

## Summary of findings from interview series and qualitative validation of webmetric analysis

Jenny Fry

Shefali Virkar

Ralph Schroeder

### *Approach and Methods*

In order to help determine whether the winner-take-all hypothesis applied to patterns of access to information we conducted a series of semi-structured interviews in a sub-sample of the original six global domains. These were Terrorism, HIV/AIDS, Climate Change, and Internet and Society. In total twenty UK-based active researchers were interviewed; five from each of the four domains. Interviewees were asked about their research background, key institutions, groups and people in their research networks, and the variety of online resources they used. Questions also focused on their online search strategies, such as the tools they used for finding information, the keywords they used and what kind of entities they tended to search for e.g. people, groups or institutions. The interviews also covered direct validation of specific aspects of the webmetric data. For this we asked respondents to comment on the Google representation of key institutions, organizations and people in their domain and the extent to which it overlapped with their own mental model of the domain e.g. their individual perception of what constitutes the core set of resources and sources (it is important to note that this is different to a particular mental model they may have at any one time in relation to a situational information need). The Google representation was derived during the webmetric data gathering and analysis by retrieving the top-thirty URLs in each domain from Google.com based on the following keywords:

|                      |                                                                 |
|----------------------|-----------------------------------------------------------------|
| Climate Change       | 'Climate change', 'Global warming', 'Ozone depletion'           |
| Internet and Society | 'Internet and society', 'Internet research', 'Internet Studies' |
| HIV/AIDS             | 'HIV/AIDS', 'HIV Infection', 'HIV prevention'                   |
| Terrorism            | 'Terrorism', 'Terrorist organisation', 'Terrorist network'      |

Any overlaps or inconsistencies between the Google representation of each domain and the participants' own mental model was further validated by coding the websites, institutions, organizations, people and other resources they reported using throughout the interview transcripts and then comparing this list with their responses to the Google representation. The interviews were recorded, transcribed in full and analysed using the Nvivo software for qualitative data analysis.

### *Main Findings*

- Geographic orientation of domain

The interviews revealed that there was only a limited overlap between the Google representation and researchers' own mental models of key networks, structures and organizations. Researchers reported that many of the key online resources in their domain were missing from the Google representation. The extent of the overlap appears to be domain dependent, with those researchers working within a more nationally orientated information environment reporting less of an overlap. For example, the HIV/AIDS researchers reported using national journals, national charity organizations, national statistics, and national public sector organizations, but none of these appear in the top thirty Google results for generic domain keywords (even when the search was repeated using Google.co.uk). Climate Change researchers, on the other hand, for whom the geographical boundaries of research were far more 'international', were able to recognise many more Google results on the Climate Change validation sheet.

- Networks of excellence

In addition to the gaps that participants identified in response to the Google representation a number of institutions, people and journals which researchers named as being leading academic institutions, people and other key organisations/resources in their field, did not appear in the top thirty Google results. This was particularly true for academic institutions and leading researchers named by the interviewees. In the cases where participants recognised some of the top sites from the list, or named key institutions, groups or people that did appear in the top thirty results, those identified were unlikely to appear in the top 10 results.

- Web-based search strategies

Despite recognition by respondents that Google is a blunt instrument in terms of seeking information, and UK-based researchers' recognition of a persistent US bias in its indexes, it was still the main tool for finding sources and resources on the web.

Though there were similarities in web-search strategies across each of the four domains, there were also important differences. For example, while respondents reported using Google almost to the exclusion of all other generalist internet search engines, the role that it played in their wider information environments varied considerably. In the HIV/AIDS and Internet and Society domains, for instance, Google is mainly used as what Beauvisage (2004) calls an "aide memoir", a locating tool for known sources. In contrast, for researchers of Terrorism, Google plays a more central role in exploring the object of research and identifying relevant sources. This may be due to the amorphous, shadowy nature of the subject matter itself – websites of terrorist groups and the message-boards, chatrooms and blogs associated with them are constantly being shut down by national intelligence agencies, only to resurface with new web-addresses, and the only way to locate these and other sources like them is for researchers to 'excavate' resources across a range of domain boundaries.

One possible explanation for differential domain patterns in the role of Google and other internet search engines as information seeking tools could be the extent to which important documents are scattered across domain boundaries (Bates). The consequence of this for web searching is that in low scatter fields, resources and sources can be found using a clearly circumscribed set of keywords and are likely to be produced by a limited number of dominant gatekeepers. Of the case studies, HIV/AIDS was the domain with the least scatter and this could explain why Google was used more as an 'aide memoir' than as an exploratory tool. Terrorism and Climate Change researchers on the other hand described their domains as scattered in terms of resources and respondents reported using Google for finding diverse sources more than in the other two domains.

- The Role of Gatekeepers

The characteristics and role of the predominant gatekeepers varied across each of the four domains. The interview responses indicate a differentiated shift towards the decentralization of gatekeepers on the Internet. For example, in climate change 'hybrid research centres' produce and disseminate important sources; and policy or academic research centres are key producers of information sources in internet and society research. In HIV/AIDS research, in contrast, although non-profit organizations were key producers and disseminators of information and played an important gatekeeping role, traditional gatekeepers such as publishers still maintain a central position in the information environment because of the continued importance of peer-reviewed articles disseminated through discipline-centric aggregated databases such as *PubMed*.

The information environment of the Terrorism researchers was similar to that of the HIV/AIDS researchers in that, while non-governmental and not-for-profit organizations play a central role in disseminating primary information resources, publishers still had an enduring role as gatekeepers to academic research. In terrorism, dissemination of research via books plays a major role in the scholarly communication system and still remains closely interrelated to the recognition and reward system. Research in terrorism is of a sensitive nature, which may account to some extent for the sustained importance of the traditional gatekeepers such as publishers. In contrast, the gatekeepers in the information environments of the climate change and the internet and society researchers were more decentralised. This meant that - rather than access to information being coordinated by a predominant gatekeeper - there were multiple gatekeepers providing specific resources in niche areas.

## Conclusions

As with quantitative webmetric results, our qualitative interview findings show that there is no uniform 'Winner-takes-all' effect in the use of online resources. Instead, there are different kinds of gatekeepers for the four topics we examined and for the types of information that are sought. It is therefore important not just to identify a concentration or democratization effect, but rather to refine under what circumstances the search for expertise will be dominated by certain results and exhibit biases, and when, instead, researchers will be led to the resources they seek and to a variety of results. Particular characteristics of a domain's information environment will determine whether Google and other Internet search engines function as a *facilitator* or as an *influential gatekeeper*.