

Social Media Metrics of Publishing Impact

Subject areas: database design & management, data visualization, academic publishing, social media, data analytics

Background

Scientific research is typically presented in the form of manuscripts published in scholarly journals. The impact and reach of a published manuscript is typically measured by the number of subsequent papers that cite it as a reference. Increasingly, this construct is seen as too narrow, and not truly representative of the overall reach of any particular paper. Counting the number of times a paper is referenced may underestimate its impact, if it is widely discussed in the Twittersphere, blogosphere, or other social media channels, or if it is mentioned in traditional media reports. Conversely, an article's impact may be overstated if it is referenced repeatedly by small, niche journals, but never discussed in a broader context.

Recently, alternative metrics for evaluating the impact of scholarly publications have been developed. Chief amongst these is [Altmetric](#), a company that specializes in aggregating article mentions in social media, and by traditional news outlets. Altmetrics may provide a more relevant and balanced view of a journal article's reach, and could be of great use to authors and publishers alike. Moreover, Altmetrics are of growing interest to researchers working in the area of "meta-research" — the study of how research is funded, conducted, and reported.

This project focuses on unlocking the immense potential of Altmetric by accessing their data repository, and cross-referencing records with additional widely used databases. Such complementary datasets include [Pubmed](#) — a registry of papers published in the biomedical literature; [clinicaltrials.gov](#) — a repository of records related to clinical trials; and [Retraction Watch](#) — a watchdog organization that monitors the published scientific literature for erroneous, fraudulent, or otherwise compromised reports.

Resources

- Altmetric provides an API (by request) that enables access to their database, which just in the last year has collected over 17 million mentions of 2.7 million different research outputs.
- Pubmed can be accessed directly, or through Entrez Programming Utilities ([E-utilities](#)), which provides server side tools for efficient query and retrieval.
- clinicaltrials.gov contains data on more than 250,000 studies. These can be [retrieved in bulk](#) as XML files for further parsing and analysis.
- Retraction Watch maintains a [database](#) of records containing detailed information
- The Queen's-affiliated Centre for Advanced Computing provides computing and storage infrastructure to support academic work at Queen's, such as the project described herein.

Deliverables

The goal of this project is to develop a comprehensive, searchable data environment merging Altmetric data with one or more of the above mentioned additional data types. The database should focus us:

- Ease and flexibility of queries
- User interface (web-based)
- Data visualizations (for aggregate data)
- Cross-referencing between data sources
- Download of individual records or groups of records

Use cases:

The merged data query and visualization tool developed in this project will support cutting edge meta-research in how scientific studies are published and covered in the media. User groups will include a small number of researchers at Queen's, as well as a much larger research community that will be connected to the resource via outreach through existing channels and professional contacts. Researchers will be able to ask and answer important questions about how research is designed, the reach and influence of discredited studies, the links between funding success and media coverage, the influence of industry sponsorship in research, and numerous other important issues.

Architecture:

Development would preferably be in python/Django using an SQL server in order to integrate with other applications in use and in development by the lab. Alternative methods will be explored in the early stages of the project.

Contacts:

Dr. David Maslove, Dept of Critical Care (david.maslove@queensu.ca)
www.conduitlab.org