

This system seeks to integrate information from the laboratories, monitoring activities, technical literature, company data, the registration process itself, and other federal activities, in addition to various contract research studies.

RESEARCH AND POLICY STUDIES

All regulatory activity results in varying degrees of cost and benefit to diverse interests within the public sector. The intent is to maximize benefit and minimize cost.

In order to make this intent as much of a reality as possible in the Pesticide Program, OPP is conducting and supporting scientific research and policy studies directed both at providing support for pesticide regulatory actions and toward establishing a firmer understanding of pest control methods as they impact man and the environment. These activities are composed of a variety of both short- and long-term investigations and are being carried out in several broad areas:

Pesticide effects

Alternative pest control methods

Laboratory methods and quality control

Pesticide wastes recovery and conversion

Cost/benefit assessment methods

Unreasonable adverse effect assessments

Monitoring optimization

Identification of new research needs

Evaluation of interrelationships between pesticides and other pollution statutes and programs.

These activities consume information from all OPP sources, channeled through the Hazard Evaluation System, or obtained directly from the sources as needed. Contractor personnel are involved in some of these studies, and draw for their work on OPP and external sources. Research and study results are usually published by OPP or are made available to the public sector under the FOI program.

Retrieval Forms and Formats for Environmental Objectives Employing Machine-Readable Biological Information†

KAY DURKIN and J. R. SMITH*

BioSciences Information Service of Biological Abstracts, Philadelphia, Pennsylvania 19103

Received November 5, 1975

The constantly changing information needs of environmental scientists can most efficiently be met by a versatile data base covering all fields of biology. This machine-readable file and the manner in which both conventional published products and customized search services are derived from it are described.

INTRODUCTION

Most large discipline-based files are broken into broad categories which usually reflect the classical classification of the scientific discipline. In biology these can be as broad as botany and zoology. Within the same discipline there can be mission-oriented approaches to classification of the literature, for example, ecology and toxicology, but these categories are equally broad. When using a machine-readable file of more than three million references covering all the fields of biology, which is growing at a rate of about one-quarter of a million references a year, it is imperative to be able to derive, on a current or retrospective basis, responses to much more specific questions. It is also important to have the facility to respond at short notice to questions which, for one reason or another, are of immediate concern. Relatively recently, "mercury pollution of rivers" became an issue upon which information was required as a matter of urgency. On the other hand, some questions are of continuing interest and require services on a continuing basis, such as "Environmental Pollution" and "Pesticides". Others require specific responses to one-time-only questions, for example, "Toxicity of Iodine Used as a Disinfectant". The ability to identify items effectively in response to any type of question depends heavily on the manner in which the material has been indexed, and efficient indexing coupled with effective retrieval is the only way in which it is possible to make purposeful use of a large machine-based file. Effective retrieval depends heavily on adequate education and

training of potential users of the file.

Different services that can be provided may be in printed form, microform, on magnetic tape, or on computer printouts. Some approaches which are currently being used are described below, as is a possible speculative extension of these.

Abstract Journals. The first and perhaps most obvious approach is to generate small abstract journals in specific areas; some typical examples of titles are shown in Figure 1. These small journals are designed not to exceed 200 abstracts a month. They are not indexed and are a fairly elaborate form of current awareness service.

Standard Profiles. A second approach is similar to the first, but comprises computer printout of citations only and is designed to respond to more specific questions. Some typical titles are shown in Figure 2. This service is generated by the building of a number of profiles which respond to the needs of at least several users. Both this service and the small abstract journals are open-ended series so that it is possible to add topics which become of interest and to delete topics which are no longer meeting a need.

Current Awareness and Retrospective Searches. A third approach and one of perhaps more immediate interest is the provision of personalized current awareness and retrospective search services either in batch or in on-line mode by Information Dissemination Centers. Examples of this kind of service will be referred to in more detail later.

Cooperative Activities. A fourth approach is a concept of a number of secondary Information Services contributing to a single service so that the coverage and indexing may be more effective. A good example is *Abstracts on Health Effects of Environmental Pollutants*. This abstract journal, which covers

† Presented in the symposium on "Information Requirements Resulting from Environmental Impact Laws", Division of Chemical Information, 170th National Meeting of the American Chemical Society, Chicago, Ill., Aug 27, 1975.

BIORESEARCH TODAY SERIES

Human & Animal Parasitology	Food Additives & Residues
Cancer A - Carcinogenesis	Population, Fertility & Birth Control
Industrial Health & Toxicology	Environmental Pollution
Food Microbiology	Pesticides
Human Ecology	

Figure 1.

Insect Chemosterilization	Reclamation of Waste Materials
Biochemical Evolution	Physiological Effects of Sound
Fungal Disease of Corn	Monitoring of Environmental Pollution
Herbicides	Air Pollution
Reactions of Insects to Light	Pesticide Residues in Soils
Phytovirology	Sewage Disposal and Sanitary Measures
Pesticide Residues in Foods	Bioclimatology and Biometeorology
Animal Communication	Plant Growth Substances

Figure 2.

between ten and twelve thousand items a year and is fully indexed, was developed by BIOSIS with the help of a contract from the National Library of Medicine, and material obtained from the National Library of Medicine data base (MED-LARS) and the BIOSIS data base has been used in selecting the content. This journal also carries additional indexing and, with the cooperation of Chemical Abstracts Service, Substance Registry Numbers have been used. This experience of three organizations working together, the National Library of Medicine, Chemical Abstracts Service, and BIOSIS, has made it possible to produce a more comprehensive and better indexed service.

A recent derivative from this publication, again achieved with the help of Chemical Abstracts Service and supported by the National Library of Medicine, is a *Chemical Index Guide* to the first three volumes of *Abstracts on Health Effects of Environmental Pollutants*. This guide contains a substance name section in which all of the substances referred to in the abstract journal, and their synonyms, have been listed alphabetically and related to the Substance Registry Number. A second section which is in Substance Registry Number order, contains the names and variants which are to be found in the Substance Registry File.

This publication is an extremely valuable tool since it is, in effect, a small subset of the Chemical Abstracts Service Substance Registry File relating quite specifically to one of the environmental issues of importance.

A Possible Future Approach. A possibility which has been under examination for some time is, in essence, an extension of the kinds of services which have been mentioned earlier. This approach envisages the possibility, not only of segmenting a large broadly based file, but of adding to a particular segment additional material (Figure 3). The criteria of selection applied by the large secondary Information Services do not always totally meet the needs of specific groups, and it is possible, provided an appropriate segment of the file can be identified, to add additional material of interest only to the specialist group. In addition, it may be possible to add a particularly appropriate type of indexing and conceivably important items of data. The achievement of any service of this kind, of course, requires a close and active relationship with that part of the scientific community concerned with the specialist topic under review.

The Data Base. The large machine-readable file which has been referred to previously is a compilation from 1959 to date of all material published in *Biological Abstracts* and *Bio-Research Index*. In 1969 a magnetic tape service was offered and has been installed in 16 information centers around the world. This service is called BIOSIS PREVIEWS. Among the 16 information centers, three are providing on-line service; they are Lockheed Information Systems, State University of New York, and the CAN-OLE Systems at the Canadian Institute of Scientific and Technical Information.

The machine data base includes citations, but not abstract text, and these bear a one-to-one relationship with *Biological*

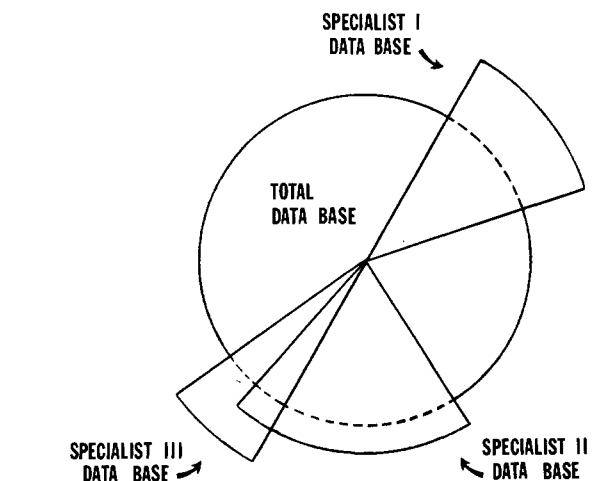


Figure 3.

<u>MERCURY UPTAKE, DISTRIBUTION, METABOLISM</u> <u>IN SOIL BACTERIA AND PLANTS</u>		
	CC51519 METABOLISM	PLANT PHYSIOLOGY
MERCURY/ AND	CC51520 TRANSLOCATION, ACCUMULATION	
MERCURIAL	CC51522 CHEMICAL CONSTITUENTS	
MERCURIC		
CC40000 SOIL MICROBIOLOGY		

Figure 4.

Abstracts and BioResearch Index. Each citation in the file includes a reference number, a CODEN and a journal name, author names, bibliographic information, and the authors' title of a paper plus enrichment keywords. It also includes CROSS and BioSystematic indexing codes.

The first major index, the Subject Index, consists of words in the authors' title and descriptors added by the BIOSIS editorial staff to enrich the authors' title. This is a natural language vocabulary and no restrictions by way of thesaurus are applied. In addition to the specific subject terms, broad concept categories such as radiation, biology, ecology, and soil science are accessible through the use of the CROSS Index included on the file. Also taxonomic affiliations or organisms mentioned in references are searchable through the Bio-Systematic Index. It is the combination of natural language keywords for specific concepts with the two structured indexes for broad concepts which provide a highly flexible and efficient means of searching the BIOSIS PREVIEWS data base. This facility, to combine the use of several indexes, is demonstrated in the following two sample searches. These searches are actual requests received by the BIOSIS search service which provides both retrospective and current awareness services.

Search 1. This search was on the topic of "Mercury Uptake, Distribution, Metabolism in Soil Bacteria and Plants". The strategy shown in Figure 4 indicates a combination of natural language keywords for Mercury with broad subject categories for metabolism, etc. In this case, it is not necessary to introduce a third parameter for bacteria and plants into the search strategy since the concept or CROSS categories already specify plant physiology and soil microbiology. If the concept categories did not exist, it would be necessary and extremely difficult to compile the extensive lists of keywords which would be necessary to satisfy the second parameter.

An example of an item derived from this search is shown in Figure 5. This reference was identified by a match between the keyword Mercury and the CROSS code for Plant Metabolism indicated in the figure by the number 51519. The BioSystematic codes, the five-digit numbers on the last line, indicate categories for mosses and grasses.

Search 2. The search strategy for the topic "Toxicity of Iodine Used as a Disinfectant", is shown in Figure 6. In this

59034244
 ATENB/ATMOS ENVIRON
 7 (7). 1974 749-754
 HUCKABEE, JR/
 MOSSES SENSITIVE INDICATORS OF AIRBORNE MERCURY POLLUTION/
 GRASSES VEGETATION
 01008 04500 06504 07504- 07506 10010 10059 10069 13010-
 22506- 37015* 51519- 51526-
 1100 21600 25305

Figure 5.

TOXICITY OF IODINE USED AS A DISINFECTANT
 IODIDE/
 IODINE/ AND CC37008 DISINFECT/VECTOR CONTROL AND CC22504 PHARM TOX
 IODINATED CC39500 DISINFECT/STERILIZATION CC22506 ENVIRON TOX
 IODO/ CC22508 VET TOX

Figure 6.

case, a combination of natural language keywords for Iodine and subject categories for Toxicity and Disinfectants was used. It was necessary in this search to have a parameter for each concept in order to minimize the identification of irrelevant material. The CROSS or concept codes have again been used for Disinfection and Toxicity.

Figure 7 shows a sample of an item retrieved from this search. Many descriptors have been added to enrich the

75001344
 BVJQA/BR VET J
 156 (2-3). 1974 145-156
 WILLINGER H/ THIEMANN G/
 CRITICAL ASSESSMENT OF STERILANT AGENTS IN THE FIELD OF
 VETERINARY HYGIENE/
 HUMAN ANIMALS FORMALDEHYDE CHLORINE COMPOUNDS PHENOL
 DERIVATIVES IODOPHOR TOXICITY
 10058 10060 10069 22501 22504- 22508- 26502 38002- 39500*
 85150 86215

Figure 7.

authors' title. This item resulted from a match between the keyword Iodophor and two CROSS codes, Disinfection and Sterilization represented by the number 39500 and Veterinary Toxicology represented by the number 22508. The Bio-Systematic codes, shown on the last line of the item, refer to general vertebrates and human studies.

The manner in which a large broadly based machine-readable file can be used to help to meet the needs of scientists concerned with environmental problems has been described. The potential for exploitation of such a file is almost limitless and can contribute significantly toward helping the scientific community to deal with some of the problems arising in a relatively specialized subfield of biology.

Bridging and Interlinking the Information Resources†

DALE C. MYERS, JOYCE A. RATHBUN,* FRED A. TATE, and DAVID W. WEISGERBER

Chemical Abstracts Service, The Ohio State University, Columbus, Ohio 43210

Received November 5, 1975

The American Chemical Society through its Chemical Abstracts Service Division plans to establish a Regulatory Registry Service (RRS), based on the CAS Chemical Registry System, to uniquely identify substances in publicly available government agency files and thereby to provide a potential interfile link for substance-oriented data in these files. Use of the RRS will avoid duplication of effort in producing and compiling information and will improve access to substance-related information contained in existing publicly available government files.

INTRODUCTION

Today's national and international concerns—energy, natural resources, ecology, food, health, consumer protection, urban development, transportation, etc.—are heavily overlapped and tightly interdependent. Likewise, governmental efforts directed at solving these problems have led to a proliferation of interrelated and often overlapping legislation and regulation at all levels of government, both in this country and internationally. Such legislation frequently assigns responsibility for compliance to more than one agency. For example, of a sample set of 40 U.S. federal regulatory laws, 13 were found to deal with the EPA, 6 with the FDA, 5 with the Consumer Product Safety Commission, 4 with the Drug Enforcement Administration, 10 with the Materials Transportation Bureau, and 16 with other agencies. These add up to more than 40 because of the substantial overlap in agency responsibility in just this one small sample.

Currently, many different government agencies or even different offices within the same agency share responsibility

or have similar responsibilities for related regulation. Of course, these interrelationships are not limited to the federal agencies. The states and many municipal and other local government activities also must deal with the same problems. The Ohio Environmental Protection Act, just one of many possible examples, includes nine regulations on water, seven on air, and one on solid waste.

The probable overlap of agency interests has vast dimensions when one considers that the multiplicities of regulatory overlap which exist within the U.S. are for just one country. Other countries have similar internal concerns and overlaps. And the flow of commerce and the universality of social problems throughout the world add still other layers of overlapping interest and regulation.

To deal effectively with this overlap of government interest as it relates to the accessing of substance-related information accumulated by government agencies, the American Chemical Society (ACS) through its Chemical Abstracts Service (CAS) Division plans to develop and implement a Regulatory Registry Service (RRS) which will be described in the balance of this paper. To simplify the following discussion, these comments will concentrate on dealing with the files of U.S. federal agencies. Also, although a fully implemented RRS would deal with the files of all U.S. federal agencies, initial attention in

† Presented in symposium on "Information Requirements Resulting from Environmental Impact Laws", Division of Chemical Information, 170th National Meeting of the American Chemical Society, Chicago, Ill., Aug 27, 1975.

* To whom correspondence should be addressed.