



Commentaries and Perspectives

Literature Forensics? Door to What Was Known but Now Forgotten

Christian G. Daughton*

Environmental Chemistry Branch, Environmental Sciences Division, National Exposure Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Las Vegas, NV 89119, U.S.A.

(Received 6 September 2001, Revised manuscript accepted 1 October 2001)

Science's all-consuming drive to make new discoveries often risks losing sight of what was already known at one time—that which already exists in the published literature. Inadequate attention to the published literature and insufficient time devoted to its mining and synthesis into new knowledge is a problem faced by all disciplines, especially highly interdisciplinary fields such as environmental forensics, whose knowledge base is fragmented across numerous disciplines. While the conduct of science applies principles of quality assurance to a wide array of its processes, how pervasive are quality controls designed to ensure that planned or ongoing research has not been undertaken before? Has sufficient energy been devoted to mining what has already been discovered and synthesizing it into a larger, more useful perspective?

This paper touches on the liabilities associated with insufficient examination of an exponentially growing published literature (“literature forensics”) and offers some suggestions for achieving a better balance between original work and capturing what has already been reported—all essential to the growing responsibility of knowledge management. By lessening the importance of the published literature, are we asymptotically approaching a point where science may be preoccupied with publishing “new” findings while few have time to assimilate what has already been published?

© 2001 AEHS

Keywords: knowledge management; knowledge mining; knowledge synthesis; science fragmentation; literature review.

The word *forensics* can conjure two seemingly unrelated images, one being public speech or debate, and the other, critical examination of evidence involved with litigation. The etymology of the word reveals, however, a close association between the two. Indeed, the final objective of both disciplines of forensics is to persuade or to reveal truth via a systematic examination of facts or principles.

Forensic and *forum* (from the Latin *foris*), among other words such as *foreign* and *forest*, have roots to an ancient word for “door”—*dhwor*. All involve aspects of being “outside the door”, e.g. “in the marketplace” (originally the enclosed space around a home). In ancient Rome, the *forum* hosted debates involving government affairs and also doubled for court trials. *Forensic* therefore, encompasses the historic practice of skilled, insightful verbal prowess, as well as the more modern biophysicochemical science that assists in resolving ethical or legal concerns; access to web resources for a broad range of sub-disciplines housed

within the legal forensics domain (not the topic of this paper) is available at a number of sites, including: <http://userzweb.lightspeed.net/~abarbour/links.htm>.

Returning to the commonality—*door*. Both types of forensics involve open examination and discourse—“out of doors”. To pursue this idea beyond the etymological linkages, consider what exists through the door. Forensics involves the extraction of pertinent information to support the argument or investigation at hand. As such, travel through the door is required to discover or uncover what is not yet known, to bring from the outside what is still (momentarily) foreign. But while forensics almost always deals with discovering or uncovering what is not yet known or accepted, it is easy to lose sight of what we had already known or found at one time but have since forgotten or lost. The analogy pertains to which side of the door we face, outward towards the future unknown, or inward toward the past, which was at one time known.

Forgotten or lost? Assuming that modern knowledge is archived in printed or electronic media, how can we lose touch with what we once “knew”. This problem is sometimes alluded to by the clichés “reinventing the wheel” and “duplicating past efforts”. While the depth of the issue is multidimensional and obviously not limited just to forensic sciences, the principles of forensics can be applied to its control across all disciplines. One aspect of this syndrome is the seemingly never-ending onslaught of new journals and new means

*Fax: 702-798-2142. E-mail: daughton.christian@epa.gov

Please note that the materials presented here represent the professional views and opinions of Christian Daughton. They do not necessarily reflect those of the U.S. Environmental Protection Agency.

Notice: The U.S. Environmental Protection Agency (EPA), through its Office of Research and Development (ORD), funded and performed the research described. This manuscript has been subjected to the EPA's peer and administrative review and has been approved for publication.

of publishing (e.g. web-based pre-prints), all of which offer up a buffet of data and knowledge that seems to defy digestion. The “problem” of an expanding universe of publications has itself been discussed in numerous articles, a recent example being a perspective written by [Reed \(2001\)](#).

The Fragmentation of Science

The topic at hand can be called “The Fragmentation of Science—Loss of the Bigger Picture: The Critical Importance of Knowledge Mining and Synthesis”. It deals indirectly with the “loss” of knowledge; a result of its being overshadowed not just by the sheer mass of competing articles but also by the obscurity or buried by a cacophony of redundant, irrelevant, poorly executed, or misleading publications. Subjectively, much of the world’s published science literature seems vastly underutilized and highly fragmented. While the information available in any research field continues to grow exponentially, proportionately less time can be devoted to “mining” or capturing this every-increasing body of knowledge, as to synthesize a larger and vastly more meaningful picture. For a wide variety of reasons, literature is often ignored or simply becomes “lost” to future investigators. Maybe the author selected a journal that reached an uninterested or inappropriate audience, maybe the paper, while technically sound, was impenetrable because it was poorly written, or because the scientific community was not yet receptive (an idea “ahead of its time”), or any number of other reasons. Is it beyond comprehension that sometime in the future we may asymptotically approach the point where science is preoccupied solely with publishing while few read what has already been published?

A major reason that the published literature does not receive the attention it should is because we as scientists often do not devote sufficient time to explore and examine what our colleagues have already done. The culture that has slowly evolved in science is one where the published literature is not as important as publishing “new” findings; whether they be truly new or seemingly new. Duplication of effort and reinvention of the wheel are symptoms of the failure to pay sufficient attention to the literature, and greatly impede the genesis of new and useful knowledge. A parallel problem is the larger picture remains obscure when the literature is not critically examined, especially for issues that cross multiple disciplines. Solutions to problems and answers to questions can simply be awaiting “discovery” amidst research that has previously been reported. However, someone must first devote time to separating the wheat from the chaff, discovering and assembling the many and varied ingredients, deciphering the proper proportions, and marshaling all into an eminently digestible morsel of new knowledge or insight.

As pointed out by [Csermely \(Semmelweis University, Budapest\) \(1999\)](#), this problem results in part from the fact there is little professional reward in attempting to distill, synthesize, and integrate what is known about a topic. Science managers tend to value publication of “original data” even if it is incremental, and if unknown to them, it happens to be merely “rediscovered

data”. Furthermore, scientists often have little time or inclination to understand the significance, relevance, impact, or outcomes of their work because they too are caught in the drive to publish; often at the expense of reading, comprehending, distilling, and synthesizing the literature.

The issue of “capture and synthesis” of fragmented knowledge has been briefly discussed by [Csermely \(1999\)](#), who argues that little attention is being paid to the fragmentation of the world’s science literature. Perhaps this is a major reason it is so difficult for individual scientists to have a broad-based appreciation for the “bigger picture”. This seems especially true for younger scientists. Csermely writes; “There is only a limited effort to achieve the appropriate balance between the discovery of new facts and finding their proper place and importance in the framework of science.” This relates partly to what can be referred to as understanding the significance, impact, or relevance of one’s work, or ultimately, being able to measure the ultimate *outcome* of our efforts. The importance of understanding the significance and outcome of published work cannot be overemphasized (e.g. see [Daughton 2001](#)).

Csermely goes on to note that “science itself is not self-integrating, and there are fewer and fewer people taking responsibility for net-making”... “Integration [of knowledge] needs time and patience...” “...greater credit should be given to those who make serious attempts to integrate their findings into the whole of human knowledge”. Synthesis is especially critical for multi-disciplinary areas such as environmental forensics. While greater effort should be expended on weaving a richer tapestry of connections and relationships that wend through wide swaths of concepts arising in the literature, countless threads for any fabric of knowledge are often left unraveled.

“Scientific conferences have turned into information ‘stock exchanges’ instead of places for evaluation and discussion”. One can immediately appreciate this simply by looking at the titles for conference presentations, even those at symposia or sessions devoted to specific topics. Often, there is little indication of conference organizers having attempted to tell an overall larger story, no thread holds the talks together.

“Grant applications follow one another almost uninterrupted; we are told it is better to publish than to perish; and a never ending stream of technically correct, but shallow, papers make excellent technicians out of our Ph.D. students instead of true scientists”. An unbalanced emphasis on publishing supplies strong selective pressure for work that adds incrementally to a long-existing and already well-developed knowledge base instead of breaking new ground.

“Integration [of knowledge] needs time and patience...”. “The term ‘competitor’ has replaced the old words ‘fellow’, ‘trusted colleague’ and ‘scholar’. We should limit our competition. Groups working on parallel projects should seek joint publications instead of the duplicate or even triplicate articles appearing in several journals...” A final excerpt from Csermely states: “...greater credit should be given to those who make serious attempts to integrate their findings into the whole of human knowledge.” These and other factors that contribute to fragmenting the literature

and which diminish our view of the bigger picture are summarized below.

Factors contributing to fragmented literature and loss of the bigger picture

Authors

- Repetitive publication (multiple versions in different journals); partly driven by ever-increasing pressures to publicize as well as “publish or perish”.
- Reduced access to journals (increasing cost and obscurity).
- Too “busy” to read, exacerbated by an ever-increasing number of journals and pages.
- Targeting lesser-known journals in order to increase chances of publication.
- Publishing in inappropriate journals (reaching the wrong audience).
- Emphasis on “incremental” work rather than value-added or original works.
- Perpetuation of misinformation e.g. failure to verify citations or contents of primary references, instead relying on secondary references; insufficiently reviewed or quality-assured material.

Publishers

- Actively proliferating new journals; further complicating, obscuring, and diluting the existing literature.
- Lessening acceptance standards; quantity versus high-quality or high-impact.
- Rapid review system—becoming more perfunctory, less critical and adding little value.
- Escalating cost; leading to narrowing accessibility by readers.
- Editors merely serving as gate-keepers; diminishing value-added advice.
- Proliferation of non-refereed, especially electronic, publications.

Reviewers

- Uneven/arbitrary standards for quality and fairness in review process.
- Focusing on the trivial aspects (e.g. grammar and format) rather than the essence of a paper.

Readers

- Reduced access to journals; increasing cost and obscurity.
- Too “busy” to read, exacerbated by an ever-increasing number of journals and pages.
- Poorly designed and incomplete literature searches.
- Insufficient appreciation and understanding of the overall state of the literature and its accessibility.

Science managers (note that here is where standards and expectations are set)

- Valuing numbers of publications rather than scholarship, significance, or potential impact.
- Solely valuing “original” work at the exclusion of “synthesis” work

In short, a revolution in the management and reward of the practice and communication of science is needed to increase the value and enhance the creation of new knowledge. After all, higher quality works are more

likely to be considered for assimilation by future knowledge miners.

Quality Assurance and Accessing the Literature

Research organizations can devote considerable resources to a variety of quality assurance activities. For research and development, rigorous quality assurance (QA) controls are usually designed and implemented for nearly the full range of activities with which a scientist/engineer is engaged. These activities include maintenance and calibration of instruments and equipment, data acquisition, data reduction, data review, and data interpretation.

A costly oversight—inadequate utilization of the published literature

Despite the progress in having established a wide array of QA in science, a major gap universally persists. How many organizations have controls in place ensuring someone else has not already completed the research planned? Enormous resources can be wasted or used inefficiently by “re-inventing the wheel”. Insufficient effort (or misguided effort) is expended on investigating (i) what previous effort has already accomplished and (ii) what work is currently underway by others regarding the topic at hand. What controls are in place to ensure that optimal and cost-effective effort is devoted to making use of what at one time known by others; and which the public has already paid for? Just because ready access can be gained to the tools for locating relevant literature (such as electronic data bases, abstracting services, and the web), this does not ensure that these tools are used in the most efficient manner. Frequently one can read a journal’s “letters-to-the-editor” section and find frustrated readers pointing out that a previous issue contained work done many years prior; many of us have experienced this with work done earlier in our careers. The work repeated today has a new author that derived no apparent benefit from our previously published data, knowledge, or insights.

For any number of reasons many prospective authors often ignore the published literature failing to communicate with peers who are working on the same (or closely related) topic. A major reason for publishing is to convey to others what has been demonstrated or discovered so that others can build or expand on this base. Even so, when initiating our own research projects, the published literature is often treated as if it were unimportant or irrelevant. Perhaps it is deemed “out-of-date”, despite its actual age and despite that much of science itself does not become obsolete or inappropriate/irrelevant. Perhaps we chose to ignore it because it was “not invented here”. Others may simply feel overwhelmed by the sheer magnitude of the published literature; new journals appear regularly, and established journals increase their pages or publication frequencies. With the exponential growth in the published literature and information in general, it is more difficult than ever to stay current in any field.

Following are some thoughts on addressing the idea of formalizing the process of literature searches/reviews and general information gathering. There are probably many parallels that could be drawn with the process of experimental research or at least with our personal experiences in “researching” a major consumer purchase such as a car, nursing home, child-care facility, private school, or house.

Purpose of QA in literature review

One purpose of QA is to minimize bias, which can be a problem with literature searches and review. For example, the searcher may subconsciously rule out certain sources of information because of parochial prejudices (e.g. an environmental chemist may ignore the pharmacology or medical literature even though it may be directly relevant). Certain journals may be routinely but unknowingly ignored because they fail to show up in the search data-base or because they are in foreign languages; the latter problem can be partially circumvented by collaborating with colleagues in other countries. Certain types of articles may routinely be ignored because they “seem” to be irrelevant. A good example is the older literature (some of which may be perceived simply as of “historical” use). Sometimes, certain pieces of knowledge literally become “lost” in the literature, waiting to be rediscovered by “library research” or worse, rediscovered by new laboratory research. This is usually a result of building upon existing literature reviews and failing to go back to the original literature. Other times, because of the frequent practice of citing secondary and tertiary sources, information becomes distorted and unfortunately becomes newly established as fact.

Type of review

Literature reviews could be classified as simple searches (electronic and manual), annotated reviews, and exhaustive critical reviews. Each serves its own purpose, and each could have different applicable QA guidelines.

Extent of review

The extent of review can range from merely cursory e.g. covering only recent literature for a project that already resides within the investigator’s area of expertise, or searching only for very recent review articles to exhaustive; when the literature review is the end in itself, representing the entire research project. A literature review is essential if the area of research is relatively new to the investigator. But how does one ensure that a supposed exhaustive review is indeed thorough? Computer searches should be done in parallel with manual, hand searches. These two approaches to searching are often complementary. Manual searching has the added advantage of permitting “browsing”; for example, additional references can be found that are in the physical vicinity of the ones located on a library shelf or in a file. Additionally relevant articles that would ordinarily be deemed irrelevant solely on the basis of their titles can be located simply by paging through journals and by reading “outside” one’s area of interest.

For example, scientists can benefit greatly by periodically visiting the current business or economic press.

Completeness of search

How do we ensure that an appropriate and sufficient literature review has been successfully concluded? Are there certain minimum standards that can be set? The completeness of computer searches must be verified by hand (manual) searches. A good knowledge of the literature as acquired by manual searching can prove essential for verifying whether an unbiased computer search strategy has been correctly formulated. One approach is to establish a suite of key or central references and to ensure that these are in turn uncovered. Another approach is to perform forward searches using key historical references as the basis for locating more recent papers.

Advice on reviews

Research organizations should consider formulating guidance documents that assist those less-versed with accessing the literature and which serve as check lists for those who are experienced. Suggestions include those key on-line data-bases or hardcopy reference materials that should be examined. As an example, an excellent recent overview of search tools specifically geared to toxicology and chemistry is available in a special issue of *Toxicology* (Wexler, 2001). Simple reminders could be helpful for many. For example, have you searched for and located any recent review articles? Dissertations are frequently overlooked as a wealth of information and often have thorough literature reviews that are not published elsewhere. The patent literature can also be helpful. If you have a key reference, have you expanded your search by locating those papers that subsequently cross-reference it? Identification of cross references (via cited reference searching, e.g. via ISI’s Science Citation Index) is an extremely powerful and useful approach for enhancing and speeding up access to literature connected by any number of threads (e.g. see: <http://www.isinet.com/isi/products/citation/sci/index.html>). Finally, a nice resource that addresses the many elements of writing review articles is Fink (1998).

Timeliness of gathered information

Has the very most recent information been incorporated in a review? Literature review involves not just the published literature (be it peer-reviewed archival or “gray” non-refereed literature), but also the process of discovering who is currently doing state-of-the-art work in the subject area and of contacting those investigators. Has information (i.e. unpublished) from peers who are currently engaged in the subject work area been gathered? This includes not only researchers in academia, government, and industry, but also from the manufacturers of the instruments, software, supplies, and materials that are used or studied in the course of our research. These current contacts can be made by referrals from others in the field, scanning abstracts and titles of papers that will be presented at upcoming conferences/symposia, locating recent letters to the

editor and errata, supporting documents and ancillary data, and by perusing current news articles (trade publications and newspapers, such as the *Wall Street Journal*). For an excellent overview on searching and accessing the news media for topics relevant to environmental toxicology/chemistry refer to South (2001) or see the links at: <http://www.epa.gov/nerlesd1/chemistry/ppcp/useful-links.htm>. Keep in mind that currency is important in literature reviews because the lags associated with publishing automatically outdate a review by 6 to 12 months (or more). Surfing the Internet with meta-search engines (even while your review article is under review) will increase your chances of locating additional current work, which you can often add with the editor's approval to the peer-reviewed version of your article. Web-scope and efficiency of searching are continually improving with each generation of Internet search engines. Use any of the ever-expanding search-engine and review sites (including general- and specialty-search engines) to learn the latest regarding web searching (e.g. <http://searchenginewatch.com/>; <http://www.searchengineguide.com/>; <http://www.leidenuniv.nl/ub/biv/specials.htm>; <http://www.mazepath.com/uncleal/net2.htm>; <http://www.metaplus.com/>).

Liabilities of Electronic Searching/Retrieval

Despite its strengths, electronic searching/retrieval is vexed by several key liabilities that warrant continued vigilance and the use of complementary "manual" searching. An obvious downside of electronic searching is it minimizes unanticipated serendipity; a chance discovery afforded by physically browsing in libraries or files. Electronic searching is extremely versed for learning about subjects that can be easily distilled into a few, unique key words. For pursuing ideas and subjects whose key words are more general and shared by many fields but with totally different meanings, electronic searching has greatly diminished power. The corollary, which serves to maximize the difficulty of electronic searching, is that different disciplines often use disparate terms for the same thing (pointing to the need for shared knowledge systems that transcend disciplines). The ease with which electronic searching can be conducted and constructed can easily give the false impression of thoroughness and accuracy. Finally, the prolific ease with which the full text of publications are compiled can lead to rapid information overload. To ensure maximal coverage of the literature, it is perhaps wisest to always doubt that one has fully revealed the extent and content of the literature on a particular subject no matter how specialized. A continuous process of updating searches; using a continually evolving and changing approach is prudent. For these reasons, asserting to be the "first" to make a claim in the literature is always fraught with peril; it is best to do so only with caveats.

Possible Growing Role for "Knowledge Management"

The bewildering expansion of the technical literature and the vast collective knowledge buried within desperately requires radically new mechanisms for its

exposition and management. Even though this is a well-entrenched problem with our archived body of literature, actions could be taken to limit its continued progression and perhaps stop its growth in the more distant future. One way to slow or limit the continuation of this trend is to begin capturing our current knowledge in more efficient ways. Combating a science's fragmentation may require its journals, societies, and research organizations to experiment with creating crucibles for "self-assembling or organizing" review articles. Web-based living drafts could take shape as a wide range of invited experts contributes to its evolution. Such ideas fall under the broad rubric of knowledge management (KM), an ill-defined and rapidly expanding body of ideas taking hold primarily within the business community. While KM can be viewed as yet another passing fad, its ramifications for all fields of endeavor are immense. For further reading about KM, refer to the countless discussions and articles available at the many KM web link sites: <http://www.knowledgemia.org/knowledgemia/knowledgemia.nsf/pages/knownlinks.html>; <http://www.kmresource.com/exp.htm>; <http://www.brint.com/km/>; <http://www.c3.lanl.gov/~rocha/lww/> [LANL's "Active Recommendation Project"]; <http://www.nature.com/nature/debates/e-access/Articles/luce.html>).

At the USEPA's Office of Research and Development, as with other research organizations, knowledge management and knowledge mining will impose increasingly greater demands for effective use of inter-disciplinary literature, especially with regard to the vagaries associated with "emerging" environmental issues (e.g. see: <http://www.epa.gov/ordinter/futures>). Knowledge mining, for example, has proved one of the major factors limiting the advancement of the emerging concern of pharmaceuticals and personal care products as environmental pollutants (<http://www.epa.gov/nerlesd1/chemistry/pharma/index.htm>).

The Door's Other Side

This discussion has focused on the multi-faceted problem of ensuring that current knowledge is captured, assembled, and disseminated but at the great risk of ignoring the importance of creativity unfettered by past thinking.

Having stressed the enormous import of the published literature, it is critical to ensure its proper place in the advancement of science. While a new and larger understanding of an issue can indeed be derived from the synthesis of smaller published pieces, attainment of new knowledge also must be gained *a priori*. In this light, it is sometimes essential to divorce oneself from the published collective knowledge and to set sail in new directions unbiased by the avenues that are already known. The seemingly paradoxical argument can be made that the best way to impede the evolution of new paradigms from research is to actually read the published literature. A balance must be struck by blending the two approaches, at times enthusiastically mining the past (literature forensics) and other times ignoring it. Both require venturing "through the door", thinking outside the box, but facing different directions.

References

- Csermely, P. 1999. Limits of scientific growth. *Science* **284**, 1622–1623. (text available: <http://www.fhsu.edu/jdurham/id150.htm>).
- Daughton, C.G. 2001. Emerging pollutants, and communicating the science of environmental chemistry and mass spectrometry: pharmaceuticals in the environment. *J. Am. Soc. Mass Spectrom.* **12**, 1067–1076. (available: <http://www.epa.gov/nerlesd1/chemistry/pharma/index.htm>).
- Fink, A. 1998. *Conducting Research Literature Reviews: From Paper to the Internet*, pp. 1–265. Thousand Oaks, Sage Publications.
- Reed, C.A. 2001. Drowning in a sea of refereed publications. *Chem. Eng. News* **79**(5), 37–38. (available: <http://cen.acs.org/isubscribe/journals/cen/79/i05/html/7905sci4.html>).
- South, J.C. 2001. Online resources for news about toxicology and other environmental topics. *Toxicology* **157**, 153–164.
- Wexler, P. 2001. *Toxicology* **157**, 1–164 (Special issue on digital information and tools) (abstracts available: <http://www.elsevier.com/locate/toxicol>).