Searching down the fisheries information web: An initial reflection

Janet Webster, Hatfield Marine Science Center, Oregon State University (United States)
Patricia Merrikin, Food and Agriculture Organization of the United Nations (Italy)
Jean Collins, Food and Agriculture Organization of the United Nations (Italy)

Abstract
The complex web of information ranges from pure science to applied techniques to management policies. All are important, yet have varying levels of accessibility and authority. Often the apex of the peer-reviewed journal article is considered the most important level discouraging scientists and managers, especially those in developing countries, from populating the other layers of the web that may be considered grey. We are particularly interested in how information produced in developing countries fits into this information web. To do so, we focused on one topic, mangroves, using it as a means to exemplify the complexity of the web of information and reveal publication patterns. Using search tools, we identified 2000-2005 mangrove-related publications regarding Africa and India, analyzed them by author's location and affiliation, publication format and availability. Our preliminary analysis suggests that grey literature remains difficult to discover, the peer-reviewed journal is the most popular means of scientific communication, and open archives are just starting to have an impact on scholarly communication. The web of information would be strengthened by improving the discovery of the grey literature through greater visibility and accessibility

<u>**Introduction**</u>

Fishing down the marine food web is an influential concept proposed by Pauly et al. (Pauly and others 1998)to describe the pattern where high trophic level fish species are over-exploited thus disturbing the ecology and economics of the system. Essington et al. (Essington and others 2006) suggests that fishing *through* the food web may better describe the situation as more trophic levels are exploited leading to a greater complexity in fisheries management than simply over-fishing. We contend that the two concepts have analogies in the complex web of information that ranges from pure science to applied techniques to management policies. All are important, yet have varying levels of accessibility and authority. Often the apex of the peer-reviewed journal article is considered the most important level discouraging scientists and managers, especially

those in developing countries, from populating the other layers of the web that may be considered grey.

Grey literature tends to be at the bottom of the information web in part because it is hard to identify and access, particularly in the increasingly digital environment. It also sinks because it is not recognized as valuable given peer pressure on people to publish in the commercial journals. Consequently, relevant information is easily overlooked and underutilized especially if it originates in a developing country. This phenomenon may devalue the work of scientists and managers in developing countries, slowing responsible management of the natural environment and its valuable resources. If information is difficult for the end user to find or access, it tends not to be used as readily (Jansen and others 2000). For those involved with collecting or monitoring grey literature, this issue is familiar. Even so, increased documentation is useful for building the case for better indexing and improved discovery tools.

We initiated the following study to test our thesis that the grey literature of fisheries needs to be more visible so it can be more viable in the information web. Researchers and managers use discovery tools that are familiar and readily available. Some of these tools are well established with sophisticated indexing and defined sources (e.g. *CAB* and *Biosis*). Others are emerging as useful tools (e.g. *GoogleScholar*). Comparing results produced using a range of current tools is one way to document what information is visible and accessible, who creates it and where it is published. It also allows some consideration of trends in coverage, although this would take more longitudinal work. This progress report on our study suggests ways to improve visibility and accessibility of the fisheries grey literature.

Approach and Methodology

In the study, we focused on information produced on mangroves in Africa and India. Mangroves are productive ecosystems for a variety of natural resources including fisheries. We originally thought to narrow our topic further by limiting mangrove information to that addressing fisheries or aquaculture, but this resulted in very small or null sets in most of the indexes used. Africa and India were chosen as geographic limiters for several reasons: they differ in their development of a scientific community; they are working on developing responsible fisheries management; and mangroves are significant habitats in each area. We also limited our searches to the time period of 2000-2005 in part to keep the results manageable, but also to see if the open access movement was having an impact on access to the literature.

Table 1 lists the search tools and the basic search strategies along with the resulting hits. From these, we chose to analyze the ones in bold. Search strategies varied depending on the capability of the search tool. *GoogleScholar* does not accommodate true Boolean logic or field searching, so the strategy was quite basic. In most of the others, we used field limiters to get more precise results. When possible, we limited to title/abstract/keyword, used geographic limiters or expanded through the thesaurus to better refine the results. On the surface, this appears to skew the overall results as the database results are more precise while *GoogleScholar* less so consequently producing more hits. To remedy this somewhat, we searched *GoogleScholar* by year (a limiter that it provides) and took the first 28 to 30 hits from each year searched to build a set

comparable to the *Biosis* and *CAB* sets. This is not a perfect comparison, but gives a

solid set of results for analysis.

Table 1:

Search Tool	Search strategy	Total hits	Total hits analyzed
AGRIS (OVID)	MANGROVE* and AFRICA	17	
	MANGROVE* and INDIA	5	
Aquatic Biology, Aquaculture and Fisheries Resources	MANGROVE* and AFRICA	590	
	MANGROVE* and AFRICA and (FISHER* or AQUACULTURE)	91	**74**
	MANGROVE* and INDIA	336	
	MANGROVE* and INDIA and (FISHER* or AQUACULTURE)	116	**112**
***Biosis* (OVID)**	**MANGROVE* and AFRICA**	186	**186**
	MANGROVE* and AFRICA and (FISHER* or AQUACULTURE)	37	
	MANGROVE* and INDIA	179	**179**
	MANGROVE* and INDIA and (FISHER* or AQUACULTURE)	24	
***CAB* (OVID)**	MANGROVE* and AFRICA	146	**146**
	MANGROVE* and AFRICA and (FISHER* or AQUACULTURE)		
	MANGROVE* and INDIA	200	199
	MANGROVE* and INDIA and (FISHER* or AQUACULTURE)		
Food and Agriculture Organization of the UN library catalog	MANGROVE* and AFRICA	2	
	MANGROVE* and INDIA	13	
GoogleScholar	**MANGROVE* and AFRICA**	3280	**176**
	MANGROVE* and AFRICA and FISH	2050	
	MANGROVE* and AFRICA and AQUACULTURE	619	
	MANGROVE* and INDIA	2550	**179**
	MANGROVE* and INDIA and FISH	1350	
	MANGROVE* and INDIA and AQUACULTURE	473	
Horizon	MANGROVE*	30	
OAIster	MANGROVE* and AFRICA	19	
	MANGROVE* and INDIA	3	
Scopus	**MANGROVE* and AFRICA**	122	**100**
	MANGROVE* and AFRICA and (FISHER* or AQUACULTURE)		
	MANGROVE* and INDIA	161	**100**
	MANGROVE* and INDIA and (FISHER* or AQUACULTURE)		
TROPAG (OVID)	MANGROVE* and AFRICA	22	
	MANGROVE* and INDIA	19	

Each result was coded for the following:

- Location of first author

- Affiliation (university, research institute, commercial, government, other)

- Format (journal, book, conference proceeding, report, newsletter)

- Location of publisher

- Available via the Internet

- Publicly accessible in electronic format

Results were analyzed using Excel. All analyzed search results were compiled into a

bibliographic database using ProCite. This enabled us to check for duplication among the

search results.

The following presents initial data from the analysis of all results by search tool.

We also analyzed the data with duplicates removed, but the findings were similar. The

results allow us to examine the efficacy and focus of various search tools. They also give

a picture of who is publishing and where. Our primary goal was to discover what

mangrove related information can be discovered by an interested user, who is writing it,

where is it published, and is it accessible.

Results

What is visible?

Table 1 presents raw data on where material is indexed, but it is difficult to truly

compare apples and oranges, *Biosis* to *GoogleScholar*. Figure 1 gives an overview of

what is visible using each tool. In general, we can say that the index, *Aquatic Biology,*

Aquaculture and Fisheries Resources (*ABAFR*), is the best access tool for the topic. This

is not unexpected as its producers draw from the relevant portions of *Aquatic Science and*

Fisheries Abstracts and enhance it with appropriate records from a variety of sources

including grey literature. It is surprising that databases such as *AGRIS* did not produce more results.

Figure 1: Type of material by search tool
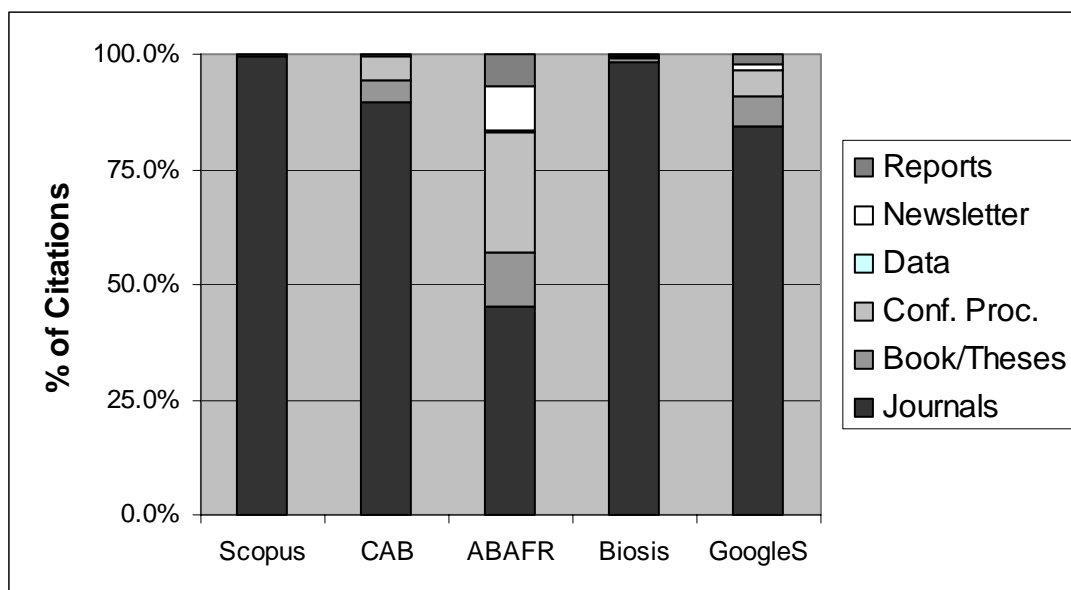
Who is writing?

Comparing India to Africa reveals that while 82.5% of the items on India are authored by scientists in India, only 47.2% of the items on Africa are authored by scientists in Africa (Table 2 &3). Further analysis of second and third authors would be useful as we noticed that many papers did have authors located in Africa, just not as the first author. Also, student authors were assigned to their institution's country and this too does not always reflect nationality.

Table 2: First author's location – India

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Africa	1		1			2	0.3%
Asia	2	3		5	4	14	1.8%

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Oceania	3	1	1	3	3	11	1.4%
Europe	14	15	9	17	26	81	10.6%
India	76	177	96	149	135	633	82.6%
North/Central America	3	2		5	10	20	2.6%
South America	0			0	1	1	0.1%
Unknown	1		3			4	0.5%
	100	198	110	179	179	766	

Table 3: First author's location – Africa

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Africa	51	85	52	82	52	322	47.2%
Asia		3	1	2	7	13	1.9%
Oceania		2	2	4	8	16	2.3%
Europe	44	46	18	82	80	270	39.6%
India		1		4	3	8	1.2%
North/Central America	3	3	1	11	26	44	6.5%
South America				1	0	1	0.1%
Unknown	2	6				8	1.2%
	100	146	74	186	176	682	

Authors are primarily affiliated with either a university or a research institute (Table 4 & 5). There is a tradition of research institutes in India that is not replicated in Africa (Krishna 1997).

Table 4: Author affiliation - India

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Commercial		2	6	3	5	16	1.9%
Government	4	8	6	11	2	31	3.7%
Other	2	20	3	2	1	28	3.4%
Research Institute	33	59	40	68	68	268	32.3%
University	60	105	53	95	103	416	50.1%
Unknown	1	5	2			8	1.0%
	100	199	110	179	179	767	

Table 5: Author affiliation - Africa

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%

Commercial		3	3	3		9	1.3%
Government	1	2	6	3	5	17	2.5%
Other	2	4	5		7	18	2.6%
Research Institute	17	37	26	59	36	175	25.7%
University	76	91	33	121	128	449	65.8%
Unknown	4	9	1			14	2.1%
	100	146	74	186	176	682	

Scientific publishing also appears stronger in India than Africa (Tables 6 & 7).
European publishing venues appear to dominate reflecting the consolidation of scientific
publishing. Over time, this trend may diminish in both Africa and India as regional
journals gain stature. Also, if commitment to open access continues, then some of the
Indian and African journals are well-poised to gain in importance.

Table 6: Location of publication – India

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Africa	0	0	0	0	0	0	0.0%
Oceania	4	8	4	9	10	35	4.6%
Australia	2	2	0	0	2	6	0.8%
Europe	33	66	21	74	81	275	35.9%
India	35	98	82	82	61	358	46.7%
North/Central America	26	21	3	14	22	86	11.2%
South America	0	0	0	0	0	0	0.0%
Unknown	0	4	0	0	3	7	0.9%
	100	199	110	179	179	767	

Table 7: Location of publication – Africa

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Africa	15	24	38	25	43	145	21.3%
Asia		1	1	4	2	8	1.2%
Australia	1	2	1	2	2	8	1.2%
Europe	75	99	26	129	100	429	62.9%
India	1	3		7	1	12	1.8%
North/Central America	8	14	1	18	21	62	9.1%
South America						0	0.0%
Unknown		3	7	1	7	18	2.6%
	100	146	74	186	176	682	

The peer-reviewed journal remains the predominant format (Tables 8 & 9). This reflects both the bias of the search tools as well as the bias of academic researchers and their institutions.

Table 8: Format - India

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Book/Book Chapter		9	13	2	10	34	4.4%
Conference Proceedings	1	8	29	1	11	50	6.5%
Data		1	1			2	0.3%
Journals	99	181	47	176	153	656	85.5%
Newsletter			12		3	15	2.0%
Newspaper					1	1	0.1%
Report			4		1	5	0.7%
Series			4			4	0.5%
	100	199	110	179	179	767	

Table 9: Format - Africa

	Scopus	CAB	ABAFR	Biosis	Google	Total #	%
Book/Book Chapter		8	9	1	9	27	4.0%
Conference Proceedings		9	19	1	9	38	5.6%
Data		1				1	0.1%
Journals	100	128	36	183	147	594	87.1%
Newsletter			5		1	6	0.9%
Newspaper							0.0%
Report			5		7	12	1.8%
Series				1		1	0.1%
Theses					3	3	0.4%
	100	146	74	186	176	682	

Most material is now available electronically, but little is truly publicly accessible (Figures 2). There is a greater open access in India (30.8% of documents versus 11.8% of African related material). *Current Science*, a journal published by the Indian Academy of Sciences, is publicly accessible on the Internet back to volume 1, 1932. This is an excellent example as it is also indexed with variability; of the 27 references found,

Scopus had 0, *CAB* 17, *ABAFR* 4, *Biosis* 17 and *GoogleScholar* 17. The longevity of the journal suggests why *CAB* and *Biosis* index it well, and its open availability is why *GoogleSchola*r brings in significant results. Additionally, other Indian journals are trending towards open access. More investigation as to the motivation would be interesting.

African Journals Online (AJOL) provides limited open access so those articles published in journals that are part of AJOL were not counted as publicly available. Researchers in low income countries may request articles for free. There is a sliding scale for others. This portal to African journals is one example of a mechanism to promote African science through both discovery and document delivery.

Figure 2: Electronic accessibility of all mangrove material
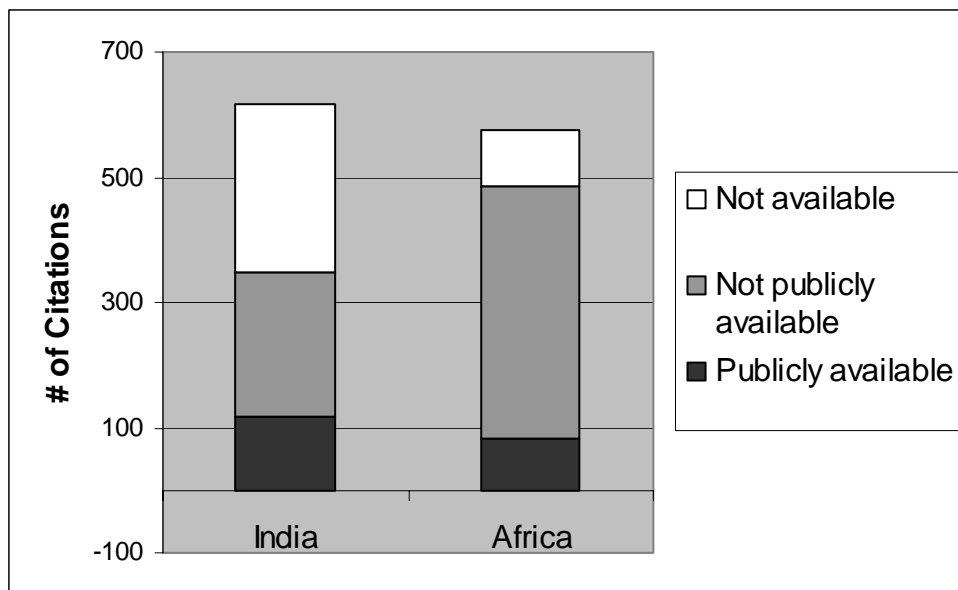
Discussion

Quality of results in terms of grey literature

We did not anticipate that all discovery tools would include much that is grey. Of course, there are definitions of grey to consider as well. From our perspective, journals published in developing countries are not grey as they are peer reviewed and distributed appropriately. We were looking for items such as conference proceedings, workshop reports, management studies and government report. We recognized that the type of coverage in commercial databases depends on the subject – pure sciences coverage seem to be almost completely journal articles (e.g. *Biosis*), while applied sciences produce more grey literature (*ABAFR*). Some cover both the pure and applied sciences (*Scopus*, *CAB* and *GoogleScholar*).

Of the tools analyzed, *ABAFR* alone appears to do a decent job of discovering the grey literature. As *GoogleScholar* harvests OAI repositories, we thought it might gather more results. The lack of diverse formats may be due to sparse holdings in OAI repositories which may be remedied as repositories grow in popularity. Another reason may be *GoogleScholar*'s relevancy ranking, and this would need more analysis of all results to see if grey literature sinks to the bottom pages of search results.

Quantity of results

Discovery tools that we expected to be productive were disappointing. For example, *Agris* did not perform well. This was surprising as it was founded to be an international, cooperative information system for the agricultural sciences and technology that utilizes 240 national, international and intergovernmental centers as contributors. In our study, 12 of the 16 hits for mangroves and Africa were from commercial journals, and the remaining four were theses from Norway and a Swedish working paper. A

similar search for India produced only five hits all from an Indian journal and *Unasylva* (FAO). Further work could be done to see if *Agris* began with deeper coverage of non-traditional information, and over time transitioned toward coverage of the journal literature.

GoogleScholar appears to search more broadly with overwhelming search results. Yet, it is still not clear what it is searching. So, while the numbers of hits are impressive, the number of irrelevant hits are also high as it does not discriminate between information items where the search terms are found in the title and abstract from those where the terms appears in the references if listed. It is limited in ways to refine search results. Duplications are a problem as it does not sort out a citation record from the record returned from a publisher's database.

We attempted to use *Scirus* as it claims to be the "most comprehensive science-specific search engine on the Internet" and promises to find results that the other search engines miss (Elsevier Ltd 2006). The search interface is more limited that *GoogleScholar*, so searches return too much with many false hits or too little if title limits are used. There is potential for improving the tool if searching could be limited to abstracts or other elements found in repositories.

Another way to look at quantity of results is examining the overlap among the results from the tools. A preliminary look reaffirms that *ABAFR* returned the most unique results with the only African duplicates being seven shared results with *GoogleScholar* (less than 10% of the 74 African *ABAFR* results) and less than 6% for the Indian results. Other African duplication between each tool and *GoogleSchola*r were 18% for *CAB* and *Biosis* and 35% for *Scopus*. A similar pattern was found in the Indian

duplicates (*Biosis* 21%, *CAB* 25% and *Scopus* 29%). While more analysis remains, it appears that there is less overlap than was anticipated indicating that each discovery tool has unique qualities, and none are comprehensive.

Access to results

Access to electronic instances of the information reflected the bias of the discovery tool for the most part: most link to the publisher's resource rather than a repository copy. The exception is *GoogleScholar* that did return a fair number of repository copies. It also revealed several electronic theses that did not appear elsewhere. *OAIster*, another free search tool that provides access to open access archives, was disappointing in the quantity of results, but all were relevant.

Access can be deceptive depending on who is doing the searching and where they located. Many believe more is publicly available as it may be for them given institutional subscriptions. Access is highly variable throughout the developing world given services such as *AGORA* that facilitates electronic access to journals from major commercial publishers at little or no cost to low income countries. Most investigation of open access is needed so we can better assess the impact of repositories and open access journals on the information web.

Language of results

We found very few non-English articles. Those few were French, German or Russian. This may be a limitation of where we searched, but probably reflects a trend towards English as the accepted language of science. That said, we remain convinced

that there is significant information in non-English languages that are not indexed due to availability, the language and the limitations of search engines. An area of further research would be to explore the modes of discovery for Arabic and Asian language material.

Conclusion

As librarians, we can collect information consistently and direct users to it in a coherent fashion if we understand what is produced and how it is indexed or discovered. We rely on our discovery tools to help us to this. However, as this preliminary study shows, these tools vary in what they find and index. For the most part, the grey literature is missing.

We suggest that Pauly's concept of declining high trophic levels and Essington's concept of greater complexity are useful in describing the web of information. Different tools influence how people exploit the web of information with some seeking out primarily the peer-reviewed, Western published material while others discover more layers of useful information. Separating the so-called grey literature from the peer-reviewed through the tools we use to discover and access information is harmful to the environment. Greater indexing of the grey literature needs to be integrated into existing discovery tools. Given the current state of discovery, we suggest that increased use of open archive repositories has potential for increasing access to grey information if they are constructed to work with discovery tools. Also, working with developers of the newer tools such as *GoogleScholar* and *Scopus* may help improve searching options and eventual discovery of more grey literature (Powell 2006). The web of information would

be strengthened by improving both the visibility and accessibility of lower trophic levels

of information or what many of us value as the grey literature of the field.

References

Elsevier Ltd. 2006. Scirus - About us [Web Page]. Located at:
http://www.scirus.com/srsapp/aboutus/. Accessed 2006 Nov.

Essington T, Beaudreau A, Wiesenmann J. 2006. Fishing through marine food webs.
PNAS 103(9):3171-5.

Jansen BJ, Spink A, Saracevic T. 2000. Real life, real users, and real needs: a study and
analysis of user queries on the web. Information Processing and Management
36(2000):207-27.

Krishna VV. 1997. A portrait of the scientific community in India: historical growth and
contemporary problems. in: Gaillard J, Krishna VV, Waast R. Scientific
communities in the developing world. New Delhi, India: Sage Publications.

Pauly D, Christensen V, Dalsgaard J, Froese R, Torres FJ. 1998. Fishing down marine
food webs. Science 279(5352):860-3.

Powell A. 2006. Pushing an OpenDOAR [Web Page]. Located at:
http://efoundations.typepad.com/efoundations/2006/10/pushing_an_open.html.
Accessed 2006 Nov.