Reviewing repository discoverability: approaches to improving repository visibility and web impact

George Macgregor, Institutional Repository Coordinator, University of Strathclyde
Web: https://purl.org/g3om4c | Twitter: @g3om4c | ORCID: 0000-0002-8482-3973

Introduction

Resource developing and exposing rich digital collections using a variety of repository technologies has been invested. In the UK this investment has resulted in unprecedented usage of institutional repositories, as evidenced by data from IRUS-UK (1). However, not all institutions have demonstrated commitment to exposing this scholarly content as optimally as possible, or rendering their repository as usable as possible. It is clear that many repositories have not enjoyed maintenance beyond establishment of the repository itself and the ongoing development of its scholarly collection. Such institutions may work hard to promote their repository content but if little is done to optimise for discovery these repositories may remain relatively unexposed (2). A significant future challenge for repositories, and the Open Access movement more generally, is therefore to ensure user expectations are better met and, in so doing, improving the index penetration of the content they wish to expose.

This poster reviews work recently undertaken on Strathprints (3), the University of Strathclyde institutional repository built on EPrints 3.3.13, to improve repository web visibility and user engagement, thereby improving usage. Expanding on previous brief reports (4) and using Strathprints as a case study, a summary of the approach adopted is provided, comparative search traffic data and usage metrics are analysed and conclusions drawn. Results are likely to positively inform repository practitioners and open scientists.

Background

To improve the usage and user engagement of Strathprints, a series of technical improvements were made to Strathprints in spring 2016 and their impact monitored during 2016/2017. Process improvements were also implemented. The principal improvements made included:

- Implementation of a refreshed Strathprints user interface (UI). Many repositories continue to demonstrate low levels of usability (5,6). Heuristic evaluation (7) of the Strathprints user interface (UI) was undertaken in early 2016 to direct UI changes intended to improve usability and user engagement [Fig.1];
- A “mobile first”, responsive re-engineering of Strathprints thus triggering important signals in PageRank and, later, heavier weighting in the “Penguin” updates (8);
- “White hat” improvements to the way Strathprints functions plus support for the Core Recommender and AltMetric API, both of which refer users to additional Strathprints content and promote user interactions;
- Implementation of a crawler friendly file-naming conventions for full-text deposits and improvements to search engine “friendliness”;
- Gradual cleaning of broken links within Strathprints thereby improving the “content health” of Strathprints and, again, triggering important signals in PageRank;
- Improved integration with social tools, including growth in social interactions which are the result of Tweets about recently deposited Strathprints content;
- Improve the repository digital object to metadata ratio. Since Strathprints exists in a hybrid IR/CRIS environment, content is fed from an institutional CRIS to Strathprints. Implementation of a “connector-lite” configuration was actioned to cultivate Strathprints as a full-text destination for users and machines alike (9).

Data and results

The impact of repository improvements was monitored and measured using a variety of metrics, including search traffic data from Google Search Console, COUNTER usage data from IRUS-UK and IRStats2, Google Analytics tracking data and routine statistical data from Strathprints itself.

The periods examined were the year up to June 2016 (2015/2016), prior to the majority of the changes being implemented, and the year up to June 2017 (2016/2017), after improvements were implemented.

Strathprints demonstrated a **15%** growth in COUNTER usage in 2016/2017 despite experiencing only a **6%** growth in full-text deposits during the same period.

Web traffic, as measured by Google Analytics (GA), grew by **144,006** visits in 2016/2017 to **445,532**, equivalent to a **48%** improvement in web traffic when compared to 2015/2016.

By exposing their content to disparate search services, and the nature of repository content itself, repositories encourage – and are conducive to – “horizontal” information seeking strategies (10). This is typically reflected in the relatively high “bounce rates” experienced by repositories.

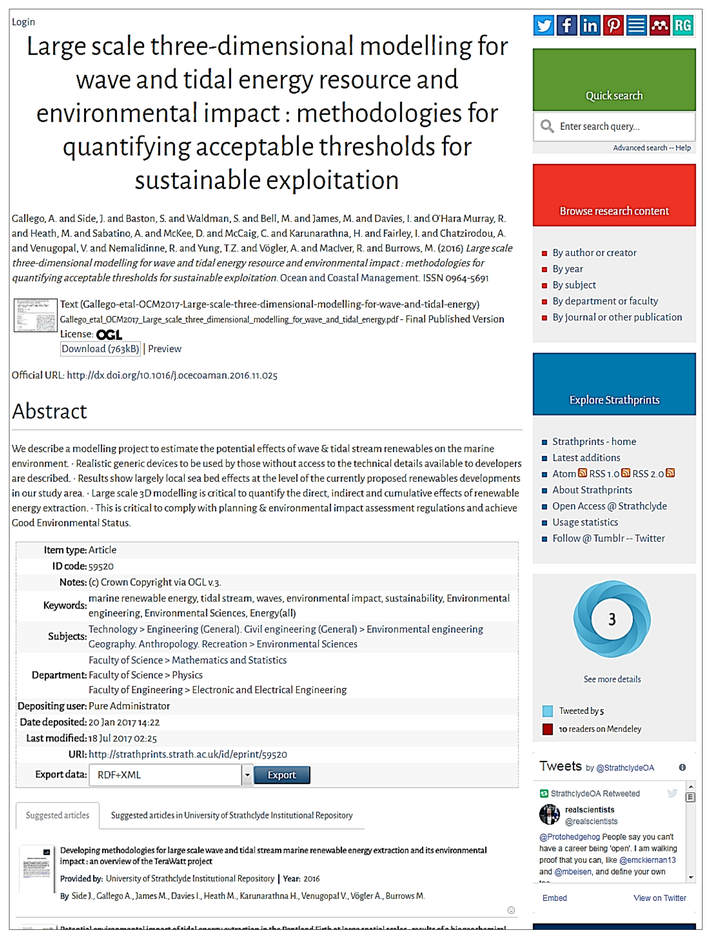

Fig.1: Strathprints UI following heuristic evaluation and refreshing.

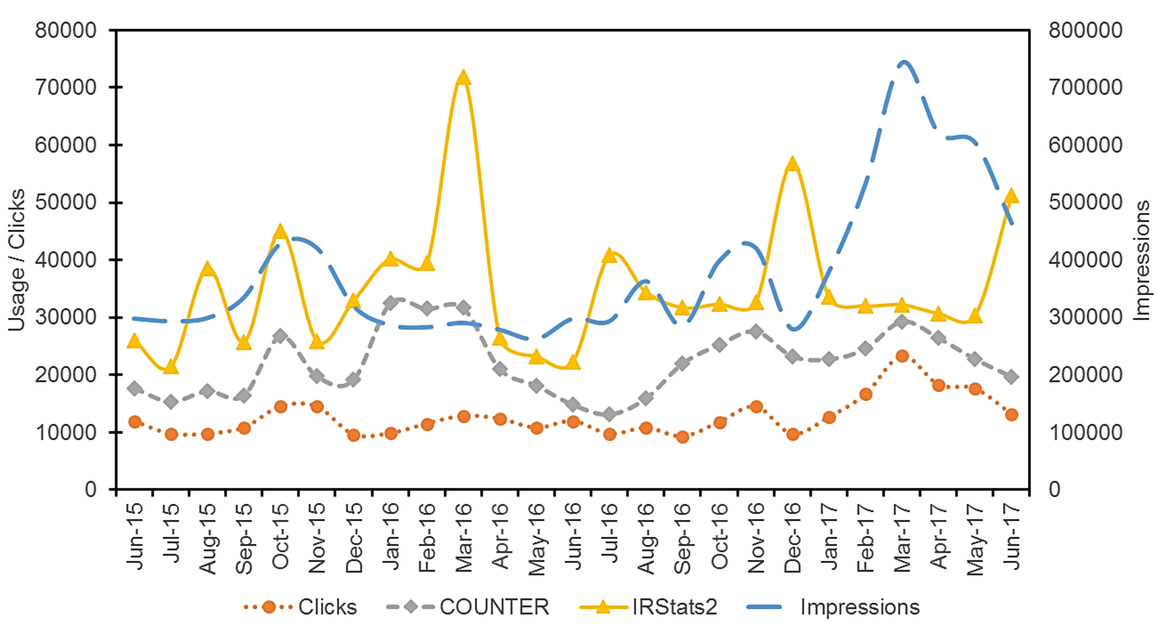

Fig.2: COUNTER usage, Google clicks and impressions, IRStats2 usage data.

The bounce rate in this study remained unchanged at **73%** across both reporting periods. However, the average time users spent on Strathprints upon arrival in 2016/2017 was **01:29**, up from **00:58** in 2015/2016 – so users, although continuing to bounce, typically spent **58%** longer on Strathprints, indicative perhaps that improvements to the UI and Strathprints functionality was enough to persuade users to defer their bounce.

A more significant measure of repository discoverability lies in search metrics. Examining the effect of the improvements on Google searching more specifically, improvements in “impressions” and “clicks” were observed at **62%** (*n* = 2,186,810) and **42%** (*n* = 287,262) respectively in 2016/2017 (via Google Search Console) compared to the 2015/2016 period, and a general upwards trend can observed in **Fig.2**. This is despite link decay resulting from a transfer of inbound links to the connected CRIS web front-end (9). Total impressions in 2016/2017 were **5,686,664** and clicks **460,993**.

To determine whether a correlation between clicks and COUNTER usage was present, Pearson’s correlation coefficient was calculated for both reporting periods. A strengthening of the correlation was detected, ranging from a weak relationship in 2015/2016 ($r = 0.24$) to a moderate-to-strong positive correlation in 2016/2017 ($r = 0.66$). This was confirmed via the t statistic ($t = 2.83$, $df = 12$, $p < 0.01$), suggesting a high level of statistical significance. However, the coefficient of determination revealed data to be more nuanced [**Fig.3 & 4**]. r^2 was stronger in 2016/2017 ($r^2 = 0.439$) than 2015/2016 ($r^2 = 0.059$); clearly a significantly higher value but indicating that only circa **44%** of the unique variance in COUNTER usage can be directly attributed to Google clicks.

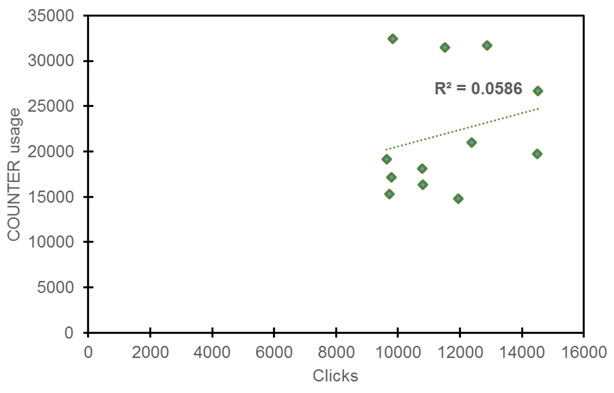

Fig.3: Association clicks and COUNTER usage with r^2, 2015/2016.

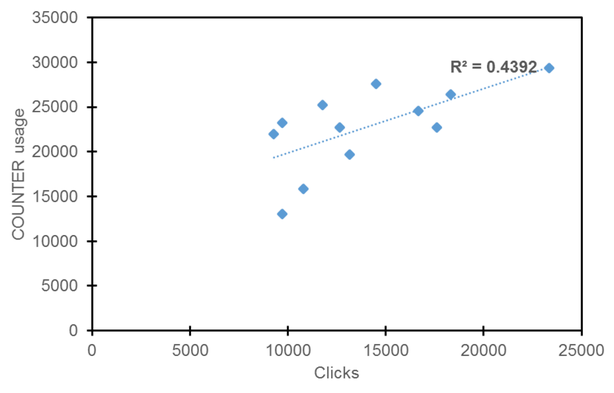

Fig.3: Association clicks and COUNTER usage with r^2, 2016/2017.

Such variability is perhaps unsurprising since not all users will use Google to discover research content – and it is significant to note that Google Search Console data pertain to search interactions within Google only. Thus the changes actioned between the two reporting periods excludes traffic generated by other discovery tools, including Google Scholar (GS). Some data are, however, available to aid understanding.

Total page views – as distinct from clicks – originating from GS grew from **55,576** in 2015/2016 to **66,671** in 2016/2017, equivalent to a **20%** increase. Repositories serving more digital content enjoy deeper indexing by GS (11). However, as a percentage of total traffic to Strathprints (based on GA data) this is actually a decline of **3%** (**15%** of total GA traffic – *n* = 445,532; **18%** of total GA traffic – *n* = 301,526); but it also suggests that it is less that visibility and ergo traffic from GS has declined and more that the changes implemented on Strathprints have yielded a far greater improvement in universal search tools *relative* to GS. As noted above, 2016/2017 witnessed a **48%** improvement in overall web traffic, a large proportion of which was generated via Google clicks as per the noted **42%** increase.

.

Conclusion and future work

This small study provides persuasive evidence that specific enhancements to technical aspects of repository, and surrounding process changes, can result in significant improvements to repository visibility, resulting in a greater web impact and consequent increases in content usage. Although indications of user interactions were less clear, it can nevertheless be concluded that both web and search traffic, and COUNTER usage, can be significantly improved on the most important search and discovery tools, with correlations observable between Google search visibility and repository COUNTER usage. Data will continue to be monitored thereby providing the basis for longitudinal analysis and subsequent analysis in 2017/2018. Current work is seeking to action a series of infrastructural improvements, including a switch to InnoDB to improve DB performance and page loading, and HTTPS. Exploration of Accelerated Mobile Pages (AMP) is also currently underway.

References

1. IRUS-UK [Internet]. 2017 [cited 2017 Jun 30]. Available from: http://www.irus.mimas.ac.uk/
2. Kelly B, Nixon W. SEO analysis of institutional repositories: What’s the back story? In: Open Repositories 2013 [Internet]. University of Bath; 2013 [cited 2017 Jul 19]. Available from: http://opus.bath.ac.uk/35871/
3. Strathprints: The University of Strathclyde institutional repository [Internet]. 2017 [cited 2017 Jun 30]. Available from: http://strathprints.strath.ac.uk/
4. Macgregor G. Reviewing repository discoverability with Strathprints [Internet]. Open Access @ Strathclyde. 2017 [cited 2017 Jun 29]. Available from: http://strathoa.tumblr.com/post/158349445900/reviewing-repository-discoverability-with
5. Zhang T, Maron D, Charles C. Usability evaluation of a research repository and collaboration website. Journal of Web Librarianship [Internet]. 2013 Jan 1; Available from: http://docs.lib.purdue.edu/lib_fsdocs/51
6. McKay D, Burriss S. Improving the Usability of Novel Web Software: An Industrial Case Study of an Institutional Repository. In: Web Information Systems Engineering – WISE 2008 Workshops [Internet]. Springer, Berlin, Heidelberg; 2008 [cited 2017 Jul 18]. p. 102–11. (Lecture Notes in Computer Science).
7. Nielsen J, Molich R. Heuristic Evaluation of User Interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: ACM; 1990. p. 249–256. (CHI ’90).
8. Continuing to make the web more mobile friendly [Internet]. Official Google Webmaster Central Blog. [cited 2017 Jul 19]. Available from: https://webmasters.googleblog.com/2016/03/continuing-to-make-web-more-mobile.html
9. Macgregor G. Feeding the beast: workloads in a hybrid IR / CRIS environment [Internet]. Open Access @ Strathclyde. 2017 [cited 2017 Jul 19]. Available from: http://strathoa.tumblr.com/post/159187095300/feeding-the-beast-workloads-in-a-hybrid-ir-cris
10. Ian Rowlands, David Nicholas, Peter Williams, Paul Huntington, Maggie Fieldhouse, Barrie Gunter, et al. The Google generation: the information behaviour of the researcher of the future. AP [Internet]. 2008 Jul 6;60(4):290–310.
11. Acharya A. Indexing repositories: pitfalls and best practices [Internet]. Proceedings of Open Repositories 2015. 2015. Available from: http://purl.dlib.indiana.edu/iudl/media/6537033b6s

Cite it: Macgregor, G 2017, 'Reviewing repository discoverability: approaches to improving repository visibility and web impact‘, Repository Fringe 2017, University of Edinburgh, Edinburgh, United Kingdom, 03/08/2017-04/08/2017,