PPCPs Literature Citation Database: Synopsis of Project

This project is designed to provide a comprehensive, publically accessible listing of literature citations that are directly or peripherally relevant to the many aspects of PPCPs as environmental pollutants. The listing is currently available in TXT and PDF versions from the US EPA's PPCPs web site: http://www.epa.gov/ppcp/lit.html. A more extensive and comprehensive version is available to EPA researchers as a full database and is compiled using bibliographic citation software.

The compiled literature citations cover all the facets of the risk assessment paradigm (including: Origins, Sources, Occurrence, Transport, Fate, Exposure, Effects, Stewardship, Monitoring, Waste & Water Treatment Technologies, Risk Assessment, Rick Communication, etc.) as well as many aspects that are peripherally related (e.g., major citations dealing with low-dose effects, mixture effects, and databases for medications, physicochemical properties, and toxicology, among others). The scope of the covered literature includes not just journal articles, but also books (and book chapters), proceedings, databases, web pages, reports, miscellaneous gray literature, and select presentations and news stories.

The major advantage of the PDF listing versus the TXT version is that it serves as a simple, but effective database. Key-word searches can almost instantly locate all the citations containing a key word or string. Users must select "use advanced search options" (located at the bottom of the basic search panel on Acrobat). The full first line of each hit is displayed in the search panel, allowing fast visual examination of the results, as well as instantly locating the citation. The TXT version is useful for importing as a delimited file directly into any of the many available commercial bibliographic citations databases. A typical entry in the public listing displays the following citation information:
Kolpin DW, Furlong ET, Meyer MT, Thurman EM, Zaugg SD, Barber LB and Buxton HT (2002). Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. streams, 1999-2000: a national reconnaissance. Environ Sci Technol 36(6), 1202-1211.

The need for a PPCPs citation database was born from the difficulty in performing comprehensive literature searches and the incompleteness of all commercial databases and inadequacies of Internet search engines. One of the major inefficiencies in the development and advancement of the PPCPs field is the difficulty in distilling the large and continually growing body of literature into useful knowledge. The published literature contains a wealth of data and insights that have never been summarized and evaluated. Lack of a comprehensive citation database is one of the reasons; another reason is that synoptic reviews and compilation of databases, although invaluable, are generally not valued in science as much as the publication of new data.

Even with commercially available scientific literature databases, or with the growing power of free search engines such as Google Scholar coupled with expanding Internet content, locating literature specific to PPCPs is extremely difficult because of the complexities in automatically filtering out the countless irrelevant hits, which are often pertinent instead to other fields such as clinical research, clinical practice, pharmacy, or human or veterinary pharmacology. Locating the relevant literature is one of the major challenges to maintaining a literature database on

PPCPs. It is not a topic conducive to key-word searches. Use of cited-citation searching (e.g., using Science Citation Index or Google Scholar's "cited by" feature) based on known, highly cited publications is perhaps the more efficient way to locate and compile relevant literature. But citation analysis is a very time consuming task and one that must be performed on an ongoing basis.

Currently, the limitations of the database are: (i) it does not have comprehensive coverage of the foreign-language literature (which is extensive), (ii) abstracts, URLs, or PDFs are missing for large portions of the citations, and (iii) lingering typographic errors and formatting problems, as well as errant duplicate entries, prove difficult to efficiently eliminate.

The EPA in-house database is updated usually on a daily basis. The public version accessible from the PPCPs web site is updated periodically (roughly every 1-2 months). New additions include not just the newly published and in-press literature, but also the many past publications that have yet to be captured and corrections to existing entries.

The database is shared with other EPA programs that also compile citations. This serves as a quality check on accuracy as well as a completeness check on the different search strategies that are employed, as no search strategy is fully comprehensive.

As of August 2008, the PPCPs literature database contained citations for 5471 references. In the complete version of the database, 66% of the entries had URLs or DOIs for direct access to full articles, 59% had complete abstracts, and 33% had PDFs of the complete publication. Although the public version only has the complete citation, the included title of the publication permits keyword searching; URLs or DOIs are often provided for direct access to full articles. We anticipate that the PPCPs citation database will be useful not just to the public (especially students) and industry, but also to scientists worldwide - - making the PPCPs literature much more easily accessible to all.

For EPA scientists having bibliographic citation software, the most recent version of the PPCPs Literature Citation database can be accessed by contacting Christian Daughton (daughton.christian@epa.gov). The advantages of having the actual database (instead of a simple PDF listing) are threefold: (1) References can be effortlessly cited and formatted in MS Word documents using on-the-fly citation capabilities. (2) References can be printed in whatever format is required (for example, to meet a journal's requirement for submitting manuscripts). (3) Abstracts and PDFs (when available) can be searched on full text and directly accessed. (4) Queries can be performed for the type of publication (e.g., all books, all journal articles, all reports, etc.). (5) Boolean searches can be performed. For example, "ethynyl OR ethinyl" finds all citations containing: ethynylestradiol, ethynyl estradiol, ethinylestradiol ethinyl estradiol, together with all the variants with oestradiol.

28 August 2008