-leading international products, Dimensions.ai and Altmeter.com. We helped UK non-profit Jisc evaluate the potential to develop commercial tools for the publishing industry and academia. We enabled international commercial software developers to deploy integrated social media analysis products through our sentiment analysis software.</p>		
2. Underpinning research <p>Since 2001, the SCRG has developed and evaluated web-based indicators for the impact of academic research and a systematic approach for applying them to evaluation contexts. We are world leaders in the number of articles published in the top journals in the field. Initially this was theoretical research, but web indicators later became commercialised in the UK and USA. We have also conducted scientometric research into academic databases and indicator robustness, investigations into academic publishing and sentiment analysis, all of which has supported impact.</p> <p>All the underpinning research in this section has the main author from the SCRG.</p> <p><u>Altmetrics1: Validated methodologies for applying altmetrics</u></p> <p>Before the founding of the first altmetrics business in 2011, Thelwall contributed more research into altmetrics (or webometrics, its earlier name) than anyone else. Of the 87 journal articles in Scopus on the topic, 20 (23%) were authored by Thelwall, with the next most common author (Ortega) authoring 12. This can be checked with the Scopus advanced query: TITLE-ABS-KEY(webometrics OR altmetrics) AND PUBYEAR < 2011 AND DOCTYPE(ar). Our integrated approach to web indicator applications developed over fifteen years is summarised in a 2017 book [R1], showing how to use altmetrics to assess groups of researchers.</p> <p><u>Altmetrics2: Web syllabus mentions are an educational impact indicator</u></p>		

Counts of citations to research from web-based syllabi were developed and shown by Kousha to be a practical and valid indicator of the educational impact of journal articles with an empirical correlation-based study and theoretical argument that a citation from a syllabus is evidence of use within education [R1].

Altmetrics3: Mendeley reader counts are robust academic impact indicators

Counts of readers in the social reference sharing site Mendeley were demonstrated by the SCRG to be robust early indicators of academic impact. We co-authored (with former PhD student Xuemei Li and librarian Dean Giustini) the first refereed publication on this indicator and have evaluated its capabilities in a series of follow-up studies, showing primarily through strong correlations with long term citations that it is a robust (early) indicator of academic impact [R1].

Altmetrics4: Altmetrics can predict long term citations

Thelwall and Nevill's large-scale systematic evaluation [R2] showed for the first time that long term citations could be predicted from early altmetric scores, enhancing the commercial value of altmetric data with this proven capability.

Scientometrics1: MNLCS is robust for citation count data

The Mean Normalised Log-transformed Citation Score (MNLCS) is our impact indicator formula that produces a single score for a set of research outputs that has a value above 1 if the average citation (or other) impact of the outputs is above average for the world, and less than one if it is below world average [R1]. These papers showed that indicators based on log-transformed citations or altmetrics are far less influenced by individual high scores and give more precise impact estimates than current standard methods, a finding of major significance for research impact assessment, which previously always used less precise measures.

Scientometrics2: Dimensions.ai has similar coverage to Scopus

A systematic evaluation of the commercial academic database and search service Dimensions.ai in comparison to Scopus was its first published scholarly evaluation, showing that its coverage of the academic literature is comparable to Scopus and hence probably wider than the Web of Science [R3], making it a credible alternative major academic database.

AcademicPublishing1: Research data is rarely shared, even in ideal conditions

Kousha developed methods to systematically evaluate academic data sharing in digital repositories, with a case study of FigShare. We built on this with a contracted study for Jisc developing and applying a new method to assess the potential for automated data quality assurance in Genomics [R4]. The study showed that research data is rarely shared, even in fields that apparently provide ideal circumstances for sharing.

AcademicPublishing2: Open peer review has national but not subsequent-reviewer biases

A contracted study for Jisc developed and applied a method to test for reviewer bias in open peer review [R5]. We found that open publication of reviews before all reports are in did not seem to induce bias, whereas same-country reviewer bias was occasionally present.

Sentiment1: SentiStrength has human-level sentiment strength detection accuracy

Our sentiment analysis software SentiStrength produces parallel positive and negative sentiment strength estimates for texts. It is designed for, and evaluated on, social web texts, and it is user-customisable, fast (14,000 tweets per second), and accurate [R6]. It is highly cited (Google Scholar: 1,700+) for its broad applicability.

3. References to the research

All the underpinning research outputs have been through rigorous peer-review in Scopus-indexed recognised journals from the host discipline except for a book in a peer-reviewed academic series, summarising and citing relevant peer reviewed SCRG research [R1]. R1 has 64 Google Scholar citations and R6 has 1,005, indicating wide academic use.

R1. Thelwall, M. (2017). *Web indicators for research evaluation: A practical guide*. San Rafael, CA: Morgan & Claypool. (<https://doi.org/10.2200/S00733ED1V01Y201609ICR052>) [this book is the culmination of the Cybermetrics Group's research into online indicators]

R2. Thelwall, M. & Nevill, T. (2018). Could scientists use Altmetric.com scores to predict longer term citation counts? *Journal of Informetrics*, 12(1), 237–248. (<https://doi.org/10.1016/j.joi.2018.01.008>) (REF 2 Output)

R3. Thelwall, M. (2018). Dimensions: A competitor to Scopus and the Web of Science? *Journal of Informetrics*, 12(2), 430–435. (<https://doi.org/10.1016/j.joi.2018.03.006>)

R4. Thelwall, M. Munafò, M., Mas-Bleda, A., Stuart, E., Makita, M., Weigert, V., Khan, N., Drax, K. & Kousha, K. (2020). Is useful research data usually shared? An investigation of genome-wide association study summary statistics. *Plos One*, 15(2): e0229578. (<https://doi.org/10.1371/journal.pone.0229578>)

R5. Thelwall, M., Allen, L. Papas, E., Nyakoojo, Z., & Weigert, V. (in press). Does the use of open, non-anonymous peer review in scholarly publishing introduce bias? Evidence from the F1000Research post-publication open peer review publishing model. *Journal of Information Science*. (<https://doi.org/10.1177/0165551520938678>)

R6. Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1), 163-173. (<https://doi.org/10.1002/asi.21662>)

4. Details of the impact

The SCRG contributed to innovation impact on commerce and the economy by providing a wide range of different types of direct support for the academic information industry in the UK and internationally, described below by type. Most impacts are achieved through direct support, but the first is indirect and collaborative in the form of establishing a research area that was subsequently exploited by start-up companies to create a new type of information industry business.

11. Establishment of altmetrics as credible for commercialisation

Altmetric providers were able to establish successful businesses through the SCRG's establishment of the credibility of altmetrics and webometrics. The SCRG led research into the development of altmetrics and webometric indicators for the impacts of academic research [Altmetrics1, Altmetrics2, Altmetrics3]. This research was commercialised from 2011, primarily through the creation of Plum Analytics (2011) and London-based start-up Altmetric.com (2011). For example, Altmetric.com's business model is based on (a) the practicality of systematically gathering altmetric indicators, and (b) the value of those indicators to end users, for both of which we played the largest role in terms of the number of refereed journal articles [Altmetrics1]. We claim impact from 2014, with our research enabling them to be credible to major academic publishers. Those signing up to Altmetric.com include Wiley (2014), Springer (2014), Taylor & Francis (2015), and the JAMA network (2017), collectively publishing 7,200 journals. "Altmetrics research from Thelwall's team helped make Altmetric.com's indicators credible to academic publishers, increasing our commercial viability" (Adie, Altmetric.com founder).

12. Development and evaluation of altmetric indicators

Altmetric.com was able to expand their indicator portfolio with new indicators pioneered by us. These include Mendeley reader counts (joint), online syllabus mentions and Wikipedia citations [Altmetrics2, Altmetrics3], which are now part of Altmetric.com's portfolio [C1]. These indicators address additional impact dimensions that enhanced the value of Altmetric.com's product. Altmetric.com were also able to demonstrate the value of their offering to potential customers through our systematic (non-commissioned, independent) study illustrating that Altmetric.com data could help predict longer term citation impact [Altmetrics4]. Thelwall gave an invited talk to an Altmetric.com webinar about this research [C2], evidencing the value that Altmetric.com placed on getting this message to their audience. The international scholarly community as a whole has benefitted from the altmetrics provided by Altmetric.com and other providers in academic literature databases (e.g., Dimensions.ai) and publisher websites, underpinned by our research, as early indicators of interest in academic publications.

In 2017 Canadian company TrendMD used the Mendeley readership indicator validated by us [Altmetrics3] as an early impact indicator to demonstrate the value of its biomedical research publicity tool, increasing their product reputation [C3].

13. Citation analysis and evaluations of Dimensions.ai

London-based Dimensions.ai (2018) developed our world leading citation impact indicator for its citation index, taking into account the citation data skewing that undermines almost all other commercial citation-based indicators [Scientometrics1], which it calls the *Field Citation Ratio* [C4]. This increased Dimensions' viability as a challenger to Scopus and the Web of Science.

Dimensions.ai become a viable alternative to the Web of Science and Scopus with the help of our published research, demonstrating that the coverage of Dimensions.ai was comparable to Scopus and hence larger than that of the Web of Science [Scientometrics2]. Thelwall gave public talks in 2018 invited by Digital Science (owners of Dimensions), online and in person, about this research, and the article was cited by a technical report by the creators of Dimensions.ai [C5]. These show the value Digital Science placed on our findings.

14. Commissioned academic publishing studies

The UK non-profit organisation Jisc evaluated the potential to develop commercial tools for the publishing industry and academia with our commissioned, Jisc-funded 2019 project. Thelwall, Kousha and Mas-Bleda used experiences of researching data sharing to research the viability of developing software to identify and quality validate data of specific types underpinning journal articles shared by authors. This project, which was research-based with its findings published as a refereed journal article, showed that non-mandatory data sharing was too rare to make such a tool viable, even for fields with apparently ideal conditions for sharing [AcademicPublishing1; the commissioned co-authored output R4 is the impact evidence]. "The evidence from the study helped Jisc to decide not to invest in automated data sharing detection tools at this stage and continues to inform discussions around future Jisc open research services and tools." (Weigert, Jisc).

Jisc was able to assess the potential for automated peer review, highlighting ethical issues but showing that there was some scope for automated peer review support, but not for fully automated peer review through our commissioned report [AcademicPublishing2]. These areas were identified by Jisc as topics that it should scope for UK-wide support and advice, with the findings enabling Jisc to advise the sector and avoid funding impractical software development. The reports form the evidence that the requested information was delivered [AcademicPublishing1, AcademicPublishing2], and Jisc used our evidence for sector advice [C6, C7].

15. Sentiment analysis software applications

Commercial software developers and artists were able to enhance their products and artworks by embedding our sentiment analysis software [Sentiment1], *SentiStrength*, into their integrated social media analytics products, enhancing their value [C8]. An artist embedded public emotions into her sculpture [C9], generating cultural impact. Our sentiment analysis software sold seven commercial licences between 2014 and 2020.

Verizon ran a high-profile cost-effective marketing campaign, saving a claimed USD4 million around Super Bowl 2014 with the help of SentiStrength [Sentiment1]. They ran a daily competition in the runup to the Super Bowl between the competing teams' fans, with the winners getting their team's colours in an evening lightshow at the top of the Empire State building. Fans tweeted for or against their teams or the opponents, with sentiment detected by SentiStrength [C10]. This initiative had a wide reach, with a Super Bowl 2014 audience of 111 million.

5. Sources to corroborate the impact

C1. Altmetric.com (2018) What outputs and sources does Altmetric track? (<https://web.archive.org/web/20201206040349/https://help.altmetric.com/support/solutions/article/s/6000060968-what-outputs-and-sources-does-altmetric-track->) and Altmetric.com (2016). What do syllabi-based altmetrics actually mean? Altmetric.com (<https://www.altmetric.com/blog/syllabi-altmetrics-meaning/>).

C2. Altmetric.com (2018). Altmetrics for impact prediction and benchmarking: the latest research. Webinar (On YouTube <https://www.youtube.com/watch?v=YZLIZfVVLm> and <https://www.altmetric.com/events/altmetrics-for-impact-prediction-and-benchmarking-the-latest-research/>).

C3. Kudlow, P., Cockerill, M., Toccalino, D., Dziadyk, D. B., Rutledge, A., Shachak, A., et al. (2017). Online distribution channel increases article usage on Mendeley: a randomized controlled trial. *Scientometrics*, 112, 1537–1556 (<http://doi.org/10.1007/s11192-017-2438-3>).

C4. Dimensions (2019). What is the FCR? How is it calculated? (<https://dimensions.freshdesk.com/support/solutions/articles/23000018848-what-is-the-fcr-how-is-it-calculated>) (Mon, 11 Nov, 2019).

C5. Hook, D., Porter, S., & Herzog, C. (2018). Dimensions: Building context for search and evaluation. *Frontiers in Research Metrics and Analytics*, 3, 23 (<https://doi.org/10.3389/frma.2018.00023>).

C6. Thelwall, M., Papas, E., Nyakoojo, Z., Allen, L. & Weigert, V. (2020). Automatically detecting open academic review praise and criticism. *Online Information Review*. 44(5), 1057-1076 (<https://www.doi.org/10.1108/OIR-11-2019-0347>)

C7. Weigart, V. (2020). Three ways to combat peer review bias (<https://www.jisc.ac.uk/blog/three-ways-to-combat-peer-review-bias-13-jan-2020>).

C8. Databahn.com. Sentiment Analysis API demo video by databahn - Part 3 of 3 (<https://www.youtube.com/watch?v=-1tRKDh3H-Y>).

C9. Mika Tajima. Human Synth (<http://mikatajima.com/human-synth/>)

C10. Forbes (2014). Verizon's Super Bowl Scheme Is To Save \$4 Million And Light Up The Sky (<https://www.forbes.com/sites/darrenheitner/2014/01/30/verizons-super-bowl-scheme-is-to-save-4-million-and-light-up-the-sky/>)