SEO Analysis of Institutional Repositories: What's The Back Story?

Brian Kelly[+] and William Nixon[*]

+ UKOLN, University of Bath, Bath, UK {b.kelly@ukoln.ac.uk}

* University of Glasgow, Glasgow, UK {William.Nixon@glasgow.ac.uk}

Abstract

This paper summarises the approaches taken to the open analysis and interpretation of findings of surveys of the SEO (Search Engine Optimisation) characteristics of three institutional repositories provided by three Russell Group universities in the UK.

1. BACKGROUND

In a paper on "*Open Metrics for Open Repositories*" [1] Kelly *et al* described how "*Metrics for repositories can be used to provide a better understanding of how repositories are being used, which can help to inform policy decisions on future investment, technical policy decisions on enhancements to the technical infrastructure*". The paper concluded by arguing that "*repository managers should be pro-active in showing a willingness to provide open access to repository metrics*" since such approaches are "*consistent with the culture of openness which underpins those involved in the provision and support of open access repositories*".

Since the paper was published the author has worked with a number of repository managers in UK institutions in order to support the open analysis and interpretation of repository metrics. This paper summarises this work.

2. SHARING FINDINGS OF SEO ANALYSES OF IRS

A blog post by Kelly "*MajesticSEO Analysis of Russell Group University Repositories*" [2] in August 2012 reported on the findings of SEO benchmarking work for the UK's major research universities. Kelly started from the premise that Google is critical in driving traffic to repositories and was intended to "*gain a better understanding of the factors which contribute to supporting the discoverability of the content hosted in institutional repositories*."

The survey described summary SEO findings for the 24 Russell Group institutional repositories, collected using the MajesticSEO service on 27-28th August 2012.

The survey results showed that two blog platforms, WordPress.com and Blogspot.com appear to be primarily responsible for driving traffic to institutional repositories. These have both high Alexa rankings together with large numbers of links to the repositories.

However it was apparent that the statistics which could be gathered from such automated surveys were insufficient to understanding the implications of the findings and their relevance to changes to operational practices or policy-making. A number of repository managers were invited to run the SEO analysis tool across their repositories and to provide information about the host institution in order to gather information about the institutional context (including factors such as the size of the institution, its portfolio of research activities and details of the institutional repository service and local policy decisions)

3. THE NEED FOR THE INSTITUTIONAL CONTEXT

This follow-up work was carried out by repository managers at the University of Warwick, the London School of Economics and the University of Glasgow. These three institutions were chosen as they are members of the UK's Russell Group which "*represents 24 leading UK universities which are committed to maintaining the very best research, an outstanding teaching and learning experience and unrivalled links with business and the public sector*" [3]. It was felt that the findings from these three institutions should represent practices which other institutions would seek to emulate or, if the findings indicated that improvements could be made, would highlight changes which should be made in order to enhance the value of the repository services.

Each of the managers worked to a consistent template to ensure commonality across the posts. These included background on the institution, details in the size/scope of the repositories, the use of any metrics (across all three cases Google Analytics featured), expectations of the survey, a summary of the results and finally some discussion in which conclusions were drawn. None of the repositories had prior experience with the MajesticSEO service and this in itself provided an opportunity to openly share experiences and expectations using MajesticSEO.

Coincidentally all three repositories used EPrints.

The open practice approach and the flexibility provided on the UK Web Focus blog to publish and disseminate the SEO findings gave a focus for sector wide discussion and engagement.

3.1 University of Warwick (WRAP)

The first guest blog post [3] was by Yvonne Budden, the University of Warwick's E-Repositories manager. Warwick was the youngest institution (1965) and the youngest repository (2008), its repository is called WRAP (Warwick Research Archives Project) and in August 2012 had over 6500 full text items and an additional 40,000 metadata records. Full text from WRAP has been downloaded over 730,000 times since 2008 and these are tracked using the EPrints download statistics plug-in IR Stats. Views and access to WRAP, averaging over 18,000 per month are tracked using Google Analytics.

The data for the MajesticSEO survey was run on the 10th September 2012 and revealed 413 referring domains and 2,533 backlinks. This was less than Budden had expected and showed a fairly low number of educational domains linking back to Warwick. There was some overlap in the top 5 backlinks (ranked by citation flow and trust) and results from Google Analytics but included others which were not expected. The top 10 backlinks included a range of resources including the UK Web Focus blog, Wikipedia and the PhilPapers repository. Budden notes that there are no mentions of any Warwick domains but that this is assumed to be because Majestic SEO excludes self-links.

The top pages shown demonstrate a trend towards full text.

In the discussion and analysis Budden asserts that the most important thing which a repository manager can do with metrics is to build stories about them while not drawing fuzzy conclusions that links X will result in downloads Y and in turn citations Z. Having sound evidence that depositing the paper in WRAP as well as posting to Twitter or using a blog post while increase the impact of their research is more challenging even with tools like MajesticSEO.

Budden also comments on the need to be aware of the impact of personalisation by search engines which increasingly will weight and rank different results based on users' previous search history. Ultimately the key challenge is to be mindful of the use our academic colleagues can make of these metrics and that we can confident in interpreting what they say and how they can be used.

3.2 London School of Economics (LSE Research Online)

The second post [4] was written by Natalia Madjarevic, the manager of LSE Research Online (LSERO). The London School of Economics is a specialist university and was founded in 1895. LSE Research Online was set-up in 2005 and as of September 2012 had over 7000 full text items and over 33,000 records. Downloads from LSERO, from Analog server statistics were over 5 million from May 2007 to September 2012. LSERO uses Google Analytics and since 2007 had received over 2.2 million visits.

Madjarevic expectations of MajesticSEO were that they would see lots of traffic from Google and backlinks from the LSE domain itself which are key routes of traffic according to Google Analytics as well as Wikipedia and Summon, a web scale discovery service had also been recently implemented at LSE.

The data for the MajesticSEO was run on the 24th September and revealed 1,285 referring domains and 8,856 backlinks. There were over 408 educational domain backlinks. Drilling down to look at the backlinks in more detail many of the backlinks are from Wikipedia in contrast to Google Analytics which shows Wikipedia as the 6th most popular domain. The top referring domains are from the Web sites provided by WordPress, Blogspot and Wikipedia.

In the discussion Madjarevic draws a number of conclusions from the results looking at the top referring domains it seems reasonable to suggest that adding links to blogging platforms like WordPress and Blogspot will lead to an increased SEO ranking. The dominance of blogging domains for referrals contrasts with Google Analytics which shows the majority of referrals come from search engines.

Like Warwick the majority of the top backlinks are to outputs with full text, or in many cases the PDF itself. The top 5 resources sorted by backlinks correlates with the most consistently popular papers according to the LSE's Analog statistics and Google Analytics. Madjarevic notes however that the difference in results between MajesticSEO and Google Analytics make it difficult in this initial analysis to draw firm conclusions from the Majestic report.

3.3 University of Glasgow (Enlighten)

The final post [5] was written by William Nixon, the manager of the University of Glasgow Institutional Repository. The University of Glasgow is the oldest of the three universities (1451) and can trace its repository work back to 2001. The university has separate repositories for published papers for theses. The MajesticSEO survey focussed on Enlighten, the published papers repository. In mid-October 2012 Enlighten had 4700 full text items and records for over 53,000 outputs. Glasgow uses Google Analytics to track views, in 2011 over 230,000 people visited Enlighten, like Warwick, the EPrints plug-in IRStats is used for downloads.

Nixon's expectations were that the most popular non-gla.ac.uk domains would feature in the MajesticSEO results, from Google Analytics these were Mendeley, Wikipedia and Google Scholar. Blogs didn't feature.

The data for the MajesticSEO survey was generated on the 22nd October 2012. The summary showed 632 referring domains and over 5,000 external backlinks - more than Warwick but less than the LSE. There were 619 educational backlinks and 54 educational referring domains.

The top five domains were from the Web sites provided by Blogspot, WordPress, Wikipedia, BBC and CNN.

Nixon comments in his summary that in the top 5 backlinks, 4 are from Wikipedia. None of the top 5 PDFs appear from MajesticSEO appears in the IRStats generated list.

In the discussion Nixon notes that while the initial work focussed on the Top 5 but extending this to the Top 5 would be more useful for further comparison and sites such as Mendeley appear in Referring Domain reports which correlate with Google Analytics. He also identifies the absence of social media sites such as Twitter and FaceBook but posits that this may because the volume is small or that it could be a breach of service.

Nixon concludes by noting that this was an interesting, challenging and thought-provoking exercise which not only provided the opportunity to explore a new tool in MajesticSEO but to also reflect on our use of Google Analytics. The results from this work provide interesting counterpoints to the existing data gathered from Google Analytics and IRStats.

4. FURTHER WORK

This work has made use of a single service, MajesticSEO. However since there is a lot of volatility in how search engines rank service (in order to minimise the risks of spam companies producing spurious results) it is felt desirable to both repeat the surveys and to make use of other SEO analysis tools. We intend to carry out such work and publish the findings.

5. CONCLUSIONS

This paper is based on the premise the Google visibility has an important role to play in maximising access to content hosted in institutional repositories. There is therefore a pressing need to gain a better understanding of the SEO characteristics of current repository services in order to identify examples of best practices and flawed approaches. However since local factors are likely to impact the visibility to search engines of content hosted in institutional repositories it will be important to ensure that such local factors are understood. The work described in this paper describes a methodology for sharing institutional findings in order to inform practices across the repository community. We therefore invite other repository managers to work in a similar fashion, critique the methodology and tools we have described and share the findings for their repository.

REFERENCES

[1] *Open Metrics for Open Repositories*, Kelly, B., Sheppard, N., Dewey, M., Delasalle, J., M., Stephens, O., Johnson, G. and Taylor, S., Open Repositories 2012, 9-13 July 2012, Edinburgh, Scotland, http://opus.bath.ac.uk/30226/

[2] *MajesticSEO Analysis of Russell Group University Repositories*, Brian Kelly, UK Web Focus blog, 29 August 2012, http://ukwebfocus.wordpress.com/2012/08/29/majesticseo-analysis-of-russell-group-university-repositories/

[3] *Russell Group*, http://www.russellgroup.ac.uk/

[4] *SEO Analysis of WRAP, the Warwick University Repository*, Yvonne Budden, UK Web Focus blog, 23 October 2012, http://ukwebfocus.wordpress.com/2012/10/23/seo-analysis-of-wrap-the-warwick-university-repository/

[5] *SEO Analysis of LSE Research Online*, Natalia Madjarevic, UK Web Focus blog, 24 October 2012, http://ukwebfocus.wordpress.com/2012/10/24/seo-analysis-of-lse-research-online/

[6] *SEO Analysis of Enlighten, the University of Glasgow Institutional Repository,* William Nixon, UK Web Focus blog, 25 October 2012, http://ukwebfocus.wordpress.com/2012/10/25/seo-analysis-of-enlighten-the-university-of-glasgow-institutional-repository/