

# HEIRNET

## **HEIRNET Web Metrics Survey:**

Summary of Results and Recommendations

*Dan Hull (CBA) and Stuart Jeffrey (ADS)*

---

In 2009, HEIRNET (the Historic Environment Information Resources Network) conducted a survey looking at how different historic environment resource providers collect and analyse web metrics. In particular, we were interested to know which Historic Environment Information Resources (HEIRs) employ web statistics as a method of gauging the popularity of their web resource, in what ways they do this, and what types of user behaviour they seek to analyse.

At present there is a wide variety of practice in web statistical techniques across the sector, making it difficult to compare data in a meaningful way. By conducting this survey, our aim was to gain a clearer collective understanding about the methods used by different web resource providers in the historic environment. The results are summarised here, and some recommendations about future practice outlined. It is hoped that by sharing good practice and, where appropriate, harmonising web metric collection techniques, results can be shared more meaningfully across the sector.

A web survey of 27 questions was circulated to all HEIRs in the HEIRNET Register, and via posts on the Britarch and HERForum lists. Responses were received from 19 different organisations, including HERs, national agencies, the CBA and PAS. HEIRNET would like to thank all of those who took part.

## **Results**

- 14 of the 19 organisations collect results as part of a regular cycle (most quarterly, some monthly, one weekly)
- Five do so on the basis of page impressions, four on the basis of hits, and others a combination of several different metrics (including user logins, database queries, session duration, and bounce rates)
- Almost all collect information on how many unique users are accessing their site(s), most on IP addresses, and a smaller number on registered users, cookies or a combination of these
- Almost all collect metrics on where their users are located: this varies between country, domain, ISP location, and information deriving from user registration stats
- Less than half collect metrics on the operating system used
- Nine HEIRs collect information on session duration

- A high percentage collect information on what the user is doing on the website, with most interested in file downloads, but also searches, adding user comments, and carrying out a transaction
- Over half don't collect information on what terms a user employs to get to a website, and less than half are interested in how a user got to the site at all
- Over half don't collect information on search terms employed once users are within a resource
- Only four HEIRs currently collect information on user pathways, though a further five intend to do so in the future
- In terms of how are the stats collected, 12 of the 19 stated that they collect them in-house, with only three using an external provider
- There is a roughly half and half split between information officers collecting metrics themselves, and a specialist IT team elsewhere in the organisation doing so
- None of the organisations who responded to the survey said that they had a specific budget for collecting and analysing stats
- Three HEIRs said they use a browser-based approach (employing Google Analytics, Nedstats, Mint or similar); eight said that they collect web logs (using a package such as Analog, Sawmill or Web Trends)
- A wide range of software is used to gather and analyse web metrics, including NetTracker, sitemeter.com, SmarterStats, Analog, Google Analytics, and Oracle functionality
- Four said that they collect metrics on robot and/or spider activity; five said that they did not, and five that they did not know
- Many organisations expressed an interest in collecting a range of further indicators in the future, including total number of visitors, the numbers 'who stay longer than seven minutes', 'user satisfaction' and other user feedback, as well as different forms of searching (such as using a map search versus a text search, specifying image returns versus non-images)
- Some also said they used non-metrical methods such as web surveys, help desk comments, and specific usability surveys

## **Draft recommendations**

1. It is recommended that access statistics are collected and analysed by HEIRs at least quarterly and that if analysis is done more frequently then results can be aggregated into quarterly reports for cross-comparison.
2. Given the difficulties of equating IP addresses with unique users due to IP pools etc., it is not recommended that metrics on IP addresses are used. Rather, more specific information collected from registered users and web surveys can be used and, where necessary, extrapolated to give a general impression of total users and user breakdown
3. In terms of user location at finer granularity than country, ambiguities imposed by ISP location make the collection of such information on the basis of metrics unreliable. It may be better to rely on user registration information, as above

4. It is surprising that less than half of HEIRs analyse information on how users reached their website. Understanding how our sites are discovered is an important tool in making them more discoverable, so logging of referring sites and (where possible/applicable) search terms is desirable.
5. It is useful to draw a distinction between information collected on access statistics and user behaviour. Information on user behaviour cannot easily be compared on a like-for-like basis, so although sharing is useful on a qualitative basis, quantitative comparisons are likely to be deceptive (with the possible exception of bounce rates).
6. Browser-based analyses using external providers (such as Google Analytics) are useful for snapshot reporting, but for the analysis of long-term trends the collection and analysis of logs internally is recommended – unless long-term access to external service providers and the data they hold can be guaranteed.
7. It is important that those responsible for analysing web metrics have a good understanding of the role of robot and spider activity since these are likely to skew statistics on user behaviour. For cross-comparison purposes, robots.txt files should also be shared.
8. Organisations in the historic environment sector could do more to compare approaches to web metrics, and also the many different non-metric approaches such as web surveys, user observation, and registration information. Shared surveys should be considered.
9. The sharing of web metrics between HEIRs should occur more regularly and openly, as a useful tool by which the sector can monitor overall take-up and use of resources. The ability to show more clearly how (and how well) historic environment information resources are used will enable us more easily to obtain the necessary resources to support and develop them into an uncertain economic future
10. In order to facilitate more widespread sharing of web metrics, common standards should be developed communally by the sector, and a straightforward (and updateable) guide to measuring and describing use of web resources produced to enable more organisations to take part in such activities

*We would welcome further feedback and information from interested parties on this issue. If you have a comment, suggestion or recommendation to make, please contact Dan Hull at [danhull@britarch.ac.uk](mailto:danhull@britarch.ac.uk)*