(18) Musen, Mark A.; Fagan, Lawrence M.; Combs, David M.; Shortliffe, Edward H. Use of a domain model to drive an interactive knowledge-editing tool. *Int. J. Man-Mach. Stud.* **1987**, *26*(1), 105-121.
(19) Neale, Ian M. First-generation expert systems: A review of knowledge acquisition methodologies. *Knowl. Eng. Rev.* **1988**, *3*(2), 105-145.
(20) Pharmacia Biosystems AB, Sweden. *A.L.F. DNA Sequencer Users Manual*, 1990.
(21) Shahsavar, Nosrat. *Knowledge Acquisition and Refinement for a Domain-Specific Expert System.* Linköping Studies in Science and Technology, Licentiate Thesis 229. Linköping University, 1990.
(22) Shaw, M. L. G.; Gaines, B. R. KITTEN: knowledge initiation and transfer tools for experts and novices. *Int. J. Man-Mach. Stud.* **1987**, *27*(3), 251-280.
(23) Shortliffe, Edward H.; Scott, A. Carlisle; Bischoff, Miriam B.; Campbell, A. Bruce; van Melle, William; Jacobs, Charlotte D. ONCOCIN: an expert system for oncology protocol management. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence, IJCAI'81*; 1981; pp 876-881.
(24) Steele, Guy L., Jr. *Common LISP: The Language*, 2nd ed.; Digital Press: 1990.
(25) Wielinga, B. J.; Schreiber, A. Th.; Breuker, J. A. KADS: a modelling approach to knowledge engineering. *Knowl. Acq.*, in press.

UTAB: A Computer Database on Residues of Xenobiotic Organic Chemicals and Heavy Metals in Plants

J. E. NELLESSEN and J. S. FLETCHER*
Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma 73019-0245

Received September 30, 1991

The UTAB Database contains information concerned with the uptake/accumulation, translocation, adhesion, and biotransformation of both xenobiotic organic chemicals and heavy metals by vascular plants. UTAB can be used to estimate the accumulation of chemicals in vegetation and their subsequent movement through the food chain. The database contains actual data from papers in the published literature dating from 1926 for organic chemicals and from 1976 for heavy metals. At present the database is comprised of more than 37 000 records pertaining to 900 different organic chemicals, 21 heavy metals, and over 350 plant species. Each record contains information on a single combination of species, chemical, and dose. Other information includes the application and destination sites, amount accumulated, rates of uptake or translocation, products and sites of biotransformation, experimental condition parameters, and the source paper. Thus, the database can be used to quickly obtain specific data pertaining to a chemical, plant species, mine spoil, etc. or it can be used for the comparative analysis of a set of data pertaining to groups of chemicals and plants.

INTRODUCTION

The accumulation of toxic chemical residues in the environment has been a well-publicized concern since the 1960's.[1] Pesticides used in agriculture, metals from sewage sludge, mining and refinery operations, and miscellaneous other chemicals from industrial processes such as PCBs (polychlorinated biphenyls) and PAHs (polycyclic aromatic hydrocarbons) may all leave toxic residues in the environment. Food chain contamination with insecticides such as DDT, resulting in deleterious effects on predatory birds, are well known.[1] In the U.S. 631 000 tons of pesticides were produced in 1989.[2] After exports and imports 416 000 tons remained for usage in the U.S. with approximately 55% as herbicides, 35% as insecticides, and 10% as fungicides.[2] Some authors suggest that a considerable quantity of these applied pesticides do not reach target organisms and may thus impinge upon surrounding ecosystems.[3] In the U.S. with the development of regulatory policies and agencies, many of the most toxic and persistent chemicals have been banned. But new chemicals are continually being developed, tested, and used. Metals in sewage sludge,[4] lead along roadsides,[5] and more recently selenium poisoning in birds[6] and its implications for human health[7] have all been topics of research. Cadmium is a toxic trace metal often present in sewage sludge, but its presence as a contaminant in phosphate fertilizers may be far greater.[8]

Since plants play an essential role in the ecosystem as primary producers by extracting resources from the abiotic environment, their potential involvment in the fate of xenobiotic chemicals in the ecosystem cannot be ignored. Plant roots may draw metals and organic contaminants out of the soil and translocate them through the stem to be accumulated in the foliage. Animals feeding on the foliage will thus become exposed. Plant species can differ in their ability to extract compounds and elements from the soil and transport them to the foliage. For example, in the southwest U.S. certain plant species in the genus *Astragalus* (milkvetch) are selenium accumulators[9] and contain enough selenium to be toxic to cattle.[10] Although selenium is an essential nutrient to animals[11] (required in very small quantities) and deficiencies are often reported,[12] certain plant species hyperaccumulate selenium and are toxic. Consideration of such differences between plant species in their tendencies to accumulate or not accumulate certain chemicals is very important when it comes to the restoration of contaminated ecosystems.

Data on chemical residues, whether toxic or not, are continually being published in a wide variety of journals. With the ever-increasing volume of published residue data it becomes increasingly hard to benefit from this pool of data when it is so widely scattered. It has been our goal to assemble this vast array of data into a computerized database called UTAB. UTAB stands for the uptake/accumulation, translocation, adhesion, and biotransformation of both organic chemicals and metals by vascular plants.

The UTAB Database is a computerized information resource that permits the rapid retrieval and comparison of chemical residue data. UTAB can be used to estimate the accumulation of applied chemicals and their biotransformed products in vegetation and the consequent exposure to wildlife pertaining to problems of food chain contamination. Data pertaining to a specific chemical, plant species, or process (e.g., uptake, translocation, etc.) can be obtained from the database. UTAB thus serves as a rapid means of accessing numerical data in minutes with no need to acquire and read the published papers from which the database was derived.

Crop, weed, native, and wild plant species exposed to metals or xenobiotic organic compounds have been considered.

0095-2338/92/1632-0144$03.00/0

UTAB BIBLIOGRAPHIC DATA FORM (BIBROW)

REFNO R05233
YEAR 82
AUTHORS MARCUS-WYNER, L., D.W. RAINS
TITLE UPTAKE ACCUMULATION AND TRANSLOCATION OF ARSENICAL COMPOUNDS BY COTTON GOSSYPIUM-HIRSUTUM
CITATION J ENVIRON QUAL 11(4):715-719

Figure 1. Sample bibliographic record from the UTAB database.

Table I. Subject Categories That Were Searched in the *Biological Abstracts* for Xenobiotic Organic Chemicals

section heading	related subheading
agronomy	(all)
ecology	plant
forestry	(all)
horticulture	(all)
plant physiology	growth substances
	metabolism
	nutrition
	respiration
	translocation and accumulation
pest control	(all)
soil science	fertility and applied studies
toxicology	environmental and industrial
	foods, food residues
	additives and preservatives

Twenty-one different metals are included. The database does not contain information on nonvascular plants (e.g., algae, fungi, mosses, etc.). All data in UTAB have been obtained from published papers. Presently, published organic chemical data cover the period from 1926 to 1987, while the metal data cover 1976–1989.

DATABASE DESCRIPTION

Design of the Database. The UTAB Database is comprised of two files: a Bibliographic File and a Fate File. The Bibliographic File is a listing of published papers about organic chemical pollutant residues and metal contaminants in or on vegetation. The Fate File contains the actual data and experimental conditions taken from the papers in the Bibliographic File. The Fate File contains information on plant species, chemicals, dosage, duration of exposure, and other coded parameters describing the type of study and plant parts exposed to the chemicals.

The UTAB Database has been designed using dBase IV software on IBM or IBM-compatible microcomputers. Each record (line) in the database contains information from several data fields (columns). Each data field has its own label, which is either mnemonic or self-explanatory. The use of records and the data fields will be described in the following sections.

Bibliographic File. The Bibliographic File contains all of the published scientific papers used to develop the Fate File. It serves as a link in identifying the sources of the data in the Fate File.

The Bibliographic File contains approximately 6550 references on xenobiotic organic chemicals and heavy metals. The literature search for xenobiotic organic compounds has been completed for the time period 1926–1987. This was a manual search of the *Biological Abstracts* using the topic headings in Table I. Approximately 4000 of the 6550 titles in the Bibliographic File are on organic compounds. Heavy metal citations were obtained through a search of BIOSIS (the Biological Abstracts Computer Database) for the time period 1976–1989. The name of a metal was linked with various concepts such as accumulation and uptake, plant physiology, mineral metabolism, etc. (Table II). Twenty metals were included in the search, encompassing all 13 on the U.S. EPA's priority pollutants list[13] (Table III). Papers were screened to exclude nonvascular plants (algae, fungi, mosses, lichens), techniques (e.g., staining), metabolism of endogenous chemicals, and review papers that did not contain original data. At present UTAB only contains papers written in or translated into English.

Table II. Search Strategy for Metals in BIOSIS (Biological Abstracts Computer Database)

	Keyword
name of metal	20 searched (e.g., lead)
	Concepts
minerals	biochemical studies and methods
	metabolism
plant physiology, biochemistry, and biophysics	accumulation and uptake
	translocation
	Exclusions
nonvascular plants (mosses, algae, lichens, fungi)	
techniques (e.g., staining)	
membrane permeability, proton pumps, ATPase	
reviews	
no appropriate data on plants or metal uptake	
non-english language	

Table III. Heavy Metals Included in UTAB

aluminum	Al	mercury[a]	Hg
antimony[a]	Sb	nickel[a]	Ni
arsenic[a]	As	platinum	Pt
beryllium[a]	Be	selenium[a]	Se
cadmium[a]	Cd	silver[a]	Ag
cesium[b]	Cs	thallium[a]	Tl
chromium[a]	Cr	tin	Sn
cobalt	Co	uranium	U
copper[a]	Cu	vanadium	V
lead[a]	Pb	zinc[a]	Zn
manganese	Mn		

[a] Currently on EPA's priority pollutant list. [b] Not included in original search, but data are being included when encountered in papers on other metals.

Table IV. Contents of a Bibliographic Record

label	description
REFNO	file number
YEAR	year of publication
AUTHORS	authors
TITLE	title of article
CITATION	journal citation

Each reference in the Bibliographic File contains standard citation information. The field labels for the information contained are listed and defined in Table IV and shown as they appear on a computer printout in Figure 1.

Fate File. The Fate File possesses information describing the fate of chemicals when they come in contact with vascular plants. Since the fate of chemicals under these circumstances is usually dependent upon one of four major plant processes, (1) uptake/accumulation, (2) translocation, (3) adhesion, and (4) biotransformation, the information in the Fate File has been catagorized according to these processes. Thus the Fate File possesses four different kinds of Fate Records, each

Table V. Data Fields Associated with Each of the Four Types of Fate Records (Uptake/Accumulation, Translocation, Adhesion, and Biotransformation) in the Fate File

data field (label name)	description
PROCESS[a]	one of four plant processes [uptake/accumulation (UA), translocation (TL), adhesion (AH), biotransformation (BT)] which distinguishes the kind of record and type of information encoded on the record
PLANT[b]	single plant taxon
CHEMICAL[b]	single chemical
CIDN[b]	chemical identifier or descriptor number
DOSAGE[b]	chemical concentration administered
TIME[b]	time elapsed between chemical exposure and endpoint measurement
PARMS[b]	series of codes describing experimental conditions
APSITE[b]	plant part exposed to chemical
DEST[c]	plant part where chemical was measured
TSITE[d]	plant part where biotransformation occurred
AMOUNT[e]	quantity or concentration measured following exposure
RATE[b]	rate of the process
PERTRANS[f]	percent of chemical translocated
PATH[f]	path of translocation
PRODUCT[d]	product of biotransformation
GRADE[b]	usefulness and/or quality of source
UTABNO[b]	record number
REFNO[b]	bibliographic source file number

[a] Each record will be coded for only one process, thereby identifying it as a specific kind of Fate Record. [b] Common to all four types of Fate Records. [c] Common to UA and TL records. [d] Only associated with BT records. [e] Common to UA, AH, and BT records. [f] Only associated with TL records.

pertaining to only one of these processes. This important distinguishing feature associated with each Fate Record is entered under the field label of PROCESS (Table V). In addition to this field, each record will possess between 12 and 14 other data fields, dependent upon the type of record. Twelve data fields are shared by all four types of records, while some fields are unique to only one of the record types, i.e., PLANT and CHEMICAL data fields are present on all records, but TSITE is only a part of biotransformation records (Table V). Nonoverlapping two- or three-letter abbreviations are used in the database to conserve computer space. The information associated with each field (label) has been entered in a precise and consistent manner throughout the database to facilitate rapid and accurate recovery of information from the computer.

Each record contains a plant name listed under the label PLANT (Table V). The standard format used in UTAB for plant nomenclature is *Phaseolus vulgaris* L. "bean" 'black valentine'. The formal Latin name is given first, then the common name in quotation marks, and lastly the varietal or cultivar name in single quotes. Each record contains a chemical name under the label CHEMICAL. The format for chemical names is "Carbofuran" 2,3-dihydro-2,2-dimethyl-7-benzofuranyl methylcarbamate 'Furadan'. The common name in quotations comes first, followed by the formal name, and lastly any additional names (e.g., tradenames) set off in single quotes. For some of the common, well-known, chemicals the common name may be all that is supplied, e.g., "DDT", "2,4-D". The label CIDN for chemical identifier or descriptor number will be used to supply Chemical Abstracts Service (CAS) Registry Numbers or SMILES numbers (Simplified Molecular Input Line Entry Systems) to further identify chemicals. Currently the CIDN field is empty. Which of these identifiers will be used in the database is pending a decision by the U.S. EPA, the project sponsor.

The concentration of the chemical to which the plant was exposed is given under the label DOSAGE. Doses are given as used in the papers with standard unit abbreviations. Occasionally descriptive words are used in order to clarify dose information. The period of time from application of the chemical to the point of measurement is encoded under the label TIME. When several doses and times (time = period from application to measurement) have been examined in a single publication, only selected doses and/or times are recorded in the database. For example, selections would include doses and times giving the highest and lowest accumulation as well as a few doses and times in-between so that one can generate a dose–accumulation curve, dose–% translocation curve, time–accumulation curve, or a time–decay curve.

Table VI. Some of the 54 PARMS Codes Used in the UTAB Database to Describe Experimental Conditions

CE	controlled environment, growth chamber
EE	actual environmental contamination
FI	field investigation
MN	mine spoil study
SG	sludge study
TC	tissue culture
DUS	plants dusted or powdered
GRO	chemical added to liquid growth medium
SPR	plants sprayed
SSL	chemical applied through soil
CHE	chelating agent used
SUR	surfactant used
DT	different distances from source of pollution
PH	different pH conditions used
SO	varying soil types tested
TE	varying temperature conditions tested

The data field with the label PARMS contains a series of codes describing experimental conditions, physical form and application mode of the chemical, use of application aids, and experimental variables studied. Examples of some PARMS codes and their meanings are shown in Table VI. If one wanted only data from sludge studies conducted in the field (as opposed to a controlled environment chamber) the PARMS field would be searched for the combination of the SG and FI codes to obtain only records on sludge studies in the field.

The data fields APSITE, DEST, and TSITE (Table V) all represent various plant parts and share the same set of codes, e.g., LF for leaf, RT for root, SH for shoot, ST for stem, etc. There are 75 different plant part codes allowing for a wide variety of both general and specific parts of plants. The field AMOUNT contains the residue concentration or quantity of the applied chemical occurring in or on the plant part measured. Units used are those given in the paper with standard abbreviations. The RATE field describes the rate of the process (e.g., uptake or translocation) and uses standard unit abbreviations. The RATE necessarily involves a time factor and is most commonly associated with translocation records (e.g., 5 $\mu g/cm^{-1}\ h^{-1}$). The fields PERTRANS and PATH (Table V) are only associated with translocation records. The PRODUCT data field contains the chemical name for the product of biotransformation. The same naming format is used as for the CHEMICAL field.

The GRADE field uses a letter code from A to F which rates the papers for data quality, quantity, and suitability for UTAB. Papers with the necessary basic information to supply most fields are graded C. Papers with more data than can be incorporated into UTAB are given A or B. Papers given a D grade are missing important information (e.g., no dosage supplied). An F grade is given to papers with no usable data for UTAB despite the fact that the title may indicate so. The UTABNO is simply a unique number assigned to each record in the Fate File. The REFNO is the bibliographic source number that links data in the Fate File to its source paper in the Bibliographic File.

Use of the Database. The database can be used to address questions of either a general nature (e.g., all available data

PLANT	CHEMICAL	DOSAGE	TIME	DEST	AMOUNT
GLYCINE MAX L. "SOYBEAN"	COPPER (560 PPM DW IN SLUDGE)	0 MT SLUDGE/HA	L001	SH	10.5 PPM DW
GLYCINE MAX L. "SOYBEAN"	COPPER (560 PPM DW IN SLUDGE)	335 MT SLUDGE/HA/5 YR	L001	SH	29.5 PPM DW
GLYCINE MAX L. "SOYBEAN"	ALUMINUM (16,080 PPM DW IN SLUDGE)	0 MT SLUDGE/HA	L001	SH	1313 PPM DW
GLYCINE MAX L. "SOYBEAN"	ALUMINUM (16,080 PPM DW IN SLUDGE)	335 MT SLUDGE/HA/5 YR	L001	SH	5988 PPM DW
GLYCINE MAX L. "SOYBEAN"	CADMIUM (12 PPM DW IN SLUDGE)	0 MT SLUDGE/HA	L001	SH	0.5 PPM DW
GLYCINE MAX L. "SOYBEAN"	CADMIUM (12 PPM DW IN SLUDGE)	335 MT SLUDGE/HA/5 YR	L001	SH	5.0 PPM DW
GLYCINE MAX L. "SOYBEAN"	MERCURY (8 PPM DW IN SLUDGE)	0 MT SLUDGE/HA	L001	SH	0.1 PPM DW
GLYCINE MAX L. "SOYBEAN"	MERCURY (8 PPM DW IN SLUDGE)	335 MT SLUDGE/HA/5 YR	L001	SH	0.9 PPM DW
GLYCINE MAX L. "SOYBEAN"	ZINC (2440 PPM DW IN SLUDGE)	335 MT SLUDGE/HA/5 YR	L001	SH	1188 PPM DW
GLYCINE MAX L. "SOYBEAN"	ZINC (2440 PPM DW IN SLUDGE)	0 MT SLUDGE/HA	L001	SH	45 PPM DW
GLYCINE MAX L. "SOYBEAN"	MANGANESE (53 PPM DW IN SLUDGE)	0 MT SLUDGE/HA	L001	SH	113 PPM DW
GLYCINE MAX L. "SOYBEAN"	MANGANESE (53 PPM DW IN SLUDGE)	335 MT SLUDGE/HA/5 YR	L001	SH	725 PPM DW
GLYCINE MAX L. "SOYBEAN"	LEAD (485 PPM DW IN SLUDGE)	335 MT SLUDGE/HA/5 YR	L001	SH	16.3 PPM DW
GLYCINE MAX L. "SOYBEAN"	LEAD (485 PPM DW IN SLUDGE)	0 MT SLUDGE/HA	L001	SH	5.0 PPM DW
GLYCINE MAX L. "SOYBEAN"	COPPER (560 PPM DW IN SLUDGE)	0 MT SLUDGE/HA	N000	LF	8.3 PPM DW
GLYCINE MAX L. "SOYBEAN"	COPPER (560 PPM DW IN SLUDGE)	335 MT SLUDGE/HA/5 YR	N000	LF	12.0 PPM DW
GLYCINE MAX L. "SOYBEAN"	ALUMINUM (16,080 PPM DW IN SLUDGE)	0 MT SLUDGE/HA	N000	LF	28 PPM DW
GLYCINE MAX L. "SOYBEAN"	ALUMINUM (16,080 PPM DW IN SLUDGE)	335 MT SLUDGE/HA/5 YR	N000	LF	83 PPM DW
GLYCINE MAX L. "SOYBEAN"	CADMIUM (12 PPM DW IN SLUDGE)	0 MT SLUDGE/HA	N000	LF	TRACE
GLYCINE MAX L. "SOYBEAN"	CADMIUM (12 PPM DW IN SLUDGE)	335 MT SLUDGE/HA/5 YR	N000	LF	0.75 PPM DW
GLYCINE MAX L. "SOYBEAN"	MERCURY (8 PPM DW IN SLUDGE)	0 MT SLUDGE/HA	N000	LF	TRACE
GLYCINE MAX L. "SOYBEAN"	MERCURY (8 PPM DW IN SLUDGE)	335 MT SLUDGE/HA/5 YR	N000	LF	0.03 PPM DW

Figure 2. Sample of uptake/accumulation records from the UTAB database displaying 6 of the 18 different data fields. Each line represents one record. Data are of sludge application to an agricultural field.[14]

on sludge) or of a more specific nature (e.g., uptake of zinc from sludge by soybean plants). Searches of the Fate File for a particular plant species or chemical require only 4–6 min of computer time (for an IBM AT 286, 8 MHz, 640K RAM, and 100 MB hard drive). Data in subfiles generated during a search can then be sorted, examined, and printed-out as necessary. Data may also be merged with its bibliographic source for reference as needed. Figure 2 presents an example computer printout of six different data fields from uptake/accumulation records. The example shows data collected on soybeans grown in sludge-amended soil under field conditions. Control data on soybeans grown without sludge are also included. For each of the seven metals examined, the shoot (all above-ground organs, SH in field DEST) concentrations after one month of growth (L001 in field TIME) increased between 3 and 26 times that of the control. When only leaves (LF in field DEST) were analyzed, the concentrations of metals were substantially less (although the time period was not specified, N000 in field TIME). Differences in how plant species mobilize and concentrate different metals between different plant parts are all important factors to be considered in assessing ecosystem contamination.

Status of the Database. The UTAB Database is currently maintained on a 100-MB IBM PC. The database itself occupies approximately 20 MB of disc space using dBase IV software. The Bibliographic File currently contains approximately 6550 references, and the Fate File contains slightly over 37 000 records. Another 16 000 records (12 000 being heavy metal data) are currently being processed. The Fate File contains data on approximately 900 chemicals and 350 plant species. The number of plant species will increase dramatically with the addition of heavy metal data. Approximately 65% of the records in the Fate File are on uptake/accumulation, 16% on translocation, 7% on adhesion, and 12% on biotransformation.

Presently, UTAB is not linked up to any computer network such as BITNET or INTERNET. Eventually it may be linked up to other chemical database systems by CAS and/or SMILES numbers as well as other biological databases on aquatics, wetlands, and animals. The ultimate utility of UTAB will be realized when its rapid ability to retrieve actual published data is used in combination with other databases and models to predict the movement and fate of chemicals in ecosystems.

ACKNOWLEDGMENT

Development of this database has been funded in part by the United States Environmental Protection Agency through Cooperative Agreement Grant CR-813585 to the University of Oklahoma. Thanks are extended to the many students who have been involved with the processing of the research papers used in preparing this database.

REFERENCES AND NOTES

(1) Moriarty, F. *Ecotoxicology: The Study of Pollutants in Ecosystems*; Academic Press: London, 1988; pp 1, 151–162
(2) United States Department of Agriculture (USDA). *The Pesticide Review, 1989*. Emergency Preparedness Branch EOLPD-ASCS: Washington, DC, May 1991; pp 2–6.
(3) Pimentel, D.; Levitan, L. Pesticides: amounts applied and amounts reaching pests. *Bioscience* **1986**, *36* (2), 86–91.
(4) Valdares, J. M. A. S.; Mingelgrin, M. G. U.; Page, A. L. Some heavy metals in soils treated with sewage sludge, their effects on yield, and their uptake by plants. *J. Environ. Qual.* **1983**, *12* (11), 49–57.
(5) Ward, N. I.; Roberts, E.; Brooks, R. R. Seasonal variation in the lead content of soils and pasture species adjacent to a New Zealand highway carrying medium density traffic. *N.Z. J. Exp. Agric.* **1979**, 7, 347–351.
(6) Ohlendorf, H. M.; Hothem, R. L.; Bunck, C. M.; Marois, K. C. Bioaccumulation of selenium in birds at Kesterson Reservoir, California, USA. *Arch. Environ. Contam. Toxicol.* **1990**, *19* (4), 495–507.
(7) Fan, A. M.; Book, S. A.; Neutra, R. R.; Epstein, D. M. Selenium and human health implications in California's San Joaquin Valley. *Toxicol. Environ. Health* **1988**, *23*, 539–559.
(8) Hansen, J. A.; Tjell, J. C. Sludge application to land—overview of the cadmium problem. In *Environmental Effects of Organic and Inorganic Contaminants in Sewage Sludge*; Davis, R. D., Hucker, G., L'Hermite, P., Eds.; Reidel Pub. Co: Dordrecht, Holland, 1983; pp 92–112.
(9) Davis, A. M. Selenium uptake in *Astragalus* and *Lupinus* species. *Agron. J.* **1986**, *78*, 727–729.
(10) Rosenfeld, I.; Beath, O. A. *Selenium*; Academic Press: New York, 1964.
(11) Lo, M. T.; Sandi, E. Selenium: occurrence in foods and its toxicological significance—a review. *J. Environ. Path. Toxicol.* **1980**, *4*, 193–218.

(12) Poole, S. C.; Bohman, V. R.; Young, J. A. Review of selenium in soils, plants, and animals in Nevada. *Great Basin Nat.* **1989**, *49* (2), 201-213.
(13) Callahan, M. A.; Slimak, M. W.; Gabel, N. W.; May, I. P.; Fowler, C. F.; Freed, J. R.; Jennings, P.; Durfee, R. L.; Whitmore, F. C.; Maestri, B.; Mabey, W. R.; Holt, B. R.; Gould, C. *Water-Related Environmental Fate of 129 Priority Pollutants: Volume I*; United States Environmental Protection Agency: Washington, DC, Dec 1979; EPA-44014-79-029a.
(14) Lutrick, M. C.; Robertson, W. K.; Cornell, J. A. Heavy applications of liquid-digested sludge on three ultisols: II Effects on mineral uptake and crop yield. *J. Environ. Qual.* **1982**, *11* (2), 283-287.

An Expert System for Analytical Data Management

J. R. LEE and T. L. ISENHOUR*

Department of Chemistry, Kansas State University, Manhattan, Kansas 66506

J. C. MARSHALL

Department of Chemistry, Saint Olaf College, Northfield, Minnesota 55507

Received October 15, 1991

Two universally important factors in research are domain expertise and data management ability. This paper describes a novel combination of a relational database management system with a rule-based expert system for research and quality control laboratories. An important advantage of this system is that it provides a user-friendly way to incorporate both domain expertise and data management abilities into a laboratory data management system. This paper presents a general strategy for archiving knowledge and data in a form immediately useful as a resource for an expert system. This strategy should facilitate the sharing of procedures among laboratories.

INTRODUCTION

Every year, there are a large number of papers concerning methods, apparatus, new techniques, and data analysis strategies from both quality control and research laboratories. Frequently, very similar work is reported by different laboratories and published in different journals. This paper will examine the use of analytical chemical techniques and expert database strategies to construct an expert analytical system. This system uses a relational data model to organize, store, and retrieve the data in such a way that it is readily and conveniently available to laboratory scientists. A novel feature of this system is the use of artificial intelligence (AI) techniques to convert the data and procedural information into a knowledge base that can be processed directly by an expert system. These techniques should help eliminate redundant work and lead to a greater standardization of methods among laboratories.

This paper will outline a knowledge base strategy that will be used to build an expert analytical system to integrate experimental data and published methods. This expert analytical system will generate and evaluate the data and knowledge base rules relating to a given procedure and then propose the best method for the analysis. With the laboratory environment standardized, the best conditions for an analysis from one laboratory can then be transferred to other laboratories.

This paper will demonstrate how this expert analytical system works by describing its application to the analysis of wastewater for soluble toxic metal ions using extraction followed by spectrophotometric analysis. Although the introduction of new instruments and new analytical methods in water analysis have had a significant impact, extraction is still a widely-used separation technique.

SYSTEM IMPLEMENTATION

System Domain. The purpose of this paper is to combine expert database strategies with traditional chemical techniques to create an expert system as an aid for the analytical research scientist. The analysis of toxic metal ions in water and wastewater is the general problem domain we have chosen.[1] This domain contains a great many standard methods frequently making it necessary for the system proposed to aid in the selection of an appropriate method from a list of alternatives.

Table I. Extraction Separation of Chelate Compounds

reagent	solvent	acidity	element extracted
DDTA	chloroform	5 N HCl	Pt(II), Hg, W, Cu
cupferron	ethyl acetate	2 N HCl	Th, V, Ti
8-H	chloroform	pH 5	Co, Ni, Pb, Cd
acetylacetone	chloroform	pH 5	Be
dithizone	chloroform	pH 9	Cu, Co, Mn, Pb, Ni

8-H is 8-hydroxyquinoline.

Analytical Techniques. The elimination of interferences for the quantitative analysis of specific metal ions is frequently necessary. Two separation techniques were used in this work to demonstrate the utility of expert database strategies in the system described.

(1) Extraction. (*a*) *Use of Chelates.*[2] The distribution of a solute between two immiscible solvents is an equilibrium process and can be written as

$$A_{aq} \xrightarrow{K} A_{org}$$

where A_{aq} and A_{org} refer to the solute concentration in the aqueous and organic solvents, respectively. $K = [A_{org}]/[A_{aq}]$ is the distribution constant. Because the distribution constant varies widely among solutes, it is possible to separate many solutes by extraction.

The extraction of a metal ion using a chelating reagent may be described as follows:

$$M^{+n}{}_{aq} + nHL_{org} \rightarrow ML_{n,org} + nH^{+}{}_{aq}$$

The extraction constant K_e is written as

$$K_e = (ML_n)[H^+]^n/[M^{+n}](HL)^n$$

0095-2338/92/1632-0148$03.00/0