Proceedings of the Nova Scotian Institute of Science (2010)
Volume 45, Part 2, pp. 91-101

2009 Graduate Student Award Winner

UNDERSTANDING INFLUENCE OF SCIENTIFIC INFORMATION IN THE DIGITAL AGE: A STUDY OF THE GREY LITERATURE OF A UNITED NATIONS ADVISORY GROUP

GREGORY R.G. HUTTON
School of Information Management
Dalhousie University
Halifax, Nova Scotia

INTRODUCTION

Today's seriously deteriorating global environmental conditions are pointing increasing attention to the urgent need for more informed use of scientific information in support of effective policy responses. Large bodies of scientific information that should inform policy decisions currently exist and accessibility is increasing in a dissemination milieu that is changing rapidly due to the evolution of the Web. Besides traditional peer-reviewed journals, researchers and information users now have greater opportunities to publish and access scientific information in a wide variety of forms, especially grey literature. Grey literature, a notable form of scientific output, is also being transformed by digital developments.

Briefly defined, grey literature is scientific information published outside of peer-reviewed journals and includes "material in print and electronic formats, such as reports, preprints, internal documents (memoranda, newsletters, market surveys, etc.), theses and dissertations, conference proceedings, technical specifications and standards, trade literature, etc." (Reitz, 2007). As authoritative scientific information, grey literature is often available from the websites of governments, non-governmental organizations, and many other publishing groups. However, greater accessibility to such scientific information due to emerging publication and dissemination methods may or may not lead to improved policy initiatives (de Alwis, 2006; Mitchell, Clark, & Cash, 2006). As a measure of whether marine science grey literature is reaching readers in research and policy communities of practice, this paper reports on a case study that is tracking use of reports of major significance on subjects of marine environmental protection. This study is aided by search tools that identify use as documented in developing digital contexts.

Email: greg.hutton@dal.ca

Unequivocal evidence shows that climate change, the most pressing environmental concern, is affecting natural systems (IPCC, 2007). Stress on the health of marine environments is one of the repercussions of climate change. For example, studies have shown that global sea levels will rise steadily as air temperatures rise and ice sheets and caps melt (Economist, 2009). In addition to problems caused by warming climates, marine systems also suffer from man-made pollution and unsustainable policies and practices, including fertilizer run off causing excessive eutrophication, pollutants such as oils and plastics entering the seas, and overfishing (Economist, 2009). Action to combat the effects of climate change and marine pollution before irrevocable damage has occurred could result from individual and public policy responses informed by authoritative scientific information.

Marine environmental concerns are the domain of many governmental and non-governmental organizations including the Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection (GESAMP), an international advisory body sponsored by the UN and seven UN-based agencies. Established in 1969 to "provide authoritative, independent, interdisciplinary scientific advice to organizations and member governments to support the protection and sustainable use of the marine environment" (GESAMP, 2008). GESAMP has produced many of its publications as grey literature in a "Reports and Studies" series (Pravdić, 1981; Windom, 1991; MacDonald, Cordes and Wells, 2004). Historically, GESAMP published its reports in print format, and since 2001 most reports have been made available as free, full-text files on the organization's website. Through thorough internal vetting and the requirement that all supporting agencies approve publications, GESAMP's technical reports are arguably more rigorously reviewed than typically occurs in peer-reviewed scientific journals. Thus, these reports, containing important findings, syntheses, and recommendations of global concern, are useful to a wide range of marine environmental policy and management initiatives.

Literature Review

Over the past decade several authors expressed positive outlooks about the influence of grey literature on the general communication of science. At the end of the 1990s Farace wrote that grey literature "in the coming century will be perceived and judged by the contributions it makes in resolving scientific and technical, as well as social problems facing the public and private sectors" (Farace, 1997, p. 73). He cited developments in electronic dissemination technologies and the "breakthrough of network publishing" as factors that will contribute to these developments (Farace, 1997, p. 73). This theme was picked up by Weintraub who pointed out the "relatively long period to effect change in a world that communicates mainly in print" (Weintraub, 2000, p. 55). The author then suggested that the increase in dissemination speeds allowed by internet technologies would "have implications for the formation of science policy and public attitudes in a more profound way than in the past" (Weintraub, 2000, p. 57). While Weintraub was no doubt correct in stating that increased speed of scientific communication has occurred because of the development of internet technologies, it is unknown whether the impact of "quick access" on scientific and public policy communities has occurred. Arguably, the speed of communication allowed by the internet has created a glut of information that has decreased users' abilities to discover relevant, timely information in the manner Weintraub envisioned.

A number of studies have attempted to show that free online access does indeed increase the use of the source. Vaughan and Shaw (2008) mentioned "mounting evidence that publications available on the web are cited more frequently than those that are less readily available" (p. 319). Vaughan and Shaw cite a 2001 study by Lawrence, wherein he argued that "the ability to locate relevant research quickly will dramatically improve communication and scientific progress" (Lawrence, 2001, p. 521). Working with conference articles in computer science, the study showed a considerably higher rate of citation for those articles available online than those that were not (Lawrence, 2001). Recently, Norris, Oppenheim and Rowland demonstrated a distinct citation advantage for open access articles, although they are unsure of why this is the case (Norris, Oppenheim, and Rowland, 2008). Even though the authors were able to establish that there were differences among the ways in which different scientific subjects are cited, they recommended more research to explain the underlying causes of the open-access advantage. While there is evidence that online availability of articles does increase their use and impact, careful consideration of how these documents are being cited needs to be carried out. Thus, we are left with the tenuous assumption that digital access does increase use which will need to be proven through direct analysis of individual incidents of citation.

Methodology

Citation analysis techniques were used to obtain data that track the use of GESAMP's grey literature. Data was collected from three sources: Thomson Reuter's Web of Science, Google Scholar, and Google. Whereas Web of Science citation data have traditionally been used to track uses of scientific publications (Thelwall, 2008), data collected from the three sources enabled building a more complete understanding of GESAMP's influence particularly with regard to publishing and dissemination made possible by developments of the web. Citations found via searches in Google Scholar and Google can track trends in web publishing and Web influence through information sources not indexed by Web of Science. Google Scholar focuses on sources in many disciplines ("peer-reviewed papers, theses, books, abstracts and articles, from academic publishers, professional societies, preprint repositories, universities and other scholarly organizations") (Google Scholar, 2009), and Google provides access to the open Web.

Detailed searches for citations were conducted in Web of Science following a methodology developed by Cordes (2004) to construct a citation dataset for GESAMP reports through to 4 February 2009. Each citing document located in Web of Science was examined to confirm citations to GESAMP reports and then a record of the citing document was entered into a Procite database where each record was coded for ease of analysis (Appendix Table 1).

From the Web of Science dataset the ten most cited and ten least cited were selected as a sample to trace citations to GESAMP's reports in Google Scholar and Google. The title of each report (entire title within quotation marks) was entered into each search engine to conduct phrase searches. When a high number of hits occurred (usually because the title was commonly used in non-GESAMP documents), the GESAMP acronym was added to the search string to narrow the focus to explicitly GESAMP grey literature reports. Only citations to the reports were collected, not citations to the small number of books or journal articles published by GESAMP members based on the reports. Preferences were set in Google Scholar and Google to maximize the efficiency of data collection. Only English language hits were retrieved, and if a hyperlink was not available in a hit to allow investigation of the source, the hit was omitted. To avoid false positive results, each link was opened and the citation to the GESAMP report located. When links were confirmed as citations to GESAMP reports, information was gathered from each citing link, coded, and entered in a Procite database for analysis.

Analysis and Discussion

Aggregate citation data collected from Google Scholar and Google searches for the ten GESAMP reports most cited in Web of Science and the ten least cited reports are presented in Tables 1 and 2. For the ten most cited reports (Table 1), the data show a strong correlation between the total number of citations in Web of Science (576) and Google Scholar (589). The total number of citations retrieved from Google (481) is not markedly different. These aggregate figures suggest that highly ranked reports based on Web of Science cita-

Table 1 Google Scholar and Google Citations: Top Ten Reports

		Number of Citations			
Report Number	**Web of Science**	**G.S. Total**	**G.S. Unique**	**Google Total**	**Google Unique**
38	86	49	11	29	24
32	83	48	10	38	24
39	80	139	55	92	86
50	74	57	27	43	35
61	51	62	40	53	46
06	45	28	16	32	28
57	42	49	18	51	42
28	41	19	9	18	16
71	37	85	52	95	87
58	37	53	26	29	27
Totals	576	589	264	481	416

Table 2 Google Scholar and Google Citations: Bottom Ten Reports

		Number of Citations			
Report Number	**Web of Science**	**G.S. Total**	**G.S. Unique**	**Google Total**	**Google Unique**
23	3	2	1	7	6
16	3	8	7	16	16
29	3	0	0	13	12
11	3	1	1	19	19
07	3	2	2	11	11
05	2	3	1	18	18
75	2	6	4	20	19
20	0	1	1	15	15
36	0	2	2	6	6
76	0	13	13	17	17
Totals	19	39	34	142	139

tions will also rank highly in Google Scholar and Google citations. However, of the 589 citations retrieved from Google Scholar, 264 were different than those found in Web of Science; similarly, of the 481 citations obtained from the Google search, 416 were different than those found in Web of Science. These unique citations in Google Scholar and Google demonstrate that GESAMP's reports are receiving wider attention than is noted by the peer-reviewed journals indexed by Web of Science.

For the ten least cited reports based on Web of Science citations (Table 2), the data show a greater number of citations in Google Scholar (15 more than in Web of Science) and in Google (about seven times more – 139 citations not found in Web of Science). This pattern suggests that GESAMP reports that traditionally obtained the least attention in peer-reviewed journals are now receiving increased notice on the Web.

Further examination of the citations retrieved from Google Scholar and Google reveal additional characteristics of the citing sources. Four reports were chosen for further analysis, based on similar publication dates, subject matter, and the total number of Web of Science citations they received. One older and one newer report were chosen from the ranked list of most cited GESAMP reports in Web of Science and matched with two of the least cited reports. Matching was based on publication dates and similar subject matter to ensure that both of the most cited GESAMP reports were paired with similar least cited reports.

The citations for all four reports obtained from Google Scholar and Google were examined and classified according to the types of documents (see Tables 3 and 4 and Appendix Table 1). The three most numerous categories found in Google (Table 3) are general bibliographies, subject specific bibliographies, and reports. Other less common types include online papers, online books and book chapters, commercial websites selling GESAMP reports, conference proceedings, meeting documents, and proposals. A similar range of citing documents was found in Google Scholar (Table 4), but a dissertation was identified and fewer citations from general bibliographies and subject bibliographies compared to Google were located.

Whereas all citing articles retrieved from Web of Science represent intellectual influence (usually from a scientific perspective due to the focus of the majority of journals indexed by Web of Science), citations retrieved from Google Scholar and Google show a spectrum of influence. Citations in journal articles, technical reports, online books, and meeting documents retrieved from Google Scholar and Google all represent direct use of GESAMP reports. For, example, the sizeable number of technical reports that cite the GESAMP reports (27 and 24 in Google and Google Scholar respectively) highlight intellectual influence. In contrast, the large number of citations in bibliographies (including library catalogues, publishers' websites, GESAMP's own website) indicate awareness of GESAMP's reports but do not confirm that the reports contributed to other scientific studies or to public policy. In the search results from Google, these more perfunctory citations represent 28% of the total (46 out of 165).

Web data must be carefully analyzed before a fair comparison is to be made. Perfunctory bibliographic citations are more prominent for the older reports (Table 5); 85% of the Web citations to report #5 are bibliographic, and 38% of the citations to report #6. By

Table 3 Google Citations to Selected GESAMP Reports

Report Number	Web Science	Google Exports	Biblio	Rep	Subject biblio	Book	Book Chap	Comm	Conf	Jour	Meet	O. P.	Other	Prop	WoS
06	45	32	12	3	1	6	4			1		1			4
71	37	95	15	21	9	9	6	2	4	1	3	7	8	2	8
05	2	18	15	2				1							
75	2	20	4	1	3				3	1	4		3		1

Table 4 Google Scholar Citations to Selected GESAMP Reports

Report Number	Web Science	G.S. Exports	Biblio	Rep	Book Chap	Comm	Conf	Diss	Jour	Meet	O. P.	Other	WoS
06	45	28	1	2	2	3				7			1
12													
71	37	85	1	20	2	12		5	1	7	3	1	
33													
05	2	3		1									
2													
75	2	6		1			1	1		1			2

Table 5 Reports analysed as per perfunctory bibliographic citation.

Report Number	GESAMP Report Title
06	Impact of Oil on the Marine Environment (1976)
71	Protecting the Oceans from Land-based Activities (2001)
05	Principles for Developing Coastal Water Quality Criteria (1977)
75	Estimates of Oil Entering the Marine Environment from Sea-based Activities (2007)

comparison, the newer reports have received fewer bibliographic citations: 16% for report #71 and 20% for report #75. This difference between older and newer reports highlights a broader spectrum of use of more recent publications.

The higher number of bibliographic citations returned in the Google search compared to Google Scholar confirm that the Web sources featured in Google Scholar are focused more towards academic types of information, rather than perfunctory sources, such as publisher's websites. As a result citations resulting from Google Scholar searches are generally indicative of more substantive use of GESAMP's reports.

The widest variety of types of citing documents is illustrated by report #71. Most of the citing documents illustrate substantive use. For example, the report is cited by "Harmful algal research and response: A national environmental science strategy 2005-2015" which is written to support the needs of American scientific and management communities in combating the proliferation of harmful algae blooms (Ramsdell, Anderson, & Glibert, 2005). Similarly, *Japan's position as a maritime nation* shows significant use of report #71 among other GESAMP reports (Hooi, 2007). The online book *Millennium Ecosystem Assessment Volume 3: Policy Responses* refers to GESAMP's scientific findings, namely, the connections between sewage and health problems (Millennium Ecosystem Assessment, 2005). Finally, "The role of policy in the conservation and extended use of underutilized plant species: A cross-national policy analysis" confirms explicit use in a public policy context (Chishakwe, 2008). In each of these examples the citing documents clearly show that report #71 is consulted for scientific and policy-oriented purposes.

CONCLUSION

Tracking the use and influence of scientific information has traditionally relied only on Web of Science as a source of citation data. However, to gain a more comprehensive understanding of the influence of scientific information in the digital age, it is necessary to examine citation evidence drawn from the web. This study illustrates a sizeable number of citations representing GESAMP's influence which are not included in Web of Science data. The significance of citations outside of Web of Science has been shown by the uses of GESAMP's reports in both science and policy contexts. The expansion of the Web in the last five years has increased the visibility and access of scientific information, particularly grey literature which many organizations produce. Evidence from citations drawn from the Web confirms both awareness and substantive uses of GESAMP's reports in conveying important scientific findings as the basis for policy recommendations. Google Scholar and Google are important tools for informing a more complete understanding of the influence of GESAMP's reports. With increasing reliance on open access via the Web, GESAMP, along with other producers of scientific grey literature, are well advised to provide unfettered access on the Web to ensure use. Similarly, to understand the influence of scientific grey literature through citation analysis, the Web must be included as a source of citations rather than relying solely on Web of Science.

Acknowledgements This paper is based on a master's thesis by Hutton, supervised by Dr. B.H. MacDonald of the School of Information Management, Faculty of Management, Dalhousie University, and supported by a research grant from the Social Sciences and Humanities Research Council of Canada to Dr. B.H. MacDonald and Dr. P.G. Wells.

REFERENCES

Chishakwe, N.E. (2008). *The role of policy in the conservation and extended use of underutilized plant species: A cross-national policy analysis*. Rome: Global Facilitation Unit for Underutilized Species (GFU). Retrieved February 15, 2009, from http://www.underutilized-species.org/Documents/PUBLICATIONS/cross_national_policy_analyses.pdf

Cordes, R. (2004). Is grey literature ever used? Using citation analysis to measure the impact of GESAMP, an international marine scientific advisory body. *Canadian Journal of Information and Library Science,* 28:45-65.

de Alwis, G. (2006). Transformation in managers' information seeking behaviour: A review of the literature. *Journal of Information Science,* 32:362-377.

Economist. (2009). Troubled waters. *Economist,* 390(8612):3-4.

Farace, D.J. (1997). Rise of the phoenix: A review of new forms and exploitations of grey literature. *Publishing Research Quarterly,* 13(2):69-76.

GESAMP. (2008). Its mission. Retrieved from http://s244621454.onlinehome.fr/common.php

Google Scholar. (2009). About Google Scholar. Retrieved from http://scholar.google.ca/intl/en/scholar/about.html

Hooi, T.K. (2007). *Japan's role as a responsible maritime nation – Enhancing global presence and commitments (with special reference to resource exploitation and conservation).* Retrieved February 15, 2009, from http://www.iips.org/07mar/07marTan.pdf

IPPC. (2007). *Climate change 2007: Synthesis Report: Summary for policymakers*. Retrieved February 1, 2009, from http://www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4_syr_spm.pdf

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature,* 411:521.

MacDonald, B.H., R.E. Cordes and **P.G. Wells.** 2004. Grey literature in the life of GESAMP, an international marine scientific advistory body. *Publishing Research Quarterly* 20(1):25-41.

Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being Volume 3: Policy responses*. Island Press. Retrieved February 15, 2009, from http://books.google.ca/books?id=dPLzt5uyNAcC

Mitchell, R.B., Clark, W.C., & Cash, D.W. (2006). Information and influence. In R.B. Mitchell, W.C. Clark, D.W. Cash & N. M. Dickson (Eds.), *Global Environmental Assessments. Information and Influence* MIT Press: Cambridge, MA, pp. 307-338.

Norris, M., Oppenheim, C., & Rowland, F. (2008). The citation advantage of open-access articles. *Journal of the American Society for Information Science and Technology,* 59(12):1963-1972.

Pravdić, V. (1981). *GESAMP, the first dozen years*. Geneva: United Nations Environment Programme.

Ramsdell, J.S., Anderson, D.M., & Glibert, P.M. (2005). HARRNESS - Harmful algal research and response: A national environmental science strategy 2005–2015. Washington, DC: Ecological Society of America. Retrieved February 15, 2009, from http://www.whoi.edu/cms/files/HARRNESS_18189_23044.pdf

Reitz, J.M. (2007). *Online Dictionary for Library and Information Science*. Libraries Unlimited. Retrieved from http://lu.com/odlis/odlis_g.cfm#grayliterature

Thelwall, M. (2008). Bibliometrics to webometrics. *Journal of Information Science,* 34(4):605-621.

Vaughan, L. & Shaw, D. (2008). A new look at evidence of scholarly citation in citation indexes and from web sources. *Scientometrics,* 74(2):317-230.

Weintraub, I. (2000). The impact of alternative presses on scientific communication. *The International Journal on Grey Literature,* 1(2):54-59.

Windom, H.L. (1991). *GESAMP: Two decades of accomplishments.* London: International Maritime Organization.

Appendix Table 1

Code	**Source of Citations**
Biblio	Bibliography (general list of publications, library catalogue)
Rep	Reports (technical reports or reports written for government or interest groups).
Subject Biblio	Subject specific bibliographies
Book	Online books
Book Chap	Specific chapters of online books
Comm	Commercial websites
Conf	Conference-related documents (proceedings, conference papers, workshops)
Diss	Master/PHD theses or dissertations
Jour	Online journal article; journal NOT indexed by Web of Science
Meet	Meeting documents
O.P.	Online paper (without evidence it is from an online journal, conference, etc)
Other	Other sources (blogs, debate notes, news stories, educational websites, and concept notes)
Prop	Proposal documents
WoS	Citing article also appears in a journal indexed by Web of Science