

Scientific Grey Literature in a Digital Age: Measuring its Use and Influence in an Evolving Information Economy

Abstract: This paper outlines methodologies to improve understanding of the influence of grey literature published in print and digital formats. The study is based on analyses of citation data regarding the UN-based Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection collected from Web of Science, Google, and Google Scholar.

Résumé:

Acknowledgements: This research was supported by a Social Sciences and Humanities Research Council of Canada grant (#410-2007-2012) awarded to Dr. Bertrum H. MacDonald and Dr. Peter G. Wells. This paper is based on research for my thesis in the Master of Library and Information Studies programme in the School of Information Management, Dalhousie University. I am grateful to my chief advisor, Dr. Bertrum MacDonald, and to my committee members, Dr. Janice Graham and Dr. Peter G. Wells, for their advice.

I. Introduction

Over the last twenty-five years the quantity of scientific literature has grown substantially. Two distinct types of publications are important components of this body of literature: peer-reviewed journals and alternative forms of publication, including documents that are considered grey literature. 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). Grey literature is often freely available from Web sources but questions about awareness and accessibility about these publications continue to occur in the current milieu of scientific communication that is characterized by an abundance of potential sources; even if grey literature is readily available, publications may still go unnoticed.

Citation analysis based on Thomson Reuters’ Web of Science has been used extensively as a tool to provide evidence of the use and influence of traditional literature (Bar-Ilan 2008). Web of Science is an important tool for collecting citations, but as its coverage focuses on peer-reviewed journals, it is limited in its ability to show the influence of grey literature in a comprehensive manner. Using a case study of a publisher of major reports, the UN-based Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection (GESAMP), this paper proposes a multi-component metric that illustrates how a more complete understanding of the use and influence of grey literature can be constructed. This metric draws on several sources of citation data, analyzed from a number of angles, including Web of Science, Google, Google Scholar, and the open web. These sources demonstrate the extent of use of grey literature by providing citation

datasets that are overlooked by traditional citation analysis techniques. Web of Science citation data are discussed in relation to other citation sources to create a more complete picture of influence. This approach increases the potential for understanding where and how grey literature is used, which Web of Science data alone cannot provide.

II. Literature Review

Traditional peer-reviewed journals and grey forms of literature have interacted in a publication milieu that experienced rapid evolution in the closing decade of the twentieth century, a pattern which shows no signs of slowing in the new millennium. Two main developments in publishing practices contributed to this change: “the computerization of the printing process, reducing costs significantly and allowing more journals and books to appear in print; and the conversion of the entire publishing cycle (submission of articles, refereeing and publication) to the internet, allowing faster and possibly cheaper communication throughout” (Thelwall 2008, 605). The resulting publication environment is characterized by an over abundance of potential sources that users must sort through to locate appropriate information resources.

Measurement of the use and influence of scientific output has relied to date on citation analysis techniques that are based on the information available from a single tool, the Science Citation Index (SCI) (Bar-Ilan 2008; Cronin 2001, 2005; Thelwall 2008). Since its invention by Eugene Garfield in 1955, the Index has developed into today’s Thomson Reuters Web of Science database, but the original intent of the system has been maintained by capturing citation data in each paper published in carefully selected journals. The SCI lists all references contained in each paper, as well as references that cite the paper itself. In explaining the purpose of the Index, Thelwall states that it “was created as a database of the references made by authors, to earlier articles, in their articles published in the top scientific journals” and argues that the “underlying idea, [which is] still highly relevant today, is that if a scientist reads an article, then s/he would benefit from knowing which articles cited it, since they may cover a similar topic and might update or correct the original article” (Thelwall 2008, 606).

Citation data available from Web of Science is no doubt an important indicator of the communication of scientific knowledge. However, this data is limited in its ability to determine the total use and influence of all published scientific output. These limitations are largely related to the evolution of Web technologies and the bevy of publication avenues and dissemination tools fostered thereby (i.e., open access online journals, institutional repositories, technical reports, authors’ homepages, etc). Grey literature is a notable exception in the list of important sources that Web of Science does not index. While it is possible to locate citations to some grey literature publications using Web of Science, the system is not designed to study this literature to the same degree as the selected journal literature. In short, the Web of Science’s indexing practices do not encompass the full scope of citation data available in the current state of scientific publication; significant quantities of citation data are not included in Web of Science, and therefore are unavailable for traditional citation analysis.

The indexing practices of Web of Science focus primarily on scientific journals. Subsequently, the tool is not as applicable in determining the use and influence of publications in other areas of study. For example, some authors have pointed out that

Web of Science cannot be used to effectively gather data about social science and humanities citations (Kousha and Thelwall 2007b). A major component of Kousha and Thelwall's argument centres on the number of sources of citations that show the influence of social sciences and humanities research not covered by Web of Science, including unpublished communications such as "conference presentations, keynote talks, e-mail lists, and panel discussions" (Kousha and Thelwall 2007b, 496). While this argument focuses on the limitations of Web of Science for measuring the influence of social science and humanities research, the same limitations can also be recognized in its coverage of scientific literature, especially grey literature.

The recent study of Web-based citations, known as webometrics, represents steps that have been taken to move beyond traditional approaches in citation analysis to gain a more holistic understanding of influence. Several researchers completed studies using Google and Google Scholar as the benchmarks for webometrics, including Vaughan and Shaw (2005; 2007) and Kousha and Thelwall (2007a). In their initial study, Vaughan and Shaw justified their use of Google by stating that it is both the "largest Web search engine" as well as the most stable (Vaughan and Shaw 2005, 1078). In another study Kousha and Thelwall gave the same reasons for using Google (Kousha and Thelwall 2007b). Both studies used phrase searching of article titles (and other indentifying information as necessary) to find web citations. Many of these recent steps are based on the conventional standards of citation analysis. As Thelwall insightfully surmised, "mainstream bibliometrics has evolved rather than undergone revolutionary change in response to the web and web-related developments" (Thelwall 2008, 607). This development is illustrated by studies that have found a direct correlation between citation searches completed within Google and Google Scholar to the results obtained via the traditional source of citations, namely Web of Science (Kousha and Thelwall 2007a; Vaughan and Shaw 2005). In their critique of Web of Science-based citation analysis, researchers have not expressed opinions that the approach should be abandoned and a new system created, but instead pay respect to previously established norms.

For well over a century, the results of scientific research have been reported in peer-reviewed journals, and publication in such journals remains among the most visible and prestigious venues available to scientists. Cronin wrote in 1984 that "scientists may be less than totally satisfied with the scholarly journal as a dissemination mechanism... [but] they are deeply attached to it as a means of preserving a faithful and reliable account of scientific progress; as a repository of accepted ideas and beliefs" (Cronin 1984, 12). However, over the last twenty years considerable shifts have occurred in publishing practices based primarily in evolving web technologies such as free accessibility of grey literature and open access materials, institutional repositories, pre-print archives, and publications placed directly online by scientists. Citation studies similar to those that attempt to measure the influence of scientific papers published in journals have been conducted for these emerging technologies, typically within webometrics.

Although grey literature has played a role in disseminating scientific knowledge for many decades, much of it is now more widely available and accessible than ever before. Whereas scholarly journal articles may have been regarded as the pinnacle of scientific communication, recent developments in publishing and attitudes about communication are turning more to open-access and grey literature in large part due to developments in accessibility. The proliferation of Web technologies and their increasingly widespread use has ushered in a new era of grey literature relevance and accessibility (Farace 1997;

Gelfand 1997; Weintraub 2000). However, even if grey literature is more accessible than ever, it cannot be assumed that its ready availability equals awareness of its production or subsequent use by researchers and decision makers. Moreover, data available in Web of Science cannot adequately illustrate use of grey literature. In contrast, citation analysis that embraces recent publication developments can be used to more fully understand the use and influence of important scientific information available primarily in grey literature.

III. Methodology

Searches for citations to GESAMP's technical reports were conducted in Web of Science, Google, and Google Scholar. Sponsored by the UN and seven UN-based agencies, GESAMP's formal title is IMO/FAO/UNESCO-IOC/WMO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection. This intergovernmental advisory body was selected for this study because it is an organization that publishes important information about marine environmental topics that can inform both scientific work and public policy decisions. It was 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" (<http://www.gesamp.org>). GESAMP's "Reports and Studies" series of publications contain important findings, syntheses, and recommendations of global concern that fulfill this mandate. Given the requirement that all supporting agencies and their technical specialists review and approve the reports, it can be argued that this review process is significantly more rigorous than typically occurs in a scholarly peer-reviewed scientific journal. GESAMP has traditionally published its documents in print format, but free, full-text files of each report have been available from its website for over five years. Some reports have been republished as journal articles or have served as the basis for books, and in some instances, the reports have been co-published by other UN agencies in similar formats, such as United Nations Environment Programme's Regional Seas series. The majority of GESAMP's publications fall into the category of grey literature.

Locating citations to GESAMP's technical reports in Web of Science required a variety of search strings to account for the numerous ways the publications have been cited in the indexed literature. Citing authors sometimes mistakenly identified the sponsoring agencies as publishers of the reports or attributed authorship to a sponsoring agency, or in some cases misspelled GESAMP's acronym in their articles. All of these deviations are indexed differently in Web of Science, necessitating separate search strategies to locate citations (Cordes, 2004). Accounting for all variables required an assortment of "Cited Author" and "Cited Work" strategies in Web of Science. Difficulties in searching were compounded by GESAMP's publication history, as the agency has published 77 technical reports to date in its *Reports and Studies* series. Titles of each report were also used during the citation searching process in order to account for misattributed citations. Once citations were located, information from the citing articles was entered into a ProCite database and coded according to the GESAMP publications they cited. Coding facilitated analysis in both the Procite database and a Microsoft Access database, where the information was exported to allow queries to specific questions. To determine whether the authors of citing documents already familiar with GESAMP through a relationship with the organization cited GESAMP reports more or less than authors without a relationship, a database of the names of individuals with some direct involvement was

compiled from names listed in each of the technical reports, meeting documents, published histories of the organization, and the organization's website. Names were entered into the database if an individual was a scientific member of GESAMP, a member of a working group that contributed to the production of a report, an observer of a meeting, a reviewer of a technical document, or a member of the secretariat staff of one of the UN agencies that sponsor GESAMP.

Citation data collected from Web of Science was used as a benchmark for Web search strategies. The Web of Science data was used to rank GESAMP's grey literature reports in terms of the number of citations each had received. The ten most frequently cited reports in GESAMP's publication history were identified and the title of each was then entered within quotation marks in Google and Google Scholar searches along with the acronym GESAMP to ensure accuracy of results. The search results for each report title were examined individually to confirm that each represented a valid instance of a citation to a GESAMP report, thus preventing the collection of false-positive hits. A citation was accepted if the title of the report was present somewhere in the resulting hit and was obviously related to GESAMP. For example, one GESAMP report is entitled "The State of the Marine Environment," which is a phrase prevalent throughout marine environmental literature and is not specific to GESAMP. Results that included this phrase but had no obvious reference to a GESAMP report were discarded. Pertinent bibliographic data for each valid result was entered in a ProCite database, including author, title of document or website, publisher, date of publication, and stable URL where available for each category. Since standard bibliographic data such as author and date of publication are often not available from sources on the Web, many records do not contain such data.

Valid hits obtained from Google and Google Scholar searches were coded according to the "type" of citation they represented as they were entered into a ProCite database. Classification types included "bibliographic references," such as cases where publishers simply listed all of GESAMP's reports or instances of records retrieved from an online library catalogue. This type of citation was considered perfunctory since no direct evidence of use was apparent. Other types of citations showed clear use or influence including citations in reports, online papers, books or book chapters, meeting documents, websites, and articles that had already been retrieved from Web of Science searches.

Searches for GESAMP's acronym were conducted in Google and Google Scholar to locate connections between information producers and GESAMP's publications not revealed by searches for citations to titles of its technical reports. For the purposes of this study, 100 search results for "GESAMP" were collected from each of Google and Google Scholar. The initial searches in Google and Google Scholar returned 36,700 and 2,440 results respectively. A sample of 100 results was selected in each case. Systematic samples were chosen from the total number of results the two search engines identified as "unique" hits (445 in Google and 988 in Google Scholar). A sampling interval was identified that would achieve a sample size of 100 in each case and which ensured websites were selected throughout the full list of "unique" results, rather than focusing on the top hits ranked by the Google and Google Scholar algorithms. A sample size of 100 was considered sufficient for the purposes of illustrating GESAMP's Web presence represented by notation of its acronym.

A Google “link” search (“link:www.gesamp.net”) was performed to locate websites that link to the GESAMP website. Since the URL www.gesamp.net is an alias for an underlying web address (<http://s244621454.onlinehome.fr>), several variations of possible URLs were tested to ensure that the resulting set included all links to the GESAMP website. Bibliographic data for each citing website was entered in a ProCite database for further analysis and to ensure that the search results would not be lost if a website were to be moved or taken down by its owner.

IV. Analysis and Discussion

Citations Obtained from Web of Science Searches

Data collected from Web of Science covering the publication history of GESAMP show that up to the end of 2008 there were 2,631 citations to the technical reports published in its *Reports and Studies* series. Each instance of a GESAMP publication listed in the references in a paper was considered a single citation, even if the publication was cited more than once in the paper. With over 2,600 citations, it is clear that GESAMP’s reports were used in some capacity, although the total figure by itself does not reveal details about patterns of use or the degree of influence. Further understanding can be obtained, however, by looking more closely at the data. As Figure 1 shows, from GESAMP’s inception in 1969 through to 1992 the number of citations fluctuated between 2 and 44 per year in a relatively flat trend line. Then, beginning in 1992, a substantial increase in the number of citations per year occurred followed by a trend that peaked in 2002 with 195 citations. Since 2002 the frequency of citations per year has leveled off at about 160, suggesting that the trend has reached a plateau.

Analysis of year by year citation data may identify cause and effect relationships. For example, a publication may be cited more frequently than others and contribute more to overall citation totals in years subsequent to its release. Publication and dissemination practices may explain the peaks and valleys of yearly citation totals. In the case of GESAMP, such a change in publishing practice occurred at the beginning of the current decade and may have had an effect on the recent yearly citation totals.

Annual citation data can be examined more closely to determine which publications received citations in a given year. By relating this information to publishing events, insights about the peaks and valleys of total citation may be discovered. Figure 2, which identifies each of the GESAMP publications cited in 1992, shows a small number of GESAMP reports received the majority of the citations and a larger number of reports received a low number of citations each. Year by year citation data laid out in the manner shown in Figure 2 can reveal which of an organization’s publications contribute the most citations to each yearly total. Citation patterns may emerge that can be explained by the subject matter of publications. As shown in Figure 2, the publications that were major contributions to the citation totals were based on GESAMP reports number 38 and 39 (which were also republished in book and journal article forms). Reports 38 and 39 were first published in 1989 and 1990, respectively. By 1992 not only had the reports been republished in other formats, but the passage of time allowed uptake in the currents of scientific communication and opportunity to gain popularity. The subject matter of the reports may also explain the higher numbers of citations, with report 38 dealing with atmospheric input of chemicals to marine systems and report 39 serving as a complete

overview of the state of marine environments. The timeliness of GESAMP's reports is reflected by their higher citation totals.

Identification of yearly citation trends may offer insights into the time lapse, from the date a document is published to when it is cited. If patterns are evident, increased understanding of the use of grey literature publications can occur. Additional trends in citation data may be determined if the citations to each publication are isolated and tracked over time. This technique will show which of an organization's publications have received the highest number of citations and if citations occur over an extended period of time, attention will be drawn to publications that may be considered more influential than others in a group's history.

Author information in Web of Science data can be probed to increase understanding about the use of grey literature. Information extracted from Web of Science citation data permits querying whether citing authors already familiar with GESAMP's publications, due to involvement with the organization, cite GESAMP reports more frequently than other authors. A comparison of the names of authors of citing documents to the names in the database of individuals associated with GESAMP shows that articles with at least one author with previous involvement contributed 627 of the 2,631 total citations. In other words, less than one-quarter (23.8%) of citations could be considered as originating from people "informed" about GESAMP because of their relationship to the organization. Conversely, more than 75% of citations were contributed by articles authored by individuals who had no direct involvement with GESAMP. This finding is particularly informative for understanding the influence that publications receive outside the "inner circles" of individuals responsible for their creation. Citation data can be partly contextualized by determining the frequency of citation from "informed" authors. Citations that originate from authors who are not so informed may be of particular interest because these citations may be clearer indicators of the extent of influence of an organization's publications.

Characteristics of citing journals provide additional insights about uses of grey literature. As shown in Table 1, the top five journals, according to the frequency of citation to GESAMP publications, are primarily scientific in nature. Fewer citations are found in policy related journals, such as *Marine Policy*, which may be a reflection of the preponderance of scientific journals covered by Web of Science's indexing practices. Citation sources can also be ranked in terms of subject, since Web of Science assigns at least one relevant subject descriptor to each journal. In the case of GESAMP, Table 2 shows that the top five most frequently occurring subjects are scientific. This ranking is similar to the findings drawn from Table 1 and suggests that GESAMP reports are most commonly used in scientific contexts. However, given the journal selection criteria used by Web of Science, which emphasizes peer-reviewed research literature, evidence from other sources may point to additional uses of the reports.

Citations Obtained from Google Scholar Searches

Search results from Google Scholar highlight the importance of using Web-based citation data for determining uses and influence of grey literature. A total of 587 citations to the ten GESAMP technical reports selected on the basis of Web of Science data were located from Google Scholar. The 587 citations were compared to citations obtained in Web of Science to identify sources duplicated between the two citation sources. Citations already

identified in the Web of Science searches were ignored. Of the 587 citations, 260 were unique to Google Scholar, meaning there was considerable overlap (55.7%) of citations available from both services. Further analysis was performed on the unique results to identify the type of citation source each represented.

Given the proprietary nature of Google Scholar's indexing practices, it was unknown at the beginning of the study what portion of the search results would provide evidence of influence. Search results were coded as perfunctory if they included an obvious reference to a GESAMP report, but were simply publication lists or records retrieved from library catalogues. Google Scholar searches for the ten most cited GESAMP publications in Web of Science only produced three perfunctory citations out of a total of 587 citations (see Table 3). This small number of perfunctory citations was represented by publication lists on either GESAMP or a related UN agency website.

The 257 citations located in Google Scholar that do show influence were represented by a variety of citation sources: reports (26.9%), books and book chapters (20.4%), online journals (22.6%), and conference or meeting papers (8.2% and 3.5%, respectively). These 257 citations include scientific sources previously undiscovered and evidence of GESAMP documents cited at policy related meetings and conferences. To gain a more complete understanding of use and influence of such grey literature, these Google Scholar citations show that moving beyond Web of Science is necessary.

More recent GESAMP publications received higher citations in Google Scholar than in Web of Science, in contrast to earlier technical reports. For example, Report # 71, published in 2001, was cited 43 times in Web of Science compared to 85 citations retrieved in Google Scholar. Of the 85 Google Scholar citations, 52 were not duplicated by Web of Science data and only one citation was considered perfunctory. This example suggests that determining document use through citation data obtained in Google Scholar may be especially relevant for newer information. This point is supported by evidence of citation totals in Google Scholar for older reports. GESAMP Report # 28, published in 1986, has received 42 Web of Science citations compared to 19 from Google Scholar, only nine of which are unique to Google Scholar.

Citations Obtained from Google Searches

Google searches were informative in highlighting how GESAMP's influence is reflected on the open Web (see Table 4). Given the extensive quantity of Web ephemera, it was necessary to carefully classify each Web result obtained in Google searches in order to confirm that it was related to GESAMP. Of the 466 results obtained in Google searches, 400 were determined to be unique to Google when compared to Web of Science data. In contrast to Google Scholar results where only 44.3% were unique to Google Scholar, 85.8% of the Google searches were unique to Google. This discovery indicates that citations from Google provide further distinctive indicators of use, a finding which is no doubt attributable to the indexing practices of the search engine as well as the nature of publishing in the wider web environment.

Google search results were much more likely to be perfunctory than those obtained in Google Scholar. Of the 400 unique hits, 117 or 29.3% were deemed to fit the perfunctory category, consisting of sources such as library catalogues or publication lists. This number of perfunctory sources indicates that citation counts obtained in Google searches

are more likely to show lower level influence than citations obtained in Google Scholar or Web of Science.

Further analysis of the 283 unique, non-perfunctory search results located in Google are required in order to understand the context in which GESAMP's reports are used. Coding of the 283 results indicates findings similar to the Google Scholar searches. Reports, books and book chapters, online papers and journals, as well as conference and meeting documents have all been identified. Analysis of such citing documents also addresses questions of use and influence, and demonstrates the importance of including Google search results in a composite metric of indicators of uses of grey publications.

Citations Obtained from Acronym Searches in Google and Google Scholar

The 100 hits obtained by searching for the acronym "GESAMP" in Google illustrate GESAMP's profile in the open web and provide additional citations to GESAMP publications. A small number (4) of the results were duplicates of citations located in Web of Science data. Beside the duplicates of Web of Science citations, which are known to show use in scholarly contexts, the remaining 96 Google search results are more unique. Perfunctory citations (e.g., bibliographies) account for 14 hits, but a number of these were from governmental sources, notably several entries from Australian government bodies. While these results are at best minor indicators of use, they still show that GESAMP's documents are being recommended and thus some degree of influence can be noted. Other Google search results are stronger indicators of use, including citations in governmental reports, workshop presentations, and educational websites that draw on GESAMP reports to identify marine environmental problems and to recommend potential solutions. Overall, most of the citations obtained by searching on the acronym in Google did show substantial use of GESAMP documents. As a consequence, understanding of GESAMP's online presence is bolstered by examining the unique results returned in this Google search.

Almost 50% of the sample of the acronym search in Google Scholar (47 out of 100) was duplicates of citations located in Web of Science. However, the remaining 53 hits included citations from law journals not indexed by Web of Science, conference papers, reports, and online papers. Since Google Scholar purports to index scholarly sources, citations from more academic subjects are to be expected. Citations from law, science, and policy contexts in a variety of forms of publications (dissertations, conference proceedings, book chapters, online papers, reports, etc.) suggest that GESAMP's influence extends into a variety of scholarly disciplines. In total the search sample from Google Scholar gives further evidence of GESAMP's web presence and illustrates an additional contribution to the metric for documenting use of grey literature.

Link Searches

In total, 19 websites were located with links to GESAMP's website (see Table 5). Links to websites signify a relationship that mirror citations in documents, since they show that website authors have made a judgment about the relevance of the linked website and a link often implies a recommendation about other web-based sources to aid information users about a topic. Understanding where such links originate and their purpose helps clarify the types of relationships present in web links.

Of the 19 links, nine originate from the websites of UN-based agencies. These agencies include FAO (three websites), IMO and UNESCO (two websites each), as well as UNEP and the WMO (one website each). These websites exhibit an interesting array of linking motivations, including a recommendation about GESAMP literature for purposes of increasing understanding of policy frameworks, acknowledging and justifying the connections between a sponsoring agency (WMO) and its continued sponsorship of GESAMP, as well as promotion of enhanced dissemination of scientific information for purposes of education, science, and policy. Four additional links that fall in the UN-related category originated from within GESAMP's own website, but these internal links were largely navigation aids for users of the website. In total, 13 of the 19 links were from websites of organizations with direct affiliation with the UN.

Link sources other than UN-based agencies include webpages of governmental bodies (such as the European Commission, the Japanese Oceanographic Data Center, and the United States Environmental Directories). These links show a connection between governmental organizations and grey literature that encompasses an understanding or trust in GESAMP's publications that transcends the basic act of citation to a single document. This citation type also includes non-governmental organizations, such as the Conservation International Marine Portal and the Large Marine Ecosystems of the World group.

V. Conclusions

Each citation or link dataset collected in this study provides unique insights into use and influence of GESAMP's publications. Each set of data supplies evidence that GESAMP's publications were used or its information was recommended in ways that cannot be determined through reliance on one source of citation data. While Web of Science provides access to a very large source of citation data, this source is limited in its ability to represent the full extent of grey literature use largely because of the restricted scope of Web of Science. Each dataset represents an informative building block for measuring overall influence of GESAMP's publications. A composite metric emerges from consideration of multiple datasets which uniquely demonstrate where influence is evident. The proposed metric would combine findings from Web of Science, Google Scholar, Google, as well as the acronym and link searches in order to demonstrate use of grey literature from several angles.

Web of Science data can be analyzed from a number of perspectives to reveal numerous insights about how publications have been cited in scholarly sources in primarily scientific contexts. As this case study of GESAMP has demonstrated, questions can be posed to the citation data collected from Web of Science while recognizing that this data only represents the citations appearing in journals indexed by the database. Such questions include who cited the literature, in which journals and what subject areas citations appear, and citation rates over time. The data can be analyzed further to determine whether citing authors are mostly individuals who have had an affiliation with GESAMP. GESAMP's publications are cited in sizeable numbers, which indicates that the group's technical reports were disseminated through a variety of channels and that the publications and the group itself are both seen as legitimate and authoritative. Web of Science citation data is the traditional standard used for citation analysis, and the wealth

of information that it includes makes it a required building block in the understanding of GESAMP's influence. However, Web of Science is not sufficient as a single source for evidence of use of grey literature.

Google Scholar and Google searches represent a shift away from a traditional source of citation data towards data that more completely account for current developments in scientific publishing. The large number of influential citations in the results from Google Scholar and Google emphasize the importance of conducting citation searches on the Web. Only three of the 587 Google Scholar search results gathered in this study were considered perfunctory, which confirms that a large majority of results are indicative of influence. Unique Google search results are more likely to include perfunctory citations, with about 30% in this category, but the majority of citations represent more substantial use. It is also important to note the rates of overlap between citations retrieved with the Google search engines and in Web of Science. Whereas 44% of the Google Scholar results were unique to the search engine (i.e., not duplicated in Web of Science), over 85% of the Google results were unique. While there is a strong commonality between Google Scholar and Web of Science results, complete duplication between the two sources does not occur. The degree of commonality between Google and Web of Science results is low. This study has also shown that most Google Scholar and Google search results indicate influence (perfunctory citations are in a minority). Further, Google Scholar and Google results also supply evidence of uses of newer information available on the Web to a larger extent than citations in Web of Science will reveal. Findings from citations to publications located via Web searches, especially of publications in the last decade, are pivotal building blocks for a metric that aims to understand the use of grey literature.

Searching for GESAMP's acronym in both Google and Google Scholar also provided insights into how the organization is represented on the Web. The sample data collected from the two search engines both provided further indicators of influence with different degrees of overlap with data from Web of Science. While searching for a term like GESAMP introduces the complication of dealing with Web ephemera, the understanding gained outweighs the time required to collect and interpret this data. The data included evidence not found in other citation datasets and therefore extended understanding of the use of grey literature. The importance of conducting searches on the name of publishers of grey literature as well as publication titles has been shown. Findings from this type of analysis become another building block in the metric designed to more fully measure influence of grey literature.

Web links demonstrate the use and influence of grey literature in a way that draws on tenets of citation analysis without relying on the traditional understanding of what constitutes a citation. Instead, this method collects data from sources that are becoming increasingly important in the global communication of information. By showing which websites link to the website of a producer of grey literature, the Web link evidence show direct connections between those who are using or recommending use of GESAMP publications and the grey literature publisher itself. Many of the websites that link to the GESAMP website are hosted by UN-based agencies, which bear a similarity to the "informed" citing authors in Web of Science data. This later relationship may suggest that GESAMP's website visibility beyond the scope of other UN agencies is limited. In an era where scientific information, especially in grey literature forms, can easily be disseminated on the Web, determining whether Web links exist and from where they

originate is an important component of understanding influence. As the evolution of publication and dissemination of grey literature on the Web continues, findings from hyperlink relationships will be a further element in the measure of grey literature use and influence.

Studies of information use and influence have, since the mid-1950s, been based on the citation data available from Web of Science. While the citation data available from the database is a crucial element in understanding the influence of information, it does not encapsulate all evidence of use, especially with regard to grey literature. Alternative forms of scientific publication such as grey literature can be excellent sources of timely, salient information, but are also often stigmatized due to their assumed lack of peer-review processes, which tarnishes their credibility. Negative assumptions about reputability are bolstered to a degree because Web of Science does not index highly regarded sources of grey literature, such as GESAMP; while citation data pertaining to the agency can be collected from Web of Science, it is not comprehensive. This lack of comprehensiveness is coupled with increasingly varied forms of online scientific publication. As such, citations are available and can be accessed in ways that transcend Web of Science's abilities; therefore, more comprehensive citation collection practices are required to best understand use and influence. By considering Web of Science, Google Scholar, Google, and Web link information together, this study has shown that a varied approach to gathering citations from online sources produces unique, relevant instances of the use of grey literature. The purpose of this study has not been to identify which source is the "best" indicator of influence, but instead to suggest that a multi-faceted approach to the collection of citation data and use indicators is required in order to fully understand GESAMP's influence. The compilation and analysis of data from multiple sources shows the wide variety of settings and sources in which GESAMP publications have been cited, a mosaic that demonstrates the overall influence of the group more effectively than consulting a single source. Further research will delve into other possible sources of citation data, including monographs, which represent another source likely to produce unique insights into the influence of grey literature.

VI. References

Bar-Ilan, J. (2008). Which h-index? A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257-271.

Cordes, R.E. (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, 49-69.

Cronin, B. (1984). *The citation process: The role and significance of citations in scientific communication*. London: Taylor Graham.

Cronin, B. (2001). Bibliometrics and beyond: Some thoughts on web-based citation analysis. *Journal of Information Science*, 27(1), 1-7.

Cronin, B. (2005). *The hand of science: Academic writing and its rewards*. Lanham, MD:

Scarecrow Press.

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.

Gelfand, J. M. (1997). Academic libraries and collection development implications for grey literature. *Publishing Research Quarterly*, 13(2), 15-23.

Kousha, K. & Thelwall, M. (2007a). Google Scholar citations and Google web/URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, 58, 1055-1065.

Kousha, K. & Thelwall, M. (2007b). The Web impact of open access social science research. *Library and Information Science Research*, 29, 495-507.

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, 605-621.

Vaughan, L. & Shaw, D. (2005). Web citation data for impact assessment: A comparison of four science disciplines. *Journal of the American Society for Information Science and Technology*, 56, 1075-1087.

Vaughan, L. & Shaw, D. (2008). A new look at evidence of scholarly citation in citation indexes and from web sources. *Scientometrics*, 74, 317-330.

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

Figure 1. Citations to GESAMP Publications by Year from Web of Science

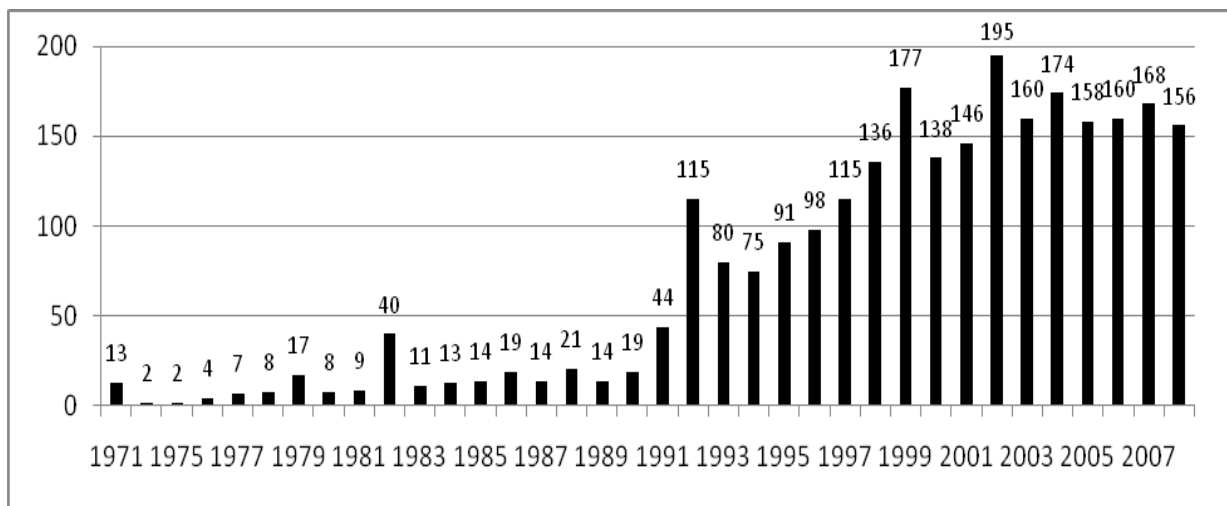
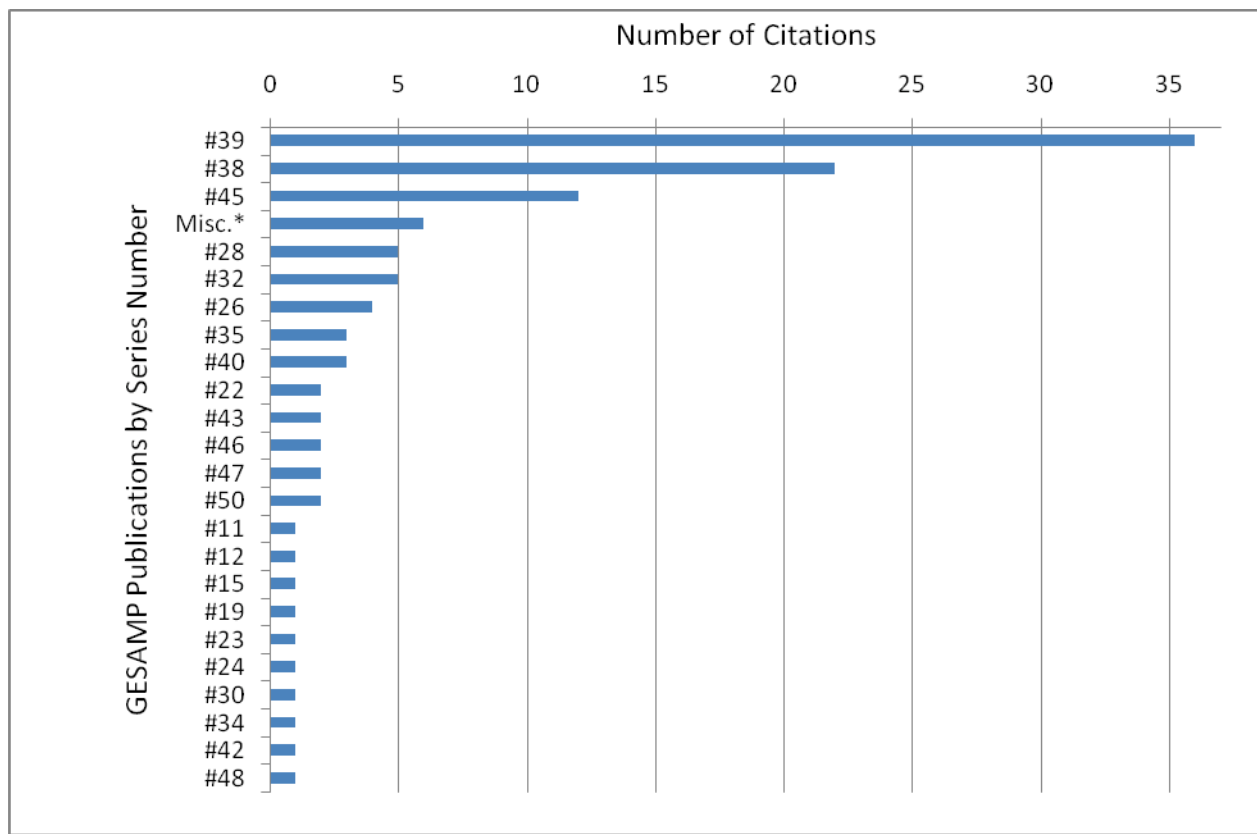


Figure 2. Citations to GESAMP Publications in 1992 from Web of Science



* Citations to GESAMP documents published outside the *Reports and Studies* series.

Table 1. GESAMP Citations by Web of Science Indexed Journals

Journal Name	Citations
Marine Pollution Bulletin	275
Journal of Geophysical Research-Atmospheres	82
Marine Chemistry	78
Marine Ecology-Progress Series	74
Science of the Total Environment	70

Table 2. GESAMP Citations by Web of Science Subject Areas

Subject Name	General Category	Number of Citations
Environmental Sciences	Science	1004
Marine & Freshwater Biology	Science	690
Oceanography	Science	398
Geosciences, Interdisciplinary	Science	325
Meteorology & Atmospheric Sciences	Science	192

Table 3. Citations to GESAMP Reports from Google Scholar

Report Number	G.S. Exports	G.S. Unique	Perfunctory Results
38	47	9	0
32	47	9	0
39	139	55	0
50	57	27	0
61	62	39	0
6	28	16	1
57	50	18	1
28	19	9	0
71	85	52	1
58	53	26	0
Total	587	260	3

Table 4. Citations to GESAMP Reports from Google

Report Number	Google Exports	Google Unique	Influential Citations	Perfunctory Citations
38	29	24	16	8
32	38	24	10	14
39	89	83	72	11
50	43	35	16	19
61	52	45	31	14
6	31	27	15	12
57	42	33	25	8
28	18	16	7	9
71	95	87	72	15
58	29	26	19	7
Totals	466	400	283	117

Table 5. Google Link Search Results

	# of Links
UN Sources (13 links)	
FAO, UNEP, WMO, etc.	9
GESAMP	4
Non-UN Sources (6 links)	
European Commission	1
Japan Oceanographic Center	1
U.S. Environmental Directories	1
Peri-urban mangrove forests as filters and potential phytoremediators of domestic sewage in East Africa	1
Conservation International	1
Large Marine Ecosystems of the World	1
Total	19