(31) Levy, G. C.; Craik, D. J.; Kumar, A.; London, R. E., submitted for publication.
(32) Berger, S.; Kreissl, F. R.; Grant, D. M.; Roberts, J. D. *J. Am. Chem. Soc.* **1975**, *97*, 1805.
(33) Levy, G. C.; Craik, D. J.; Norden, B.; Phan Viet, M. T.; Dekmezian, A. *J. Am. Chem. Soc.* **1982**, *104*, 25.
(34) Beierbeck, H.; Martino, R.; Saunders, J. K. *Can. J. Chem.* **1979**, *57*, 1224.
(35) Axelson, D. E.; Holloway, C. E. *Can. J. Chem.* **1980**, *58*, 1679.
(36) Somorjai, R. L.; Deslauriers, R. *J. Am. Chem. Soc.* **1976**, *98*, 6460.
(37) Levy, G. C.; Kumar, A.; Craik, D. J., unpublished data.
(38) Levy, G. C.; Wang, D., unpublished data.

Summary of the History and Status of the System Development for the Environmental Chemicals Data and Information Network (ECDIN)

JUDITH M. HUSHON,*† JOHN POWELL,‡ and WILLIAM G. TOWN‡

Enviro Control Division, Dynamac Corporation, Rockville, Maryland 20852, and Commission of the European Communities, Joint Research Centre, I-21020 Ispra (Varese), Italy

Received June 17, 1982

The Environmental Chemicals Data and Information Network (ECDIN) has been developed over the last decade by the European Communities to provide information on all aspects of chemicals to the member countries through EURONET DIANE. This paper describes the evolution of the software for ECDIN from a homemade management system to the adoption of ADA-NATURAL. The expansion and structuring of the chemical files are also discussed, including the recent addition of the European Core Inventory (ECOIN).

The Environmental Chemicals Data and Information Network (ECDIN) developed by the European Communities provides a key source of data on chemical substances to many Europeans through EURONET DIANE. This paper is designed to document ECDIN's origin, development, and expansion to its present status as the central focal point for handling the European Inventory of Existing Commercial Chemical Substances (EINECS).

HISTORICAL BACKGROUND

The European Communities have adopted an environmental policy supported by a research program which provides technical and scientific backup to that policy and to the activities of the Member Countries (states). The research program consists of four main parts one of which is the management of environmental information.

Under this research program, the European Communities in 1973 decided to identify the problems involved in collecting and accessing data on environmental chemicals and to work toward their solution. The scope of the pilot phase was as follows:[1] to determine the availability of data; to refine the data structures in light of the available data; to assess the cost of data collection; to develop an input format; to locate data sources and potential contributors to a "network"; to demonstrate the computerized system; to test the reaction of users; to design a computer system for ECDIN.

The decision was also made to locate this effort at the EC Joint Research Centre at Ispra, Italy, which already had a large computer operation. This pilot operation was named ECDIN, the Environmental Chemicals Data and Information Network.

During this developmental phase, some basic principles were established:[2] (1) ECDIN would store relevant data on any chemical compound produced in sizeable quantity (>500 kg/year) irrespective of its toxicity or harmfulness. (2) ECDIN would be designed for management and retrieval of "hard" data (not document retrieval). (3) The organization would be chemically oriented. (4) The basic language of the system would be English with other languages being considered for later addition. (5) ECDIN would be available on a time-shared system. (6) The system would eventually be networked from its central location at Ispra, Italy, to at least one location in each of the EC Member states.

Table I. ECDIN Data Categories

no.	category
1	identification of the chemical
2	codified structure information
3	physical and chemical properties
4	chemical analysis data and methods
5	supply: production and trade
6	transport, packing, handling, storage, and hazards
7	use and disposal
8	dispersion and transformation in the environment
9	effects of the chemical on the environment (including toxicity)
10	regulatory data

During the next several years, the elementary data structure of ECDIN was evolved. The structure adopted was basically hierarchical, which permitted the organization of the data elements into a tree structure. Access to the data structure would be by inverted file, and numerical data could be accessed through data ranges. Data were to be stored on a chemical-by-chemical basis in the ten data categories shown in Table I. Data structures within these categories were subsequently elaborated. The major initial emphasis by the ECDIN staff was on the development of structures in the chemical identification, physical/chemical properties, and effects on man and his environment categories.

During the pilot phase, the number of chemicals in the system was limited to 5000. In addition, a subset of 100 chemicals was selected for which more complete data sets were assembled. Following the pilot phase, expansion of the ECDIN system has been very rapid due to the merger of several existing files: compounds of interest to the Administration of the Customs Union in Brussels (16 000); compounds in the 1976 edition of the *Registry of the Toxic Effects of Chemical Substances* (RTECS) published by the U.S. National Institute

† Dynamac Corp.
‡ Commission of the European Communities.

0095-2338/83/1623-0038$01.50/0

of Occupational Safety and Health (19 000); the BIOKON "black list" of chemicals (document ENV/599/76) (1500); SRI Directory of Chemical Producers—Western Europe (15 000). With the addition of substances from the European Core Inventory (ECOIN) and the compendium of known substances, the total size of the data base now consists of 65 000 substances.

At the time that the decision was made by the Council of the European Communities to proceed with the development of a chemical data system, the decision to locate it at the Joint Research Centre of the Commission of the European Communities at Ispra, Italy, was strongly influenced by the fact that this computer center had already developed the SIMAS software system for storage and retrieval of data on their IBM 370/165. SIMAS was originally designed to service a library of computer programs and was subsequently updated to meet the requirements of the ECDIN data bank. SIMAS allowed the ECDIN system designers to set up a number of categories each containing objects that could be cataloged. Key words and searchable identifiers could be assigned to the objects. In the ECDIN system, the objects were individual chemical compounds.

To enable the data stored in SIMAS to be retrieved, it was necessary to develop a thesaurus for ECDIN. SIMAS already contained procedures for thesaurus production and maintenance. The thesaurus was arranged as a hierarchy of key-word groups. One improvement in the SIMAS system as a result of the incorporation of the ECDIN data base was the ability to associate numerical values with key words and to search them by using operators such as equals, less than, greater than, and between. An additional feature permitted the translation of input units to a standard searchable unit.

The structure of the SIMAS system demanded the organization of the data into logical records, one for each compound. Each logical record could contain data for over 100 data elements where the data could be a distinct numerical value such as a melting point or a free-text summary of a field.

Although the availability of SIMAS on the Ispra comupter had strongly influenced the Council's decision to locate the ECDIN system there, it had always been recognized as a temporary measure until ECDIN's needs could be more fully defined. Since SIMAS was eventually to be replaced, efforts were made to keep the ECDIN data base independent of SIMAS. Data were, therefore, stored in a generalized input format and converted as needed into a format compatible with SIMAS. Because both the entry format and the SIMAS format involved a hierarchical structure, it was not possible, for example, to indicate the manufacturer of a chemical by a code cross-referenced to a master file of manufacturers. This resulted in continual repetition of names and addresses of major manufacturers and an added storage expense.

A decision was made in early 1977 that other data-base management systems (DBMSs) should be reviewed as alternatives to SIMAS to meet the software needs of ECDIN and the other data bases at the Joint Research Centre. ECDIN was by far the largest program, so it had the biggest impact on the selection. Since the way the ECDIN system would be used would affect its data structures and content, two user studies were commissioned.[3,4] The following major user groups were identified: governmental authorities at all levels, international organizations and institutions involved in environmental protection and control (establishment of quality objectives, regulatory activities, etc.); police and other public services (countermeasures in case of accidental contamination, inspections, law enforcement); water management authorities; industry (decisions on product development and marketing, research priorities, etc.); designers of monitoring networks; all kinds of scientists involved in environmental research, in laboratory as well as in field work (bibliography, definition of priorities for toxicity testing, identification and determination of pollutants in environmental samples, identification of disease factors in epidemiological studies, modeling of pollutant dispersion, etc.).

Also fundamental to the selection of a new software system was the reexamination of the structure of the ECDIN data base. The new software requirements were defined by the ECDIN staff assisted by representatives from the scientific data management section of the Joint Research Centre (JRC) computer department. This group considered four options: (1) extension of SIMAS, (2) development of a new DBMS, (3) use of an existing inverted file system (e.g., ADABAS, INQUIRE, or SYSTEM 2000), and (4) postponement of the decision pending new software developments.

The following observations were made regarding ECDIN's relationship to SIMAS and to any other DBMS: SIMAS requires repetition of elements related to more than one data entity; more flexibility was needed to introduce new elements in the form of descriptors, decode tables, etc.; the SIMAS four-character coded key words were inconvenient to use and would present particular user problems when the system became networked; under SIMAS the numerical searching procedures were not particularly sophisticated; SIMAS provided little flexibiltity for changing data structures; all updating required copying the file (under SIMAS interaction correction or addition was inconvenient); there is need to call external subroutines during machine–user dialogues in the user command language which was very complex with SIMAS.

The DBMS evaluation group then scored the ECDIN system and the other JRC data bases with regard to their requirements for text handling, general retrieval, and data management. ECDIN was the most demanding of the five systems scored. The following requirements were found to be key to ECDIN: key word indexing, thesaurus structuring, indexing on formatted fields, user retrieval language, structures for multiple data types, and data administration (i.e., facilities to help the person responsible for the data base maintain the contents and structures of the data base). Of secondary importance were text searching, programming language interface, report generation, access control, user accounting, direct updating, and user working format.[5,6]

The group then scored the existing DBMS alternatives against these primary and secondary requirements. The following DBMSs were considered: SIMAS, IMS, ADABAS, INQUIRE, SYSTEM 2000, TOTAL, 204, and RAMIS. Following this exercise, they recommended that the detailed selection process be limited to SYSTEM 2000, INQUIRE, and ADABAS. They also recommended that the software vendors be approached directly and that demonstrations be arranged as part of the detailed selection process.

For the detailed evaluation, the group then developed a list of required features and a list of desirable features based on the following scenario of use. The evaluation criteria are presented below:[5,6]

Required Features: (1) ease of use for nonprogramers (the system should allow for progressive refining of searches and should not depend on a detailed knowledge of the data structure), (2) multiple file/hierarchy access (must be able to link data in different hierarchies), (3) macro facility (this will provide a shorthand for commonly used searches, formats, and reports; it permits extension of the command set, and it permits menu selection, help, etc.), (4) user exits (must be able to call prewritten routines in a programing language and return), (5) processing of lists of selected records (used to refine searches and combines with Boolean expressions), (6) multiple occurring fields, (7) repeating groups of fields, (8) network data structure, (9) on-line updating of an existing record (a

Table II. Scoring of ADABAS, INQUIRE, and SYSTEM 2000

essential requirements	ADABAS	INQUIRE	SYSTEM 2000
ease of use	8	8	7
multifile access	6	8	5
macro facility	9	8	8
user exits	7	6	5
processing of lists	7	7	6
multiple fields	8	8	7
repeating groups	8	7	9
networks	9	9	6
on-line update	8	8	8
TSO etc. interface	8	8	8
subtotal	78	77	69
important features	ADABAS	INQUIRE	SYSTEM 2000
search criteria	4	5	5
listing options	4	5	5
prog lang int	5	5	5
standard language	4	4	4
floating point	4	5	3
variable-length fields	5	4	3
on-line add/delete	4	5	5
add field	5	3	3
upgrade field to key	5	4	5
add new record type	5	5	5
thesaurus handling	4	5	3
core usage	4	4	4
speed of retrieval	5	4	5
software support	4	3	4
cost	4	5	3
subtotal	66	66	62
grand total	144	143	131
cost of purchase (1977), AU	~80 000	~60 000	~120 000

Table III. Comparison of Key Features of ADABAS and INQUIRE for ECDIN Application (1977)

ADABAS	INQUIRE
Advantages	
good macro facility	comprehensive command language
expandibility through programming language interface	multiple file flexibility
Disadvantages	
too much left for do-it-yourself macros	less easy user exits, therefore less flexibility
multiple file listing inconvenient (ADASCRIPT)	communication problems with software support

feature not required for ECDIN), (10) interface with TSO or IMS/DC (these are the current monitors available at Ispra).

Desirable Features: (1) comprehensive search criteria, (2) comprehensive listing options, (3) programming language interface to the data base, (4) adaptability to a command language standard (European), (5) floating point option, (6) variable-length records, (7) on-line addition/deletion, (8) addition/deletion of fields, (9) upgrading of a field to key, (10) introducing a new record type, (11) thesaurus capability, (12) core requirement, (13) speed of response, (14) software support, (15) software purchase cost.

The abilities of the three systems to meet these requirements were then evaluated. The scores are shown in Table II. Table III contains a summary of the key advantages and disadvantages for ADABAS and INQUIRE on the basis of the features of the systems in 1977. The scores for ADABAS and INQUIRE were extremely close, but ADABAS was recommended as service was available nearby in Milan, Italy. The decision to adopt ADABAS also meant that the ECDIN DBMS would be compatible with that of DG XIII of the European Communities and also with a number of other European data bases.

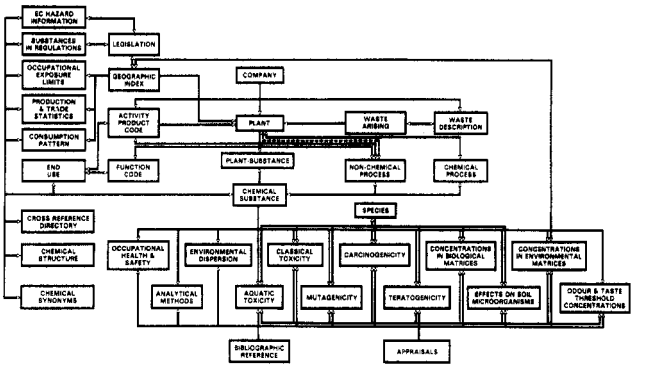

Figure 1. Current ECDIN file structure.

OPERATION OF ECDIN USING ADABAS

In late 1977 following the Council of the European Communities' agreement to support the Joint Research Centre's request for the purchase of an ADABAS system, a design study was undertaken.[7] The design study related to implementing ECDIN under ADABAS was performed by Scicon, a London-based firm, and several members of the ECDIN staff. The expressed goals of this design study were as follows: to analyze the data to be included in ECDIN based on both current holdings and projected additions, to design an ADABAS structure onto which the various ECDIN data can be mapped and specification of all required file and record structures, to recommend appropriate methods for converting current ECDIN data to the new ADABAS format, to design a strategy for executing the conversion, to recommend suitable updating procedures, to consider the access and display aspects of the proposed ECDIN system and how these are met by ADABAS, and to estimate the amount of effort needed to implement an operational ECDIN system by using ADABAS.

Pressure was being placed on the ECDIN staff to discontinue use of SIMAS as soon as possible and to rapidly establish an operational system to focus on the core components. The goal was adopted that service to users under ADABAS must be at least equal to that under SIMAS.

Figure 1 shows the proposed overall layout of the ADABAS data base. The boxes represent ADABAS files, and the lines between represent the ADABAS logical relations. The existing files are described in Table IV. The other files shown in Figure 1 will be added as they are designed and implemented.

During the design study, a number of decisions were made that were relevant to more than one file.[7]

Formating and Indexing. Whenever possible, data were to be incorporated into well-defined ADABAS fields. To facilitate this, descriptor fields were developed wherever such a need was anticipated.

Text. The standard method for holding text would be in a multiple-value field, limited to one line of text (65 characters). ECDIN, at least initially, will not be designed for text searching. Text in ADABAS is to be stored in the file to which it relates rather than in a central, cross-referenced file.

Name Searching. The first 30 characters of a normalized name will be stored in a descriptor field. Several names may be stored for the same chemical substance. Phonetic descriptors can also be used.

Precision of Numerical Values. Since ADABAS does not support floating-point rotation for searching, all fields with large numbers will be stored both directly and as the logarithm, with searching being performed on the logarithmic value.

Table IV. Descriptions for Existing ECDIN Files

file	type of data	estimated no. of records (12/31/81)
chemical substance	registration data and P/Chem properties	65 000
plant-substance	data about a plant relevant to production of a particular chemical substance	2 500
plant	name, address, location	1 100
company	name of company	600
standard industrial classifications	SIC codes	
chemical process[a]	data on process, byproducts, etc.	2 500
chemical structure	WLN data using CROSSBOW; will eventually contain data from the DARC file	25 000
chemical synonyms	synonyms and preferred name	350 000
occupational exposure limits	threshold limit values	12 000
production and trade statistics	quantities produced, consumed, imported, etc.	6 000
consumption patterns	use statistics	1 000
uses[a]	descriptions of uses, releases, exposures	450
occupational health and safety	safety measures and recommendations for treatment	500
analytical methods	descriptions of methods	350
EC hazard information	chemical specific hazard descriptions	900
dispersion and transformation in the environment[a]	environmental transport and fate data	
aquatic toxicity	aquatic toxicity data	18 000
carcinogenicity	carcinogenicity data	2 300
mutagenicity	mutagenicity data	500
classical toxicity	toxicity data	50 000
species index	thesaurus for identification of organisms	25 000
effects on microorganisms	biodegradation data	1 350
concentrations in environmental matrices	environmental monitoring data	20 000
odor and taste threshold concentrations	concentrations of a chemical necessary for detection or recognition	1 500
bibliographic references	references for citations in remainder of data base	19 500

[a] File being redesigned.

Source Codes. Source codes indicating the source of the data will be recorded at the record level.

Range Searches. A standard methodology will be used to facilitate searches of numerical data.

Initially, only a modest attempt was made to estimate disk space requirements. In 1979, when ECDIN consisted of 300 000 card images or about 6 megabytes of data, it was recommended that at least 15 megabytes be allowed for ECDIN and that a 100% increase in volume should be expected within the next year (i.e., 30 megabytes). This estimate was extremely low since as of December 1981 there were 200 megabytes, and 400 megabytes are expected by the end of 1982.

The ECDIN update system was developed to cope with batch updating of the data base in a manner which would guarantee data-base integrity. Rather than use a piecemeal approach, a table-driven solution to the design problem was found. By use of this technique, the input format, the checks to be performed, and the action to be undertaken with respect to updating the data base can be specified as entries in a table which itself is a file of ADABAS. Thus, changes in the input formats can be accommodated merely by changing the table, and the update program itself will only require modification when new types of checks on the data are required. Two levels of checks are allowed in the program: (1) syntax checks or other validity checks at the level of a field in an input format (e.g., checking that a numerical field contains only digits, signs, or decimal points); (2) validity checks requiring access to the ECDIN ADABAS data base (e.g., ensuring that a compound being added as a new compound is not already present in the data base).

The general program is thus independent of the ECDIN ADABAS data base. The preparation of the tables and the subroutines which perform the checks and updates are, of course, highly specific to the data base. The update system is now in routine operation.

One of the capabilities that is essential to a multipurpose, large-scale chemical data base is the ability to search on structure as well as nomenclature. The ECDIN system is acquiring access to two such systems, both of which were developed elsewhere and will merely interface with ECDIN. The CROSSBOW interface being developed by Fraser Williams (Scientific Systems) Ltd. consists of two parts: a preprocessor which accepts input in the form of a Wiswesser line notation (WLN) or a molecular formula and a fragment-search routine. A similar access to the DARC substructure search system is being developed by ARDIC.

At the present time, the system does not contain data actually designated as confidential, but this is expected within 2 years when the chemicals laws of the Member Countries become effective.

ECDIN DISPLAY SYSTEM

When ADABAS was installed at JRC, Ispra, in 1977, it was clear that none of the existing ADABAS command languages was ideal for ECDIN. The inherent complexity of the ECDIN data base led to the concusion that if ECDIN were to be made available to the casual user through EURONET DIANE, it would be necessary to develop a user-friendly conversational language capable of dealing with the more common types of questions. The ECDIN display system is the result of this decision.[8]

The ECDIN display system is still developing both in terms of the data file coverage and the types of search available. The ECDIN display system currently consists of four major segments: (1) access segment, access point for a search is chosen; (2) search segment, search parameter is chosen; (3) select segment, data files to be displayed are selected; (4) report segment, selected data are displayed and reviewed.

Certain commands have effect only within one of the segments while others have effect at any point in a user session. The main groups of commands are the following: SUPPORT commands which help the user to modify the dialogue, CONTROL commands which may be used to change the current search or initiate a new one, and REPORT commands

which may be used to modify the report being displayed.

The ECDIN display system is designed to be used by casual inexperienced users. ECDIN will, however, be accessed regularly by other users who will become expert in the use of the system. The user can determine for himself the level of dialogue to be used. These commands are as follows: PROMPT can be used by the experienced user to eliminate the messages which remind the casual user of the options available at various points in the user dialogue. SHORT removes the prompts and at the same time abbreviates all messages sent by the ECDIN display system. TITLE is used to eliminate titles from reports. This command is useful for teletype-compatible printing terminals (i.e., terminals which do not have video display screens). MENU is used to eliminate the phase in which a selection of data from a file is made before the file is displayed. When the menu is suppressed, display begins with the first record in the file. STATUS is used to determine which type of user dialogue is in use at any given time. It also gives information on the number of lines per page. LINES is used to set the number of lines per page to the number of lines that can be accommodated by a VDU terminal.

Once a search is complete, the user can then select how he wishes the data displayed. During the display of data, there are several commands which can modify the report being displayed or alter the sequence of reports. These commands are as follows: BIBLIO is used to display bibliographic references for the record currently being displayed. It is not available in all files. DISP allows the user to redisplay data from the current file and to make a different selection of records to be displayed. If the menu option has been suppressed, the DISP command causes the first record to be redisplayed. END causes the display of data from the current file to be terminated. Display of data continues with the next file selected. EXPAND is used to obtain explanations of codes shown in the display of the current record. The EXPAND command is not available in all files. NEXT causes the display of data for the current entry to be terminated. Display of data continues with the next entry retrieved. Naturally, the reports may also be modified by the LINES, MENU, and TITLE commands mentioned previously.

In August of 1980, the ECDIN system was transferred from the IBM 370/165 to an AMDAHL 470/V7A, a larger and faster machine which was upgraded to a 470/V8 in August of 1981. This was needed to prepare for the EURONET interface which was inaugurated in May of 1981. Trial use of the system by a selected set of users at CEC headquarters in Brussels and in each EC member country is currently underway.

ADVANTAGES OF NATURAL OVER ADASCRIPT+ FOR THE ECDIN APPLICATION

One final feature of ADABAS which influenced its selection as ECDIN's DBMS was the availability of ADA-NATURAL, a more comprehensive query language than ADASCRIPT+. NATURAL was first introduced in 1979 by Software a.g., the developers of ADABAS. It was advertised as "a consistent extension of ADASCRIPT+ and ADACOM". NATURAL is a stand-alone, user-oriented, program-development language. It can be used to create ad hoc queries, generate on-line file-description queries, and provide terminal-independent mapping. According to the vendor, "NATURAL saves about eighty percent of the original programing time". It does this by freeing the user from consideration of such areas as the technical functioning of the DBMS concerning control blocks, user IDs, and opening and closing of these areas.

For the ECDIN application, it proved to have many advantages, and the initial ADASCRIPT+ coding was rapidly converted when NATURAL became available. These advantages are summarized below.

(a) For a network arrangement, NATURAL lets the user process and print data from several files simultaneously. Under ADASCRIPT+ one could couple data from two files for searching, but could not combine the data for printing purposes. By use of NATURAL any file can be dynamically linked to as many as 80 others.

(b) NATURAL facilitates program development by permitting more efficient coding. The user does not have to become involved in the mechanics of conventional languages such as are usually required for file definition and program compilation. The coding from NATURAL is stated to be of comparable efficiency to that obtained under COBOL. In short, NATURAL is more of a programing language than ADASCRIPT.

(c) NATURAL permits the storing of source and object programs in a system file for subsequent execution or editing.

(d) NATURAL is easy to use for personnel with limited data-processing experience. The user does not have to have host language coding experience such as COBOL to be able to access the data files by using English-like commands.

(e) NATURAL can print multiple groups or periodic groups vertically for report writing. This was not possible with ADASCRIPT+ and made the construction of print routines extremely contrived.

(f) NATURAL offers an increased command set. For example, the LIST command can be used to list programs, modules, and files. This then can serve as a help device to a programer or other system user who wants to know what options exist.

(g) With NATURAL, because of the ease of accessing macros and other commands, the user can use the standard utility sort or the internal ADABAS sort. For large sorts, the standard utility sort is more efficient.

(h) NATURAL accomplishes the task of both ADASCRIPT+ and ADAWRITER. In fact, the vendor is not updating ADAWRITER anymore but is developing NATURAL instead.

(i) With ADASCRIPT, it is difficult to repair errors once they are loaded into the system. With NATURAL, one can check for syntax errors at the time of file creation by electing the syntax check mode and specifying the checks to be made.

(j) NATURAL provides a powerful mapping facility which is independent of the teleprocessing monitor and terminal used. By use of NATURAL, teleprocessing applications can be created much more readily than by using standard programing languages and teleprocessing mapping systems.

The conversion from ADASCRIPT+ to NATURAL did rectify some of the early user complaints about the software package. Despite these improvements, however, there are still some problems with implementing the ECDIN data base by using ADABAS. The following are examples: (1) The ADABAS file structure allows for only 2.5 levels of hierarchy since one can have multiple values within a periodic group; this results in the need to include redundant data and to set up a large number of files. (2) Linking of values belonging to the same occurrence of a repeating group can be achieved, but by roundabout means. (3) Use of the OR operator in search criteria is limited to alternative values of the same field. (4) Computed expressions are not allowed in search criteria. (5) Previous search criteria can only be referred to by processing lists of selected records. Special software has had to be written to compensate for this on ECDIN for the EURONET user interface. (6) Limits on field size require division of long fields such as text into lines. (7) ADABAS leaves a lot to do-it-yourself macros. ECDIN has had to develop its own user command facilities, including dialogues. These include HELP routines, menu selection, and preset screens. (8)

ADABAS requires additional software for thesaurus structuring. Should ECDIN ever want to add a directory function with a built-in thesaurus, it would require a great deal of work. (9) ADABAS is not a text-handling system. ECDIN currently has no plans to incorporate a bibliographic file with character-string searching. Rather its bibliographic file is merely an adjunct to the data files. (10) If ECDIN grows rapidly, updating and searching of the file can become very cumbersome and complex due to the ADABAS pointer system.

Should future changes in ADABAS rectify some of the problems, it is likely that these new changes will be rapidly incorporated by the ECDIN facility. Until then, they are forced to accept redundancy, proliferation in the number of files, and the necessity for preparing numerous macros.

ECDIN'S ROLE IN THE EUROPEAN CHEMICAL INVENTORY

As part of the Council of the European Communities Sixth Ammendment to the 1967 Council Directive on the Classificaiton, Packaging, and Labeling of Dangerous Substances, it was determined that an inventory of chemical substances manufactured and used in the Member Countries was required as a means of defining those new chemicals that were subject to special reporting provisions. The ECDIN facility has been responsible for constructing the European Core Inventory (ECOIN), a compendium of about 35 000 chemicals now on the European market. This became official when it was printed in the *Official Journal of the European Communities* on March 31, 1982. Over the next 9 months, chemical companies are expected to submit data on an estimated 25 000–30 000 additional chemicals not included in ECOIN. The ECDIN staff will be examining these submissions and using them to prepare the final European Inventory of Existing Commercial Chemical Substances (EINECS). This will have the net result of increasing ECOIN's size to about 60 000 unique chemical substances.

REFERENCES AND NOTES

(1) Geiss, F.; Town, W. G. "Presentation at a Seminar Held on Environmental Chemicals Data and Information Network (ECDIN)"; Research Department of the Environment: London, England, July 6, 1977; Report No. 29.
(2) Boni, M.; Geiss, F.; Petri, J. H.; Town, W. G. "The Development of a Data Network on Chemicals and their Effects on the Environment. The Environmental Data and Information Network (ECDIN) of the European Community". In "Proceedings of Eurim II, A European Conference on the Application of Research in Information Services and Libraries"; RAI International Congrescentrum: Amsterdam, 1976.
(3) Bletchly, J. D. "A Report on the Apparent Needs of Possible Customers in the United Kingdom for an Operational ECDIN"; 1979.
(4) Fourneau, J. P. "Les Utilisateurs Potentiels de l'Environmental Chemicals Data and Information Network (ECDIN)"; Paris, 1979.
(5) Tagg, R. M. "A Study of Software Requirements for Scientific Data Bases"; Scicon: London, 1977; Part a.
(6) Tagg, R. M. "Evaluation of Alternative Software for Scientific Data Bases"; Scicon: London, 1977; Part b.
(7) Huitson, P. R. "ECDIN Input Format Manual"; Scicon: London, 1977.
(8) Town, W. G.; Powell, J.; Huitson, P. R. *Inf. Process. Manage.* **1980**, *16*, 91-108.

Conformation Specification of Chemical Structures in Computer Programs

ANDREA L. FELLA,* JAMES G. NOURSE, and DENNIS H. SMITH

Department of Chemistry, Stanford University, Stanford, California 94305

Received September 15, 1982

A computer implementation of conformation specification of chemical structures is described. The implementation produces a canonical name for the conformation given an arbitrary specification of the conformation of the structure. A compact or coded form of the conformation name for use in data bases is also described.

INTRODUCTION

Complete representation of a chemical structure or substructure within a computer program requires that constitution, configuration, and conformation all be specifiable. The DENDRAL group at Stanford University has implemented this representation in three successive stages. Constitutional representations have been dealt with by the GENOA[1] and CONGEN[2] programs which are capable of constrained generation of all constitutional isomers for any molecular formula. Inclusion of stereochemistry was accomplished by the STEREO[3] program which is capable of constrained generation of all possible stereoisomers. The structural representation within the programs is in the form of connection tables augmented with stereochemical information on the chiral centers.

Recently a computer program was written by our group which extends this representation to include conformation. While others have mentioned that their structural representations could be extended to include conformations,[4] a program that actually implements this is not, to our knowledge, available. Note that this program is not a generator of conformations, it expects that defined constitution and configuration are provided, and it also expects information about a specific conformation. From this information, the program produces a canonical (unique) representation of the conformation of the structure (or substructure).[5] This is not an explicit geometric representation, although given standard bond lengths and angles, along with the torsional angles around appropriate bonds, the approximate geometry of the structure could be determined. The uses of such a nongeometric representation include structure registry, search and retrieval systems, and data bases where the conformation of the structure is important, such as proton NMR data bases.

DESCRIPTION OF NAMING SCHEME FOR CONFORMATIONS

In order to determine a canonical name for the conformation of a structure, one must have a naming scheme.[6] The first aspect of the naming scheme deals with the torsional angle about a single bond. This continuous variable needs to be converted into a discrete variable for our purposes. For an example of this property, consider the six possible staggered and eclipsed positions around an sp^3–sp^3 bond (see Figure 1). The positions are described uniquely in terms of the atom numbers by always basing the conformation definition on the

0095-2338/83/1623-0043$01.50/0