

# Aggregated Erevnametrics: bringing together alt-metrics through Research Objects

Patrick McSweeney, Rikki Prince, Charlie Hargood, David E. Millard and Leslie A. Carr  
{pm5, rfp07r, cah07r, dem, lac}@ecs.soton.ac.uk  
ECS, University of Southampton, SO17 1BJ.

*This position paper examines the spectrum of artefacts that contribute to the research process, in the context of bibliometrics and errevnametrics<sup>1</sup>. We examine how existing research evaluation techniques could be enhanced with richer information about the research process and research outputs.*

Currently the key metric for ranking research is citation count. This contributes to journal impact factor, or “JIF” (Garfield, 1972), and by proxy author impact factor (AIF). These measures are primarily of benefit to the journals themselves and the research councils looking to fund the best researchers. However, due to the time delay before they become useful measurements (in the order of years from when a paper is published), they are not a useful measure of whether the research is worth examining at the immediate time of publication.

To determine the true value of modern research we need to analyse more than just the papers produced; we need to be able to measure and aggregate the impact of all output produced and how it is reused in future research. The JISC programme entitled “Managing Research Data” (MRD)<sup>2</sup> investigated this topic of “data citation”, and in 2010, Bechhofer et al proposed a new kind of publication: one that is an augmented aggregation of the resources used to produce the research. This publication has been dubbed a “Research Object”. In this paper we explore the benefits Research Objects offer to calculating existing metrics. We then look at which metrics are useful to researchers and how Research Objects enable these metrics to be calculated in future.

To frame our discussion let us consider an example of a piece of computer science research, which has produced a piece of software, conducted a user trial that created some usage data, a research paper submitted to a conference, the presentation slides used, a video recording of their presentation at the conference and links to the blog posts the research team made during the project. We now examine the metrics that would be available for each of these artefacts and how might they inform the stake holders in the research process.

## Benefits to calculating existing metrics

The semantically connected Research Object gives rich and machine readable information about research at its point of release. Papers written and metadata about the papers it cites are among the items collated. This makes processing the citations within a Research Object a reliable machine processable task. This radically reduces the cost and potentially increases the accuracy of calculating citation count when compared to calculating it by hand. The increased ease and decreased cost of calculating citation count also means a much broader range than the current 85% (Brody, 2006; Thompson ISI, 2004) could be calculated. This could then be used to calculate JIF and AIF more accurately.

## Existing metrics applied to Research Objects

As well as enhancing the metrics we already use, Research Objects give us an opportunity to re-purpose existing metrics in a new context. For example, published research data can be cited as a data source. This could lead to data, presentations and other components of the Research Object each having a citation count. These citations, when aggregated, are a Research Object’s citation count or Research Objects Impact Factor (ROIF). To a researcher looking for data to augment, re-evaluate or compare, the citation count of the data is far more important than the citation count of papers analysing that data.

---

<sup>1</sup>research metrics

<sup>2</sup><http://www.jisc.ac.uk/whatwedo/programmes/mrd/clip/>

## New metrics enabled by Research Objects

Static analysis of computer source code produces metrics such as measures of cohesion and coupling, Martin's (2002) software package metrics and Halstead's complexity measures (1977) which give a researcher some insight into the quality of a piece of code and how understandable it will be to someone looking to build on it. If this code is hosted on a suitable open source repository (e.g. Sourceforge or Github), this gives another set of metrics we can harvest, such as the number of open and closed bugs, the number of projects forking from this project and the number of commits.

Datasets in a Research Object can be rated on Berners-Lee's (2010) Open Linked Data scale. This ranges from 1 star for any format of data, up to 5 stars for data that is structured, non-proprietary, uses URLs and links to other data. This metric gives researcher valuable indication of whether the data is in a format which supports reanalysis or merging.

A video of the conference presentation hosted on a video/slide sharing website (such as SlideShare or YouTube), will have a count the number of times the presentation has been viewed, commented on and marked as favourite.

## Conclusion

Research Objects are a model for the potential future of research publishing which offers a lot of opportunity for evaluation metrics. We have evaluated the benefits Research Objects offer to calculating existing metrics but note that currently these metrics are focused on enabling funding councils to measure the impact of research. We also discuss potential metrics which may become available as a mechanism of evaluating a broader range of research artefacts available. These new metrics offer value to the researcher looking for good research to build on, not just as mechanism for funding councils to rank research. These new metrics offer an opportunity to improve scientific practices while also enabling funding councils to offer more flexible or specialised fund calls.

Further more by establishing a publication method of a structure of related research objects encourages the release of research artefacts beyond the traditional paper. This allows for a greater spread of alternative metrics to be measured and potentially facilitates their aggregation into a central space where all metrics relevant to a particular piece of research can be gathered. This way traditional metrics and alternative metrics may be used together to get a fuller picture of the characteristics of a piece of research.

## References

Bechhofer, S., Ainsworth, J., Bhagat, J., Buchan, I., Couch, P., Cruickshank, D., Roure, D. D., Delderfield, M., Dunlop, I., Gamble, M., Goble, C., Michaelides, D., Missier, P., Owen, S., Newman, D., & Sufi, S. (2010). Why linked data is not enough for scientists. In *2010 IEEE Sixth International Conference on e-Science*, (pp. 300-307).

URL <http://dx.doi.org/10.1109/eScience.2010.21>

Berners-Lee, T. (2010). Open, Linked Data for a Global Community. In Gov 2.0 Expo, Washington, DC, May 25-27, 2010.

URL <http://www.gov2expo.com/gov2expo2010/public/schedule/detail/14247>

Brody, T. (2006). *Evaluating Research Impact through Open Access to Scholarly Communication*. Ph.D. thesis, University of Southampton.

URL <http://eprints.ecs.soton.ac.uk/13313/>

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science* , (178), 471-479.

URL <http://www.garfield.library.upenn.edu/essays/V1p527y1962-73.pdf>

Halstead, M. H. (1977). *Elements of Software Science (Operating and programming systems series)*. New York, NY, USA: Elsevier Science Inc.

URL <http://portal.acm.org/citation.cfm?id=540137>

Martin, R. C. (2002). *Agile Software Development, Principles, Patterns, and Practices* . Prentice Hall, 1st ed.

URL <http://www.worldcat.org/isbn/0135974445>