United States
Environmental Protection
Agency

Council on
Environmental
Quality

November 1980

EPA
CEQ

First Report of the Interagency Toxic Substances Data Committee

ITSDC

MEMBER ORGANIZATIONS:
(*Lead Organizations)

- Arms Control & Disarmament Agency
- Consumer Product Safety Commission
- Council on Environmental Quality*
- Department of Commerce
- Department of Energy
- Department of Health & Human Services
- Department of Transportation
- Environmental Protection Agency*
- Fish & Wildlife Service
- Food & Drug Administration
- International Trade Commission
- National Cancer Institute
- National Center for Health Statistics
- National Institute of Environmental Health Sciences
- National Institute of Health
- National Institute for Occupational Safety & Health
- National Library of Medicine
- Naval Medical Research & Development Command
- Nuclear Regulatory Commission
- Occupational Safety & Health Commission
- Office of Science & Technology Policy

FIRST REPORT

OF THE

INTERAGENCY TOXIC SUBSTANCES

DATA COMMITTEE

November 1980

INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

Council on Environmental Quality
722 Jackson Place, N.W.
Washington, D.C. 20006

U.S. Environmental Protection Agency
Office of Toxic Substances
Washington, D.C. 20460

This is the first report of the Interagency Toxic Substances Data Committee (ITSDC). Under sections 10 and 25 of the Toxic Substances Control Act (TSCA) of 1976, the ITSDC is responsible for federal chemical information system development and coordination. The goal of the committee is systematized retrieval of toxicological and other scientific data which should be useful to those with responsibilities in the areas of research, risk analysis, and decisionmaking. This capability is expected to contribute to a better understanding of potential chemical hazards and ultimately to prevention of adverse chemical effects on human health and the environment. The report summarizes progress since enactment of TSCA. Although much has been accomplished, the coming years will be crucial to further development of projects initiated by the ITSDC.

GUS SPETH
Chairman
Council on Environmental Quality

DOUGLAS M. COSTLE
Administrator
U.S. Environmental Protection Agency

Preface

More than a decade ago, a panel of experts from the President's Science Advisory Committee (PSAC) identified an urgent need for a coordinated, complete, computer-based file of toxicological information which would be generally available to all those needing it in the public and private sectors. Much of the computerized capability called for by PSAC is now available at the National Library of Medicine, the Oak Ridge National Library, and other sites, but the quantity of information about toxic and hazardous chemicals and the corresponding need for it have expanded. To organize and understand the universe of existing chemicals and those not yet developed, society needs to know not only the toxicological effects of each agent but also its actual or potential exposure to humans and the environment. A coherent means is needed to gather and analyze testing data; chemical, physical, and biological properties data; monitoring data; and chemical production and use information.

Today there are over 220 federal systems containing information relevant to toxic substances. Partly because their primary purpose is to support the missions of differing agencies, the systems are dispersed, contain duplicate information, and are not coordinated for optimal use. They use dissimilar equipment and computer languages and programming, are often difficult to use, and can contain highly specialized data. Data-sharing problems are compounded because agencies generally do not know what data have already been collected by whom.

The Toxic Substances Control Act (TSCA) recognized the breadth of the information required for rational decision-making. In its directives to the Council on Environmental Quality under section 25(b) and to the Environmental Protection Agency in section 10(b), TSCA also explicitly recognized the importance of improving data management and interagency coordination.

This is the first report of the Interagency Toxic Substances Data Committee (ITSDC) on its efforts to coordinate classification, storage, and retrieval of chemical information. Since its inception in 1978, the ITSDC has focused its efforts on the construction of a Chemical Substances Information Network (CSIN). The preliminary version of the network, in use since January 1980, takes advantage of significant advances made in the information and computer sciences during the 13 years since the PSAC report, such as development of minicomputers suitable for simple onsite processing, new methods of storing vast quantities of information in a small physical space, growth in the information industry, expanded capabilities and use of time-shared networked computers, and decreasing costs per user as the number of users has increased. With such advances it has become possible to establish an effective system for retrieval of chemical data through a coordinated network of specialized information systems.

The Chemical Substances Information Network is designed to facilitate efficient and widespread use of many independent sources of chemical information. By providing analytic capabilities and interconnections between various kinds of data, it streamlines collection and integration of information from diverse sources, reducing the time and paperwork required to reach informed decisions about chemical hazards and benefits. CSIN will ultimately contribute to better decisionmaking by all sectors of society as they grapple with the known and unknown risks of chemicals.

For additional information, readers are encouraged to contact the Office of Network Administration at the Environmental Protection Agency, Office of Pesticides and Toxic Substances, Office of Toxics Integration (TS-777), 401 M Street, S.W., Washington, D.C. 20460.

CONTENTS

I. INTRODUCTION

Several years ago the following events occurred. A worker in a small Northeastern chemical plant was asked to clean a set of gears. He knew that other workers who had done so had developed a rash, perhaps caused by breathing a substance, Compound 1189, as they cleaned. He refused to perform the task and was subsequently fired. Under the section of the Occupational Safety and Health Act dealing with discriminatory firing, the worker sought recourse by contacting a compliance officer in the Occupational Safety and Health Administration (OSHA) regional office.

Because the compliance officer's first concern was that the plant might be unsafe, he invited his office's one industrial hygienist to sit in on a discussion with the worker. The hygienist's subsequent check of the office library showed that Compound 1189, a pesticide, had a very low toxicity rating; in fact, in its application dilution it was one of the safest pesticides on the market. A call to the state Department of Industrial Hygiene, where the hygienist had worked earlier, confirmed his findings. The worker seemed satisfied when the regional compliance officer told him that he would try to get his job back for him.

Later, the compliance officer learned through a newspaper article that the chemical plant had been discharging into the municipal waste water system, decimating the bacteria that are crucial to sewage treatment. The city had asked the Environmental Protection Agency (EPA) for help; EPA found that the entire river basin was contaminated. Workers blamed their work environment for tremors, eye trouble, sterility, and other disorders. Further investigation revealed high pollution levels around the plant. The state banned commercial fishing and shellfish harvesting in the area and ordered the plant closed. The National Cancer Institute had just completed a study that identified the substance feared by the complainant as carcinogenic.

The OSHA compliance officer then learned that Compound 1189 was first registered as a pesticide in 1959, when the Department of Agriculture was responsible for pesticide control. Toxicity information included with the registration showed that Compound 1189 caused "DDT-like tremors." Other information on file at EPA, which has been responsible for the pesticide program since 1971, showed that Compound 1189 disrupts growth and reproductive functions and induces tumors in experimental animals.

The director of OSHA was required to explain why a regional officer had failed to follow up on the original complaint. The officer could have taken more effective action had he known that Compound 1189 is also "Kepone" and that the information was available in EPA files. But the regional officer could not have known. In seeking data through an informal network of experts, the officer followed the usual procedure available to him and thus did not turn up the information in EPA files.

Only a few details have been changed in this account. Fortunately, since then steps have been taken to help prevent such incidents from occurring. OSHA has changed its procedures for processing complaints, industry has increased efforts to prevent such situations, and Congress has improved the institutional framework for controlling chemicals by passing the Toxic Substances Control Act (TSCA) and the Resource Conservation and Recovery Act (RCRA). Nevertheless, the success of such preventive efforts will require altering the logistics of finding and obtaining information when it is needed.

The purpose of this document is to report the progress that has been made since 1976 toward developing a comprehensive and efficient system to provide data and information relevant to chemicals. Beginning in February 1981, a Chemical Substances Information Network (CSIN) will offer a powerful computerized tool for scanning government and commercial files for the existence of all types of data relating to chemical substances--nomenclature, composition, structure, properties, toxicity, effects, production, use, regulation, and other aspects.

CSIN is a user-oriented network designed to make data on toxic substances rapidly and easily accessible. The system will ultimately serve not only government agencies with regulatory responsibilities, but also scientists, the industrial community, academia, public interest groups, and others concerned with chemical substances. A telecommunications network will link CSIN users to the information stored in the system's independent component resources (computer-stored sets of information made accessible through the network). Two major directories will help users find the information they need. A variety of other CSIN components will meet specialized user needs.

Development of the Chemical Substances Information Network has been scheduled to progress in three stages: 1) preprototype, 2) version I prototype, and 3) version II prototype. In January 1980, the preprototype CSIN was implemented on an intelligent terminal (a computer terminal with limited storage and programming capabilities) to serve the information needs of the TSCA program at EPA. Under TSCA, the Environmental Protection Agency must evaluate notices of intent to manufacture new chemicals within 90 days of their receipt. The National Library of Medicine of the National Institutes of Health also uses the preprototype version of CSIN.

The preprototype assists the user in executing multi-system searches (searches initiated in one system in order to find information contained in more than one system) and can be used to search for information contained in five different information resources. Use of the preprototype terminal has helped to clarify the kinds of chemical information needed and to provide helpful experience with establishing connections to automated components.

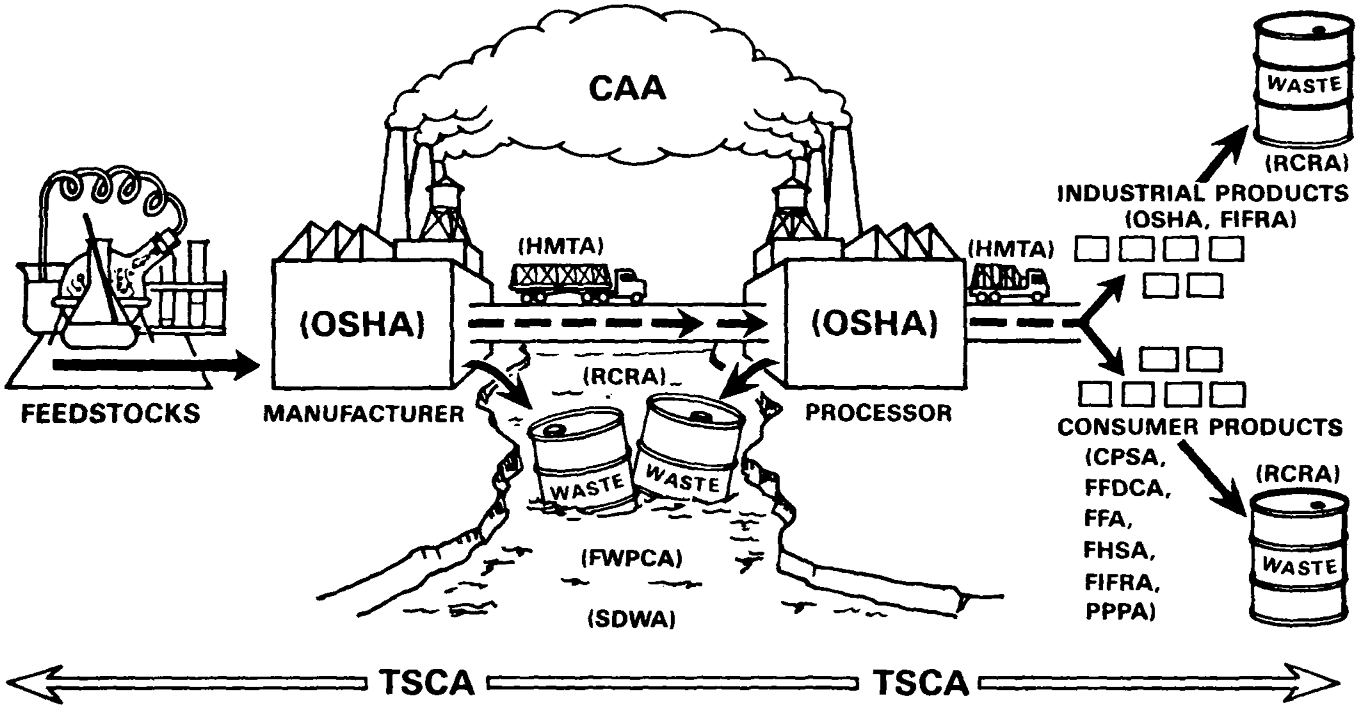
The version I prototype is expected in February 1981, and the version II prototype approximately one year later. Each stage in the development of CSIN will achieve greater integration of databases (sets of information stored in computers or in manual files) and a more powerful overall system.

CSIN is a project of the Interagency Toxic Substances Data Committee (ITSDC), which was formed in February 1978 as a permanent body for managing interagency cooperation in the use of data relating to chemicals. Under at least a dozen major federal laws, agencies collect data about almost every conceivable aspect of chemicals, including their production and use (see Figure 1). More than 220 federal data systems in a variety of agencies contain such information. Most of these systems have been designed to meet the needs of their home agency, and provide, for example, information on occupational exposures or on the transportation of hazardous substances. Many systems contain varied, often highly specialized, and sometimes duplicative information. Some of the content is confidential business information, which complicates its use. Moreover, the information requirements posed by chemical regulation, toxicological research, chemical manufacture, and other related activities have grown steadily and dramatically.

This first annual report of the Interagency Toxic Substances Data Committee begins with a description of the organization of the committee. It continues with a discussion of the committee's accomplishments since 1978, focusing on the development of CSIN and its major components. The report concludes with a section on the role of chemical information as the tools for making use of such data continue to grow in number and sophistication.

FIGURE 1

LEGISLATIVE AUTHORITIES AFFECTING THE LIFE CYCLE OF A CHEMICAL*



• KEY •

CAA	= CLEAN AIR ACT	HMTA	= HAZARDOUS MATERIALS TRANSPORTATION ACT
CPSA	= CONSUMER PRODUCT SAFETY ACT	OSHA	= OCCUPATIONAL SAFETY & HEALTH ACT
FFDCA	= FED. FOOD, DRUG, & COSMETIC ACT	PPPA	= POISON PREVENTION PACKAGING ACT
FFA	= FLAMMABLE FABRICS ACT	RCRA	= RESOURCE CONSERVATION & RECOVERY ACT
FHSA	= FED. HAZARDOUS SUBSTANCES ACT	SDWA	= SAFE DRINKING WATER ACT
FIFRA	= FED. INSECTICIDE, FUNGICIDE, & RODENTICIDE ACT	TSCA	= TOXIC SUBSTANCES CONTROL ACT
FWPCA	= FED. WATER POLLUTION CONTROL ACT		

* from EPA Journal, July/August 1979

II. THE INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

Congress included in the Toxic Substances Control Act two provisions that directly address the problem of information identification and management. Section 25(b) directs the Council on Environmental Quality (CEQ), in consultation with other departments and agencies,

> to coordinate a study of the feasibility of establishing a standard classification system for chemical substances and related substances and . . . a standard means for storing and for obtaining rapid access to information respecting such substances

and to report to Congress on the results. Section 10(b) directs the Environmental Protection Agency to establish and be responsible for an interagency committee

> to design, establish and coordinate an efficient and effective system . . . for the collection, dissemination to other Federal departments and agencies and use of information submitted under [the] Act,

and of a system for the retrieval of toxicological and other scientific data which could be useful for carrying out the purpose of the law.

To fulfill their responsibilities under Sections 10 and 25 of TSCA, EPA and CEQ formed the ITSDC in February 1978. The ITSDC has a permanent charter, reports to the Administrator of EPA and the Chairman of CEQ, and is co-chaired by the two agencies. Membership includes the Departments of Agriculture, Commerce, Defense, Energy, Interior, Labor, State, and Transportation; the Department of Health and Human Services and several of its component agencies; the Consumer Product Safety Commission; and the National Science Foundation. In keeping with its intent to maximize coordination and usefulness of federal information resources, the committee holds public, bimonthly meetings of which detailed minutes are kept.

The committee is currently guiding the development of the Chemical Substances Information Network on the basis of a survey of existing chemical information systems prepared for CEQ by the Mitre Corporation. The ITSDC has also undertaken to coordinate federal information gathering activities and the use of information outside the federal government, and to find methods of transferring trade secrets and other confidential information among agencies without breaching secrecy. Accordingly, the committee has organized three subcommittees.

The Chemical Substances Information Network Subcommittee was formed in April 1978 to conduct studies to develop information systems to capture, store, and allow rapid access to relevant data and information. Chaired by the National Library of Medicine, this group steers the design, policies, implementation, management, and uses of the developing CSIN. In February 1979, the committee appointed a Network Administrator to oversee the day-to-day activities of CSIN development. The Environmental Protection Agency is currently responsible for design, development, coordination, and the administration of the network. The CSIN Subcommittee has also enlisted the help of a Technical Review Panel composed of experts from Lawrence Berkeley Laboratories, the University of Florida, and the National Bureau of Standards. These experts meet semi-annually to discuss their evaluation of CSIN design and implementation with the subcommittee. The subcommittee has made considerable headway in the design and construction of the Information Network. Section III of this report will describe these developmental efforts in detail.

A Trade Secrets Subcommittee has been appointed to handle issues relative to the use of confidential data. Both statutory provisions concerning data confidentiality and practices of federal agencies currently act as barriers to the sharing of confidential data, even on a need-to-know basis, between federal agencies and among the states. These obstructions have led to duplication of effort and inefficient data collection. The public, in order to participate in decision making and in independent evaluation of data, also requires access on a good-cause basis to confidential data. The Trade Secrets Subcommittee, chaired by CEQ, is currently working on legislative proposals which address each of these issues as recommended in May 1980 by the Toxic Substances Strategy Committee.

The Public Liaison Subcommittee was established to provide a link to state and local government, academic institutions, industry, environmental and consumer groups, international organizations, and other interested parties concerned about federal actions on chemical substances. It also serves these non-federal groups as an avenue for suggestions and reactions.

III. THE CHEMICAL SUBSTANCES INFORMATION NETWORK

The CSIN Concept

If people and the environment are to be protected from toxic substances, many things must be learned about such chemicals. Knowing the chemistry of a compound is a first step in determining what dangers it may pose. To comprehend or control its life-cycle it is necessary to investigate the substance's movement in commerce and through society. A variety of questions may arise:

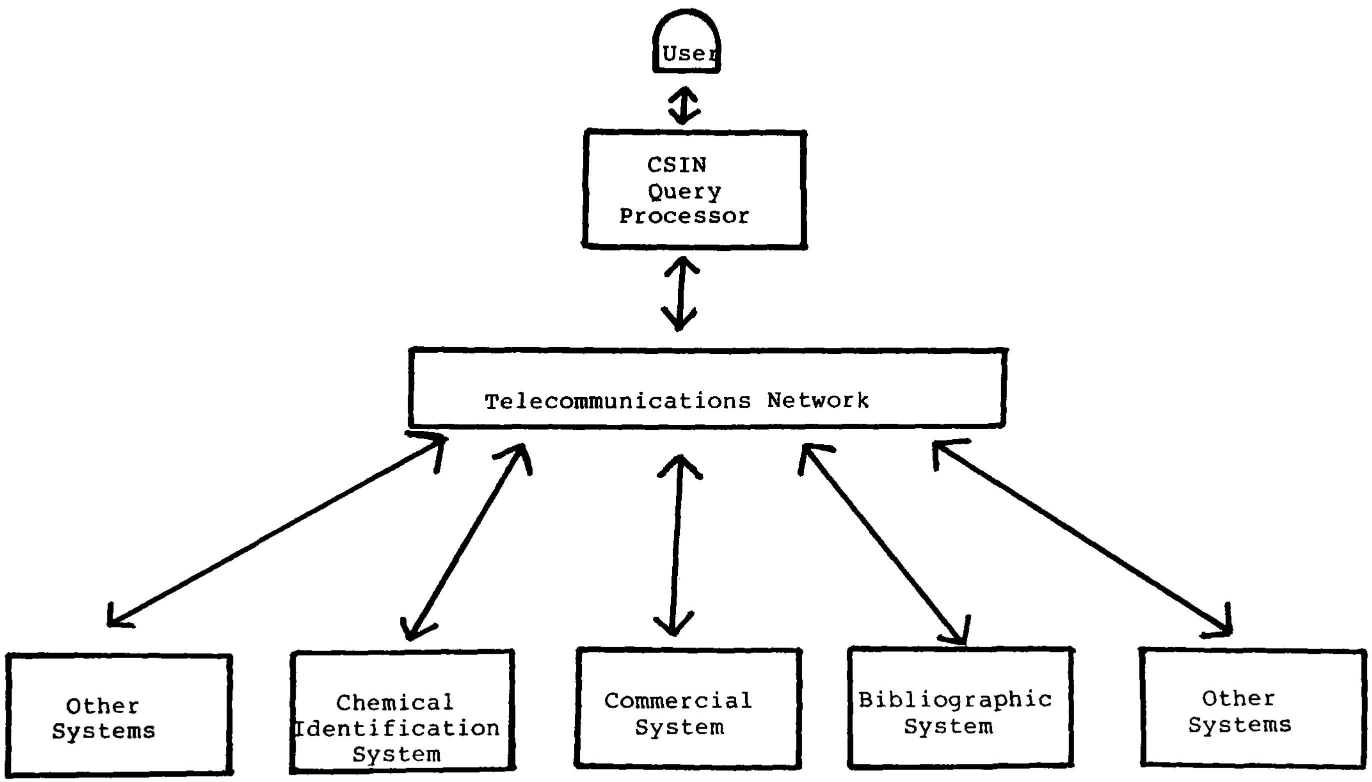
- Where is a chemical made? How is it transported and used?
- What characterizes the site of its release and the forces that act upon it--i.e., is it released to the water, land, or air?
- What effect will the action of the sun, the wind, or of other chemicals have upon its fate?
- Who or what is exposed to it?
- How will it influence biological systems upon which it impinges?
- Is the chemical toxic?
- Is its use regulated?

Individual databases usually provide only a portion of the data required to answer these questions. Users must traditionally locate and then examine many databases, one by one, in order to ensure thorough treatment of a subject. Some of the databases are automated, but others must be manipulated manually. Only a frequent user of a particular information system is apt to be familiar with its idiosyncracies. CSIN, however, coordinates many disparate and independent information systems into a network, providing access to more capabilities and data than any single system could provide.

The hub of CSIN is a query processor (management computer) that links the user to the many information systems where relevant data are available (see Figure 2). It connects independently owned and operated systems, and

FIGURE 2

QUERY PROCESSOR (MANAGEMENT COMPUTER) OPERATION



provides for their virtually simultaneous use. The processor is essential because even an apparently simple query (information search) may require interaction with a variety of sources. For example, to investigate the threat posed to human health by a given chemical, it typically would be necessary to:

a. search a chemical identification system to learn the chemical's composition, molecular structure, full name, and possible synonyms or trade names;

b. search a bibliographic system for reference to and abstracts of published material on the chemical;

c. search a management tracking system for information on government regulatory activity related to the chemical;

d. search a data management and analysis system for data on laboratory experiments that have been performed using the chemical;

e. search a management information system for data on production of the chemical.

Without the use of CSIN, access to each of these information systems would be possible only in a way unique to the specific system, requiring the user to master the "language" needed to communicate with each system. As a rule, the process would be repeated for each chemical about which information was desired, even when the query concerned related chemicals. The repetition of common operations for different chemicals and for a variety of systems makes the gathering of information a time consuming, expensive process.

CSIN streamlines and coordinates existing procedures and thus increases productivity. For example, the pre-prototype version of CSIN was used to gather information on the effects induced by 1,150 chemicals on the genetic composition and functioning of biological systems. Using traditional methods of information gathering, the search of four databases was expected to take 15 weeks. Much of this time would have been spent re-entering the data retrieved from one system into another system, and then waiting for output (computer produced responses). Such work is error-prone as well as time consuming.

Instead, CSIN enabled those responsible for the task to retrieve information from databases, combine the data with other information, transform data into the correct formats for the different databases, and submit transformed data in subsequent queries. Because there is a charge for use of the Information Network, the searches were conducted during less expensive off-hours.

The government users needed six days to enter the necessary information. Within two weeks, the results of the searches for all 1,150 chemicals were gathered. Personnel credit CSIN with saving them three months of work.

CSIN Components

CSIN will be only as useful as the information in the component systems. The concept for a Chemical Substances Information Network, including its potential components, was first described in a June 1977 MITRE Corporation report to CEQ, EPA, and the Department of Health, Education and Welfare (now HHS). The report's recommendations were based on a survey of potential users of chemical information, including federal, industrial, academic, and consumer action groups. Some of these potential users require only very general information on a large number of chemicals to aid in policy planning. Others may need in-depth information on only a few substances to identify substitute chemicals for a particular industrial application, or to study the relationship between one feature of chemical structure and a particular biological effect. The report suggested that CSIN should encompass a collection of databases for each of the major categories of user needs. The "core" or primary components, would include:

- Data gathered under the authority of TSCA, in a confidential file for EPA use.

- A portion of the TSCA file containing only non-confidential data available for public use.

- A directory of detailed information (Chemical Information Resources Directory) about each component file, which will identify those files meeting the users' specific needs.

- A chemical structure and nomenclature catalog (Chemical Structure and Nomenclature System) with identifying labels for an estimated one million chemicals, including a cross-reference to other files containing information about each substance.

- A file of data and information on the biological effects of chemicals, gathered from research and toxic testing conducted by government, industry, and academia.

- A facility for the collection, monitoring, evaluation, and reporting of unevaluated test data.
- A bibliographic system for articles from toxicological and biomedical journals.
- Data about laboratory animal strains for use in designing test systems.
- Information about government actions concerning chemicals or classes of chemicals.

Additional study confirmed that the components listed above will satisfy initial user needs, and efforts are well underway to incorporate such systems into CSIN at the earliest possible date. Some of these components are already marketed commercially, some are maintained by federal agencies, and others are under development within agencies or within groups of agencies working cooperatively. In addition, the Chemical Structure and Nomenclature System (CSNS) and the Chemical Information Resources Directory (CIRD) are being developed under the aegis of the ITSDC. They are described in more detail in Section V; those components not directly administered by the ITSDC are discussed in Section VI.

IV. CSIN DEVELOPMENT: PROGRESS TO DATE

A Chemical Substances Information Network development contract for phased implementation was awarded in September 1977. Since then, the following aspects of the CSIN project have been completed:

- Comprehensive study of user requirements.
- Sketch for design of CSIN architecture.
- Installation of a preprototype terminal, functioning as the preprototype CSIN query processor since January 1980.
- Detailed design of a prototype CSIN query processor, expected to be in use as of February 1981 by organizations representing federal and state government agencies, industry, universities, and public interest groups.

Analysis of User Requirements

Implementation of the Chemical Substances Information Network demanded a more detailed understanding of the needs of potential users. The development contractor, the Computer Corporation of America (CCA), completed an analysis in June 1980 which not only detailed the required functions of the CSIN query processor, but also described potential uses of network components. The results provided the ITSDC with a solid conceptual framework and information base for guiding further network development.

The CCA study included:

- 140 interviews;
- An analysis of more than 100 information products (research papers and management reports);
- Examination of computerized databases and numerous non-automated information sources in regular use at the surveyed organizations.

The surveyed organizations have a broad range of responsibilities (see Table 1). In addition to six departments of the Executive Branch and several independent federal agencies and commissions, the study included organizations representing the chemical industry and public interest groups. The breadth of the sample group and the strength

TABLE I

ORGANIZATIONAL MISSIONS

ORGANIZATION	ORGANIZATION TYPE	RELATIONSHIP TO OTHER ORGANIZATIONS	MISSION WITH RESPECT TO CHEMICAL SUBSTANCES
Center for Disease Control (CDC)	Service Research Regulatory	Advises other component agencies of the Department of Health and Human Services; advises state, local, and international authorities concerning clinical illness of an infectious nature. Coordinates Clinical Lab Improvement Program with DoD. Indirectly affiliated with NTP.	Investigate and assist in the control of endemic and epidemic disease on a national and international level.
Conservation Foundation (CF)	Public advocate Research	Advises EPA, FWS.	Improve the quality of the environement and promote the wise use of the earth's resources.
Consumer Product Safety Commission (CPSC)	Research Investigatory Regulatory	Advises NTP	Protect the public against unreasonable risks from consumer products.
Chemical Industry Institute of Toxicology (CIIT)	Industry advocate Research	Advises EPA, FDA, NIEHS, and NCI.	Conduct scientific, objective studies of toxicological problems involved in the manufacture, handling, use and disposal of commodity chemicals.

TABLE I (CONTINUED)

Department of Defense (DoD)	Military	Contributes funding and support to ORNL.	Investigate military applications of chemicals and protect military personnel and environments from chemical hazards.
Environmental Protection Agency - Office of Research and Development (EPA-ORD)	Research	Performs testing and evaluation for other EPA offices. Advises NTP.	Perform monitoring and research in support of other EPA divisions.
Environmental Protection Agency - Office of Toxic Substances (EPA-OTS)	Research Regulatory	Refers cases to other offices of EPA and other agencies for investigation and regulation. Coordinates EPA actions concerning toxic substances.	Protect human health and the environment through the control of certain chemical substances.
Federation of the American Societies of Experimental Biology (FASEB)	Coordinating Research Contracting	Contracts to DoD, FDA.	Represent and support the six professional biological societies that constitute its membership.
Fish and Wildlife Service (FWS)	Research Investigation Regulation	Approves permits and Environmental Impact Statements for EPA. Assesses fish and wildlife impacts of actions by DoD, NASA, DoE, and the Department of Justice. Advises state, local, and international authorities.	Conserve and manage fish and wildlife sources and their habitats.

TABLE I (CONTINUED)

Food and Drug Administration - Bureau of Drugs (FDA-BD)	Research Regulatory	Advises Public Health Service agencies. Indirectly affiliated with NTP.	Ensure that drugs, biological products, therapeutic devices, and diagnostic products are safe, effective and properly labelled.
Food and Drug Administration - Bureau of Food (FDA-BF)	Research Regulatory	Advises Public Health Service agencies. Indirectly affiliated with NTP.	Ensure that food is wholesome and safe, that cosmetics are safe, and that both are honestly labelled.
National Bureau of Standards (NBS)	Research Regulatory	Advises EPA and OSHA.	Provide a national laboratory for physical standards.
National Cancer Institute (NCI)	Research	Member of NTP.	Develop the means for reducing the incidence, morbidity, and mortality of cancer in humans.
National Center for Toxicological Research (NCTR)	Research	Member of NTP. Advises EPA, CPSC, and OSHA through NTP board affiliations.	Develop a better understanding of the adverse health effects of potentially toxic chemicals on living organisms.
National Institute for Occupational Safety and Health (NIOSH)	Research Investigatory	Advises OSHA on occupational standards. Advises Department of Interior on mining health standards.	Reduce morbidity and mortality due to occupational toxicological hazards. Conduct research to assist OSHA in developing standards to protect American workers.

TABLE I (CONTINUED)

National Institute of Environmental Health Sciences (NIEHS)	Research	Advises EPA. Member of NTP. NIEHS and NTP share a Director. Administrates Environmental Teratology Info Center, an ORNL database.	Conduct and support biomedical research that will identify, characterize and prevent adverse effects of environmental agents on human health.
National Toxicology Program (NTP)	Coordinating	NIH, NIEHS	Develop scientific information that can be used to protect the health of the American public from damage by exposure to environmental chemicals.
Oak Ridge National Laboratory (ORNL)	Service Research	Contracts to DoE. Provides resources for EPA, NIEHS, and NCI.	Perform basic and applied research on energy-related topics.
Occupational Safety and Health Administration (OSHA)	Regulatory	Receives input on occupational standards from NIOSH, EPA, FDA, NIEHS, NCI, and CPSC. Advises NTP.	Make the workplace a safe environment for employees.
Stanford Research Institute (SRI)	Research Consulting	Contracts to NIOSH, NCI, FASEB, and ORNL	Perform basic and applied research, with a strong emphasis on problem solving, under contract to clients in business, industry, and government in the U.S. and abroad.

of key conclusions support expectations that the study results will apply to the total community of prospective CSIN users.

The study confirmed that more effective integration of and access to already available chemical information is the most pressing need. Although the community of prospective users is diverse, the study found that a relatively small number of operations (activities performed by a computer to answer information requests) will support most organizational missions. These activities, which the study called transactions, are executed several hundred times every day. The following hypothetical example illustrates the concept of a transaction.

A woman who had worked in the carding room at a cotton mill for 25 years developed persistent respiratory trouble. Her doctor noticed shadows on her most recent chest x-rays, and suggested that the respiratory problem might be occupationally related. To find out if her doctor's surmise was correct, she called a regional office of the National Institute for Occupational Safety and Health (NIOSH).

By engaging in activities called processes, NIOSH researchers tried to determine what, if any, harmful substances the woman might have been exposed to at work. They explored the effects of any such exposure and related them to those that she had experienced.

One of the processes that they used was called "Identification and Quantification of Effects." The goal of the "Effects" process was to determine whether exposure to cotton dust or other substances in the work environment might have produced the described adverse effects. This process can be supported by a number of information gathering tasks, called transactions. The ten transactions that were used in support of the "Effects" process are shown in Figure 3. The two most critical transactions, "Retrieval of Chemicals Causing Specific Effects" and "Retrieval of Biological Effects," involved direct searching for links between substances and their effects.

In this search, the NIOSH officials posed the following question: What substances have been associated with chronic respiratory illness which is neither bacterial nor viral in nature? To find an answer, they performed a transaction--"Retrieval of Chemicals Causing Specific Effects." As part of this transaction, the computer provided the list of databases and bibliographic sources listed in Figure 4. The subsequent automated search of these sources resulted in a long list of references to written material concerning the causes of various types of chronic

FIGURE 3

TEN TRANSACTIONS SUPPORTING THE "IDENTIFICATION AND QUANTIFICATION OF EFFECTS" PROCESS

1. Retrieval of Chemical and Physical Properties
2. Derivation of Chemical and Physical Properties
3. Retrieval of Chemical Relationships
4. Derivation of Chemical Relationships
5. Retrieval of Effects (Biological)
6. Retrieval of Effects (Physical Environment)
7. Retrieval of Chemicals Causing Specific Effects
8. Retrieval of Methods to Be Applied
9. Retrieval of Transport (Biological) Information
10. Retrieval of Transport (Physical) Information

FIGURE 4

SOURCES OF INFORMATION ON CHEMICALS CAUSING SPECIFIC EFFECTS

Sources of Bibliographic References

APTIC FILE - Air Pollution Technical Information Center File

BIOSIS Previews - Biosciences Information Service Previews

CANCERLIT - Cancer Literature

CHEMABSTRACTS - Chemical Abstracts

EXCERPTA MEDICA

MEDLINE - Medical Literature and Retrieval Service Online

NIOSHTIC - National Institute of Occupational Safety and Health Technical Information Center Database

POLLUTION ABSTRACTS

SCISEARCH - Science Citation Index

TITUS - Textile Information Treatment User's Service

TOXLINE - Toxicology Information Online

WTA - World Textile Abstracts

Sources of Data (Not Specifically Limited to Biographic Citations)

CICIS - Chemicals in Commerce Information System

HEEDA - Health Effects and Environmental Data Analysis System

TDB - Toxicology Data Bank

respiratory problems, including smoke, dust (coal and cotton), pollens, asbestos, and fiberglass.

Another approach was to search for the potential effects of substances found in the cotton industry working environment. From a study of the woman's work environment and from trace compounds discovered in her medical examination, a list of substances to which she had probably been exposed at work was compiled. A search for the chronic biological effects of those substances was then made. Accordingly, the next transaction to support the "Effects" process was called "Retrieval of Biological Effects."

The researchers posed this question: What chronic respiratory effects have been identified with cotton, cotton dust, smoke, and pollens? To find an answer, the computer searched the databases and bibliographic reference sources shown in Figure 5. The result was a list of several hundred references concerning the chronic respiratory effects of the substances mentioned. Among the effects were black lung, emphysema, asthma, and byssinosis.

An analysis of the results of the two transactions indicated a possible link between the woman's disease, byssinosis, and the cotton industry work environment. The information was uncovered through a combination of extensive analysis by professionals and performance of relatively straightforward information gathering activities--called transactions--that are subject to automation. Each transaction used data obtained from two or more sources. The data was then automatically collected, organized, and integrated into a single, unified output.

Transactions are complex operations, subject to considerable variation from one instance to the next. The study of user requirements suggested that after mapping the transactions of prospective users, a set of specific databases that contain required information could be selected. The following related conclusions were drawn from the study; some are discussed later in greater detail:

- There is a heavy and accelerating demand for chemical information in the surveyed community.
- Because the described transactions are needed to satisfy central goals and support the regular business of these organizations, a chemical information network should be available to users on a frequent and regular basis.
- The surveyed user community can be expected to execute the 20 studied transactions at least tens of thousands of times per year.

FIGURE 5

SOURCES OF INFORMATION ON

BIOLOGICAL EFFECTS OF CHEMICALS

Sources of Bibliographic References

CHEMABSTRACTS - Chemical Abstracts

EXCERPTA MEDICA

MEDLINE - Medical Literature and Retrieval Service Online

NIOSHTIC - National Institute for Occupational Safety and Health Technical Information Center Database

POLLUTION ABSTRACTS

SCISEARCH - Science Citation Index

TOXLINE - Toxicology Information Online

WTA - World Textile Abstracts

Sources of Data (Not Specifically Limited to Biographic Citations)

CICIS - Chemicals in Commerce Information System

HEEDA - Health Effects and Environmental Data Analysis System

TDB - Toxicology Data Bank

- Transactions originate at many different locations across the country, and the data required by the transactions are dispersed throughout the United States.

- It is essential that the design of CSIN provide for efficient control of confidential data so that they are not accessible to unauthorized personnel.

- The quality of information obtained is of great importance to surveyed users. Such information should meet the criteria of timeliness, consistency, completeness, and verifiability.

- A chemical information network must be able to support a sequence of functions. It should be able to select the relevant databases and language required for gaining access to them, translate a user's question into database terminology, gather information from many resources, and assemble the information into a response stated in the user's terminology, using an optimal search strategy throughout.

The CSIN Design

ITSDC efforts to date have centered on the most innovative aspect of CSIN--the design of the query processor. The most basic function of the CSIN query processor or management computer is to serve as a communications facility capable of connecting users to the component information systems. The processor also helps integrate the resources of the individual components. By submitting a single request for information, a user can gain access to data contained in more than one of the component systems.

CSIN will also facilitate information searching in the three ways discussed below.

I. User-defined queries are helpful in performance of repetitive tasks. A preliminary search of bibliographic systems for literature references is one of the first steps in the investigation of most substances. This is a mechanical procedure that might proceed as follows:

1. Obtain a chemical identifier, such as the formal chemical name, for the substance;

2. Query a prespecified set of bibliographic systems using the chemical identifier as well as other names as search terms (keywords);

3. Collect and print the references.

This type of request is routine, but time consuming if conducted by the user individually interacting with each of several bibliographic systems. The operation requires access to chemical identification system for step #1 and several bibliographic resources for step #2. To retrieve the information, the user must gain access to each component, invoke the desired subsystem (portion of a component), and issue the correct commands (words used to obtain information from the computer). All of these operations take time. Moreover, because the syntax or search strategy required varies from system to system, the user must know a variety of command languages, error correction procedures, and other system-dependent details. These problems can be eliminated through use of the "query definition" facility, which enables the user to select a prepackaged sequence of transactions--a so-called SCRIPT--which can be invoked with a single command. CSIN will then execute the entire set of transactions automatically and deliver the results to the user. Repetitive searches can thus be performed with far greater speed and accuracy than is possible using a manual procedure.

II. The query facility reduces hurdles which normally might prohibit multi-system searching. It will aid users who wish to browse through several systems as well as those with ad hoc queries. Ad hoc queries are less predictable in pattern than the repetitive operations embodied in other query procedures. Though the actual question may be simple, ad hoc queries often range over more than one component system, each with its own procedures, terms, and syntax. When either browsing or making an ad hoc query, users need to decide as they search which component systems they would like to examine next. The query facility gives them that flexibility.

However, only selected data will be available through the query facility. At times a user may need to gain direct access to component systems. In this case, the user will need the third CSIN capability.

III. Certain special functions will be available through direct interaction with CSIN components. For instance, a user might wish to know which chemicals are structurally related to a particular substance. The resulting interaction with a chemical substructure search system could require graphical display (pictorial depiction of information at a computer terminal) that might best be implemented at the level of the component system. CSIN will ease this kind of search in two ways. First, it will provide facilities to streamline direct user communication with a specialized component system. Second, CSIN will make it possible to use data retrieved from the specialized component in a subsequent query involving other systems.

V. THE CSIN DIRECTORIES

As we have seen, the first step in retrieving information is locating potential sources. Two CSIN components, the Chemical Information Resources Directory and the Chemical Structure and Nomenclature System, will assist in this process. These key systems are being developed under the aegis of the ITSDC. Together they will provide a comprehensive and coordinated set of indexes to the information available throughout the network.

Chemical Information Resources Directory (CIRD)

The CIRD describes to the user what types of data are available in the different network components. To enable users to identify the databases relevant to a specified subject, the CIRD will use subject and descriptive catalogs, designed to function together.

The subject catalog provides a means of selecting information resources appropriate to a subject area. The descriptive catalog contains information pertinent to the use of a particular database, including its availability, sponsor, scope of information covered, and other characteristics. The design requirements for the CIRD were completed in August 1979. Eventually, the directory will be an interactive component of CSIN. Further development of this feature of CSIN is underway.

Use of CIRD has further ramifications. Consider the case of a glass manufacturer who, under the provisions of TSCA, must substantiate the safety of the cadmium salt used to color glass yellow. CIRD will help him efficiently uncover the needed data, first by identifying the relevant component data bases. In this case there would be two: TOXLINE (Toxicology Information Online) which contains references to more than one million scientific journal articles; and a commercial system, which contains data on recent production and uses of cadmium salts.

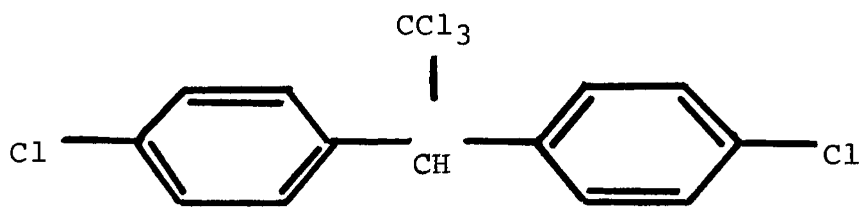
An article referenced in TOXLINE, entitled "Comparative Toxicity of Cadmium Compounds," could indicate that cadmium salt might be toxic. Turning to the commercial system, the glass manufacturer would find that the cadmium salt has been used as a photo conductor in the solar cell industry. By contacting several of the 21 solar cell manufacturers cited in the commercial system, the glass manufacturer possibly can determine what research has already been conducted on cadmium salts. Without CSIN, the glass manufacturer might not have known that extensive analyses on the cadmium compounds of interest had already been performed under the auspices of manufacturers in an industry unrelated to his own.

Chemical Structure and Nomenclature System

As it processes each request for information, CSIN must be able to definitively identify a unique chemical substance and unambiguously communicate that identify to other systems in the network. The Chemical Structure and Nomenclature System (CSNS) will help meet this need. The CSNS will enable an information network user to determine whether a particular substance or a set of structurally related chemicals of interest exists in the network, and, if so, which component may contain the desired information.

For CSNS to work, each component file must denote the identity of individual chemical substances by at least one common convention. Only then can diverse information systems identify chemicals and exchange information about them. The ITSDC will use the Chemical Abstracts Registry Service (CAS) number of the American Chemical Society as the standard chemical identifier in CSIN. Because the CAS number is critical to the information network, its origin and use are explained here.

Chemical substances can be identified clearly by a structural diagram; for example, the structural diagram for a molecule of the predominant compound of DDT (p,p'-DDT) is:



DDT

However, although a computer system can print such diagrams, in most files that contain chemical information it is awkward if not impossible to carry structure diagrams for identification purposes. Instead, most files identify chemical substances by name. Names vary in accuracy from formal chemical nomenclature, which can properly represent the structure, to generic names, codes, or numbers which have little or no information content. If the same substance is identified differently in various files, ambiguities and mistakes may occur in attempting to transfer or link information concerning chemicals from one file to another.

The solution to this problem is to assign a unique number to each chemical structure, and to use at least that "social security number" to identify the same chemical each time information about it is stored in a file. A computerized numbering system for chemicals is presently maintained by CAS, which has assigned numbers to more than five million substances since it began registering chemicals in 1965. In its Report to Congress, required by section 25 of TSCA, CEQ recommended that "Government agencies should obtain numbers from this 'dictionary' for all chemicals in their files, so it will be feasible and possible to retrieve accurate information quickly from diverse sources." The ITSDC endorses this recommendation and is working with its member agencies and with CAS to implement this resolve.

Another important feature of CSNS is its ability to identify chemicals by their structural or substructural features. This will allow the users to search for chemicals which have common structural characteristics. It will facilitate grouping of chemicals for structure activity correlation (comparison of chemical activities based on structure) and toxicity profiling (ranking of chemicals according to the severity and type of threat they pose to humans and the environment).

CSNS will also allow users to search for chemicals by name or name fragments. This capability will help users who are unable to specify compounds by their structures or who find it easier to refer to chemicals by name. CSNS will make it possible to search for materials, such as tars, with ill-defined or variable structures.

CSNS may build upon several existing systems that perform a substance identification function: the Chemical Information Services Index (CIS), CHEMLINE (Chemical Dictionary Online) of NLM, the National Cancer Institute's Inquire information system, and the Structure Search System of CAS. Each of these has some of the capabilities and some of the data needed for the Chemical Structure and Nomenclature System. The Committee expects to fund a contract for CSNS development in fiscal year 1981, so that the CSNS will be operational by mid-1982.

VI. OTHER CSIN COMPONENTS

This section discusses the existing component systems already included in the preprototype version of CSIN. They are all administered by organizations other than the ITSDC. Additional systems will be added to the information network as the need arises, but it will include at least:

-- Bibliographic Systems

-- Chemicals in Commerce Information System

-- Chemical Information System

-- Laboratory Animal Data Bank

-- Toxicology Data Management System

Bibliographic Systems. Many CSIN users will need to find existing literature on chemical and biomedical subjects. The following publicly available bibliographic systems are current or prospective components of CSIN:

-- Bibliographic Retrieval Services

-- Lockheed Corporation

-- MEDLARS of the National Library of Medicine

-- System Development Corporation

Chemicals in Commerce Information System. Of the 55,000 chemicals now in commercial use, there are many about which little is known. The Toxic Substances Control Act requires manufacturers, processors, exporters, and importers to provide information on new chemical substances proposed for production and on selected existing chemicals. CICIS (Chemicals in Commerce Information System) will be the repository for almost all the information submitted to EPA under TSCA and some that is generated by EPA in the TSCA program. The system will provide access to information concerning production, distribution, uses, industry studies, and process information.

Among the subsystems of the completed CICIS will be:

-- Chemical Information Database: (contains data in the current EPA inventory and additional data from submissions)

-- Chemical Inventory System: (contains industry-reported data on chemical substances currently manufactured, processed, or imported into the U.S. and data on the manufacturers, processors, and importers)

-- Health and Safety Subsystem: (used for tracking health and safety studies submitted to EPA; some actual data from the studies may also be stored in this data base)

-- Organization/Site Databases: (contains data on industry submitters of data)

-- Premanufacture Submission Module: (used for display of selected portions of data submitted in notices of intent to manufacture a new chemical)

-- Test Results Subsystem: (stores test result data received under TSCA and from other sources; results will be linked to chemicals, documents, and other data in the database)

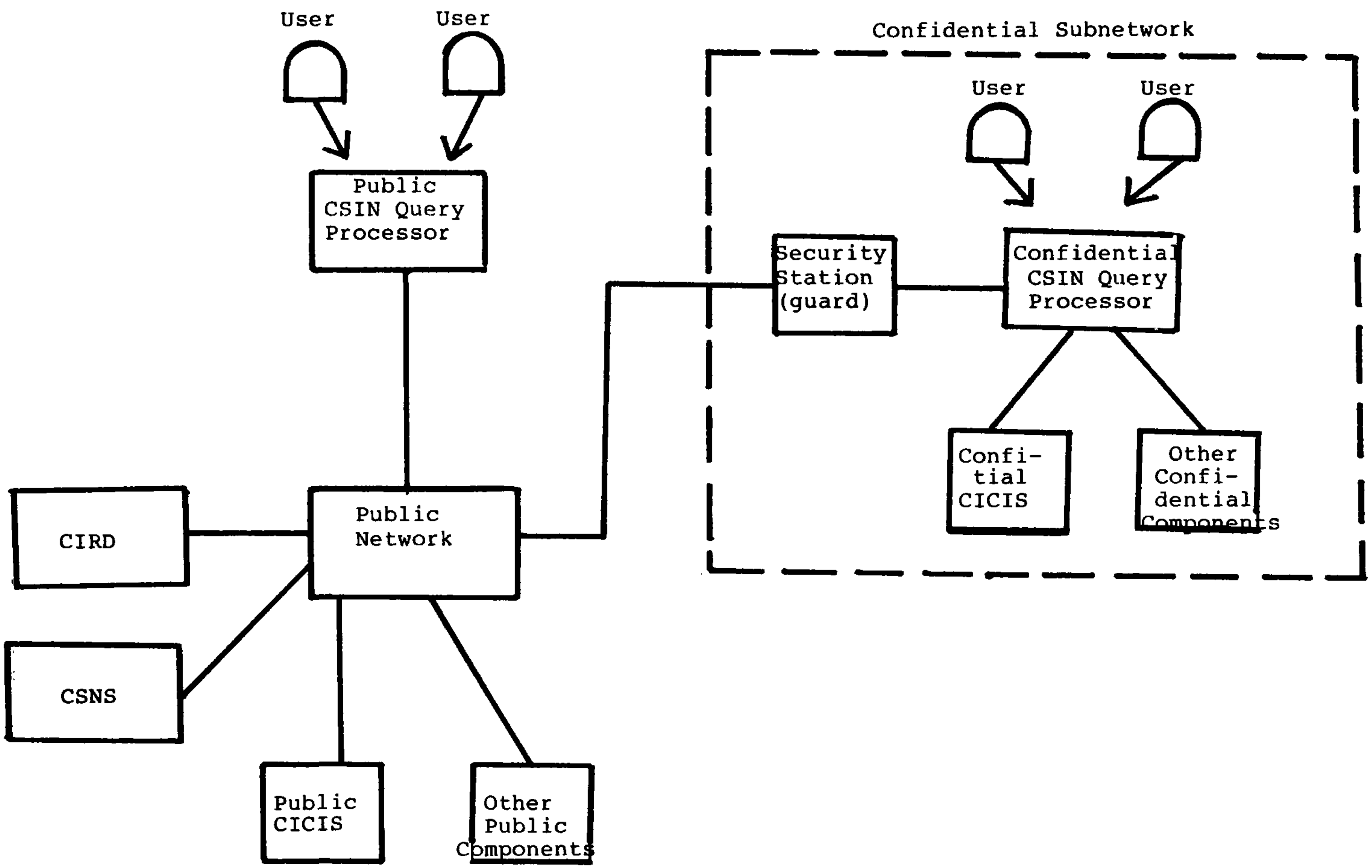
CICIS introduces special complexities into the design of CSIN. Because it will contain confidential business data, such as production and marketing information, it requires a secure operating environment. On the other hand, the nonconfidential TSCA data will be sought by CSIN users, and must be publicly available according to the provisions of TSCA. To ensure availability of some data and protection of others, two CICIS systems will be constructed and two CSIN processors: one confidential, the other public (see Figure 6). Access to the confidential system will be restricted to authorized users. The confidential users will reside in a secure installation and employ data encryption (coding) and other measures to protect the confidential information. Users of the confidential system will also have access to the rest of the network.

Chemical Information System. The Chemical Information System (CIS) is a repository of physical and chemical information on chemical substances. The system can be used to retrieve data on the known properties of substances and to estimate properties for which determinations have not yet been made.

Laboratory Animal Data Bank. The Laboratory Animal Data Bank (LADB) is a comprehensive source of information on laboratory control animals. LADB maintains normal values (e.g., average body weight at two months) and pathology data on species, strains, and colonies of animals. These data can be helpful in developing new tests as well as in evaluating previous test results, and in the selection of animal species for future tests. LADB is already operative and expanding.

FIGURE 6: CSIN CONFIGURATION

PROPOSED TO ENSURE PROTECTION
OF CONFIDENTIAL INFORMATION



The Toxicology Data Management System. Section 4 of TSCA authorizes EPA to require industry to test chemicals for health and environmental effects. Test data are also reported to EPA under federal pesticides laws and to FDA under its authorities. Research institutes around the country generate additional test data for other purposes. Such data are produced in staggering quantities, especially during complex long-term studies concerned with the carcinogenic potential of a chemical substance.

The Toxicology Data Management System (TDMS) will maintain raw data from toxicological experiments. Because the system requires less paperwork than manual procedures, it can also serve as an efficient management and collection agent for ongoing experiments. As a core component of CSIN, TDMS will allow CSIN users, including testing laboratories and chemical manufacturers, to report and monitor the status of ongoing experiments, to examine raw data and procedures used in different experiments, and to find out if experiments are being performed on a given chemical.

Other Related Systems. Many other information systems will be important to the CSIN user community, and will be added to the system over time.

- The Chemical Regulation and Guideline System. The Chemical Regulation and Guideline System (CRGS) maintains information on the status of chemical regulations and guidelines as they develop. CRGS will cover federal regulations, U.S. statutes, state laws and regulations, court decisions, international and foreign regulations, government standards and guidelines, and other relevant documents. The documents will not be fully reproduced, but summary records will be available for each citation. Relevant records will be obtainable through the use of chemical names, product/trade name, CAS registry number, product code, date of issue, document category code, responsible agency or controlled subject terms. It will be available in 1981.

- CHEMTRAX includes information on the status of analyses for regulatory purposes. It provides data on who is examining a substance, why they are studying it, and where they are conducting their analysis.

- The Health Effects and Environmental Data Analysis (HEEDA) system is currently under development. It will provide information on the results of tests and evaluations of chemical effects on simple and complex animals (amoebas to rodents).

- UPGRADE is a system for analyzing data about natural resources, health effects, the environment, and related issues. The UPGRADE facilities include all the statistical capabilities found in SAS (Statistical Analysis System), combined with facilities to manipulate data.

VII. CONCLUSION

Through the development of CSIN, the ITSDC has made progress toward fulfillment of its TSCA mandate to establish an efficient system for the coordination of chemical information. Analysis has shown that the information network will serve the demands of a wide range of users and will greatly improve the usefulness of retrieval systems. The CSIN query processor or management computer will be the hub of the completed system. Through a telecommunications network, it will link the user to the many component resources. The Chemical Information Resources Directory will direct users to the proper database, and the Chemical Substances and Nomenclature System will assure positive identification of all chemical substances. A range of specialized components will cater to more particular user needs.

The ITSDC hopes for rapid and broad expansion of the user community once the second prototype version of CSIN is made available. However, several problems lie ahead. Few are technical; rather, they are institutional, legal, and economic. The variety of users and component systems will complicate the procedure for formulating accounting and billing procedures for components accessed through CSIN. Each component system must prepare itself for the increased volume of queries generated by CSIN. Among prospective users, priorities must be established until the information network is fully operational. Users must learn to take full advantage of the CSIN query capabilities, but also to understand the powers of its components, their limitations, and the utility of the information retrieved.

The long term benefits of CSIN are at least threefold. First, CSIN will serve to facilitate efficient and effective information retrieval. And it will go beyond its prescribed role to meet a second need. By tapping existing information and uncovering information gaps, the CSIN framework will enable users to plan future research and other efforts more efficiently. Thirdly, CSIN can be thought of as the forerunner to similar systems in other subject areas that will undoubtedly be developed during the coming decade. The diversity and dispersion of automated information resources has necessitated development of coordinating networks such as CSIN, which facilitate optimal use of existing resources.

The ITSDC is pleased to report the progress that has been made toward simplifying the retrieval of chemical information. For the CSIN user, the intricate process of searching multiple information sources will be immensely simplified. Faster retrieval of more complete data promises to improve both the long-term decisionmaking capabilities and emergency responses of all users.

BIBLIOGRAPHY

Bracken, Marilyn C., et al. Chemical Substances Information Network. Two volumes. McLean, Virginia: The Mitre Corporation, 1977. [Volume I: PB 274 637, A10, A01; Volume II: PB 274 638, A13, A01]

Council on Environmental Quality. The Feasibility of A Standard Chemical Classification System and a Standard Chemical Substances Information System. Washington, D.C.: GPO, 1978. (041-011-00039-4)

Eastlake III, Donald E., et al. Design of the Version I Prototype Chemical Substances Information Network--draft. Cambridge, Massachusetts: Computer Corporation of America, revised 1980.

Graf-Webster, Erika. The Chemical Substances Information Network Directories: Their User Functions and CSIN Functions. McLean, Virginia: Graf Associates, 1980.

Hororwitz, Aren J., et al. Design of a System for Pre-Prototype Use of the Chemical Substances Information Network. Cambridge, Massachusetts: Computer Corporation of America, 1979. [PB 80 178 320, A04, A01]

"Interagency Toxic Substances Data Committee Notice of Charter, Membership and Meetings." Federal Register 43: 22776-8. May 26, 1978.

Lefkovitz, David, et al. Requirement for a Chemical Structure and Nomenclature System. Philadelphia: University of Pennsylvania, 1979. [PB 105 216, A11, A01]

Levitan, K., et al. Chemical Information Resource Directory Two volumes. McLean, Virginia: The MITRE Corporation, 1979. [Volume I: PB 105 208, A99; Volume II: PB 105 190, A05, A01]

Note: The information given in parentheses indicates that the item is available from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. That given in square brackets indicates availability (accession number and price code, first for paper and then for microfiche) from the National Technical Information Service, U.S. Department of Commerce, Springfield, Va. 22151. Because it is not feasible for the Superintendent of Documents to correct prices manually in all publications stocked, the prices charged may differ from the prices printed in the publications.

Marks, Bennett D., et al. A Sample Analysis for Requirements for the Chemical Substances Information Network. Cambridge, Massachusetts: Computer Corporation of America, 1979.

Overview of the Chemical Substances Information Network. Sigma Data Computing Corporation, 1979. [PB 284 030, A10, A01]

"Toxic Substances Control Act." Public Law 94-469.

Toxic Substances Strategy Committee. Toxic Chemicals and Public Protection. Washington, D.C.: GPO, 1980. (0-318-148/6294)

Wilson, Gerald A., et al. An Analysis for the Chemical Substances Information Network. Two volumes. Cambridge, MA: Computer Corporation of America, 19. [Volume I: PB 80 222458, A06, A01; Volume II: PB 222 466, A10, A01]

__________. Immediate Requirements for a Chemical Substances Information Network. Cambridge, Massachusetts: Computer Corporation of America, 1979.

Winter, Richard A., et al. Immediate REquirements for a Chemical Substances Information Network: Supplementary Analysis. Cambridge, Massachusetts: Computer Corporation of America, 1979.

__________. A Prototype Chemical Substances Information Network. Cambridge, Massachusetts: Computer Corporation of America, 1979. [PB 108 319, A07, A01]

*U.S. GOVERNMENT PRINTING OFFICE: 1981-0-341-082/225