A Unique Method of Computerizing Chemical Process Information

J. M. HUSHON,* E. CLARENCE-SMITH, J. SCOPEL, J. A. KING, J. POWELL, and W. G. TOWN

Research Systems Division, Bolt Beranek and Newman, Inc., Arlington, Virginia 22209

Received December 2, 1983

Traditionally, organic chemical process data have been organized on a chemical-by-chemical basis. The unit process/product process approach to organizing chemical process data that has been developed for the European Commission's ECDIN data base has advantages in that it permits a linking of functionally similar reactions, it facilitates a hierarchical organization of data, and it simplifies the query procedure. Unit processes can be defined as single-step chemical reactions. With application of this method to ECDIN, the processes for making the highest volume chemicals have been evaluated and the sequences of unit processes identified. In addition, the chemical process files added to ECDIN will permit access to information on process conditions and the process, as well as the commonly reported feedstocks and products/byproducts.

INTRODUCTION

Over the last 10 years, the European Commission has been developing the Environmental Chemicals Data and Information Network (ECDIN) at its Joint Research Centre at Ispra, Italy. The ECDIN data base is designed to provide access to the many types of data on chemical substances that are needed to evaluated their impacts on man and his environment. The types of data in ECDIN include chemical identification information, structure information, physical and chemical properties, analytical chemical methods, production and trade data, transport, packing, storage and hazard data, use and disposal data, information on dispersion and environmental transformation, effects of the chemical in the environment (toxicity), and regulatory data. These data are currently organized into 26 files, part of which can be accessed by both government and private users through EURONET DIANE, TYMNET, and other networks.

To evaluate chemical contaminants, releases, and exposures, data are required on organic chemical processes. However, because there are 40 000–60 000 organic chemicals in commerce at any time, the difficulty of obtaining and organizing this information is great. The approach to organizing organic chemical process data in EDCIN is unique and can provide a viable model for other systems and applications.[1]

USER NEEDS FOR CHEMICAL PROCESS INFORMATION

During ECDIN's developmental phase, the European Commission requested that consultants perform several user-needs surveys and that they include in these surveys the following user groups: governmental authorities at all levels and 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; designers of monitoring networks; 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.).[2,3]

Through these surveys and subsequent contacts with chemical engineers, the following types of questions were identified to which ECDIN is being designed to provide a response. What are the identities and, if possible, concentrations of the pollutants predicted to be in the wastewater from product process X or from the manufacture of Y? Which of the chemicals in the waste from X process are the most toxic, persistent and hazardous? Which plants use a given process and make a given product or byproduct? Are the plants making a particular chemical concentrated geographically? Which chemicals at a facility should be included in a monitoring program? Chemical X has been detected in a river water sample, which facilities upstream may discharged it? Which processes form chemical X as a product, a byproduct, or a waste? Is there a process for producing chemical X that does not involve producing chemical Y?

ORGANIZING ORGANIC CHEMICAL PROCESS DATA

To decide what files need to be created, it is necessary to have available a model of real-life chemical processes that is simple enough to allow a practical description of the most important functions but that does not lose significant information.

A chemical process can, in general, be considered as being made up of four basic units: preconditioning units, where the feedstocks are brought to the necessary conditions of temperature, pressure, and physical state; reaction units, where both the desired products and the undesired products or byproducts are formed; separation units, where the desired products are separated from each other and from other chemicals present (byproducts, unreacted feedstocks, entrained catalysts, etc.); storage units, which act as buffers between upstream supply and downstream demand.

Because most chemical processes use several reaction steps to transform the initial feedstocks to the final desired products, these four units can be repeated many times in the same process, and the relationship between them can be complex. Nevertheless, each reaction unit in a chemical process, which corresponds to a step in the reaction pathway, is generally preceded by a preconditioning unit and followed by a separation unit and, sometimes, a storage unit (Figure 1). We have called such a set of units a "product process". Thus, a chemical process can be considered as being made up of a string of such product processes, with the products of one or more product processes becoming the feedstocks for the next product process. The description of each product process and the order in which they are linked define a chemical process.

The advantage of describing a chemical process as a string of product processes is that this strictly links the chemicals involved in that process to the reaction step where they are (intentionally or unintentionally) consumed or produced. This allows the information gathered on reaction outcomes of chemical processes to be used to predict the reaction outcomes of chemical processes on which little or no information is

0095-2338/84/1624-0148$01.50/0

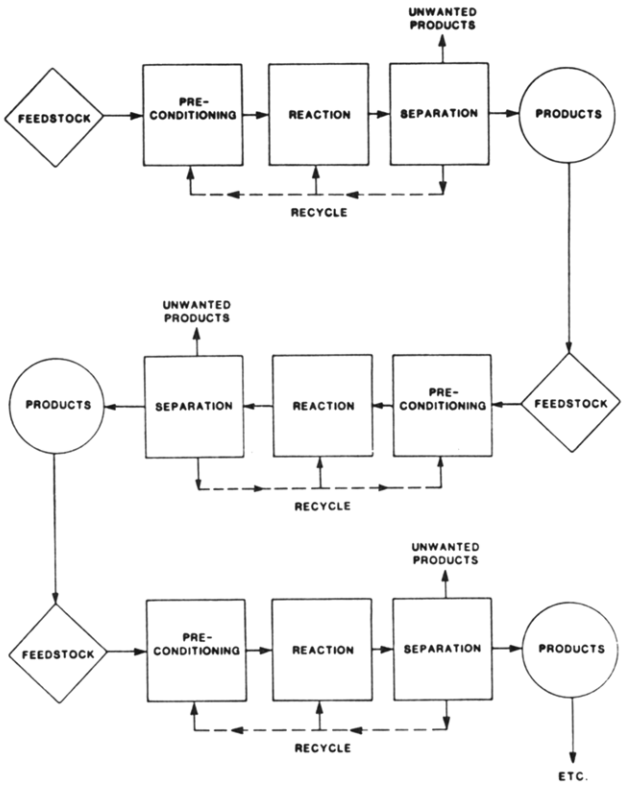

Figure 1. Conceptual model of a chemical process.

available. This is of greatest value for determining byproduct formation in chemical processes on which relatively few data exist but for which environmental contamination poses a potentially serious problem.

This predictive ability rests on the assumption that chemicals belonging to the same chemical family will react in similar ways; in this way, data on the reactions of one member of a chemical family can be used to understand the reactions of other members of the family. This predicitive ability is reinforced where commercial chemical reaction pathways are concerned because many patented reaction processes are designed to be applicable to a family of chemicals rather than to single chemicals. While this approach is empirical, it is more economical to use on a large scale and in a computer environment than a more scientific approach based on an understanding of the fundamental thermodynamics and kinetics of each reacting system.

To build this predictive ability into ECDIN, entries in one file listing the specific carbon-based chemicals consumed or produced (intentionally or unintentionally) in product processes will be paralleled by entries in another file listing these chemicals in their generic form. For example, a specific product entry saying that ethanol is reacted commercially with ammonia to form ethylamine with morpholine, piperizine, and nitrogen-containing polymers as pollutants in the discharge streams will generate an entry in the generic reaction file saying that an aliphatic alcohol is reacted with ammonia to form an alkylamine with alkylglycols, nitrogen-containing heterocyclics, nitrogen and oxygen containing heterocyclics, alkylamines, and nitroalkyl polymers as pollutants. The two file entries will be linked. Thus, if a user cannot find the information he requires about chemical reactions in a particular product process, he can look for similar information at the generic level by using information on similar reactions of chemicals from the same family. We have called such a set of generic information a "unit process"; and each unit process relates to multiple specific product processes.

To allow a more logical organization of the information on chemical processes, use has been made of the fact that only a limited set of about 40 reaction types are used commercially to synthesize organic chemicals (the major ones are listed in Table I). Each unit process and product process are then classified under one of these reaction types. The use of this classification facilitates information storage and retrieval.

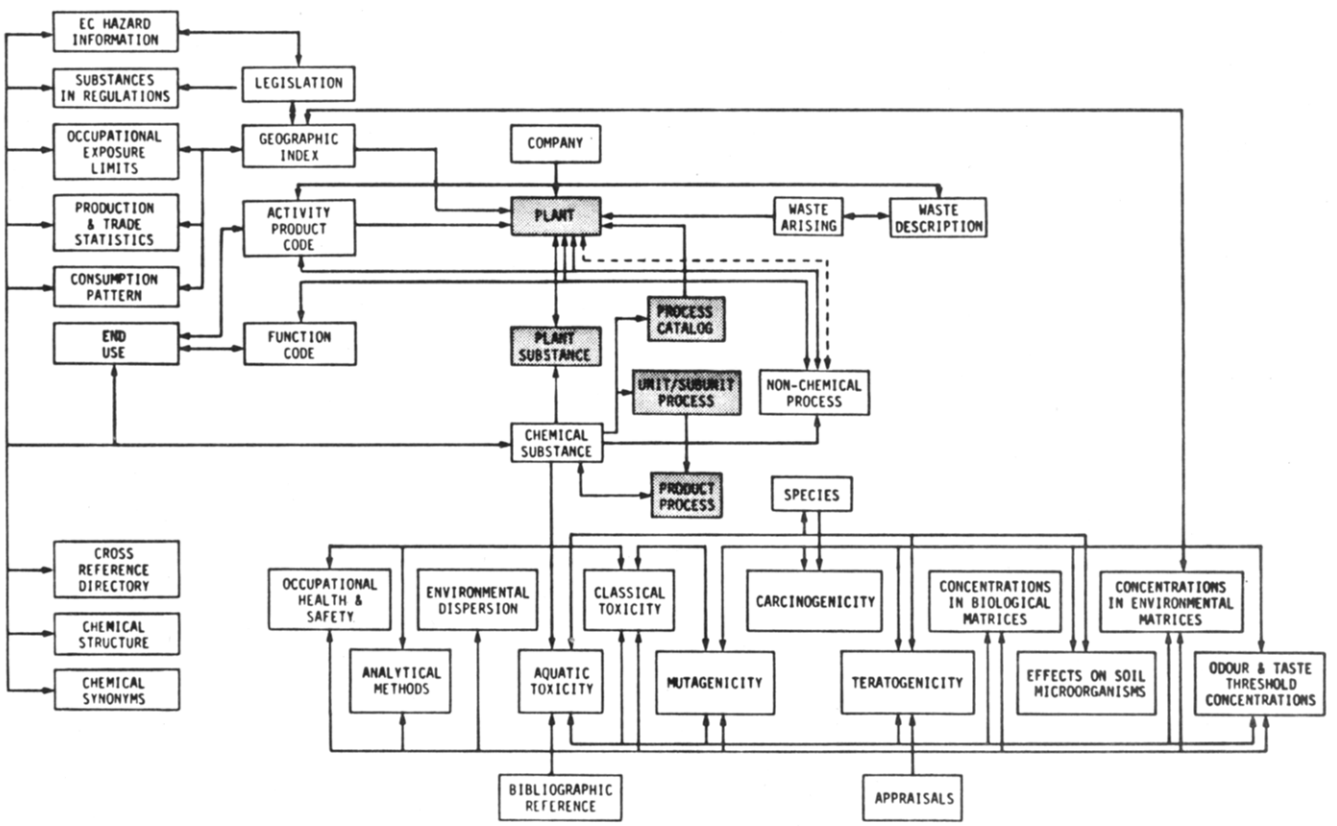

Figure 2. Current ECDIN file structure.

Table I. Major Unit Processes

(1) alkylation	(17) halogenation
(2) ammonolysis/ amminolysis	(18) hydrodealkylation/ transalkylation
(3) ammoxidation	(19) hydrohalogenation
(4) carbonylation (oxo)	(20) hydrolysis hydration
(5) chlorohydrination	(21) isomerization
(6) condensation	(22) nitration
(7) cracking (catalytic)	(23) oxidation
(8) cyanation	(24) oximation
(9) cyclization	(25) oxyhalogenation
(10) dehydration	(26) phosgenation
(11) dehydrogenation	(27) polymerization
(12) dehydrohalogenation	(28) rearrangement
(13) dimerization	(29) reduction
(14) epoxidation	(30) salt formation
(15) esterification	(31) sulfonation
(16) ether formation	

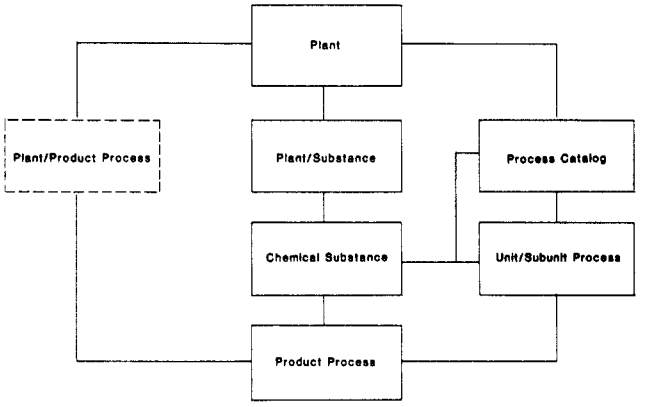

Figure 3. File linkages for chemical process information.

FILE ORGANIZATION TO ACCOMMODATE THE PRODUCT PROCESS/UNIT PROCESS APPROACH

The ECDIN data base is implemented by using the ADABAS data base management system. This strongly influences how the files must be structured. Figure 2 shows the overall existing ECDIN file structure and the relationship of the chemical process files to the rest of the ECDIN structure. As can be seen from this figure, the CHEMICAL SUBSTANCE file occupies a central position and provides chemical identification information for use in a number of the ECDIN files, including all of the organic chemical process files. Figure 3 shows those new files that relate to storing and retrieving chemical process information. The data elements included in three of the new files are shown in Table II.

The use of the ADABAS data base management system has definitely affected the design of the ECDIN system. Because ADABAS only supports two levels of hierarchy within each file, the ECDIN designers have often been forced to create more files than would be necessary under a more hierarchical system. This means that files must be created with redundant data elements that make possible the bridging (indicated in Figure 3 by the lines and arrows) between related files. In more hierarchical systems, for example, the SUBUNIT PROCESS and PRODUCT PROCESS files and maybe even the PROCESS CATALOG file could have been combined. ADABAS was selected for this application, however, because of its record maintenance utilities, its speed of retrieval, its networking potential, its macrofacility, its expandability through a programming language interface, and its ease of use.

The PRODUCT PROCESS file contains information on individual product processes derived from a number of sources in the open literature. This file contains a listing of the chemicals consumed and produced in the product process including descriptions with reaction conditions, media to which waste components are discharged, and the approximate quantities of each waste. Figure 4 provides a sample of the output from this file.[4]

The UNIT/SUBUNIT PROCESS file contains a description of the unit process in more general terms than the related entries in the PRODUCT PROCESS file. Some unit processes can be broken down one step further into subunit processes. For example, there are three kinds of ammonolysis reactions (dehydrochlorination, dehydration, addition), and each is treated as a subunit process. Figure 5 contains a sample unit process description.[5]

Table II. File Contents

UNIT/SUBUNIT PROCESS File
- specific subunit process code[a]
  - unit process code[a]
  - unit/subunit indicator[a]
- unit/subunit process name
- number of subunit processes code
- number of product processes code
  - product process codes[a]
- lines of unit/subunit process description
  - unit/subunit process description

PRODUCT PROCESS File
- process process code[a]
  - unit process code[a]
  - unit/subunit process indicator[a]
- number of lines of process description
- description of specific operating conditions
- number of wastes
  - ECDIN number of waste[a]
  - predicted/measured indicator[a]
  - level[a]
  - medium[a]
  - means of separation[a]
- number of feedstocks
  - ECDIN numbers of feedstocks[a]
- number of products
  - ECDIN numbers of products[a]

PROCESS CATALOG File
- process catalog number[a]
- number of feedstocks
  - ECDIN Number[a]
  - CAS Registry Number
- number of products
  - ECDIN Number[a]
  - CAS Registry Number
- number of unit processes
  - unit process code[a]
- number of owner companies
  - owner/licensor of process
- number of plants
  - plant numbers of specific locations using this product process[a]

[a] Key data elements, indexed, can be used for searching.

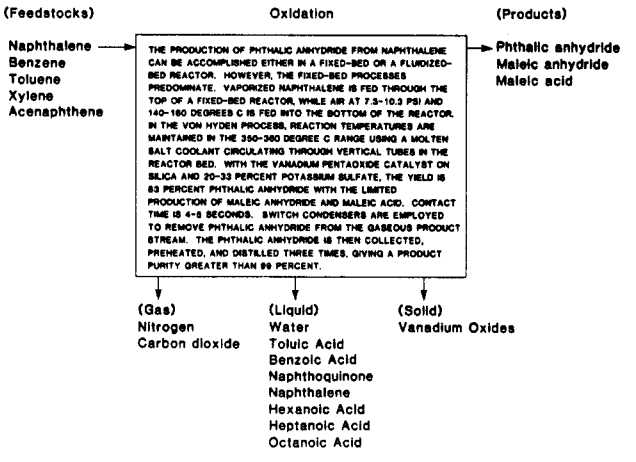

Figure 4. Sample product process.

```
OXIDATION
0024
FOR THE ECDIN SYSTEM, OXIDATION IS DEFINED AS THE PROCESSES
WHEREBY OXYGEN (OR ANY OTHER ELECTRON ACCEPTOR) IS ADDED TO
A COMPOUND, OR HYDROGEN IS REMOVED FROM A MOLECULE VIA THE
USE OF OXYGEN (OR ANY OTHER ELECTRON ACCEPTOR). OLEFINS MAY
BE OXIDIZED TO ALCOHOLS, THEN TO ALDEHYDES, OR KETONES AND
THEN TO CARBOXYLIC ACIDS, ETC.
EXAMPLES OF SOME OF THE TYPES OF OXIDATIONS ARE:
     A. TO SHOW AN INCREASE OF OXIDATION STATE, ETHANOL CAN
        BE REACTED WITH OXYGEN IN THE PRESENCE OF A
        FERROMOLYBDENUM OXIDE CATALYST AT 300 DEGREES CENTI-
        GRADE TO FORM ACETALDEHYDE AND WATER.
     B. AS AN EXAMPLE OF THE INTRODUCTION OF OXYGEN INTO A
        NON-OXYGEN CONTAINING MOLECULE, METHANE AND OXYGEN
        ACT TO FORM FORMALDEHYDE.
     C. OXIDATION CAN ALSO BE USED TO EFFECT AN INTRODUCTION
        OF OXYGEN PLUS CRACKING. FOR EXAMPLE, NAPHTHALENE
        PLUS OXYGEN CAN FORM PHTHALIC ANHYDRIDE PLUS CARBON
        DIOXIDE.
     D. IN AMINO COMPOUND OXIDATION, IF OXYGEN IS ADDED TO
        AN AMINO COMPOUND, AN AZO COMPOUND PLUS WATER IS
        FORMED.
     E. AS AN EXAMPLE OF SULFUR COMPOUND OXIDATION, IF SOME
        FORM OF OXYGEN IS ADDED TO THIOETHANOL, ETHYL
        SULFONIC ACID IS PRODUCED.
```

Figure 5. Sample UNIT/SUBUNIT PROCESS file entry.

Finally, a PROCESS CATALOG file has been created. This file serves as an index to cross-link the UNIT/SUBUNIT PROCESS and PRODUCT PROCESS files by providing, for each unit process, a list of related product processes. It also contains information on who owns the process, the number of licensees, and the number of commercial plants. This catalog extends some of the initial work done by MITRE Corp. for the U.S. Environmental Protection Agency. In the ECDIN catalog, all entries represent single-step chemical reactions. This is an improvement over the older version in which processes were occasionally grouped and it was impossible to determine the reaction sequence. This improvement was necessary, however, before a machine file could be constructed. Figure 6 provides an example of a typical set of PROCESS CATALOG entries.[6,7]

This arrangement of files and data elements preserves the logical integrity of the system by permitting related elements to be stored together and facilitating accessing of several related files with carefully constructed cross-links.

APPLICATIONS OF THE ECDIN UNIT PROCESS/PRODUCT PROCESS APPROACH

It is possible to envision a number of possible applications of the ECDIN approach and data base to real problems confronting governments and industry. First, they can be used to predict waste-stream components from organic chemical processes to select those of greatest concern as possible regulatory or monitoring targets. By taking the ECDIN list of predicted wastes from the various processes at a facility, the combining of the pollutant data with an indication of toxicity allows the regulators to select streams or chemicals that are of greatest potential regulatory concern for the facility. This then may help them target control strategies.

Another potential application may occur when a particular waste is identified or a mixture of toxic chemicals is found in a river. The ECDIN system can be used to identify what processes might have produced those wastes and then, by use of the plant data file, to select those nearby plants that may have been responsible for the release.

In general, the applications include the following: identifying potentially hazardous chemicals in the discharges from organic chemical processes; identifying the chemicals in wastes as a first step in annalyzing whether those wastes are likely to be toxic; estimating the impact of wastes from production of organic chemicals on the environment, occupational population, and general population as a function of production trends and subsequent releases to all three media; predicting quality and quantity of wastes from specific processes, plants, or industry as a whole.

The unit process/product approach also obviously has direct applications for storing organic chemical process data in other computerized data systems. Historically, chemical process data have been organized by product chemical, which permits minimal indexing. The approach suggested here permits the

OXIDATION (continued)

PROCESS CATALOG NUMBER	PRODUCT	FEEDSTOCK	OWNER OF PROCESS	NO. LICENSED COMMERCIAL PLANTS	TOTAL NO. OF U.S. COMMERCIAL PLANTS
0030-03	Acetaldehyde	Ethyl Alcohol, O_2	BP Chemicals Int. Ltd.	4	(2)
			Veba Chemie AG	2	-
2844-01	Pelargonic acid Caproic acid Azelaic acid	Oils (tall, red, soy bean) Ozone, O_2	Welsbach Corp.	U	-
2844-02	Pelargonic acid Undecanoic acid Tridecanoic acid	alpha-Olefins Ozone, O_2	Welsbach Corp.	N	-
1070-01	Cumene hydroperoxide	Cumene, O_2	Allied Chemical	U	(17)
			BP Chemicals International Ltd. and Hercules Inc.	22	(17)
			Rhone-Poulenc S.A.	5	-
			UOP Process Div.	3	-
2960-01	Phthalic anhydride	Naphthalene, O_2	Nippon Steel Chemical Co.	U	(11)
			Scientific Design Co., Inc.	8	-
			Sherwin Williams Chemicals/Badger Co., Inc.	14	-
			United Chemicals & Coke, Ltd.	U	-
			Von Heyden/Wacker	U	-
2060-02	Phthalic anhydride	o-Xylene, O_2	Alsuisse	3	(1)
			BASF AG	20	-
			Rhone-Poulenc S.A.	7	-
			Scientific Design	4	-
			Von Heyden/Wacker	65	-
3132-01	Pyromellitic acid	Durene (1,2,4,5-tetra-methyl benzene), O_2	Gelsenberg-Chemie GmbH	N	-
3239-01	Suberic acid Dodecanoic acid	Cyclic olefins Ozone, O_2	Welsbach Corp.	N	-
3280-01	Terephthalic acid	p-Xylene, O_2 Methanol	Eastman Kodack Co.-Tennessee Eastman Div.	1	(8)
			Institut Francais du Petrole	N	-
			Lummus Co.	U	-
			Standard Oil Co. (Indiana)	26	-
			Toray Ind. Inc.	2	-
			Uni-Hills	1	-

Figure 6. Sample catalog of unit process entry. (U) At least one plant exists; total is unknown. (N) No commercial plants exist.

more logical primary organization by type of unit process than by product process, which is directly relatable to a unit process. This permits the creation of logical hierarchies and minimizes the sizes of the indexes required for most operations.

PLANNED EXPANSION OF THE ECDIN SYSTEM IN THE CHEMICAL PROCESS AREA

At the present time, two of the three primary files are completed (UNIT/SUBUNIT PROCESS and PROCESS CATALOG files), and the PRODUCT PROCESS file has been filled with data on 76 out of an estimated 382 product processes. Additional work is therefore needed to complete the Product Process file and to develop a user-dialogue display system to permit a user friendly interface to the data. It is expected that outside users will be limited to the use of a set of prestructured queries as they are in the other areas of ECDIN. This permits the system developers to limit access to the actual system. Future efforts are also required to develop an algorithm to permit creation of production/waste trees that will support the display system.

Later, it is envisioned that a PLANT PRODUCT/PROCESS file will be implemented to provide locations for the production facilities employing the product processes by using preestablished map grids. Another anticipated file is that on CONTROL TECHNOLOGY that will be linked to the UNIT/SUBUNIT PROCESS file. The unit process approach is particularly well suited to this logical addition since most control technologies are designed to be applied to sets of similar reactions such as constitute a unit process. It is expected that the CONTROL TECHNOLOGY file will contain data on the efficiency, cost, and manufacturer(s) of the technology and where facilities that use this approach are located.

CONCLUSIONS

The advantages to the user from storing and manipulating process data as developed for implementation on the ADABAS include the ability to consider as a group those chemical processes that operate under similar conditions, the ability to access data on selected chemical processes in a logical fashion, and the ability to combine waste data for an entire production stream. Though the file system described here is only partially implemented, when it is completed, it will provide a powerful accessible, interactive tool for indentifying and solving chemical process waste problems.

REFERENCES AND NOTES

(1) Hushon, J. M.; Powell, J.; Town, W. G. "A Summary of the History and Status of the System Development for the Environmental Chemicals Data and Information Network (ECDIN)". *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 38.
(2) Bletchly, J. D. "A Report on the Apparent Needs of Possible Customers in the United Kingdom for an Operational ECDIN"; EC Joint Research Centre: Ispra, Italy, 1979.
(3) Fourneau, J. P. "Les Utilisateurs Potentiels de l'Environmental Chemicals Data and Information Network (ECDIN)"; Paris, 1979.
(4) Dynamic "PRODUCT PROCESS File"; prepared under Contract 1857-81-05 EP ISP for the EC Joint Research Centre at Ispra, Italy, 1982.
(5) Dynamic "UNIT/SUBUNIT PROCESS File"; prepared under Contract 1857-81-05 EP ISP for the EC Joint Research Centre at Ispra, Italy, 1982.
(6) Herrick, E. C.; King, J. "Catalog of Organic Chemical Industries Unit Processes"; MITRE Corp.: McLean, VA, 1979.
(7) Dynamac "Revised Catalog of Organic Chemical Industries Unit Processes"; prepared under Contract 1857-81-05 EP ISP for the EC Joint Research Centree at Ispra, Italy, 1982.

An Interpretation of Chemical Abstracts Service Indexing Policies†

RUSSELL J. ROWLETT, JR.
Caropines, Myrtle Beach, South Carolina 29577

Received December 5, 1983

Throughout the early years of Chemical Abstracts Service, its document analysts selected index-access points that coincided with where a searcher would be most likely to look first. The growth of the literature, the broadening interests of chemists, and the increased complexity of the science demanded more systematization. Accordingly, six basic indexing philosophies have evolved from the beginning and are still followed today. They are (1) selection of subjects, not words, from original documents, (2) use of molecular formulas as central building blocks, (3) a high degree of specificity, (4) inverted names for organic molecules, (5) continuity, and (6) use of highly trained scientists–analysts for preparation of abstracts and index entries.

Much documentation exists of how Chemical Abstracts Service (CAS) indexes chemical substances and general subjects. The numerous indexes produced over the last 76 years are described thoroughly. The policies that guided index production are well-known and detailed in several publications. I will not refer specifically to any of these excpet as they are needed to understand an illustration.

My goal is to interpret six of the major philosophies that have guided CAS indexing for 66 years and continue to guide this indexing today. An experienced searcher will have encountered these philosophies one by one. I know of no collection and interpretation of their interrelationships.

† 1983 Herman Skolnik Award address, presented before the Division of Chemical Information, 186th National Meeting of the American Chemical Society, Washington, DC, Aug 30, 1983. R.J.R. is a former Editor and Director of Publications and Services of Chemical Abstracts Service.

I said "philosophies that have guided CAS indexing for 66 years". *Chemical Abstracts* (CA) is 76 years old, but indexes as we know them today did not exist during CA's first 9 years, 1907–1916. In the 1952 history of the American Chemical Society,[1] the 43-year Editor of CA, E. J. Crane, wrote, "In a journal as extensive as is *Chemical Abstracts*, information can be buried as far as retrospective searching is concerned unless a really effective index key is provided. With many other developmental problems to handle, the editors apparently did not give special attention to indexing of *Chemical Abstracts* during the first few volumes. Accordingly, when the First Decennial Index was authorized, reindexing by subjects of the first nine volumes was properly undertaken. To provide a model for the First Decennial Index, and for future annual indexes, a special study of subject indexing was made before Volume 10 was indexed. This resulted in a number of changes

0095-2338/84/1624-0152$01.50/0