Promoting content discovery of open repositories: reviewing the impact of optimization techniques (2016-2019)

OR2019

George Macgregor
https://purl.org/g3om4c
@g3om4c

Thursday 13/06/2018

Paper overview

- Conference themes 1 & 2
  - Understanding user needs & user experience
  - Discovery, use & impact
- Defining the problem space for repositories
- Repository enhancement experiments
  - Longitudinal dataset & ‘sequel’ to prior work
- Using Strathprints, University of Strathclyde institutional repository, as case study
  - Relevance for all repository software platforms
- Detailed data analysis provided in accompanying paper

A brief institutional history!

- University of Strathclyde – Established 1796 as "place of useful learning" by John Anderson
  - Now 21,470 FTE students & 3,200 staff
- Among 20 top research-intensive universities in the UK
- Research income 2016: £60 million

Assessed by UK Research Excellence Framework (REF) to have #1 research in physics

- 3rd in Electrical & Electronic Engineering (1st in Scotland)
- 4th in Engineering (Aerospace, Mechanical, Marine, etc)

Strong outside science & engineering…

Repository problem space: motivation

- Relevance of open repositories as nodes within open science infrastructure
  - ‘Existential threats’: proprietary ‘solutions’ which are less open & less discoverable
- Repositories as exemplars of content exposure and resource discovery
  - User [author] expectations surrounding search
  - Repositories not static systems; the ‘rotting repository’
  - COAR Next Generation Repositories synergy [1]
  - Several contributions at OR2019 about visibility
- Gathering evolving evidence on discovery a necessity going forward
- The ‘Strathclyde’ motivation…

Repository discovery context

University of Strathclyde Glasgow

- Importance of repositories in promoting open scholarly communication & discovery widely noted [2, 3, 4, 5, 6]
  - Importance of SEO, 'white hat' improvements in promoting Google & Google Scholar indexing
  - Recommendations from Acharya [7]
- Few studies have codified or evaluated visibility or discovery strategies
- Wider understanding about repository discoverability remains embryonic
- Macgregor [8, 9] – addressed this
  - Encouraging evidence about positive impact of repository enhancements on web impact & discovery. But…

Repository

Running EPrints 3.3.13: Strathprints

Strathprints home > Open Access

Strathprints

Logged in as Mr George Macgregor | Manage deposits | Manage records | Profile | Saved searches | Review | Admin | Reports | Edit page | Logout

Strathprints: The University of Strathclyde institutional repository

The Strathprints institutional repository is test a digital open archive of University of Strathclyde research outputs. It has been developed to disseminate Open Access research outputs, expose data about those outputs, further the goals of open science, and enable the management and persistent access to Strathclyde's intellectual output. Explore Strathprints by searching and browsing.

Enter your search query...

Advanced search | Browse by subject | Latest additions

About Strathprints | Usage statistics | Open Access @ Strathclyde

Latest additions

1. Kitaev, Sergey and Enright, Jessica (2019) *Polygon-circle and word-representable graphs.* In: TCDM 2018 – 2nd IMA Conference on Theoretical and Computational Discrete Mathematics, University of Derby. Electronic Notes in Discrete Mathematics, 71 . UNSPECIFIED, pp. 3-8. Item not available from this repository.
2. Turnbull, A. and Carroll, J. and Koukoura, S. and McDonald, A. , ed. (2018) *Prediction of wind turbine generator bearing failure through analysis of high frequency vibration data and the application of support vector machine algorithms.* In: The 7th International Conference on Renewable Power Generation, 2018-08-26 - 2018-09-27, DTU, Lyngby.
3. Tsioumpri, Eleni and Stephen, Bruce and Dunn-Birch, Neil and McArthur, Stephen D.J. , ed. (2018) *Data analytics to support operational distribution network monitoring.* In: IEEE PES Innovative Smart Grid Technologies Conference Europe 2018, 2018-10-21 - 2018-10-25, Sarajevo.
4. Rana, Shuvendu and Sur, Arijit (2018) *View invariant DIBR-3D image watermarking using DT-CWT.* Multimedia Tools and Applications. pp. 1-29. ISSN 1380-7501

Browse research conte

- By author or creator
- By year
- By subject
- By department or faculty
- By journal or publication

Explore Strathprints

- Strathprints - home
- Latest additions
- Atom RSS 1.0 RSS 2.0
- About Strathprints
- Open Access @ Strathclyde
- Usage statistics
- Follow @ Tumblr -- Twitter

Contact us

Open Access enquiries: openaccess@strath.ac.uk
Repository enquiries: strathprints@strath.ac.uk

Login

Making modelling count - increasing the contribution of shelf-seas community and ecosystem models to policy development and management

Hyder, Kieran and Rossberg, Axel G. and Allen, J. Icarus and Austen, Melanie C. and Barciela, Rosa M. and Bannister, Hayley J. and Blackwell, Paul G. and Blanchard, Julia L. and Burrows, Michael T. and Defriez, Emma and Dorrington, Tarquin and Edwards, Karen P. and Garcia-Carreras, Bernardo and Heath, Michael R. and Hembury, Deborah J. and Heymans, Johanna J. and Holt, Jason and Houle, Jennifer E. and Jennings, Simon and Mackinson, Steve and Malcolm, Stephen J. and McPike, Ruaraidh and Mee, Laurence and Mills, David K. and Montgomery, Caron and Pearson, Dean and Pinnegar, John K. and Pollicino, Marilena and Popova, Ekaterina E. and Rae, Louise and Rogers, Stuart I. and Speirs, Douglas and Spence, Michael A. and Thorpe, Robert and Turner, R. Kerry and van der Molen, Johan and Yool, Andrew and Paterson, David M. (2015) *Making modelling count - increasing the contribution of shelf-seas community and ecosystem models to policy development and management.* Marine Policy, 61. pp. 291-302. ISSN 0308-597X

Text (Hyder-etal-MP-2015-Making-modelling-count-increasing-the-contribution-of-shelf-seas-community)
Hyder_etal_MP_2015_Making_modelling_count_increasing_the_contribution_of_shelf_seas_community.pdf
Accepted Author Manuscript
License:
Download (7MB) | Preview

Official URL: https://doi.org/10.1016/j.marpol.2015.07.015

Abstract

Marine legislation is becoming more complex and marine ecosystem-based management is specified in national and regional legislative frameworks. Shelf-seas community and ecosystem models (hereafter termed ecosystem models) are central to the delivery of ecosystem-based management, but there is limited uptake and use of model products by decision makers in Europe and the UK in comparison with other countries. In this study, the challenges to the uptake and use of ecosystem models in support of marine environmental management are assessed using the UK capability as an example. The UK has a broad capability in marine ecosystem modelling, with at least 14 different models that support management, but few examples exist of ecosystem modelling that underpin policy or management decisions. To improve understanding of policy, and management issues that can be addressed using ecosystem models, a workshop was convened that brought together advisors, assessors, biologists, social scientists, economists, modellers, statisticians, policy makers, and funders. Some policy requirements that can be addressed without further model development were identified including: attribution of environmental change to underlying drivers, integration of models and observations to develop more efficient monitoring programmes, assessment of indicator performance for different management goals, and the costs and benefit of legislation. Multi-model ensembles are being developed in cases where many models exist, but model structures are very diverse making a standardised approach of combining outputs a significant challenge, and there is need for new methodologies for describing, analysing, and visualising uncertainties. A stronger link to social and economic systems is needed to increase the range of policy-related questions that can be addressed. It is also important to improve communication between policy and modelling communities so that there is a shared understanding of strengths and limitations of ecosystem models.

36 | 36 Total citations | 25 Recent citations | 8.06 Field Citation Ratio | n/a Relative Citation Ratio

33 | Referenced in 1 policy documents | Tweeted by 42 | On 1 Facebook pages | 114 readers on Mendeley | 0 readers on Connotea | 0 readers on CiteULike

See more details

Item type: Article
ID code: 54075
Keywords: ecosystem models, marine policy and management, UK environmental assessment, management, and monitoring, Probabilities. Mathematical statistics, Aquaculture. Fisheries. Angling, Aquatic

Quick search

Enter search query...

Advanced search -- Help

Browse research content

- By author or creator
- By year
- By subject
- By department or faculty
- By journal or publication

Explore Strathprints

- Strathprints - home
- Latest additions
- Atom RSS 1.0 RSS 2.0
- About Strathprints
- Open Access @ Strathclyde
- Usage statistics
- Follow @ Tumblr -- Twitter

Contact us

Open Access enquiries: openaccess@strath.ac.uk
Repository enquiries: strathprints@strath.ac.uk

Adjusting & improving: Strathprints case study

University of Strathclyde Glasgow

- Repositories with OOTB discovery support
  - Aggregation, Google Scholar, SEO, etc.
- Wide variation in relative visibility of repositories, even across similar or same software
- Case study involved:
  - 'Improvements' = substantive modifications to repository functionality
  - 'Adjustments' = refinement of existing aspects of the repository
  - Quasi repeated measure approach
- Full details of technical changes documented in prequel paper at Code4Lib Journal [8]

Mobile first indexing – PageRank [10]

`@font-face`

`async="async"`

University of Strathclyde Glasgow

Key technical 'adjustments'

Modification of file-naming conventions.
"Minification" of all relevant repository files.
Rationalisation of all CSS and Javascript (JS) files in order to remove unused rules and variables.
Asynchronous loading of JS resources & CSS loading optimizations
GZIP compression
Image optimization
Migration to InnoDB as the MySQL storage engine
Deployment of Google Data Highlighter

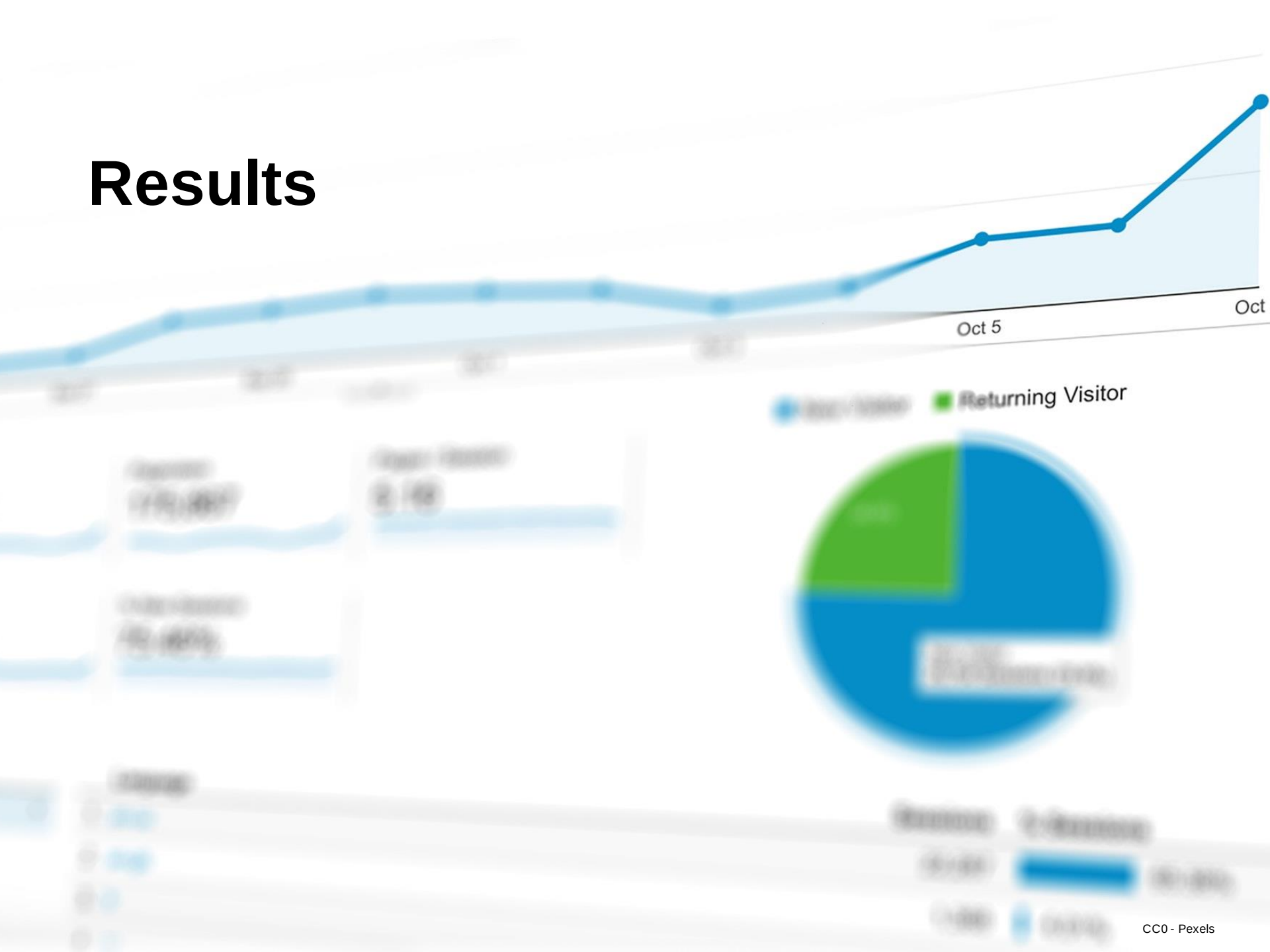
Key technical 'improvements'

Repository user interface (UI) improvements.
"Mobile first", responsive re-engineering of repository.
"White hat" improvements to improve to internal linking (e.g. navigation, hyperlink labels, etc.) and content improvements to promote user interaction.
"Connector-lite" ecosystem implemented within repository-CRIS interactions.

Data capture

- Metrics monitored to measure influence of 'adjustments' & 'improvements'
  - Google Analytics (GA) (web tracking data)
  - Google Search Console (search traffic data)
  - IRUS-UK (COUNTER compliant usage data)
  - Routine data from Strathprints itself
  - Digital deposits volumes during reporting period
- Data monitored on annual basis*:
  - Year up to **March 2016 (Y1): baseline year**
  - Then: **Y2 (2016/2017)**, **Y3 (2017/2018)** & **Y4 (2018/2019)**
- Use of alternative data snapshot compared to prior work
- Unique dataset for interrogation about visibility and discovery

Results



Web traffic

- Web traffic & unique traffic measured using GA
- Variation on annual % traffic & unique traffic increases compared to prior work but…
  - 65% increase in web traffic
  - 69% increase in unique web traffic
  - Tendency for less variation: higher mean traffic, lower standard deviations
- Exceeds results reported in [8, 9]
- Google as the ‘traffic funnel’ for repository usage: no change!
  - 82% traffic coming from Google services: Google & Google Scholar
  - 26% from Google Scholar (big increase but lower than OBrien et al. [5])

Web traffic: digging deeper

- Google traffic (inc. unique traffic) increase = **1500%** (approx.)
- Google Scholar = **1920%**
- Attributable to low – but real – baseline traffic in Y1?
- Exclusion of outlying data, large increase remains observable:
  - **74%** and **70%** increase in referral and unique Google Scholar traffic respectively
  - **67%** and **69%** increase in vanilla Google
- Exceeds growth rates observed in wider pool of referral sources

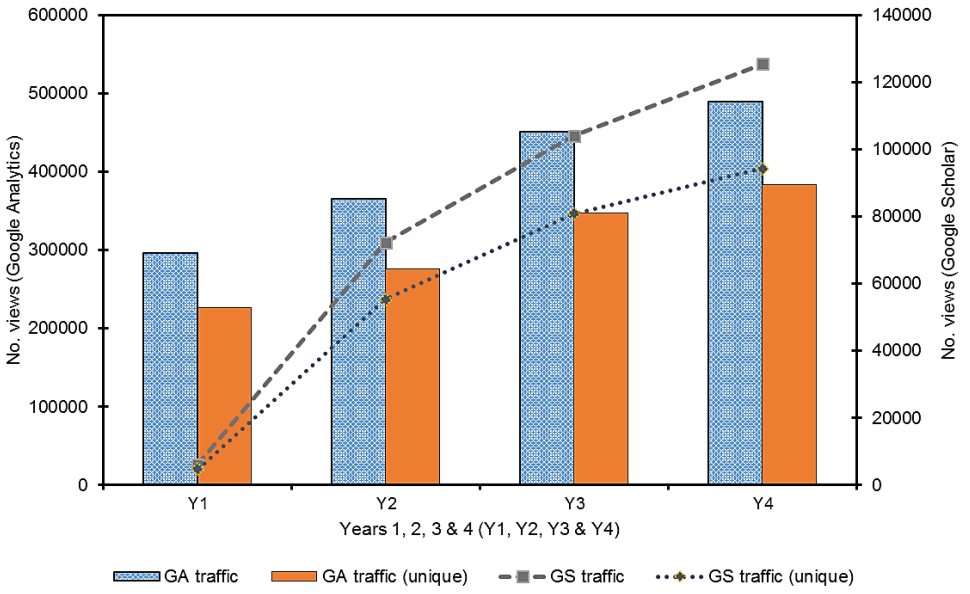


Fig. 1 Volume of Google and Google Scholar referral traffic , including unique traffic in Y1, Y2, Y3 & Y4.

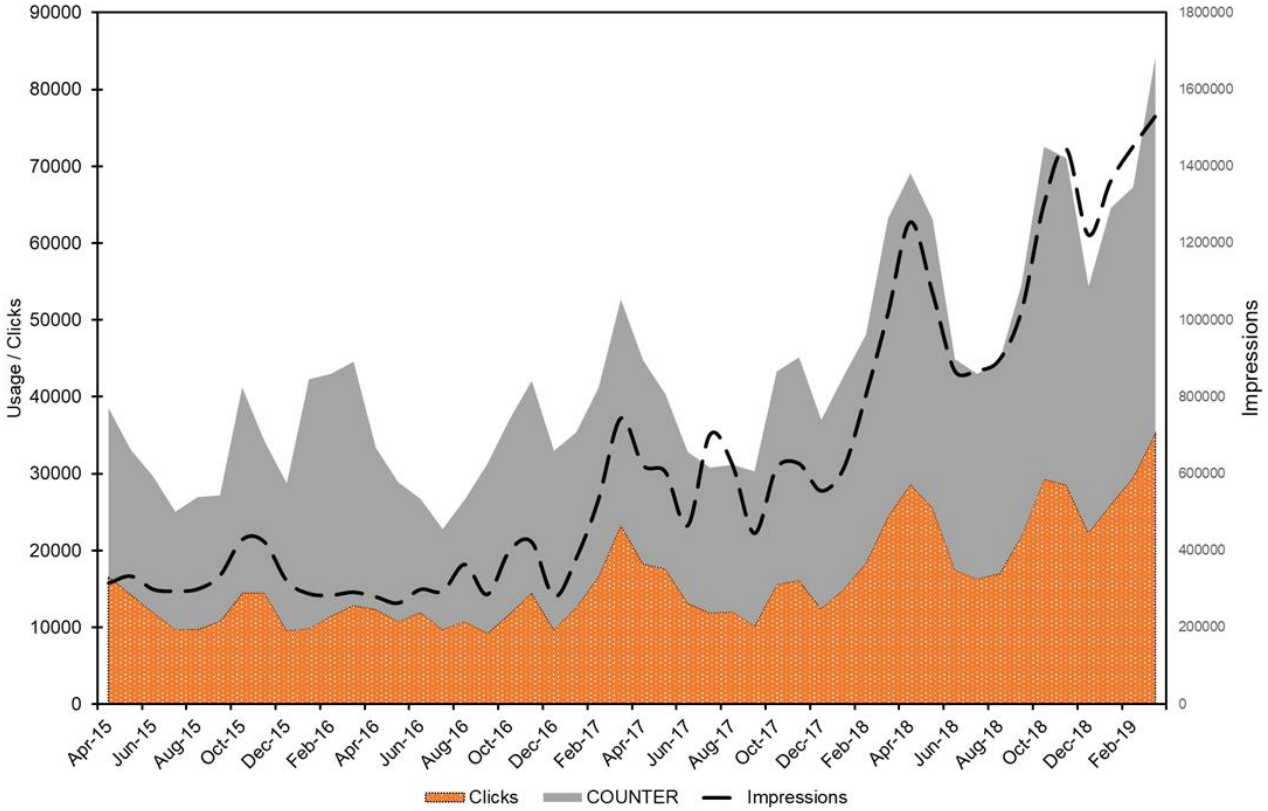


Fig 2 Strathprints COUNTER usage during Y1-Y4 alongside Google clicks and impressions during the same period.

University of Strathclyde Glasgow

Repository content: discovery & usage

- Search metrics offer appropriate measure of repository content discoverability
- 'Impressions' & 'clicks'
  - **Y2** at **16%** (*n* = 4,537,744) and **23%** (*n* = 153,539) respectively when compared to the Y1 period.
- Acceleration in Y3 and Y4:
  - **Y3** = **69%** (*n* = 7,687,550) and **21%** (*n* = 185,232) increase in impressions and clicks
  - **Y4** = **86%** (*n* = 14,290,059) and **61%** (*n* = 298,020) increase

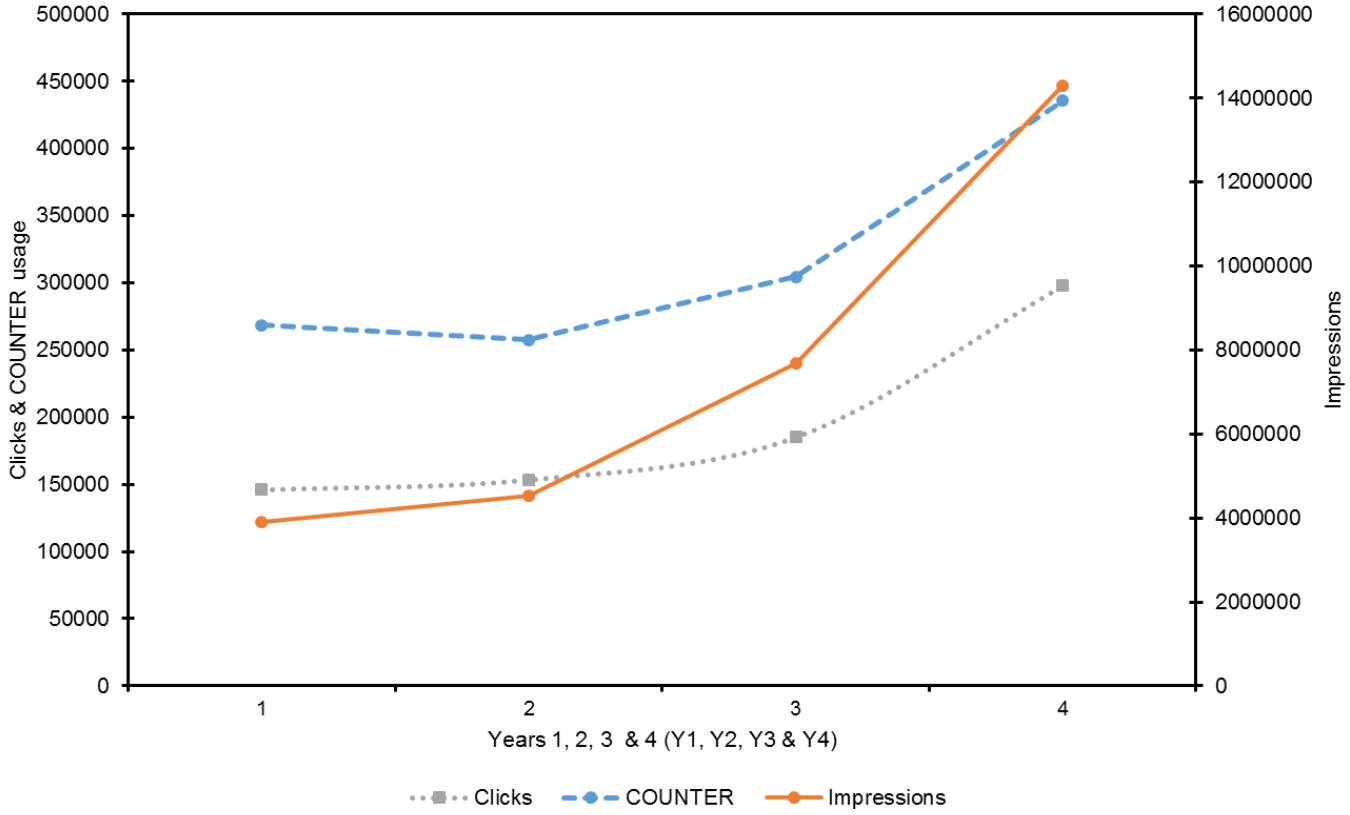
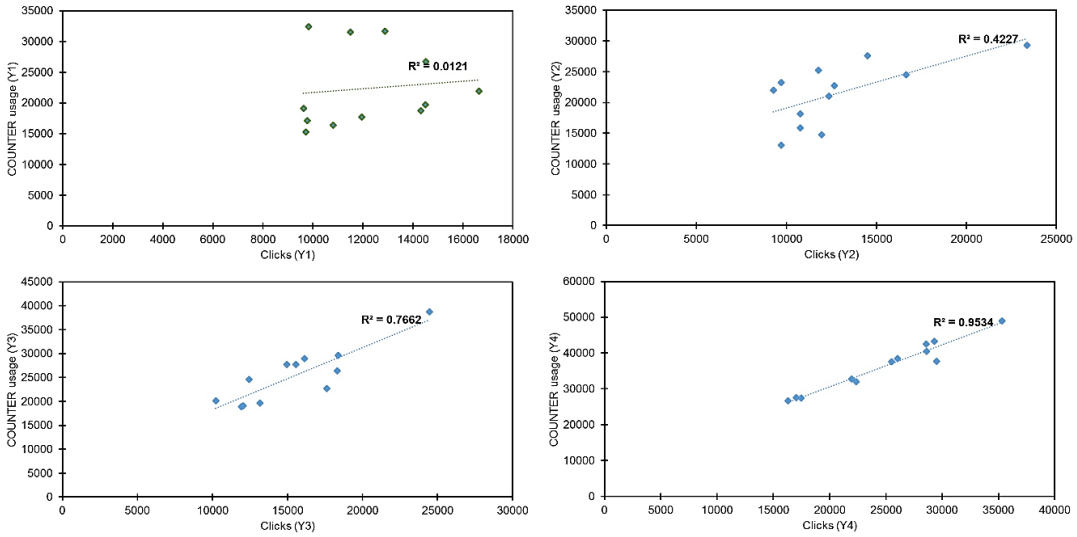


Fig 3 Charted data on observed clicks, impressions and COUNTER usage during Y1, Y2, Y3 & Y4.

- Total search metrics for impressions & clicks = **266%** & **104%** increase respectively
- **62%** growth in COUNTER compliant downloads (Y1-Y4)
  - Percentage of users referred increased at higher rate than full-text deposit
  - Only **23%** growth in full-text deposits (Y1-Y4)
- COUNTER growth variable but **43% between Y3 & Y4**
  - Cumulative effect of full-text content…?



Coefficient of determination (r squared) for Y1, Y2, Y3 & Y4 between clicks and COUNTER usage.

Clicks & COUNTER usage: what relationship?

- Pearson's correlation coefficient calculated for each reporting year
  - Weak in Y1 (r = 0.11)
  - Moderate positive in Y2 (r = 0.65)
  - *Strong* by Y3 (r = 0.87) and Y4 (r = 0.97)
- Positive correlation confirmed via t statistic
  - **Y3** (t = 5.72, df = 11, $p < 0.0005$)
  - **Y4**, at a far higher level of statistical significance (t = 14.30, df = 11, $\mathbf{p < 0.0005}$)
- Coefficient of determination, r^2 stronger in Y2 than Y1 (r^2 = 0.423) – **42%** of variance attributable to clicks but…
- Variance narrows in Y3 (r^2 = 0.766) & Y4 (r^2 = **0.953**)

Some limitations…

- Effecting change on 3rd party systems
  - Controlling every variable hypothesised to influence web visibility impossible
  - Experimental design impossible given the 3rd party constraints
- Search metric data – Search Console a necessary data compromise
- Additional analysis of data possible but limited by nature of conference paper format
  - Possible expansion in any proceedings publication…?
- Conference preprint: https://doi.org/10.17868/67963
- There is always more that could be done to improve discovery!
  - WebSub, ResourceSync, default support for schema.org, etc.

Conclusions & wrap up…

- Corroborates previous evaluative studies
- Persuasive steer on development of repositories in coming years
  - Users and user needs…
- Importance of technical enhancements as a way to secure significant web impact & usage gains
  - Large web traffic gains (even where outlying data removed)
    - Especially large Google and GS traffic growth
  - Traffic from GS demonstrated considerable growth – but as a proportion of total traffic small (26%?)
    - OBrien et al. found **48%-66%** of traffic from GS [5]
  - Promotion of traffic from competing platforms diminishes the size of GS traffic improvements in this case study…? [8, 9]
- 62% growth in COUNTER usage despite lower rates of full-text deposit
- 266% & 104% growth in Google impressions & clicks

Predictive potential of Google clicks – further work, replication with different repositories

References

1. COAR. “Next Generation Repositories: Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group.” Göttingen: COAR, November 2017. https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf.
2. Lee-Hwa, Tan, A. Abrizah, and A. Noorhidawati. “Availability and Visibility of Open Access Digital Repositories in ASEAN Countries.” *Information Development* 29, no. 3 (August 2013): 274–285. https://doi.org/10.1177/0266666912466754.
3. Tonkin, Emma L., Stephanie Taylor, and Gregory J. L. Tourte. “Cover Sheets Considered Harmful.” *Information Services & Use* 33, no. 2 (January 2013): 129–137. https://doi.org/10.3233/ISU-130705.
4. Arlitsch, Kenning. “Driving Traffic to Institutional Repositories: How Search Engine Optimization Can Increase the Number of Downloads from IR.” presented at the COAR Webinar and Discussion Series, September 18, 2017. https://doi.org/10.5281/zenodo.894564.
5. OBrien, Patrick, Kenning Arlitsch, Leila B. Sterman, Jeff Mixter, Jonathan Wheeler, and Susan Borda. “Undercounting File Downloads from Institutional Repositories.” *Journal of Library Administration* 56, no. 7 (October 2016): 1–24. https://scholarworks.montana.edu/xmlui/handle/1/9943.
6. Kelly, Brian, and William Nixon. “SEO Analysis of Institutional Repositories: What’s the Back Story?” In *Open Repositories 2013*. University of Bath, 2013. http://opus.bath.ac.uk/35871/.
7. Acharya, Anurag. *Indexing Repositories: Pitfalls and Best Practices*, 2015. https://media.dlib.indiana.edu/media_objects/9z903008w.
8. Macgregor, George. “Improving the Discoverability and Web Impact of Open Repositories: Techniques and Evaluation.” *The Code4Lib Journal*, no. 43 (February 14, 2019). https://journal.code4lib.org/articles/14180.
9. Macgregor, George. “Reviewing Repository Discoverability : Approaches to Improving Repository Visibility and Web Impact.” In *Repository Fringe 2017*. John McIntyre Conference Centre, University of Edinburgh, 2017. https://strathprints.strath.ac.uk/61333/.
10. Wang, Zhiheng, and Doantam Phan. *Using Page Speed in Mobile Search Ranking*, 2018. https://perma.cc/8QKP-NE5S.

Photo by Emily Morter on Unsplash