METREK DIVISION

MTR-7558
Vol. I

Chemical Substances Information Network

Volume I
User Requirements and Systems Development Options

M. BRACKEN
J. DORIGAN
J. HUSHON
J. OVERBEY, II

JUNE 1977

MITRE

MITRE Technical Report
MTR-7558
Vol. I

Chemical Substances Information Network

Volume I

User Requirements and Systems Development Options

M. BRACKEN
J. DORIGAN
J. HUSHON
J. OVERBEY,II

CONTRACT SPONSOR	Council on Environmental Quality Environmental Protection Agency National Library of Medicine
CONTRACT NO.	CEQ7A010
PROJECT NO.	15360
DEPT.	W-56

JUNE 1977

THE MITRE CORPORATION
METREK DIVISION
McLean, Virginia

Department Approval: Lydia A. Kirmao

MITRE Project Approval: Marilyn C. Backus

ABSTRACT

Under a joint contract with the Council on Environmental Quality, the Office of Toxic Substances, Environmental Protection Agency, and the National Library of Medicine, METREK has surveyed potential users of chemical substance information and has analyzed the ability of existing data bases, both Federal and private, to meet these expressed needs. In order to provide information on chemical substances in an optimal manner, METREK has proposed the development of a Chemical Substances Information Network. This network is designed to contain both core component systems, and external systems locatable through a directory. Modifications to existing systems which meet user requirements are suggested as well as general specifications for new systems. Strategies for management and implementation of the various components of the network are also presented.

EXECUTIVE SUMMARY

Under a joint contract with The Council on Environmental Quality, the Office of Toxic Substances of the Environmental Protection Agency and the National Library of Medicine, METREK has surveyed potential users of chemical substance information within EPA, the Federal establishment and other industry, educational and consumer action group users. User requirements for information have been characterized with respect to subject matter and application to existing legislated authorities and new mandates under the Toxic Substances Control Act (TSCA).

In addition, METREK has collected data on various systems which could supply some of this requested chemical substance information in the categories of Substance Identification, Production Aspects, Marketing, Exposure, Epidemiology, Biological Effects, Environmental Effects, and Standards and Regulations. Information collected by the Council on Environmental Quality as a part of their inventory of Federal Chemical Data Bases was supplemented with information on private as well as additional Federal data bases related to chemical substances. The data bases were then considered with regard to their relevance and breadth of coverage in each of the information categories. Those data bases determined to be most important were then labeled primary systems and were evaluated more fully in those categories where they might supply useful information.

Other METREK efforts involved the performance of a detailed analysis of alternative approaches for satisfying user requirements. The information available in each category was compared to the needs expressed during the interviews and the applicable existing data bases and need for additional data bases identified.

Based on those existing primary systems identified above as best able to supply essential information, the specifications for a Chemical Substance Information Network are presented. The Chemical Substances Information Network is designed to contain both core component systems and external systems, locatable through a directory file. The core systems include The Chemical Data Bases Directory, the Chemical Structure/Nomenclature System, the TSCA Chemical Data Systems (Proprietary and Public), the TSCA Reports Management System, the Toxicology Data System, the Chronic Testing Support System, the Bibliographic Literature Scanning System, the Laboratory Animal Data System and the Regulated Chemicals Standards System. The content, management and time-phased implementation of these core components are considered. Where existing data systems can be used directly or modified to provide the basic needs of these core systems, they are presented. In those areas essential to chemical evaluation and regulation, where information does not exist or is inadequate, new systems are recommended for development. In addition, where existing systems are found to be useful but redundant, consolidations are suggested.

To investigate alternative strategies for establishing the Chemical Substances Information Network and its member data systems, three

scenarios for systems development are examined. These scenarios are based on various TSCA implementation strategies and differ with regard to the nature of the information requested from industry and the timing of these requests. For each scenario, different systems development options are presented due to the variance in dependence on external files to supply data potentially obtainable under TSCA.

Volume II of this report contains the appendices to Volume I, including detailed documentation of the user requirements interviews and background data for each of the primary data systems.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Concluded)

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (Concluded)

1.0 INTRODUCTION

The Toxic Substances Control Act (TSCA) was signed by the President on October 11, 1976 and became effective January 1, 1977. This Act provides EPA with the authority to regulate chemicals in commerce not covered by existing Federal regulatory authorities. One of the main thrusts of the Act is that it provides for a vital source of new data with which to assess the possible risks and benefits of chemicals in the environment. Under TSCA, manufacturers, processors, exporters, and importers are required to report on: (1) information on new chemical substances proposed for commercial production and selected new uses of existing substances, (2) annual production activities for selected existing substances as listed in EPA reporting regulations, and (3) health and safety data. "Trade secrets" and other confidential information may be included and must be protected against unauthorized disclosure.

In addition to the specific reporting requirements of the Act, elements of EPA will require supporting information from a variety of external existing sources for decision-making purposes, particularly in developing regulations calling for testing or restrictions on the manufacture, use or distribution of certain substances.

Furthermore, EPA has stated in its draft strategy document, Assessment and Control of Chemical Problems, EPA, February 1977, that information obtained under the Act will be made available as promptly and widely as possible to enable other Federal, state and

local agencies as well as the private sector to be utilized as fully as practical in meeting the purposes of the Act.

Section 25(b) of TSCA requires the Council on Environmental Quality to coordinate a study within 18 months of the feasibility of establishing (1) a standard classification system for chemical substances and related substances, and (2) a standard means for storing and for obtaining rapid access to information on these substances.

This study was undertaken in support of CEQ's responsibilities as stated above and the responsibility of the Information Management Unit of the Office of Toxic Substances, EPA to design and establish an effective system for the retrieval of toxicological and other scientific data as called for by the Toxic Substances Control Act. This effort is also supporting the National Library of Medicine in its requirements to assemble and make available information concerning chemical substances.

1.1 Scope of Work

Task 1 involved the identification and characterization of groups that have regulatory responsibilities to control toxic substances and/or concern with the general goal of protecting human health and the environment from unreasonable risks presented by chemical substances. Those groups considered included:

1. The Office of Toxic Substances (OTS) in EPA
2. Other EPA Headquarters Offices
3. EPA regional offices
4. EPA laboratories

5. Other Federal agencies and departments
6. State and local government agencies
7. International organizations
8. Other interest groups (industry, universities, public and private interest organizations)

METREK characterized the information requirements of these users with respect to:

1. Subject matter (e.g., physical properties, production data, toxic effects, etc.).
2. How the information relates to the TSCA mission.
3. Who would use the information and for what application (preparation of the regulations, creation of criteria documents, etc.).
4. Characteristics of data required to satisfy the need (e.g., timeliness, volume, accuracy, quality).
5. When the information is needed and how rapidly.
6. The kinds of manipulations of the data required to produce useful information.
7. The form in which the information need would be expressed (telephone query, written request, etc.).
8. The form in which the need could be satisfied (ad hoc report, annual report, on-line interactive retrieval, etc.).

For each user or user group, METREK considered the importance of each requirement with respect to (1) the particular use or user for which it is intended (e.g., early warning, research, monitoring, etc.) and (2) EPA priorities under TSCA. Section 2 of this report presents the findings concerning the user requirements task.

Task 2 involved the identification and evaluation of potential information sources to satisfy user requirements. Specifically, METREK created an inventory of existing files, both Federal and private, containing applicable information concerning chemical substances. The results of these findings are discussed in Section 3.

METREK characterized the information activities of each source with respect to:

1. Ownership (agency, public domain, private, foreign).
2. Who uses that information and for what purpose or application (preparation of regulations, creation of criteria documents, etc.).
3. Types of information -- bibliographic or numeric.
4. Mode of retrieval (batch, on-line interactive, manual).
5. Subject matter (physical properties, production data, etc.).
6. Characteristics of data (timeliness, volume, accuracy, quality).
7. Kinds of manipulation of data available.
8. Form in which requests for information must be expressed.

9. Form in which information is disseminated.

10. Maintenance (Is the file being maintained now or is it a "dead" file that still contains useful information? Does maintenance involve "updating" or "rebuilding"? Who maintains the file, and who pays for maintenance? How often is it done?).

Task 3 matched the user requirements identified in Section 2 with the evaluated existing systems identified in Section 3, and clarified those areas where user requirements are not being met by existing files. In Section 4, METREK demonstrated the need for new files as a result of their user requirements analysis and characterized those files which should be established to satisfy TSCA's requirements.

Task 4 involves the analysis of the results of the first three tasks and the development of an integrated information systems plan from a user requirements point of view. METREK inventoried existing and proposed systems for linking the files identified in Tasks 2 and 3, evaluated their strengths and weaknesses in the context of the user requirements analysis of Task 1, and made recommendations as to various system development options. These systems are or are envisioned as on-line interactive retrieval systems that could (1) link directly with a series of computerized information files, and (2) direct the on-line user to other external information sources, with or without on-line access, that are not physically linked to the central file.

Several levels of systems development are presented which depend on the TSCA implementation strategy and the timing and nature of data requests from industry. These scenarios provide the basis for a discussion of location and structure of those files required by EPA to fulfill its mandate under TSCA.

1.2 Limitations of the Study

The time frame for completion of the study was constrained by previously scheduled activities of EPA which were dependent on the output of this effort. The user requirements study, the identification of existing systems, and their evaluation (Tasks 1, 2 and 3) were completed in two months and Task 4, two months later. Consequently, the number of interviews that could be conducted was dependent on the available time and the funding allocated to this task by the project officer.

The selection of the groups to be interviewed was determined by the Government project officer and the METREK project officer. In addition, the quality of the interview was dependent on the representatives chosen by the various agencies/institutions to discuss their respective user requirements and use of existing data systems. In some cases, the representatives felt they could not address the total needs of this organization due to the size and diversity of programs. In several cases, additional interviews were held or further clarification was sought through telephone interviews.

The same limitations existed with respect to the quality of information obtained concerning existing data systems. CEQ conducted a survey of Federal data bases and some agencies contacted by CEQ failed to return their questionnaires and knowledge of their systems had to be gained through telephone interviews. The information that was provided on the data systems varied greatly in its degree of completeness. Again, efforts were made to obtain more information about relevant systems.

2.0 USER REQUIREMENTS FOR INFORMATION CONCERNING CHEMICAL SUBSTANCES

2.1 Introduction and Approach

In this Section user requirements are first discussed with regard to their need for information on chemical substances and second the importance of each requirement is ranked with respect to both the particular use of the data and EPA priorities for information under their strategy for implementing the TSCA authorities. In ranking the relative importance of information requirements, particular emphasis has been given to that information needed to support the testing and pre-manufacturing activities under TSCA section 4 and section 5 respectively.

To identify the requirements for these types of data, a large number of face-to-face interviews were conducted in a structured manner with representatives of the EPA Office of Toxic Substances, other EPA Headquarters Offices, EPA Regional Toxic Coordinators, EPA Laboratories, other Federal agencies and departments, international organizations, and other interest groups representing the viewpoints of industry, universities and other groups in the private sector. A list of the specific organizations contacted is presented in Appendix A.

During the interviews, the representatives of each organization were asked to characterize their specific responsibilities and to describe on going or anticipated actions or programs in response to their respective existing responsibilities and/or their responsibilities

as mandated by the passage of TSCA. It was felt that by initially obtaining a comprehensive understanding of the organizations' responsibilities, we could better discriminate between solicitations for information justifiable by specific identifiable functions performed by the organization and those which were less relevant. On this basis, specific information requirements and the characteristics of these requirements could then be identified for each potential user. Additionally, during the interviews, specific information sources were identified which are used currently to satisfy the need for data. The currently unmet information needs, and which of these needs could be fulfilled by the authorities for information collection mandated by TSCA were also discussed.

User requirements were divided into nine general categories. These include Substance Identification, Production Aspects, Marketing, Exposure, Epidemiology, Biological Effects, Environmental Effects, Standards/Regulations, and Managerial/Administrative. Within the broad categories, requirements for specific data elements were defined. The particular elements of each category are listed in Table 2-1. For each of these categories of data, it is also necessary to determine the characteristics of both actual or anticipated usage, including the data's accuracy and currency, access frequency, access mode, retrieval mode, application or purpose, relationship to TSCA mission and manipulations required to enhance the data's utility.

TABLE 2-1

CHEMICAL INFORMATION REQUIREMENTS
FOR ENVIRONMENTAL AND HEALTH HAZARD ANALYSIS

I. Substance Identification
  A. Descriptive Identification
    1. Nomenclature
      a. CAS Registry Number
      b. CAS Preferred Name
      c. Synonyms
      e. Trade Names
      f. Wiswesser Line Notation
      g. Other Codes
    2. Chemical Structure/Form
      a. Chemical Structure
      b. Molecular Formula
      c. Formula Weight
    3. Composition
      a. Methods of Determination
      b. Impurities
        (1) identification (same as I.A.1. and I.A.2.)
        (2) detection limits
        (3) percent
        (4) source

  B. Chemical Properties
    1. pH
    2. Reactivities
      a. With Water
      b. Oxidation-Reduction
      c. With Acid
      d. With Base
      e. Photoreactivity
      f. Nucleophilicity
      g. Electrophilicity
      h. Thermal
    3. Dissociation Constants
      a. Organic Bases
      b. Organic Acids
  C. Physical Properties
    1. State/Color/Texture
    2. Density
    3. Index of Refraction
    4. Melting Point
    5. Boiling Point
    6. Freezing Point
    7. Flash Point

TABLE 2-1 (Continued)

8. Volatility
    a. Vapor Pressure
    b. Vapor Density
9. Solubility
    a. Water
    b. Organic Solvents
    c. Octanol/Water Partition Coefficient
10. Spectral Properties
    a. Absorption Spectroscopy
        (1) ultraviolet range
        (2) visual range
        (3) infrared (IR) spectroscopy
    b. NMR Spectroscopy
    c. Fluorescence Spectroscopy
    d. Optical Rotation, Optical Rotatory Dispersion or Circular Dichroism
    e. X-Ray Diffraction
    f. Mass-Spectroscopy
11. Persistence (half-life)
    a. Hydrosphere
    b. Atmosphere
    c. Lithosphere
    d. Shelf-life

D. Methods of Identification
    a. Suitable Analytical Techniques
    b. Standard Protocols
        (1) AOAC methods
        (2) ASTM methods
        (3) other methods

II. Production Aspects

A. Production Source
    1. Name and Location of Manufacturers
    2. Amount Produced by Site
    3. Fraction of Production Lost
    4. Process
    5. Control Technology
    6. By-Products
        (1) identity
        (2) amounts
        (3) disposal methods
    7. Impurities

B. Commerce
    1. Annual U.S. Production
    2. Annual U.S. Imports
    3. Annual U.S. Exports
    4. Annual U.S. Consumption

TABLE 2-1 (Continued)

- C. Shipping Procedures
  - 1. Handling
  - 2. Storage
  - 3. Transport
  - 4. Fire Danger Rating

III. Marketing
- A. Uses
  - 1. Amounts by Use
  - 2. Trend Data
- B. Users
  - 1. Amounts by Use
  - 2. Place of Use
- C. Substitute Chemicals
- D. Economic Information

IV. Exposure
- A. Occupational
  - 1. Total Work Force
  - 2. Occupational Group
  - 3. Duration and Frequency
  - 4. Route of Exposure
- B. Consumer
  - 1. Food
  - 2. Drugs
  - 3. Cosmetics
  - 4. Pesticides
  - 5. Other Products
  - 6. Exposure Rate and Duration by Route
- C. Environmental
  - 1. Air
  - 2. Water
    - a. Surface
    - b. Ground
    - c. Marine or Estuarine
    - d. Drinking Water
  - 3. Soil
  - 4. Plants
  - 5. Wildlife

V. Epidemiology
- A. General Population
- B. Occupational Population

TABLE 2-1 (Continued)

VI. Biological Effects
- A. Clinical Studies
  - 1. Exposed Populations
  - 2. Procedures
  - 3. Results
- B. Toxicology (Human/Animal)
  - 1. Acute Toxicity
    - a. Study Characteristics
  - 2. Sub-chronic Toxicity (experimental conditions)
    - a. Study Characteristics
  - 3. Chronic Toxicity
    - a. Carcinogenicity
    - b. Teratogenicity
    - c. Mutagenicity
    - d. Other
- C. Metabolism (Human/Animal)
  - 1. Blood and Other Organ Levels
    - a. Parent Compound
    - b. Metabolites (with CAS numbers)
  - 2. Excretion Rates
    - a. (as above)
  - 3. Absorption (gut, skin, respiratory tract)
    - a. (as above)
  - 4. Distribution
    - a. Organ/Tissue Sites
  - 5. Chemical Interactions

VII. Environmental Effects
- A. Degradation
  - 1. Biodegradation
    - a. Organism
    - b. Products (with CAS number)
  - 2. Chemical Degradation
    - a. Rates
    - b. Products
- B. Environmental Transport and Fate
- C. Ecological Effects
    - a. Effects on Vertebrates (birds, fish, amphibians, and reptiles)
    - b. Effects on Invertebrates (annelids, arthropods, and crustaceans)
    - c. Effects on Plants
    - d. Effects on Microorganisms
- D. Materials Effects
- E. Weather and Atmospheric Modification
- F. Bioaccumulation/Bioconcentration

TABLE 2-1 (Concluded)

VIII. Standards and Regulations
- A. Federal Standards and Regulations
- B. State Standards and Regulations
- C. Local Standards and Regulations
- D. Non-U.S. Standards and Regulations
- E. International Standards and Regulations

Upon the completion of the interviews, it was necessary to determine whether the information requested by an individual was actually required to fulfill his responsibilities. There was the additional task of determining if the justifiable requirements could be satisfied by some legislative authority other than TSCA so that one could effectively prioritize these requirements with respect to EPA priorities under their strategy for implementing TSCA.

2.2 Scope and Limitations of the User Requirements Study

So that the results of this user requirement analysis can be viewed in the proper context, it is necessary to highlight some specific considerations which were associated with this task:

- In determining user information requirements, the types of information which could be obtainable under authorities in addition to TSCA were considered. This was done so that a more comprehensive characterization of requirements for data on chemical substances could be developed to aid CEQ in performing its requirement under section 25 of TSCA.
- The time period within which the requirements analysis was to be completed was constrained by the timing of other EPA on-going, related studies. For example, the results of the requirements analysis were not only to provide input to the second phase of this effort but also to EPA's activities associated with developing the information system to handle data being provided in response to TSCA.

- In order to obtain an assessment of the user requirements of industry and other interest groups including consumers, the Government project officer directed METREK to meet with representative groups such as industrial trade associations, select public interest groups and a representative of the university community. In some cases, the groups were responsive and provided representatives who had considerable knowledge of the concerns and user requirements of the group they represented, and in other cases, less information was obtained during the interview situation. The list of interviewed groups is not meant to be comprehensive and, in fact, could not be, due to funding limitations, time constraints and the impracticability of meeting with large numbers of similar groups. The groups selected are representative of their constituents and have provided a valid assessment of user requirements and existing sources of information which they presently use.
- The specific policies, actions and assignment of responsibilities of EPA were evolving during the time period in which the interviews were conducted. The impact of this circumstance is that the relative priorities of the identified user requirements while representing the most accurate determination at this point in time, must not be considered as static but, rather, might be subject to changes in their relative emphasis.

It is unlikely, however, that major changes in user requirements will occur.

The remaining portions of this Section contain a summary discussion of the major features of the integrated requirements. A summary of the content and conclusions from each of the individual interviews is presented in Appendix B of Volume II.

2.3 Legislative Authority of Regulatory Agencies in Controlling Chemical Substances

Although a number of Federal regulatory agencies are involved with controlling chemical substances, their legislative mandates vary in terms of the specific chemical substances involved, the stage during the chemical life cycle (e.g., pre-manufacturing, production, transportation, use, disposal) or the application (e.g., industrial, consumer, commercial). It is difficult to set forth a definitive list of specific jurisdictional involvements of the relevant agencies since there are a number of overlapping jurisdictions. Moreover, policies for implementing the legislative authorities change as the agencies analyze and clarify their positions.

Table 2-2 presents an indication of the legislative responsibilities for various types of chemicals by agency. Several of the legislative authorities impose provisions requiring manufacturers, processors, and distributors to maintain records and report various types of information. Production information, health and safety data and environmental effects data are examples. The passage of the Toxic Substances Control Act provides for imposition of additional research, record-keeping and

TABLE 2-2

LEGISLATIVE RESPONSIBILITIES OF AGENCIES IN THE CONTROL OF CHEMICALS

AGENCIES	TYPES OF CHEMICALS						
	FOODS	DRUGS	COSMETICS	PESTICIDES	OTHER CONSUMER PRODUCTS	INDUSTRIAL	RESEARCH
FDA	X	X	X				
CPSC		X[1]			X		
OSHA[2]	X	X	X	X	X	X	X
ERDA							X
DOT[3]				X	X	X	X
EPA[4]				X		X	
USDA	X						
DOD							X

1) Child Resistant Packaging Regulations
2) Concerned with protection of workers exposed to all chemicals
3) Transportation regulations
4) Also responsible in terms of plant emissions and effluents for all types of chemicals

reporting requirements which extend the Federal government's information-gathering and regulatory authorities. The implementation of TSCA provides the opportunity to coordinate the collection of information among the Federal agencies regulating similar areas of chemical substances so timely and accurate information can be obtained with the least possible burden on business and industry.

One purpose of this study is to identify specific, justifiable Federal and private sector requirements for chemical substance information which may or may not be addressable under existing legislation.

2.4 Integration and Prioritization of Individual Requirements

To aid in characterizing user requirements, a comprehensive understanding of the potential applications of this information is necessary. When examining these applications and examining the budget categories of the regulatory agencies, common functional responsibilities and their chronological sequence could be identified.

2.4.1 Identification of the Functional Areas

Within EPA, and also to a large extent within other Federal regulatory agencies, the functional responsibilities of individual offices fall among ten general categories. These functional categories include:

- Hazard Identification/Prioritization and Early Warning of Potential Risks
- Hazard Analyses
- Research/Development
- Development of Decision Packages (Criteria Documents)

- Preparation of Regulations and Guidelines
- Monitoring/Testing
- Enforcement/Compliance
- Information/Education
- Support to Other Agencies/Organizations
- General Administration and Management

The typical decision-making pattern involves initially identifying a hazard, followed by a "hazard analysis" which in some cases must be conducted within a short time period. In other cases, development of testing protocols and research are necessary to adequately assess the hazard to humans and the environment from exposure to chemicals. A "decision package," examining alternative regulatory options, is prepared once the hazards are clearly identified. This package is then forwarded to an action group for decision-making concerning regulatory resolution of the problem. Monitoring data may need to be collected and analyzed to determine the extent of the exposure. Should the decision be made to regulate the substance/item (be it label, ban, limit or control the manufacturing, etc.) a comprehensive data gathering activity occurs which includes a more thorough economic analysis of the impacts associated with individual regulatory actions. In some cases more research, monitoring and data analyses are required to support the regulation preparation stage. Subsequently, compliance and enforcement of the regulation is the primary functional activity in conjunction with an evaluation component to determine the

effectiveness of the regulation in reducing the risk to the public and the environment from exposure to that chemical substance.

For characterizing and integrating the requirements of individual offices according to their application of the data, the above mentioned functions have been used. No single office performs each of these functional activities. Some offices, depending on their respective mandate, perform several of the functions (e.g., research and hazard analysis) in support of other agencies or groups. In Table 2-3, the specific responsibilities of individual EPA offices and other agencies are identified.

Hazard Identification/Prioritization involves selection from the universe of chemicals of those with which the agency or group will be concerned. This category includes the function of early warning which attempts to restrict the total number of chemicals by calling attention to those which may have significant potential for risk. The types of chemicals are different depending on the particular mandate of the agency or group. For example, the Consumer Product Safety Commission examines chemicals used in consumer products; the Food and Drug Administration focuses on chemicals used in foods, drugs, and cosmetics. Typically, the agency, using various criteria and various types of data, selects a subset of chemicals for which there is greater concern about the risk of exposure.

Hazard analysis includes surveying the literature and analyzing health and environmental effects test data submitted in response to

TABLE 2-3

OFFICES/AGENCIES AND THEIR FUNCTIONAL ACTIVITIES RELATED TO TOXIC SUBSTANCES

	Hazard Identification/ Prioritization	Hazard Analysis	Research/ Development	Decision Package Development (Criteria Document)	Monitoring/ Analysis	Preparation of Regulation and Guidelines	Enforcement/ Compliance
OTS							
Regulation		x				x	
Testing	x	x	x			x	
Coordination*							
Hazard Assessment	x	x		x			
Special Actions		x		x	x		
Pre-manufacturing		x					
Information Management*							
Program Management*							
OTHER EPA							
Enforcement						x	x
Water and Hazardous Materials	x	x	x	x	x	x	
Air and Waste	x	x	x	x	x	x	
Research and Development		x	x				
Regions	x	x			x	x	x
Laboratories		x	x	x	x		
OTHER FEDERAL							
FDA	x	x	x	x	x	x	x
OSHA	x	x			x	x	x
NIOSH	x	x	x	x	x		
CPSC	x	x	x	x	x	x	x
DOC*							
DOI	x	x	x		x		
DOD	x	x	x		x	x	x
NIEHS		x	x				
NCI	x	x	x		x		
ERDA		x	x				
Interagency Testing Committee	x	x					
NAS		x					
DOT	x	x	x	x	x	x	x
NLM		x	x				
INDUSTRIAL/TRADE ASSOCIATION/CONSUMER							
SOCMA*							
CSMA*							
CIIT	x	x	x				
MCA	x	x	x				
NRDC	x	x					
Cons. Found.		x	x				
Labor Unions*							
UNIVERSITY							
NYU			x		x		

*Perform information support and/or administrative functions

mandates requiring testing of selected chemicals. It includes a preliminary hazard assessment in response to a citizen's petition or a substantial risk notification as well as the assessment made with respect to a pre-manufacturing notice under TSCA. It further includes a limited economic analysis of the impact of alternative regulatory options. In performance of the activities of hazard analysis, similar compounds are often structurally compared.

Research/Development includes conducting the fundamental research necessary to define, measure and control the effects of chemicals, to understand their biological interactions, and to provide a basis for the elimination or reduction of the exposure to those chemicals which are deleterious to human health and the environment. Test method development and research on the applicability of various control technologies are also included in this function.

Development of "Decision Packages" (Criteria Documents) includes the development of comprehensive documentation which serves as the basis for a decision concerning the need to regulate a chemical. No single group has complete responsibility for the entire function but rather contributes component parts of such a package.

Preparation of Regulations, Standards, and Guidelines is defined here to include the analysis of and selection from the various regulatory options available to the different agencies, (e.g., banning, seizing, labeling, packaging requirements, controlling the exposure limits, etc.). The function of the development of regulations is

consistent with the need for a comprehensive data package which will document and substantiate the recommended regulatory strategy.

Monitoring/Analyses includes monitoring and subsequent raw data analysis of chemical concentrations in the air, water, and soil. Also included in this functional activity is the analysis of epidemiological studies to identify the effects of exposure on human health and other species*.

Enforcement/Compliance involves enforcing compliance with the particular laws that the agencies administer. It includes compliance monitoring to identify violators, laboratory analysis to substantiate violations, and compilation of evidence to support legal action when violators are found.

It is further recognized that three other functions exist for which requirements for information could be identified. These include the function of Information and Education**, Support to Other Agencies/ Organizations, and General Administration and Management. However, the decision was made not to include these functions in this effort since they did not impose unique data requirements separate from those already identified for program responsibilities.

2.4.2 Functional Groupings

When these functions are analyzed, they can be grouped into three categories which have common data requirements or data

*It is recognized that many of the epidemiological studies and/or systems are also used for purposes of hazard identification and establishing program priorities.

**The function of Information and Education incorporates the critical activity of making the chemical information data bases available to the scientific and academic communities for further enrichment and confirmation.

attributes, and common characteristics with respect to the time frame within which actions are required.

The first category includes the function of Hazard Identification/Prioritization of chemical substances. When conducting these activities, all chemical substances must be considered, the time frame is typically long (or not a constraining parameter) and the information need not be highly specific or detailed.

The second category of functions are associated with actions which often occur in response to an external stimulus such as notifications of intent to manufacture a new chemical, substantial risk, imminent hazard or citizens petitions. Typically, the identity of the chemical substance cannot be anticipated and the time frame within which actions occur is generally short. The data must be sufficiently specific and accurate to permit a fairly comprehensive assessment of risk. It must also be defendable with respect to possible resulting litigation. The functions in this category include Hazard Analysis, Preparation of Decision Packages, Monitoring/Testing, Preparation of Regulations, and Enforcement/Compliance.

The third category includes the same functions as the second category with the additional activity of conducting Research and Development. However, the characteristics of the data needed to support functions for this category differ from the characteristics of Category II. The distinction is that the particular chemicals for which these functions are being performed are those which were

identified either as a result of the Hazard Identification/Prioritization process or those chemicals which were identified through imminent hazard notifications or pre-manufacturing notices for which additional assessment and review is required. The time frame available for actions with respect to Category III functions is considerably longer that that associated with Category II functions. The data developed for supporting these functions must also be more defendable (i.e., accurate) than that required for Category II functions.

The functions of Information and Education, Support to Other Agencies, and General Administration are not included in this effort, although vital and essential, since they are not a direct part of the regulatory chain of events.

The differences between these three categories are illustrated conceptually in Figure 2-1. At the highest level, Category I, there is a requirement for the least specific information for the largest number of chemical substances. At lower levels, Categories II and III, information is required for fewer chemical substances but the need arises for data in additional categories and for more specific and accurate information within each category.

It should also be noted that there is a normal progression from Category I to Category III activities. There may also be a progression from Category II to Category III depending on the type and adequacy of the regulatory action selected in the Category II regulation step.

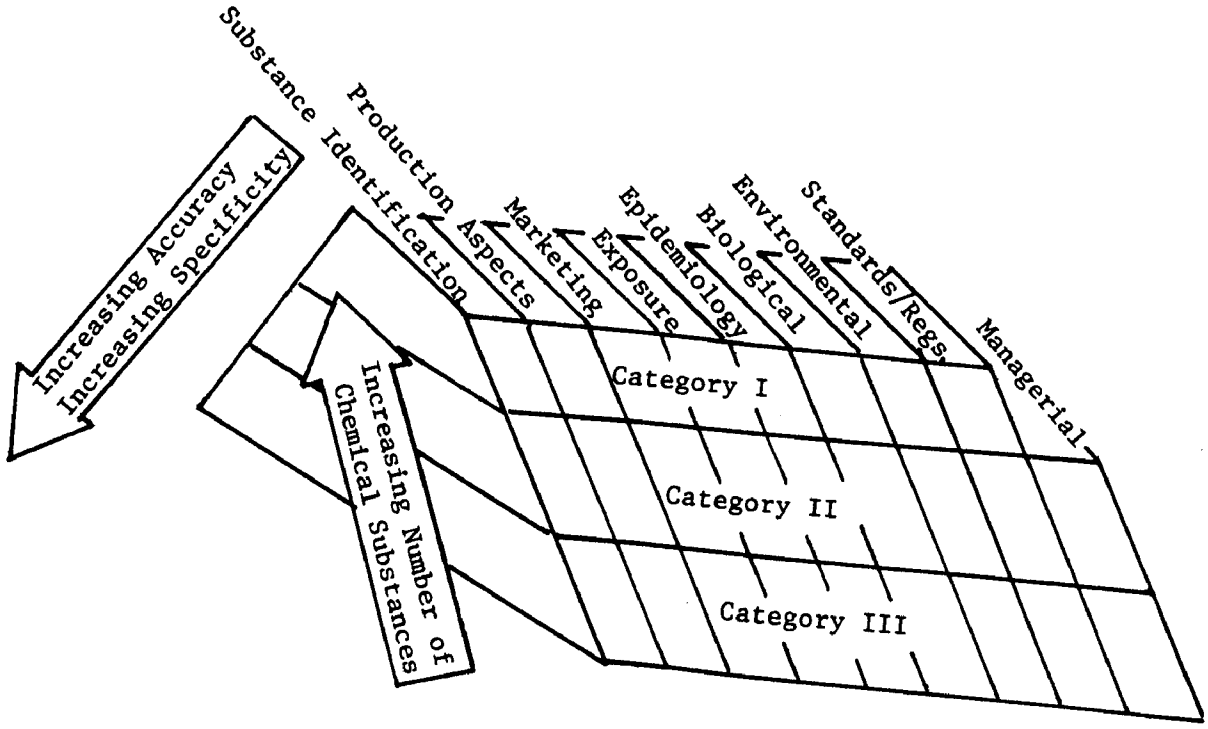

FIGURE 2-1
ILLUSTRATION OF DATA REQUIRED AND ASSOCIATED ATTRIBUTES

2.5 Analysis and Integration of User Requirements

It is within this context that the requirements for information concerning chemical substances are discussed. First, the categorized information requirements are integrated across all users. Next, they are integrated across EPA users. Finally, the categorized requirements are integrated across EPA users according to specific priorities identified in the EPA strategy for implementing TSCA as reflected in Assessment and Control of Chemical Problems - "An Approach to Implementing the Toxic Substances Control Act"; Environmental Protection Agency, February, 1977.

2.5.1. Prioritization of Requirements Integrated Across All Users

In Table 2-4, the requirements for information within each category, together with the requisite attributes of the data, are listed as they relate to each function of responsibility for all users. The scoring shown on this chart reflects the responses obtained from representatives of these agencies during the interviews. When a category or item is blank, there is no justifiable requirement cited. In a few select cases, certain data elements listed in Table 2-1 have been either eliminated or combined to form those listed in Table 2-4. The source of the requirement is identified in Table 2-5.

2.5.1.1 Requirements Associated with Category I Functions In conducting an initial screening of all chemical substances to identify a restricted set of substances for which a more detailed examination

will be conducted, there is a consensus among the regulatory agencies in their requirement for selective Substance Identification information for a large number of chemical substances updated on an annual basis.

As can be seen by examining the left column of Table 2-4, the number of data elements required within this category is limited. However, those which are required must be available for a large number of chemicals. To have this information on only a restricted set of substances would severely restrict their ability to conduct meaningful hazard identification, early-warning, and prioritization in any systematic manner. Without this information, substances of similar molecular structure could not be grouped.

Beyond Substance Identification data, requirements for additional categories of data vary somewhat according to the mandate of the requestor and the approach employed in conducting initial screenings of all substances. The most frequently cited requirements are those for Production, Use, and Exposure information. This consensus was supported by EPA, NCI, OSHA, NIOSH and the Interagency Testing Committee.

In conducting initial screening, the above agencies have requirements for data on the quantity of each substance produced. For most, this need can be adequately met by range type of data, indicating the total amount produced. For this reason we have indicated that there is a requirement for summary data (least specific). This does not imply, however, that a lack of accuracy is acceptable. In the instance where a large number of manufacturers are engaged in the

production of a specific substance, highly accurate information must be obtained from each manufacturer to avoid a highly imprecise total when the individual production quantities are aggregated. Information regarding the amount produced by "small manufacturers"* must also be obtained to ensure the accuracy of aggregated statistics.

Within EPA and OSHA there is a justifiable requirement to obtain site specific production data to be used in the initial prioritization of substances. The EPA regions, in particular, stated a requirement for this data for establishing resource allocation priorities in a predictive rather than reactive manner. Aggregations of the amount produced on a geographical or corporate entity basis will not satisfy their requirement. Similarly, OSHA requires site specific production information for establishing priorities for executing its responsibilities.

General indications of changes in the production process or technologies for controlling emissions and effluents resulting from chemical production are required as the state-of-the-art evolves. This information is used as an early warning indicator of a potential new hazard.

Information regarding the usage of substances is required in general categories sufficiently specific to permit the identification of new usages. Baseline usage data with amounts are needed to assess significant new usages.

*Currently, EPA is engaged in developing a quantifiable definition of this term.

In addition to production and usage information, data on the workforce exposed to substances during their manufacture, and environmental and consumer exposure are required as an initial indication of the extent to which humans are exposed to the substance. An aggregated national figure updated annually will satisfy the requirements cited by EPA, OSHA, NCI, and the Interagency Testing Committee.

With the exception of information on changes in production processes, control technologies and site specific production information, it is required that the above information be accessible in an interactive mode to facilitate the screening process. Non-automated access to information regarding changes in production processes and control technologies is adequate. However, due to the large amounts of data associated with site specific production information, it is recommended that this data should also be automated to facilitate updating and maintaining the currency of that information.

Having restricted the total number of chemical substances from the thousands which exist to a limited number of perhaps a few hundred, additional information is required for the remaining substances to enable a secondary screening to identify particular substances for which a detailed hazard analysis will be conducted. For substances selected as a result of the initial screening, both additional and more detailed data are required.

In the Substance Identification category, information (in addition to that previously cited) is required on the chemical and physical

properties of each substance. There is no requirement, however, that any of the information be automated - standard reference handbook texts are adequate for physical properties data. Chemical property data, however, is not currently available in easily retrievable and updated form. Automation of these two types of data, however, would greatly facilitate its access. Composition data for chemical substances is also required and, except for product composition data required by certain regulatory agencies, as mentioned above, does not need to be automated.

General descriptions of the particular production process employed, control technologies available and resulting by-products is required in addition to the information previously cited in the Production Aspects category. It is required that this information be updated as significant changes occur. Automation of this data is not required, but might be desirable to facilitate access.

The total quantities associated with each use and user category are needed along with summary economic information from the Marketing category. Specific workforce exposure by occupational group and consumer, and environmental exposure data are required together with data on media-specific concentrations, environmental persistence, and transport and fate to further assess potential exposure threats. Summary Epidemiology and Biological Effects information are needed for determining human health effects. Finally, information regarding existing Standards and Regulations is required. Except for biological data, it is not required that this additional data be automated.

2.5.1.2 Requirements Associated with Functional Categories II and III. When dealing with chemical substances in Category II whose identity is unanticipated until a request such as a pre-manufacturing or imminent hazard notification is received, or even the priority chemicals (Category III), requirements for substance identification are generally similar to those cited for Category I activities. For both Categories II and III, there is a requirement for information regarding impurities present in the marketed grade of the substance to aid in the evaluation of potential human health and environmental effects. For the same purposes, there is a requirement to know the place of use of the substance. Substance substitute information is necessary for identifying the condequences of alternative regulatory options from health and economic aspects. The requirement for interactive access capability is much stronger for Category II than for Category III, however, due primarily to the shorter time within which these functions must be performed*. The need for interactive systems for Category II chemicals can be further justified by the increased requirement for data manipulation and correlation capabilities to facilitate hazard analysis and decision making.

However, it is important to realize that while, with the above exceptions, no major difference occurs for the data required for Categories II and III as opposed to Category I, a major difference does

*Normally, policy decisions for pre-manufacturing are required within 10 days according to the EPA/OTS Strategy Document. This can be extended up to 90 days when a detailed analysis is required. In this instance the remaining functions would be conducted as Category III functions.

exist with respect to the chemical substance for which those data are required. This difference is illustrated by a Venn diagram, Figure 2-2. The large circle represents all chemicals. The circle labeled A represents those Category I chemicals for which secondary screenings of hazard identification are performed. The circle labeled B represents those unanticipated chemical substances (i.e., Category II) identified through pre-manufacturing notices and substantial risk notifications. The circle labeled C represents selected priority chemicals (i.e., Category III) for which detailed hazard analyses are performed. While systems can be designed to handle the large set of data (i.e., the union of circles A, B, and C) which are responsive to time frames associated with these functions, resources must be provided for analyses and interpretation of the data to adapt it for the purpose of regulatory decision-making.

2.5.1.3 Summary of Requirements By and Across Functional Categories. In developing an information system to maintain the required data and to be responsive to the access characteristics of all users, it is unnecessary to consider the functional application of information, given that all applications are considered as being valid and must be satisfied. The implication of this is that the characteristics of any data category or specific item which represent the most stringent requirement or demand on a system's capability become the system design parameter.

Table 2-6 was developed to aid in evaluating the degree to which existing data sources and systems could be utilized for satisfying

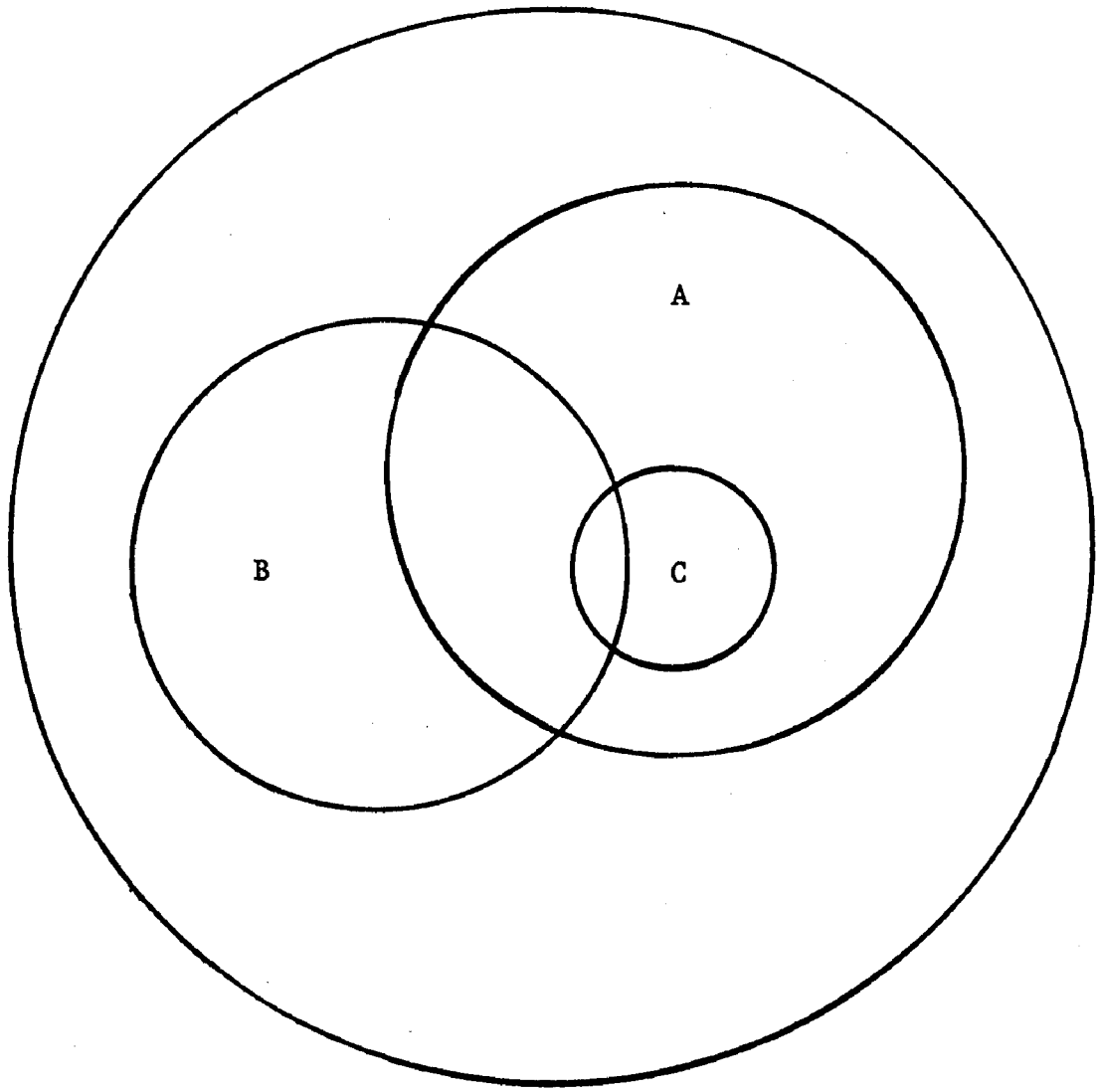

FIGURE 2-2
VENN DIAGRAM OF CHEMICAL SUBSTANCES

information requirements of all users. The data characteristics of this table represent, both the summary of individual functional categories and the system design parameters (integrated across functional categories), to identify the most stringent requirement.

For certain data categories there is a requirement for accessing data in an interactive mode. This requirement exists for Substance Identification (molecular formula through chemical structure), Production Aspects (site specific production quantity), Marketing (users with amounts and uses with amounts), Exposure (workforce, air and water, environmental and consumer), and Biological Effects data. In general, the greatest degree of specificity is required for these items. These requirements arise partially from the need for a capability to manipulate, within short periods of time, large volumes of data associated with many chemical substances. The requirement also arises from the need to review, assess, and summarize biological activity data indicating tests conducted, method of testing utilized and summary abstracts of the results.

For several other data categories, computerization of data, although unjustifiable by cited requirements, would enhance the utility of the data functions associated with developing pre-manufacturing decisions and responding to unanticipated substantial risks. Such categories include physical and chemical property data and environmental degradation and bioaccumulation. For example, it would be useful to develop computer files of baseline information on chemical and physical

property data so that correlations with biological activity data can be assessed for use in predicting biological effects of new substances. The EPA Strategy Document has stated that response to pre-manufacturing notices must be made in a very short time. Therefore, systems for assessing the completeness of a pre-manufacturing notice must be developed as well as a system to assist in the analysis of the data submitted. The development of similar analytical techniques will be required to assist in the review of testing data.

2.5.2 Prioritization of Requirements with Respect to TSCA Authority

Since the results contained in Table 2-5 were derived by integrating requirements from all offices with EPA with those from other Federal agencies, it is possible that requirements not directly related to TSCA functional responsibilities are the main driving force in determining the data items and their associated characteristics. To aid in examining the extent to which this situation had occurred, Table 2-7 was constructed by integrating over only those EPA offices which have, or will have, a direct connection with implementing TSCA responsibilities.

As can be seen by comparing these two tables, relatively few differences exist either in the data required, the functional usage of it, or in the attributes of the individual items: the requirement for workforce exposure by occupation, the requirement for economic information in Category I, and the request for Biological Effects data in support of the Research and Development function are eliminated from

the EPA table. Nowhere in EPA was there a cited requirement for information on the identify of individuals capable of providing expert witness testimony. It would appear, however, that this would be a justifiable requirement.

Table 2-8 was developed to aid in evaluating the degree to which existing data sources and systems could be utilized for satisfying information requirements prioritized with respect to the TSCA authority. As before, the characteristics of the data items represent, both for each functional category and across categories, the most stringent requirement in terms of systems capabilities. In comparing Table 2-6 and 2-8, no major differences can be found.

2.5.3 Prioritization of Requirements With Respect to EPA's Strategy for Implementing TSCA

As stated in its approach to implementing TSCA, EPA has divided the activities it will be conducting under TSCA into four major functional areas and several supporting areas, all of which are interrelated. The major functional areas are:

1. Acquisition of Information and Assessment of Risks to Health and the Environment;
2. Necessary Control of New Chemicals through TSCA Authorities;
3. Necessary Control of Existing Chemicals through TSCA Authorities; and
4. Dissemination of Information and Assessments to Other Programs and Interested Parties.

Supporting activity areas include the conduct of research, assistance to interested parties and implementation of TSCA procedural aspects.

During the initial three years of TSCA implementation, EPA has assigned top priority to the following operational activities: establishment and implementation of a Pre-manufacturing Review System; establishment of initial testing requirements; regulatory actions to control a limited number of environmental problems associated with existing chemicals; and assessment and control of unanticipated problems of urgent concern. With respect to collecting information in support of these top priority activities, it is the policy of EPA to gather data on a highly selective basis to serve specific purposes. Confidentiality considerations are to be a major factor influencing data collection, use and dissemination activities and strategies. In selecting priorities among the potential environmental problems, EPA has established the following principles:

- National or global toxic substance problems receive priority over localized problems,
- Human health effects of toxic substances receive special attention*, and
- Discharges into the environment of substances in significant quantities or those wich persist and/or bioaccumulate are of particular concern.

In light of these priorities, as set by EPA in its implementation strategy for TSCA, requirements for data to support pre-manufacturing review, development of testing requirements and regulatory actions for

* Recognition is given to ecological impacts that affect human health.

priority selected chemicals and unanticipated problems could be ranked by relative importance. When this is done, there is no change in the data items or their characteristics from that of Table 2-7. This finding should be differentiated from any determination regarding the provisions of that implementation strategy to satisfy these needs.

3.0 EXISTING FILES APPLICABLE TO TSCA

3.1 Introduction

METREK has attempted to assemble complete information on as many files containing data relevant to toxic substances assessment as possible.

Under their mandate in section 25(b) of TSCA to study the feasibility of establishing a standard means for storing and for obtaining rapid access to information concerning toxic substances, the Council on Environmental Quality (CEQ) conducted a survey of Federal data bases. In order to locate these data bases, CEQ combined the results of two previous environmental data system surveys: the "Study of Environmental Quality Information Program" prepared by EPA in 1971 but never published, and the "Survey of Environmental Data Systems: prepared in 1974 by GAO. In early 1977, the heads of the relevant agencies were then sent lists of systems attributed to their agency along with a two-page questionnaire to be completed on each system. Extra questionnaires were also included to cover new systems. When necessary, a follow-up was performed by CEQ. It was discovered that some of the systems for which information was sought were no longer in existence or had been incorporated into other existing systems.

Two hundred twenty four completed questionnaires were made available to METREK by CEQ. Where information on a given system proved to be inadequate, a telephone call was placed to the person designated as a contact for that system to provide additional clarification.

More Federal data systems were also uncovered during the interviews described in Section 2 of this report. These were then followed up and a questionnaire completed for them. The Directory of the Congressional Referral Center, Library of Congress "Federal Information Sources and Systems" also provided information on approximately twenty additional Federal data systems, bringing the total number of Federal systems to 239.

METREK included 55 private and foreign as well as the Federal systems in its inventory. A number of files applicable to toxic substances are available on private systems and are heavily used by both the Federal and private sectors. Many of the private data systems contain large numbers of files covering varied subject areas. This means that the data held in these systems generally provides a broader spectrum of information than that in the Federal data bases. They also have the advantage of being available to anyone willing to pay for the services.

Through this searching, METREK with the aid of CEQ, has attempted to assemble the maximum amount of data on all aspects of chemical substances. Much of the material collected in this initial compilation was duplicative or very highly specialized in nature. The subsequent sections of this chapter describe how a narrowed list of files was selected which it was felt would fulfill the information required to support TSCA-dictated activities and similar activities in other Federal agencies with mandates to regulate toxic chemicals. The types

of information required to fulfill the various TSCA-related data needs identified in Section 2 are discussed and the data systems most capable of supplying that information are identified.

3.2 Criteria Used to Select Files of Maximum Usefulness

In order to design an efficient data management system for information concerning chemical substances, it was necessary first to determine which of the existing Federal and private data systems could provide useful data. This condensation of files was accomplished in several stages.

First, the 260 files containing information in one or more of the eight toxic substances categories explained in detail in Section 2 were segregated from those 34 files described in Appendix C which were considered irrelevant. All files containing information pertinent to toxic substances were retained.

In order to further limit the number of files needed to supply relevant information, a dual scoring methodology was developed to better characterize the individual data files.* The first element of the score denotes the importance of the information to toxic substances research and regulation. This "importance factor" varies from a high of "1" to a low of "4".

---

* The methodology was subjective, and in some cases scoring was based on insufficient information about the system. Efforts were made to obtain adequate knowledge of systems in order to make valid judgment, and when in doubt systems were included until the second stage of the project when more specific attention will be given to the feasibility of systems integration.

The second element of the score is a measure of the value of the data and is determined by the following criteria:

- The number of records contained in the system;
- The specificity of the information;
- The extent to which the data were evaluated;
- The ease in accessing the data by both system and subject; and
- The breadth of coverage by the information in the data system.

The "value factor" was scored from a high of "a" to a low of "d". For example, BIOSIS, a bibliographic file, received the highest possible score, (1a), for information on Exposure, Epidemiology, Biological Effects, and Environmental Effects. This was predicated on the relevance of these categories of data to toxic substances assessment (earning a "1") and the exceedingly large number of records in the system, the extensive and in-depth coverage of all biological topics, and the ready availability of referred journal citations in a computerized file (adding up to a score of "a"). Substance Identification was given a score of 2b for BIOSIS, because chemical and physical characteristics of compounds or information other than common names are not expected to be found in the BIOSIS files. On the other hand, the Chemical Information System contains extensive data on Substance Identification, making chemical identification possible from varied inputs. The score of "1a" was based on these highly specific and comprehensive records, including mass spectrometry data, CAS registry numbers, Wiswesser line notations, X-ray diffraction patterns, CNMR values, and two-dimensional representations of molecules.

Each data base was ranked according to this methodology in each of the eight subject categories. The binary scores awarded to each data system are included in Table 3-1. Some systems received a high rating in several areas, some in only one, and some not at all. A low score implies that the system does not contain data of the highest value to TSCA-related activities.

Two additional columns have been included in Table 3-1 which provide supplemental file characteristics. One shows whether the system is manual or automated, while the other indicates the data base ownership. These facts are useful in determining the ease of accessing the information.

Based on the binary ranking scheme, it was possible to select those data bases of highest applicability to TSCA-related activities. All data systems receiving a minimum score of "1b" in any data category were selected for further consideration. These files are considered to be of primary importance and are designated by an asterisk in Table 3-1.*

---

* In a number of cases it was discovered that individual files were completely contained and accessed from a major system. For example, AEROS contains NEDS, SAROAD, EDS and HATREMS all of which contain information applicable to toxic substances. Only AEROS was designated on Table 3-1 as a primary system because the subsystems were available through it. It was also discovered that some identified systems were merely specific subject area subfiles of other systems. For example, CANCERLIT and CANCERPROJ are subfiles of CANCERLINE. As above, only CANCERLINE was designated as primary.

TABLE 3-1 DATA SYSTEM SCORING

			DATA TYPE								
			SUBSTANCE IDENTIFICATION	PRODUCTION ASPECTS	MARKETING	EXPOSURE	EPIDEMIOLOGY	BIOLOGICAL EFFECTS	ENVIRONMENTAL EFFECTS	STANDARDS AND REGULATIONS	(C) COMPUTERIZED OR (M) MANUAL
SYSTEM	ACRONYM	OWNER	I	II	III	IV	V	VI	VII	VIII	
*Advisory Center on Toxicology		NAS/NRS	1b	-	3c	2b	1a	1a	1a	1a	M
*Aerometric and Emission Reporting System	AEROS	EPA	2b	1b	-	1a	-	-	-	-	C
*Agricultural On-Line Access	AGRICOLA	NAL/USDA	2b	2b	1a	1a	-	-	1a	-	C
Agricultural Research Service		ARS/USDA		2c	-	3c	-	-	-	-	C
*Air Pollution Technical Information Center	APTIC	EPA	2b	3b	-	2a	-	-	1a	1a	C
Air Quality Implementation Planning Program		EPA	3c	2b	-	2b	-	-	-	2b	C
American International Traders Index Register	AITR	DOC	2c	2b	2b	-	-	-	-	-	C
American Statistical Index	ASI	Cong.Info.Serv.	2b	-	-	-	-	-	-	-	C
Animal History Data System		FDA/HEW	3c	-	-	-	-	3c	-	-	C
*Annual Survey of Injuries and Illness		BLS/DOL	-	1b	-	-	1a	-	-	-	M
*Annual Survey of Manufacturers		DOC	2c	1b	3c	-	-	-	-	-	C
APILIT		Am.Petrol.Inst.	3c	2b	2b	2b	-	-	-	-	C
Army Chemical Information and Data System		Army/DOD	2b	-	-	-	-	-	-	-	C/M
Association of Data Base Producers	ADP	Asso. DBP	-	-	-	-	-	-	-	-	C
*Astro-4 Drug Information System		FDA/HEW	1a	1b	2a	2c	-	-	-	2a	C
*Atlas of Cancer Mortality		NCI/NIH/HEW	-	-	-	-	1b	-	-	-	M

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
*Biological Sciences Information Service	BIO-STORET	EPA	2b	-	-	-	-	3c	1a	-	C
*Biological Sciences Information Service	BIOSIS	Biosciences Info.Service	2b	-	-	2a	1a	1a	1a	-	C
*Biomedical Studies Group		EPA	2b	1a	1a	1a	1a	1a	1a	-	M
Bird Toxicity & Repellency Data Base		FWS/DOI	2b	-	3c	-	-	3c	-	-	C
Boston Collaborative Drug Surveillance Program		Boston Univ.	3c	-	-	2a	-	2b	-	-	C
*Cancer Information On-Line	CANCERLINE	NCI/NIH/HEW	2a	2b	-	1a	1a	1a	-	-	C
CANCERLIT		NCI/NIH/HEW	2b	-	-	-	-	2b	-	-	C
Cancer Projects	CANCERPROJ	NCI/NIH/HEW	2b	-	-	-	-	1a	-	-	C
Carbon-13 Nuclear Magnetic Reasonance Spectral Search System	SNMR	NIH/EPA	1a	-	-	-	-	-	-	-	C
Carcinogen Use Registry		NIH/HEW	2b	-	-	-	-	-	-	-	C
*Carcinogenesis Bioassay Data System	CBDS	NCI/NIH/HEW	1a	-	2c	2c	-	1a	-	-	C
Catalog of Information on Water Data		USGS/DOI	3c	-	-	-	-	-	2c	-	M
*Census Bureau Foreign Trade Statistics		Census/DOC	2b	1a	-	-	-	-	-	-	M
*Centre Information de Securite	CIS	Intl. Labor Office,Zurich	3c	-	-	-	2a	-	-	2b	M
*Census of Manufacturers		Census/DOC	2b	1a	-	-	-	-	-	-	C
CG-388 Chemical Data Guide for Bulk Shipment by Water		USCG/DOT	2c	-	-	3c	-	-	-	-	M
Chemical Abstracts Condensates	CA-CON	Amer.Chem.Soc.	2b	2b	-	-	-	3c	-	-	C

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
*Chemical Abstracts Service Chemical Registry System		Amer.Chem.Soc.	1a	-	-	-	-	-	-	_	M
*Chemical Abstracts Service Information System		Amer.Chem.Soc.	1a	2a	2b	2b	-	-	-	3b	C
Chemical-Biological Data Base for Herbicidal Information		Army/DOD	2a	-	-	-	-	3c	2b	-	C
Chemical Data Center		Chem.Data Ctr.	1b	2b	-	-	-	-	-	-	C
*Chemical Dictionary of the U.S. International Trade Commission		U.S. ITC	1b	1a	2b	-	-	-	-	1b	C
*Chemical Dictionary On-Line	CHEMLINE	NLM/NIH/HEW	1a	-	-	-	-	-	-	-	C
*Chemical Economics Handbook		SRI	2b	1a	1a	-	-	-	-	-	M
Chemical Hazard Response Information System	CHRIS	USCG/DOT	2b	-	-	3c	-	-	-	-	C/M
Chemical Industry Notes	CIN	Predicasts/ Chem.Abs.Serv.	-	-	1a	-	-	-	-	-	C
*Chemical Information Data System	CIDS	Army/DOD	1a	-	-	-	-	-	-	-	C
*Chemical Information System	CIS	NIH/EPA	1a	-	-	2b	-	-	-	-	C
*Chemical Monograph Referral Center	CHEMRiC	CPSC	1a	-	-	-	-	-	-	-	C
Chemical Mutagenesis: A Survey of the 1971 Literature	ORNL/EMIC-2	EMIC/ORNL	2b	-	-	-	-	2b	-	-	M
*Chemical Names File		NCI/NIH/HEW	1b	-	-	2b	-	1c	-	-	C
Chemical Structure Index	CSI	ISI	2a	-	-	-	-	-	-	-	M
Chemical Toxicological Data Retrieval System		FWS/DOI	4d	-	-	-	-	4c	-	-	C

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
*Chemical Transportation Emergency Center	CHEMTREC	Mfg.Chem.Asso.	1b	-	-	-	-	2b	2b	-	M
Chemistry and Effects of Biocides in Aquatic Systems		ESIC/ORNL	2b	-	-	2a	2b	2b	2a	-	C
Chemistry Data System		FDA/HEW	3c	-	-	2c	-	-	-	-	C
Chick Embryo System		FDA/HEW	3c	-	-	-	-	3b	-	-	C
*Clinical Toxicology of Commercial Products	CTCP	U. of Rochester	2b	2c	-	-	-	1b	-	-	C
Clintox Literature System		CDC/HEW	2c	-	-	-	2c	2c	-	-	C
Combination Chemotherapy Master File		NCI/NIH/HEW	3c	-	-	-	-	3c	-	-	C
Compendium of Toxicology		AFIP/DOD	2b	-	-	2a	-	2b	-	-	C
Compliance Data System		EPA	3c	3c	-	2c	-	-	-	2a	C
*Component Information for Chemical Consumer Products		CPSC	1a	3c	-	3b	-	-	-	-	C
Computerized Engineering Index	COMPENDEX	Eng. Index,Inc.	-	3b	-	-	-	-	-	-	C
Comprehensive Dissertation Index	CDI	Univ. Microfilms International	2b	-	-	2b	2b	2b	2b	-	C
Conformational Analysis of Molecules in Solution	CAMSEQ	NIH/EPA	1c	-	-	-	-	-	-	-	C
*Congressional Information Service Index	CIS Index	Cong.Info.Serv. Inc.	-	-	-	-	-	-	-	1a	C
*Congressional Record Abstracts	CRECORD	Capitol Services	-	-	-	-	-	-	-	1a	C
Cosmetics Information System		FDA/HEW	2b	2c	2c	-	-	2b	-	3b	C

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
Cosmetics, Toiletry and Fragrance Association Ingredient Dictionary		CFTA	2a	2b	2b	-	-	-	-	-	M
CPSC Chemical Abstracts		CPSC	1c	-	-	-	-	-	-	-	C
*Current Abstracts of Chemistry and Index Chemicus	CAC & IC	ISI	2b	2b	-	-	-	-	-	-	M
Current Employment Statistics		BLS/DOL	-	1b	-	1a	-	-	-	-	M
Current Energy Research Information Retrieval System	CERIRS	ANL/ERDA	-	-	-	-	-	-	-	-	C
*Current Industrial Reports		DOC	3c	1b	3c	-	-	-	-	-	C
*Data Base of U.S. International Trade Commission		U.S. ITC	2b	1a	1a	-	-	-	-	1b	C
Database on Teratogenic Effects of Selected Chemical Substances		Swedish Government	2b	-	-	-	-	2a	-	-	M
Data Bases for Energy Systems		LLL/ERDA	2a	-	3c	-	-	-	2a	-	C
Data Extraction & Analysis		ORNL/NLM	2a	2b	2a	2a	2a	2a	2a	2b	C
*Defense Documentation Center	DDC	DLA/DOD	1a	2c	-	-	-	-	1b	-	C
Diagnostics Subsystem		FDA/HEW	3c	-	-	2c	-	3c	-	-	C
*Directory of Chemical Producers	DCP	SRI	2b	1a	-	-	-	-	-	-	M
*Distribution Register of Organic Pollutants in Water	WATERDROP	EPA	2c	-	-	-	-	-	1a	-	M
Document Reference System		NCI/NIH/HEW	2b	-	-	2b	2a	2a	-	2b	C
Drug Distribution & Inventory System		NCI/NIH/HEW	2b	-	-	2b	2a	2a	-	2b	C
Drug Efficacy Study Implementation		FDA/HEW	2b	-	-	2c	-	2c	-	-	C

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
Drug Experience Information System		FDA/HEW	2b	-	-	2c	-	2c	-	-	C
Drug Experience Reports		FDA/HEW	3b	2b	-	3c	-	3c	-	-	C
*Drug Registration & Listing System		FDA/HEW	1a	2b	-	-	-	-	-	-	C
Drug Research & Development Biological Data		NCI/NIH/HEW	2b	-	-	-	-	3c	-	-	C
Drug Research & Development Chemical Information Bibliography File		NCI/NIH/HEW	2b	-	-	-	-	-	-	-	C
*Drug Research & Development Chemical Information System		NCI/NIH/HEW	1a	-	-	-	-	-	-	-	C
*Dun's Market Identifiers	DMI	Dun & Bradstreet	3c	2b	1b	1b	-	-	-	-	C
Effluent Data System	EDS	EPA	2c	2b	3c	2b	-	-	-	-	C
EIS Industrial Plants		Predicasts	-	1a	2b	-	-	-	-	-	C
Emissions Data System	EDS	EPA	3c	2b	-	-	-	-	2b	3c	C
Energy Data System		EPA	3d	3c	-	3c	-	-	-	-	C
Energy Information		ERDA	2a	-	-	2b	2b	2b	2a	-	C
Energy Line		Env.Info.Ctr.	-	-	2b	-	-	-	2b	2b	C
Energy Research and Development Inventory		ORNL/ERDA	-	-	-	-	-	2c	2c	-	C
Environmental and Health Aspects of Selected Organohalide Compounds		ERC/ORNL	3c	-	-	-	-	-	-	-	C
Environmental Chemical Data and Information Network	ECDIN	OECD	-	-	-	-	-	-	-	-	
Environmental Contaminant Evaluation Program		FWS/DOI	3b	-	-	2b	-	-	-	-	M
*Environmental Contaminant Monitoring Program		FWS/DOI	3b	-	-	2b	-	2c	1b	-	M

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
Environmental Data Index	ENDEX	NOAA/DOC	2b	-	-	1a	-	-	1a	-	C
Environmental Data System	EDS	NOAA/DOC	2b	-	-	2b	-	-	2a	-	C
Environmental, Health, and Control Aspects of Coal Conversion		ERC/ESIC/ORNL	3c	3c	-	-	-	3b	3c	-	C
Environmental Information System	EIS	Swedish CEI	3b	2b	-	2b	2b	-	2b	-	C
*Environmental Mutagen Information Center	EMIC	NIEHS/NIH/HEW	1b	-	-	-	-	1a	-	-	C
Environmental Mutagen Information Center Agent Registry File	EMICARD	EMIC/ORNL	2b	-	-	-	-	2b	2c	-	C
Environmental Pollution Effects on Aquatic Resources		NOAA/DOC	-	-	-	-	-	-	2a	-	M
*Environmental Reports Summaries		EPA	2c	-	-	-	-	-	-	1b	C
Environmental Residual Information System		EPA	3c	2b	2b	-	-	-	-	-	C
Environmental Resource Center		ORNL/ERDA	-	4c	-	2b	-	3b	2b	-	C
Environmental Science Information Center	ESIC	NOAA/DOC	3b	-	-	2a	-	3c	2b	-	C
*Environmental Teratology Information Center	ETIC	NIEHS/NIH/HEW	1c	-	-	-	-	1b	-	-	C
Environmental Teratology Information Center Agent Registry File	ETICARF	ETIC/ORNL	2b	-	-	-	-	2b	2c	-	C
EPA Reports System		NTIS/DOC	2b	2a	1b	1c	1a	1a	1a	-	C
Epidemiological Studies Program System		EPA	3c	-	-	-	2b	-	-	-	C
Establishment/Product Licensing System		FDA/HEW	2b	2b	-	-	-	3c	-	3b	C
Establishment Registration Support System	ERSS	EPA	2c	2a	-	-	-	3b	-	-	C

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
Excerpta Medica		Information System	2b	-	-	2c	2c	2a	-	-	M
Export Monitoring and Control System		OEA/DOC	3c	2c	2c	-	-	-	-	-	C
*Exposure Dictionary for National Occupational Hazards Survey	EDNOHS	NIOSH/HEW	1a	-	-	-	-	-	-	-	C
*Federal Inventory on Environmental and Safety Research		ERDA	2c	-	-	2b	2b	2a	1b	-	C
*Fish Control Laboratory - Data Base Information		FWS/DOI	2b	-	2c	-	-	2b	1a	-	M
*Fish-Pesticide Research		FWS/DOI	2b	-	-	2c	-	1b	1a	-	M
Food Information Storage and Retrieval		FDA/HEW	2b	-	-	-	-	2b	-	-	C
Foreign Trade of Member Countries of the OECD	Date Base	ERS/DOC	3b	2b	2b	-	-	-	-	-	C
Foreign Traders Index	FTI	DOC	3c	2c	2c	-	-	-	-	-	C
Fuel Additive Registration		EPA	2b	2b	2c	3b	3b	2b	3c	2b	M
Funk & Scott (F&S) Indexes		Predicasts	3c	1b	1a	-	-	-	-	-	C
Geophysical Monitoring for Climate Change		NOAA/DOC	-	-	-	-	-	-	2b	-	C
Global Environmental Monitoring System	GEMS	UNEP	3c	-	-	3c	-	2c	2b	-	M
Graphical Interactive NMR Analysis Program		NIH/EPA	3c	-	-	-	-	-	-	-	C
Great Lakes Environmental Contaminant Survey	GLECS	FWS/DOI	3c	-	-	3c	-	-	2b	-	C
Great Lakes Fishery Information		EPA	3c	-	-	-	-	2b	2a	-	C
Hazardous and Trace Emissions System	HATREMS	EPA	2b	-	-	1b	-	-	-	-	C

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
*Health Hazard Evaluations		PHS/CDC/HEW	4d	2a	2a	2b	2a	3c	-	1b	M
Heavy Metals		TVA	3c	-	-	2c	-	2b	2b	-	M
Heavy Metals and Related Trace Elements in Aquatic Environments		ERC/ORNL	3b	-	-	3c	-	2b	2b	-	C
*Index Chemicals Registry System	ICRS	ISI	1a	-	-	-	-	-	-	-	C
Industrial Hygiene Automated Data System		TVA	2b	2b	3c	1c	2b	3b	-	2a	C
Industry Surveys		Standard & Poor's	4d	-	3c	-	-	-	-	3c	M
*Industrywide Studies		NIOSH/CDC/HEW	3b	3a	3a	1b	1b	3b	-	2b	M
Information Analysis Centers		DSA/DOD	-	-	-	-	-	-	-	-	M
Information and Documentation System for Environmental Planning	UMPLIS										
*Information Bulletin of the Survey of Chemicals Being Tested for Carcinogenicity		WHO	1b	-	-	-	-	1b	-	-	M
Information Center for Energy Safety		ORNL/ERDA	2c	-	-	-	-	-	2b	2b	C
*Information Storage and Referral Section		NIEHS/NIH/HEW	-	-	2b	2b	1b	1a	-	-	C
*Inorganic Chemical Computer Toxicology Parameter Data Base		EPA	3b	1a	-	-	-	1b	1a	-	C
INSPEC Science Abstracts		Inst. of Elec. Engineers, U.K.	2b	2b	-	-	-	-	-	-	C
*International Cancer Epidemiology Clearing House		ICRDB/IARC/CCR	2b	-	-	-	1a	1b	-	-	C

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
International Classification of Diseases for Oncology	ICD-O	WHO	-	-	-	-	-	2b	-	-	M
International Joint Commission Coordinated Program on Fish Contaminants		FWS/DOI	-	-	-	-	-	-	2b	-	-
International Referral System for Sources of Environmental Information		UNEP	-	-	-	-	-	-	-	-	M
International Registry of Potentially Toxic Substances		UNEP	-	-	-	-	-	-	-	-	C
Investigational New Animal Drug Index		FDA/HEW	2b	2b	3c	2b	-	-	-	-	C
Iowa Drug Information Service	IDIS	U. of Iowa	2b	-	-	-	-	2b	-	-	M
*IPC Chemical Data Base		IPC Industrial Press, U.K.	2b	1b	1a	-	-	-	-	-	C
Isotopic Label Incorporation Determination		NIH/EPA	3c	-	-	-	-	-	-	-	C
*Kirk-Othmer Encyclopedia of Chemical Technology		Interscience Publishers	2b	1a	1b	-	-	-	-	-	M
Laboratory Analysis Data Base		CPSC	2c	-	-	-	-	2b	-	2b	C
*Laboratory Animal Data Base	LADB	NIH/HEW	2c	-	-	-	-	1a	-	-	C
Laboratory Management System		EPA	3c	-	-	3c	-	-	-	-	C
Lower Lakes Reference Group		FWS/DOI	3c	-	-	2b	-	3c	2a	-	M
*Mammal Toxicity and Repellency Data Base		FWS/DOI	2c	-	3c	-	-	1b	-	-	C
Marine Ecosystem Analysis Program	MESA	NOAA/DOC	2b	-	-	2b	-	1a	1a	-	C
Mass Spectrometry Data Centre		Atomic Weapons Res.Estab.U.K.	1a	-	-	-	-	-	-	-	C

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
Mass Spectrometry Bulletin Search		NIH/EPA	2a	-	-	-	-	-	-	-	C
Mass Spectral Identification		NIH/EPA	2a	-	-	-	-	-	-	-	C
Mass Spectral Search System		NIH/EPA	1a	-	-	-	-	-	-	-	C
Masters Theses in the Pure and Applied Sciences		Plenum Publ.	3c	-	-	3c	3c	3c	3c	-	M
*Meat & Poultry Inspection Monitoring Program		PAHIS/USDA	2b	-	-	1b	-	-	-	2b	M
*Medical Literature Analysis and Retrieval System On-Line	MEDLINE	NLM/NIH/HEW	2a	-	-	-	1a	1a	-	-	C
Medical Subject Headings Vocabulary	MESH	NLM/NIH/HEW	1c	-	-	-	-	-	-	-	C
*The Merck Index Text Editing System		Merck	1b	-	-	-	-	2b	-	-	C
Michigan Dept. of Natural Resources Fisheries Division		Michigan State	2c	-	-	2c	-	3d	3c	-	M
*Microconstituents in Fish and Fishery Products		NOAA/DOC	3c	-	-	2b	-	-	3b	-	C
MI-KOM Environmental Information Services		Swedish CEI	-	-	-	-	-	-	-	-	M
*Military Entomology Information Service	MEIS	Army/DOD	2b	-	-	2a	1a	1a	1a	-	C
*Mineral Commodity Survey System		BOM/DOI	1b	1b	1b	-	-	-	-	-	C
Multilateral Trade Negotiations Data Base	MTNDB	DOC	2c	1b	1b	-	-	-	-	-	C
Multistation Atmospheric Pollution from Power Production Study	MAPPPS	ERDA/NOAA	3d	-	-	-	-	-	3c	-	C
The Mutagenicity and Teratogenicity of a Selected Number of Food Additives	ORNL-EMIC-1	EMIC/ORNL	2b	-	-	3b	-	2b	-	-	M

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
*NASA Scientific and Technical Information System		NASA	1b	-	-	2b	-	2b	2b	-	C
National Air Surveillance Network	NASN	EPA	2b	-	-	1b	-	-	-	-	C
National Cancer Institute (NCI) Carcinogenesis Program File		NCI/NIH/HEW	3c	-	-	3c	-	2b	3c	-	C
*National Center for Health Statistics	NCHS	HEW	-	-	-	-	1a	-	-	-	C
*National Center for Toxicological Research (NCTR) Integrated Research Support System		FDA/NCTR	3c	-	-	-	-	1b	-	-	C
National Clearinghouse for Mental Health Information		NIMH/NIH/HEW	2b	-	-	-	-	2c	-	-	C
*National Electronic Injury Surveillance System	NEISS	CPSC	2b	-	-	1a	1a	-	-	2b	C
National Emissions Data	NEDS	EPA	2b	1b	2b	2a	-	-	-	-	C
National Fire Data Center		DOC	3c	-	-	-	-	3c	-	-	M
National Index of Energy and Environmental Related Data		ERDA	3d	-	-	3c	-	2b	3b	-	C
National Index of Energy and Environmental Related Models		ERDA	4d	-	-	3b	3b	3b	3b	-	C
*National Occupational Hazard Survey File	NOHS	NIOSH/CDC/HEW	2b	2b	1a	1a	-	-	-	-	C
National Park Service (NPS) Pest Control System		NPS/DOI	3c	3c	3b	4d	-	-	3c	-	C
National Pollutant Discharge Elimination System	NPDES	EPA	3c	2b	-	3c	-	-	-	2b	C
National Referral Center		Library of Congress	3b	-	-	-	-	-	-	-	C

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
*National Technical Information Service	NTIS	DOC	2b	2a	1a	1a	1a	1a	1a	1a	M
National Water Data Exchange	NAWDEX	USGS/DOI	3c	-	-	-	-	-	2c	-	M
Navy Environmental Protection Support Services		Navy/DOD	2b	-	2b	2b	-	2b	-	-	C
Nevada Applied Ecology Information Center		ERDA	2b	3c	-	2a	-	-	1b	-	C
New Animal Drug Applications		FDA/HEW	2b	2b	3v	2b	-	-	-	-	C
New York Times Information Bank		New York Times	3c	3c	3c	3c	3c	3c	2c	2b	C
*NIOSH Technical Information Center	NIOSHTIC	NIOSH/HEW	3c	2b	3c	3b	1a	1a	4d	-	C
Occupational Safety and Health		OSHA/DOL	3c	2c	-	3c	2c	-	3c	2b	C
*Oceanic Abstracts		Data Courier, Inc.	2b	-	-	2b	-	-	1a	-	C
*Oceanic and Atmospheric Scientific Information Service	OASIS	NOAA/DOC	2c	-	-	2b	-	-	1b	-	C
*Office of Standard Reference Data Chemical Files		NBS/DOC	1b	4d	-	-	-	-	-	-	C
*Oil & Hazardous Materials Technical Data System	OHM-TADS	EPA	2a	2b	2b	2b	-	1b	2a	-	C
*Organic Chemical Producers Data Base		EPA	2c	1a	2c	2c	-	1b	-	-	C
PaperChem		Inst. of Paper Chemistry	3c	2b	2b	-	-	-	-	-	C
Parklawn Health Library	Kwik Index	PHS/HEW	3c	-	-	-	-	2b	-	-	C
Pathology Data System		FDA/HEW	3c	-	-	-	-	2c	-	-	C
Permit Compliance System (Water)		EPA	3c	3c	-	2c	-	-	-	-	C

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
P/E News		Amer. Petrol Institute	3c	2b	2b	-	-	2b	-	-	C
*Pesticide and Industrial Chemicals		FDA/HEW	1b	-	-	-	-	3c	-	-	C
*Pesticide Enforcement Management System	PEMS	EPA	2b	2b	2b	-	-	-	-	1b	C
Pesticide Import File Region X		EPA	2b	2b	2c	3b	-	-	-	2b	M
Pesticide Registration Systems	(now PARCS)	EPA	2b	2c	-	-	-	-	-	-	C
*Pesticide Reporting System		FDA/HEW	2d	2c	3c	2b	-	-	-	1a	C
Pesticide Sampling Information System - Region X		EPA	2c	2c	2c	2b	-	-	-	2b	C
*Pesticides Analysis Retrieval and Control System	PARCS	OPM/EPA	1a	1a	1b	-	-	2b	-	-	C
Pharmaceutical News Index	PNI	Data Courier, Inc.	-	2b	2b	-	-	-	-	2a	C
Pilot Data Base for Hazardous Substances		CPSC	2c	-	-	-	-	2b	-	2b	C
*POISINDEX		Micromedex	1b	-	-	-	-	1b	-	-	M
Poison Control Centres of Canada		Consumer & Corp Affairs, Canadian Govt.	2a	-	-	3c	2c	2a	-	-	C/M
*Poison Control Online Inquiry System		FDA/HEW	1b	-	-	-	1b	1b	-	-	C
*Pollution Abstracts		Data Courier, Inc.	-	-	-	2b	-	-	1a	-	C
*Population Studies System		EPA	2c	-	-	1a	1a	1b	-	-	C
Predicasts Domestic Statistics		Predicasts	3c	2b	1a	-	-	-	-	-	C

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE I	II	III	IV	V	VI	VII	VIII	(C)(M)
Predicasts Federal Index		Predicasts	-	-	-	-	-	-	-	2b	C
Predicasts International Statistics		Predicasts	3c	2b	1a	-	-	-	-	-	C
Predicasts Market Abstracts		Predicasts	-	2a	1a	-	-	-	-	-	C
*Predicasts Marketing Systems		Predicasts	3c	2a	1a	-	-	-	-	-	C
Product Safety Indexed Document Collection		CPSC	-	3b	2b	3b	3b	3b	-	2b	C
Program for Toxicology of Combustion Products		NBS/DOC	2a	2b	-	-	-	2c	2a	-	C
Proton Affinity Retrieval		NIH/EPA	1b	-	-	-	-	-	-	-	C
Psychological Abstracts		Am. Psych. Assn.	-	-	-	-	-	3d	-	-	C
*Registry of Toxic Effects of Chemical Substances	RTECS	CDC/HEW	1a	-	-	4d	-	1a	2b	1a	M
*Reporting of Economic Data for Negotiation of International Transportation Conventions	REDNITRAC	DOC	2b	1a	1a	-	-	-	-	-	C
Research Information Services for the Agricultural Sciences		SSIE	2b	-	-	-	-	3c	3b	-	C
Research Materials Information Center		ORNL/ERDA	2a	-	-	-	-	-	-	-	M
*Research Program of Chemicals That Impact Man		NCI/NIH/HEW	1a	1a	1a	1b	1a	1a	1a	-	C
Retirement History Study		OPP/HEW	-	-	-	2c	2b	2c	-	-	C
RINGDOC		Derwent Publ.	2b	-	-	-	-	2b	-	-	C
Science & Technical Division		Lib. of Cong.	2c	-	-	-	-	-	-	2b	C
Science Citation Search	SCISEARCH	ISI	2b	-	-	2b	2c	1b	-	-	C
Scientific Manuscript Bibliographic System		FDA/HEW	3c	-	-	4d	-	4d	-	-	C

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
Scientific Reference Services Branch		CDC/HEW	4c	-	-	-	-	2b	2b	-	M
Selective Dissemination of Information On-Line	SDILINE	NLM/NIH/HEW	2b	-	-	-	1b	1b	-	-	C
Single Drug Master File		NCI/NIH/HEW	3c	-	-	-	-	3b	-	-	c
*Smithsonian Scientific Information Exchange	SSIE	SSIE	2b	-	-	1a	-	2b	1a	-	M
Soil, Water, Estuarine Monitoring System	SWEMS	EPA	2a	-	-	2b	2a	2a	2a	-	M
*Solid Waste Information Retrieval System	SWIRS	EPA	2b	-	-	-	-	-	1b	-	C
Special Reports - Grant Supported Literature Index	GENIUS	NCI/NIH/HEW	2b	-	-	-	2a	2a	-	-	C
*Special Trade Representatives Centralized Data Bank	STRCDB	Off. of Spec. Representative for Trade Neg.	2b	1a	1b	-	-	-	-	-	C
*Standards Completion Program		NIOSH/CDC/HEW	4d	-	-	1b	1b	4d	-	1b	C
State Implementation Plans	SIPS	EPA	3c	-	-	2c	-	-	-	-	C
Statistical Center for the Tyler Texas Asbestos		NCI/NIH/HEW	4d	-	-	4d	4d	4d	-	-	C
Strategic Environmental Assessment System	SEAS	EPA	4d	-	-	-	-	-	2c	-	C
*Storage and Retrieval for Water Quality Data	STORET	EPA	2b	-	-	1a	-	-	-	-	C
Storage and Retrieval of Aerometric Data	SAROAD	EPA	2b	-	-	2a	-	-	-	-	C
*Subject Content Oriented Retriever for Processing Information On-Line	SCORPIO	Lib. of Cong.	2c	-	-	-	-	-	-	2b	C
Substructure Searching System	CIS-SSS	NIH/EPA	1b	-	-	-	-	-	-	-	C

TABLE 3-1 (Cont'd.) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
*Supplementary Data System		BLS/DOL	2b	-	-	1b	1a	1a	-	-	C
*Survey of Compounds Which Have Been Tested for Carcinogenic Activity	PHS-149	NCI/PHS/HEW	2b	-	-	-	-	1b	-	-	M
Swedish Register of Environmental Research		Swedish CEI	3c	-	-	2b	2b	2b	2b	-	C
*Technical Data Center	TDC	OSHA/DOL	1b	-	-	1a	1a	1a	-	1b	M
Technical Files		TVA	2b	3c	3c	2b	2a	2a	3c	2b	M
Technical Library Information Office		TVA	-	-	-	-	-	-	-	-	M
The Environment Information Retrieval System	TEIRS	Army/DOD	2b	-	-	-	-	-	-	-	C
*Thermophysical Properties Research Center		Purdue U.	1a	-	-	-	-	-	-	-	M
Toxic Materials Information Center		ERDA/NSF	2b	2b	3c	2a	2a	2a	2a	-	C
Toxic Substances Information Act		Virginia State	2b	3b	-	-	-	-	-	-	C/M
Toxicological Studies		NIOSH/CDC/HEW	2a	2c	2b	2b	-	2a	-	2b	M
*Toxicology Data Bank	TDB	NLM/NIH/HEW	1a	1a	2a	2a	1a	1a	2a	-	C
*Toxicology Information On-Line	TOXLINE	NLM/NIH/HEW	2b	2b	2b	1a	1a	1a	1a	-	C
Toxicology Information Response Center	TIRC	ERDA	2b	-	-	2b	2b	2b	2b	-	C
Toxicology Research Projects Directory		NLM/NIH/HEW	2b	-	-	1a	1a	1a	1a	-	M
*Toxicology Testing In Progress	TOX-TIPS	NLM/NIH/HEW	2a	-	-	2a	2a	1a	2b	-	C
Toxline Backfile	TOXBACK	NLM/NIH/HEW	2b	2b	2b	1a	1b	1a	1a	-	C
Trace Contaminants Abstracts	TCA	TMIC/ORNL	3c	-	-	4d	-	3c	3c	-	M

TABLE 3-1 (Concluded) DATA SYSTEM SCORING

SYSTEM	ACRONYM	OWNER	DATA TYPE								(C)(M)
			I	II	III	IV	V	VI	VII	VIII	
*Trade Name Ingredient Clarification	TNIC	CDC/HEW	1a	2c	-	-	-	-	-	-	C
Upper Lakes Reference Group		FWS/DOI	3c	-	-	2b	-	3c	2a	-	M
USDA-ERS Use of Pesticides		ERS/USDA	3c	-	2b	3c	-	-	-	-	M
VIOLOG		EPA	3c	2c	-	3c	-	-	2b	2b	C
Walter Reed Army Institute of Research Biological Data System		Army/DOD	2b	-	-	-	-	2a	-	-	C
Walter Reed Army Institute of Research Chemical Inventory System		Army/DOD	2b	-	-	-	-	-	-	-	C
Walter Reed Army Institute of Research Chemical Structure System		Army/DOD	2a	-	-	-	-	-	-	-	C
Walter Reed Army Institute of Research Index File		Army/DOD	2a	-	-	-	-	-	-	-	C
Water Quality Data Base		TVA	2b	-	-	2b	-	-	-	-	C
Water Resources Scientific Information Center	WRSIC	DOI	2b	-	-	2a	-	2b	2b	2a	C
Water Storage Data and Retrieval System	WATSTORE	USGS/DOI	2b	-	-	2b	-	-	-	-	C
X-Ray Crystal Data Retrieval System		NIH/EPA	1a	-	-	-	-	-	-	-	C
X-Ray Crystal Structure Retrieval System		NIH/EPA	1a	-	-	-	-	-	-	-	C
X-Ray Powder Diffraction Retrieval System		NIH/EPA	1a	-	-	-	-	-	-	-	C

This narrowed the list of Federal and private files under consideration to 100. These primary files will be used in the second stage of the effort under this contract which calls for recommending a basic methodology for accessing and linking existing toxic substances information and the identification of new files needed. Due to the short time frame of this project, the initial survey of potentially useful data files could not be exhaustive.

3.3 Characterization of Selected Systems

When selecting files for inclusion in an information system, it is necessary to compare those systems containing data in similar subject areas. Tables 3-2 through 3-9 contain descriptions of the selected data files by subject category. The eight information categories used in Table 2-1 are broken down into subcategories to permit a rapid comparison of these systems containing data in each data category. The primary systems designated on Table 3-1 by a "1a" or "1b" in a given data category are included on these category-specific tables. A column is also included for comments. If a more in-depth comparison is desired, all primary data systems are described in detail in Appendix D.

In addition, the primary systems were examined to determine (1) whether data were generated internally or whether they were merely compiled from external sources of information, (2) if they contain proprietary information, and (3) if they are collected as a result of a mandatory solicitation. This information is included in Table 3-10.

TABLE 3-2
DATA SYSTEMS APPLICABLE TO SUBSTANCE IDENTIFICATION

SYSTEM	CAS #	NOMENCLATURE	STRUCTURE	CHEMICAL/PHYSICAL PROPERTIES	COMPOSITION (INCLUDING IMPURITIES)	CHEMICAL ANALYSIS*	BIBLIOGRAPHIC (TOTALLY)	COMMENTS
Advisory Center on Toxicology	X	X		X	X		X	Manual card file
Astro-4 Drug Information System	X	X	WLN	X	X			Drug production and registration information
Carcinogenesis Bioassay Data System	X	X	WLN	X	X			Lab experiment data
Chemical Abstracts Service Chemical Registry System	X	X	X					
Chemical Abstracts Service Information System				X			X	
Chemical Dictionary of the U.S. ITC	X	X					X	Tariff information
Chemical Dictionary On-Line	X	X	WLN					
Chemical Information & Data System	X	X	X					CIDs registration system
Chemical Information System	X	X	WLN SSS	X				Also x-ray CNMRs and Mass spec.
Chemical Monograph Referral Center		X		X				Referral system to monographs with these data
Chemical Names File	X	X	WLN					Compounds tested for carcinogenicity
Chemical Transportation Emergency Center		X		X				File used in case of accidental spills
Component Information for Chemical Consumer Products	X	X			X			Formulation of 15,000 products to 0.1% level
Defense Documentation Center	X	X		X	X		X	
Drug Registration and Listing System	X	X		X	X		X	
Drug Research & Development Chemical Information System	X	X	SSS	X				

WLN = Wiswesser Line Notation
SSS = Substructure Searching

* Not on original questionnaire

TABLE 3-2 (CONCLUDED)

SYSTEM	CAS #	NOMENCLATURE	STRUCTURE	CHEMICAL/PHYSICAL PROPERTIES	COMPOSITION (INCLUDING IMPURITIES)	CHEMICAL ANALYSIS*	BIBLIOGRAPHIC (TOTALLY)	COMMENTS
Environmental Mutagen Information Center	X	X	WLN				X	May be expanded for data sources of mutagen information
Exposure Dictionary for NOHS	X	X			X			12,000 chemical names
Index Chemicals Registry System		X	WLN					
Information Bulletin of the Survey of Chemicals Being Tested for Carcinogenicity		X						
IPC Chemical Data Base		X						Imports/Exports
Mineral Commodity Survey System		X			X			Survey of mineral industry
NASA Scientific and Technical Information Center		X		X			X	Environmental information
Office of Standard Reference Data Chemical Files		X		X				
Pesticides and Industrial Chemicals		X		X	X	X		Pesticide chemistry
Pesticides Analysis Retrieval and Control System		X			X			New system, use and formulation
Poison Control On-Line Inquiry System		X			X			Contains 10,000 household products and drugs
POISINDEX		X			X			Contains 160,000 entries
Registry of Toxic Effects of Chemical Substances	X	X	WLN					Basic toxicology of 22,000 chemicals
Research Program of Chemicals that Impact Man	X	X	X		X			(3,200 chemicals by SRI)
Technical Data Center	X	X					X	Documentation on occupational safety and health
Thermophysical Properties Research Center				X				
Toxicology Data Bank	X	X	WLN	X	X		X	New system, on-line access to toxicology data
Trade Name Ingredient Clarification		X			X			

* Not on original questionnaire

TABLE 3-3

DATA SYSTEMS APPLICABLE TO PRODUCTION

SYSTEM	MANUFACTURER*	PRODUCTION QUANTITY	PLANT LOCATION	PRODUCTION PROCESS	BY-PRODUCTS/IMPURITIES	CONTROL TECHNOLOGY	BIBLIOGRAPHIC ONLY	COMMENTS
Aerometric and Emission Reporting System	X	X	X	X				NEDS
Annual Survey of Injuries and Illnesses	X	X	X					All establishments >111 employees by SIC code
Annual Survey of Manufacturers	X	X	X					By SIC code
Astro-4 Drug Information System	X	X						Drug producers and amounts
Biomedical Studies Group	X	X	X	X	X	X		For 14 compounds
Census Bureau Foreign Trade Statisticx		X						Imports/exports
Census of Manufacturers	X	X	X					By SIC code
Chemical Economics Handbook	X	X	X					
Current Industrial Reports	X	X	X					By SIC code
Data Base of the U.S. International Trade Commission	X	X						Manufacturers and importers in summary form
Directory of Chemical Producers	X	X	X					Manual
Employment and Earnings	X		X					Size of workforce
Inorganic Chemical Computer Toxicology Parameter Data Base			X	X	X	X		172 inorganics (new system)
IPC Chemical Data Base	X	X	X					Imports/exports on 100 chemicals
Kirk-Othmer Encyclopedia of Chemical Technology				X	X	X		Manual
Mineral Commodity Survey System	X	X	X					200 mineral industries
Multilateral Trade Negotiations Data Base		X						Imports/exports
Organic Chemical Producers Data Base	X	X	X	X	X	X		400 chemicals
Pesticides Analysis Retrieval and Control System	X	X						Formulation information by producer
Predicasts Marketing Systems	X	X	X	X		X		F & S, FIS of Predicast
Reporting of Economic Data for Negotiation of International Transportation Conventions		X						Import/export
Research Program of Chemicals That Impact Man	X	X	X		X			On 3,200 chemicals (SRI)
Special Trade Representatives Centralized Data Bank		X						Imports/exports
Toxicology Data Bank	X	X	X	X				1,000 chemicals (new system)

*Not on original questionnaire

TABLE 3-4
DATA SYSTEMS APPLICABLE TO MARKETING

SYSTEM	USAGE	USERS	SUBSTITUTES	ECONOMICS	PLACE OF USE	BIBLIOGRAPHIC ONLY	COMMENTS
Agricultural On-Line Access	X	X		X		X	Agricultural chemicals
Biomedical Studies Group	X	X	X	X	X		On 14 chemicals
Chemical Economics Handbook	X			X			Manual
Data Base of U.S. International Trade Commission	X			X			8,000 chemicals - some manufacture, some imports
Dun's Market Identifiers				X			
IPC Chemical Data Base				X			Import/export on 100 chemicals
Kirk-Othmer Encyclopedia of Chemical Technology	X		X				
Mineral Commodity Survey System	X	X	X	X			Survey of 200 industries
National Occupational Hazard Survey	X				X		Workplace uses
National Technical Information Service	X	X	X	X	X	X	Government reports
Pesticide Analysis Retrieval and Control System	X		X				Pesticides
Predicasts Marketing Systems	X	X	X	X	X	X	All systems
Reporting of Economic Data for Negotiation of International Transportation Conventions				X			Imports/exports
Research Program of Chemicals that Impact Man	X	X	X	X	X		SRI file on 3,200 chemicals
Special Trade Representatives Centralized Data Bank				X			Import/export

TABLE 3-5

DATA SYSTEMS APPLICABLE TO EXPOSURE

SYSTEM	OCCUPATIONAL	CONSUMER	ENVIRONMENTAL	MONITORING*	BIBLIOGRAPHIC ONLY	AIR	WATER	COMMENTS
Aerometric and Emission Reporting System	X		X	X		X		Includes NEDS, SAROAD, HATREMS, EDS, NASN
Agricultural On-Line Access			X		X			
Biomedical Studies Group	X	X	X			X	X	14 chemicals only
Cancer Information On-Line	X	X	X		X	X	X	
Current Employment Statistics	X							
Dun's Market Identifiers	X							
Industrywide Studies	X							100 occupational studies
Meat and Poultry Inspection Monitoring Program		X						Levels of pesticides, drugs, metals and residues
National Electronic Injury Surveillance System		X		X				Emergency room injuries associated with consumer products
National Occupational Hazard Survey	X			X				
National Technical Information Service	X	X	X	X	X	X	X	Government reports
Oceanic and Atmospheric Scientific Information Service			X	X		X	X	ENDEX
Population Studies Program				X				
Research Program of Chemicals that Impact Man	X	X	X			X	X	3,200 chemicals - SRI file
Smithsonian Scientific Information Exchange	X	X	X	X	X	X	X	Research in progress
Standards Completion Program	X							
Storage and Retrieval for Water Quality Data			X	X			X	400 chemicals - includes WATSTORE, ECMS, NPDES, LAM
Supplementary Data Center	X							
Technical Data Center	X				X			5,000 chemicals
Toxicology Information On-Line	X	X	X	X	X			Including Toxback

*not included on original questionnaire

TABLE 3-6

DATA SYSTEMS APPLICABLE TO EPIDEMIOLOGY

SYSTEM	OCCUPATIONAL	GENERAL	BIBLIOGRAPHIC ONLY	COMMENTS
Advisory Center on Toxicology	X	X		Manual (minimal added data)
Annual Survey of Injuries and Illnesses	X			BLS biannual survey
Atlas of Cancer Mortality		X		
Biomedical Studies Group	X	X		14 Chemicals
Biological Sciences Information Service	X	X	X	
Cancer Information On-Line	X	X	X	
Industrywide Studies	X			100 studies performed by NIOSH
Information Storage and Referral Section	X		X	New system
International Cancer Epidemiology Clearinghouse	X	X		
Medical Literature Analysis and Retrieval System On-Line	X	X	X	Includes SDILINE
Military Entomology Information Service	X	X	X	
National Center for Health Statistics		X		Baseline information
National Electronic Injury Surveillance System		X		Consumer epidemiology
National Occupational Hazard Survey File	X			Plant profiles
National Technical Information Service	X	X	X	Government reports
NIOSH Technical Information Center	X	X	X	8,000 chemicals
Poison Control On-Line Inquiry System		X		Procuring incidence reports
Population Studies System	X			CHESS
Research Program of Chemicals that Impact Man	X	X		SRI
Standards Completion Program	X			Surveillance re. 400 chemicals with standards
Supplementary Data System	X			State Unemployment Insurance Records
Technical Data Center	X	X	X	OSHA Data Bank
Toxicology Data Bank	X	X	X	New system - now covers 1,000 chemicals and drugs
Toxicology Information On-Line	X	X	X	Including TOXBACK

TABLE 3-7

DATA SYSTEMS APPLICABLE TO BIOLOGICAL EFFECTS

SYSTEM	CLINICAL STUDIES	ACUTE TOXICOLOGY	CARCINOGENICITY	MUTAGENICITY	TERATOGENICITY	METABOLISM*	TESTING METHODOLOGY*	BIBLIOGRAPHIC ONLY	COMMENTS
Advisory Center on Toxicology		X	X	X	X	X			Manual card index
Biomedical Studies Group	X	X	X	X	X	X			14 chemicals
Biological Sciences Information Service	X	X	X	X	X	X		X	
Cancer Information On-Line	X	X	X	X	X		X	X	Cancerproj.
Carcinogenesis Bioassay Data System	X	X	X				X		
Clinical Toxicology of Commercial Products	X	X							20,000 trade names with toxicity
Environmental Mutagen Information Center			X	X				X	
Environmental Teratology Information Center					X			X	
Fish Pesticide Research		X					X		
Information Bulletin of the Survey of Chemicals Being Tested for Carcinogenicity			X				X		
Information Storage and Referral Section	X	X	X	X	X	X		X	New system
International Cancer Epidemiology Clearing House			X				X		
Laboratory Animal Data Base		X	X				X		50,000 animals
Mammal Toxicity and Repellency Data Base		X	X						
Medical Literature Analysis and Retrieval System On-Line	X	X	X	X	X	X		X	
Inorganic Chemical Computer Toxicology Parameter Data Base		X							172 Inorganics

*not included on original questionnaire

TABLE 3-7 (CONCLUDED)

SYSTEM	CLINICAL STUDIES	ACUTE TOXICOLOGY	CARCINOGENICITY	MUTAGENICITY	TERATOGENICITY	METABOLISM*	TESTING METHODOLOGY*	BIBLIOGRAPHIC ONLY	COMMENTS
Military Entomology Information Service	X	X	X	X	X			X	
National Technical Information Service	X	X	X	X	X	X	X	X	Govt. Reports
National Center for Toxicology Experiment Integrated Research Support System		X	X				X		
NIOSH Technical Information Center	X	X	X	X	X	X	X	X	
Oceanic and Atmoshperic Scientific Information Service		X					X		MESA sub
Oil and Hazardous Materials		X	X	X	X				
Organic Chemical Producers Data Base		X	X						400 Chemicals
POISINDEX		X							
Poison Control On Line Inquiry System	X	X							10,000 products
Population Studies Program									
Registry of Toxic Effects of Chemical Substances	X	X	X	X	X				
Research Program of Chemicals That Impact Man	X	X	X	X	X				
Smithsonian Scientific Information Exchange	X	X	X	X	X	X	X	X	TRPD Subset
Supplementary Data Base		X							Injuries, Illness
Survey of Compounds That Have Been Tested for Carcinogenicity		X	X						
Technical Data Center	X	X				X		X	
Toxicology Data Bank	X	X	X	X	X	X		X	
Toxicology Information On-Line	X	X	X	X	X	X	X	X	Including TOXBACK
Toxicology Testing in Progress		X	X				X		

TABLE 3-8
DATA SYSTEMS APPLICABLE TO ENVIRONMENTAL EFFECTS

SYSTEM	BIOACCUMULATION	ECOLOGICAL EFFECTS	PHYSICAL EFFECTS	DEGREDATION	MONITORING AND ANALYSIS TECHNOLOGY*	BIBLIOGRAPHIC ONLY	COMMENTS
Advisory Center on Toxicology		X		X			Manual file
Agricultural On-Line Access	X	X		X		X	
Air Pollution Technical Information Center	X	X	X	X	X	X	
Biomedical Studies Group	X	X	X	X	X		For 14 chemicals
Biological Sciences Information Service	X	X	X			X	
Biological Data Storage and Retrieval System	X	X					Biological effects of water quality(new system)
Defense Documentation Center				X		X	
Distribution Register of Organic Compounds in Water					X		New system
Environmental Contaminant Monitoring System				X	X		Fish bioaccumulation studies
Federal Inventory on Environmental Safety and Health Research	X	X	X	X	X	X	2,466 projects
Fish Control Laboratory Data Base Information	X	X		X			1,500 chemicals in 8 species manual
Fish Pesticide Research	X	X		X			500 chemicals in 100 species manual
Military Entomology Information Center	X	X		X		X	
National Technical Information Service	X	X	X	X	X	X	Government reports
Oceanic Abstracts	X	X		X		X	
Oceanic and Atmospheric Scientific Information Service	X	X		X	X		
Pollution		X				X	
Research Program of Chemicals that Impact Man	X	X		X			SRI file of 3,200 chemicals
Smithsonian Scientific Information Exchange	X	X	X	X	X	X	Research in progress
Solid Waste Information Retrieval System				X		X	
Toxicology Information On-Line	X	X		X		X	
Inorganic Chemical Computer Toxicology Parameter Data Base		X		X			172 inorganics (new system)

* Not on original questionnaire

TABLE 3-9
DATA SYSTEMS APPLICABLE TO STANDARDS AND REGULATIONS

SYSTEM	FEDERAL	STATE	LOCAL	INTERNATIONAL	BIBLIOGRAPHIC ONLY	COMMENTS
Advisory Center on Toxicology	X	X				Manual refers to allowable concentration
Air Pollution Technical Information Center	X	X	X		X	
Congressional Record Abstracts	X				X	Same as below
Congressional Information Service Inc.	X				X	All Congressional publications
Data Base of U.S. International Trade Commission	X					
Environmental Reports Summaries				X		Foreign regulations
Health Hazard Evaluations	X					Compliance file
National Technical Information Service	X	X			X	Government reports
Pesticides Enforcement Management System	X					Support to pesticide monitoring and enforcement
Pesticide Reporting System	X	X				Sampling results
Registry of Toxic Effects of Chemical Substances	X					NIOSH testing of TWAs
Standards Completion Program	X					400 compounds with TLVs
Technical Data Center	X					OSHA - File of existing standards - 20,000 chemicals, 400 compounds with TLVs

TABLE 3-10

SOURCE OF DATA AND THE PROPRIETARY STATUS OF THE PRIMARY SYSTEMS

PRIMARY SYSTEMS	ACRONYM	INTERNALLY GENERATED DATA	EXTERNALLY GENERATED DATA	PROPRIETARY INFORMATION	MANDATORY SOLICITATION DATA
Advisory Center on Toxicology		X	X	X	
Aerometric and Emission Reporting System	AEROS	X	X		
Agricultural On-Line Access	AGRICOLA		X		
Air Pollution Technical Information Center	APTIC		X		
Annual Survey of Injuries and Illnesses			X	X	X
Astro-4 Drug Information System			X	X	X
Atlas of Cancer Mortality			X		
Biological Data Storage and Retrieval System	BIO-STORET	X	X		
Biological Sciences Information Service	BIOSIS		X		
Biomedical Studies Group			X		
Cancer Information On-Line	CANCERLINE		X		
Carcinogenesis Bioassay Data System	CBDS	X		X	
Census Bureau Foreign Trade Statistics			X	X	X
Census of Manufacturers			X	X	X
Chemical Abstracts Service Chemical Registry System		X			
Chemical Abstracts Service Information System			X		
Chemical Dictionary of the U.S.ITC		X			
Chemical Dictionary On-Line	CHEMLINE		X		
Chemical Economics Handbook			X		
Chemical Information and Data System	CIDS	X	X		
Chemical Information System	CIS		X		
Chemical Monograph Referral Center	CHEMRIC		X		
Chemical Names File	PHS-149		X		
Chemical Transportation Emergency Center	CHEMTREC		X		
Clinical Toxicology of Commercial Products	CTCP	X	X		
Component Information for Chemical Consumer Products			X	X	X
Congressional Information Service Index	CIS INDEX		X		
Congressional Record Abstracts	CRECORD		X		
CPSC Chemical Abstracts		X			
Current Employment Statistics			X		X
Data Base of the U.S. ITC			X	X	X
Defense Documentation Center	DDC	X	X	X	
Directory of Chemical Producers	DCP		X		
Distribution Register of Organic Pollutants in Water	WATERDROP	X	X		
Drug Registration and Listing System			X	X	X
Drug Research and Development Chemical Information System	DR&D CIS	X	X	X	

TABLE 3-10 (Continued)

SOURCE OF DATA AND THE PROPRIETARY STATUS OF THE PRIMARY SYSTEMS

PRIMARY SYSTEMS	ACRONYM	INTERNALLY GENERATED DATA	EXTERNALLY GENERATED DATA	PROPRIETARY INFORMATION	MANDATORY SOLICITATION DATA
Dun's Market Identifiers	DMI		X		
Environmental Contaminant Monitoring Program		X			
Environmental Mutagen Information Center	EMIC		X		
Environmental Reports Summaries			X	X	
Environmental Teratology Information Center	ETIC		X		
Exposure Dictionary for the National Occupational Hazards Survey	EDNOHS	X			
Federal Inventory of Environmental and Safety Research					
Fish Control Laboratory-Data Base Information		X			
Fish-Pesticide Research		X			
Health Hazard Evaluations			X		
Index Chemicals Registry System	ICRS		X		
Industrywide Studies		X			X
Information Bulletin of the Survey of Chemicals Being Tested for Carcinogenicity			X		
Information Storage and Referral Section		X			
Inorganic Chemical Computer Toxicology Parameter Data Base			X		
International Cancer Epidemiology Clearinghouse			X		
IPC Chemical Data Base			X		
Kirk-Othmer Encyclopedia of Chemical Technology			X		
Laboratory Animal Data Base	LADB	X	X		
Mammal Toxicity and Repellency Data Base		X			
Meat & Poultry Inspection Monitoring Program		X		X	X
Medical Literature Analysis and Retrieval System On-Line	MEDLINE		X		
Microconstituents in Fish and Fishery Products		X			
Military Entomology Information Service	MEIS		X		
Mineral Commodity Survey System			X		
NASA Scientific and Technical Information Service		X	X		
National Center for Health Statistics	NCHS	X	X		
National Center for Toxicology Integrated Research Support System		X			
National Electronic Injury Surveillance System	NEISS		X		
National Occupational Hazard Survey File	NOHS	X			
National Technical Information Service	NTIS		X		
NIOSH Technical Information Center	NIOSHTIC		X		

TABLE 3-10 (Concluded)

SOURCE OF DATA AND THE PROPRIETARY STATUS OF THE PRIMARY SYSTEMS

PRIMARY SYSTEMS	ACRONYM	INTERNALLY GENERATED DATA	EXTERNALLY GENERATED DATA	PROPRIETARY INFORMATION	MANDATORY SOLICITATION DATA
Oceanic Abstracts			X		
Oceanic and Atmospheric Scientific Information Service	OASIS	X	X		
Office of Standard Reference Data Chemical Files		X	X		
Oil & Hazardous Materials Technical Data System	OHM-TADS	X	X		
Organic Chemical Producers Data Base			X		
Pesticide and Industrial Chemicals			X		
Pesticide Enforcement Management System	PEMS	X			
Pesticide Reporting System		X			
Pesticides Analysis Retrieval and Control System	PARCS		X	X	X
POISINDEX			X		
Poison Control On-Line Inquiry System			X		
Pollution			X		
Population Studies System		X			
Predicasts Marketing Systems			X		
Registry of Toxic Effects of Chemical Substances	RTECS		X		
Reporting of Economic Data for Negotiation of International Transportation Conventions	REDNITRAC		X	X	
Research Program of Chemicals That Impact Man		X	X		
Smithsonian Scientific Information Exchange	SSIE		X		
Solid Waste Information Retrieval System	SWIRS		X		
Special Trade Representatives Centralized Data Bank	STRCDB		X	X	
Standards Completion Program		X	X		
Storage and Retrieval for Water Quality Data	STORET	X	X		
Subject Content Oriented Retriever for Processing Information On-Line	SCORPIO	X			
Supplementary Data System			X	X	
Survey of Compounds Which Have Been Tested for Carcinogenic Activity			X		
Technical Data Center	TDC		X		
Thermophysical Properties Research Center		X	X		
Toxicology Data Bank	TDB		X		
Toxicology Information On-Line	TOXLINE		X		
Toxicology Testing In Progress	TOX-TIPS		X		
Trade Name Ingredient Clarification	TNIC		X	X	X

4.0 IDENTIFICATION AND EVALUATION OF DATA FILES CONSISTENT WITH USER REQUIREMENTS

4.1 Introduction

This section presents a summary of user requirements for information concerning chemical substances and compares these with the capabilities of existing files. The primary files identified in Section 3 are evaluated with respect to their characteristics and attributes (e.g., accuracy of data, specificity of data, degree of mechanization and access). The characteristics of these files are compared with those characteristics associated with the functional categories in the User Requirements Analysis (Section 2).

Following a discussion of the primary files applicable to each subject area, those primary files best able to supply the information requirements are presented. The strong points and inadequacies of each primary file are then analyzed. In the following sections of this report, these applicable files are combined with new files which must be created because the primary files are inadequate to meet the user needs. The result is an integrated systems plan for supplying information on chemical substances.

4.2 Substance Identification

The discussion of systems applicable to substance identification data is divided into five sections. These are Basic Identification Data, Chemical and Physical Properties, Composition Data, Compound Impurities, and Chemical Analyses Techniques.

4.2.1 Basic Identification Data

Basic identification data for chemical substances include molecular formula, chemical structure, CAS registry number, CAS-preferred name, and synonyms. Molecular formula and chemical structure are required for all chemicals in commerce for all three functional categories (Categories I, II, and III). They are required to be available on an interactive basis, updated annually and possess a high degree of specificity.

There are a number of files which contain varying amounts of this information. The NIH/EPA Chemical Information System (CIS) now has the "candidate list" on-line through the TYMSHARE System, which provides access to the CAS registry number, the preferred name, the chemical structure, and the molecular weight. CIS is searchable by chemical structure, substructure and CAS number. CIS can be used to search for every occurrence of a complete structural formula or fragment in its file, as opposed to a molecular formula. This procedure is termed substructure searching and involves a search through a file of connection tables for the part that has been specified by the user. A number of additional externally generated files (e.g., OHM-TADS, Merck Index, etc.) have been registered and are structurally searchable through the CIS substructure searching system. CIS will update this file when the final inventory is published by EPA, thereby providing access to these data elements for all chemicals on the inventory. In order for this file to maintain its currency, the file will have to be

updated on an annual basis to include changes made in Chemical Abstract numbers, names, etc. Changes in CAS numbers will impact on all systems maintaining CAS numbers as an access key. Manufacturers who may be required to report on an annual basis such items as changes in production, use, etc., should be aware that CAS numbers do change as new information about chemical structure is reported.

CHEMLINE is another file which provides basic identification data for a large number of chemicals (100,000) and, in addition, provides a locator designator which points to other files in the NLM system which have information on this chemical. Where applicable, each CAS number record in CHEMLINE contains ring information. At the present time, the CHEMLINE system can be searched by this ring information or by name fragments. NLM is considering loading the candidate list into CHEMLINE, thereby providing access to the large numbers of users already having access to the NLM data bases.

The Systems which will be discussed in more detail in Sections 5 and 6 include:

(1) CHEMLINE
(2) Chemical Information System (CIS)
(3) Army's Chemical Information Data System (CIDS)

4.2.2 Chemical/Physical Properties

The user analysis study indicated that chemical and physical property data were not necessary for first level screening of chemicals. However, for second level screening and Category II and Category III functions they were necessary but there was no justifiable requirement for an interactive system.

The following are existing data systems which may be able to supply relevant chemical property data: Chemical Information System (CIS, Chemical Abstract Services Information files (e.g., CACondensates, CBAC), NASA Scientific and Technical Information Data Base, and the Toxicology Data Bank (TDB). Physical property data are available from the Pesticides and Industrial Chemical File, Toxicology Data Bank, the Office of Standard Reference Data Chemical Files and the Thermophysical Properties Research Center.

The Chemical Information System contains extensive files of mass spectral data, x-ray diffraction, and CNMR data which are available on-line through a commercial system making it widely accessible.

The toxicology Data Bank presently contains selected chemical and physical data on approximately 1000 chemicals. These data have been extracted from various handbooks and published sources and have been evaluated before being entered into the system. TDB provides a potential focal point for physical and chemical data. The anticipated file is expected to contain data on 4000-5000 chemicals. Selected chemicals (Category III) for which hazard analysis, criteria documents, and/or regulations are planned by various agencies, could be primary candidates for inclusion into TDB, thereby enlarging the file and centralizing such information.

Relevant data found in the Standard Reference Chemical Data file and the Thermophysical Property Research Center already serve as sources for much of this information, but for purposes of establishing a centralized file, TDB provides an established mechanism for such data.

Pre-manufacturing data, substantial hazard notifications, etc., received by OTS which fall into Category II will be handled by the EPA Reports Management System. Plans to coordinate these data with data in CIS and TDB will be addressed in a later report.

CAS files such as Chem Condensates and NASA files can serve as sources of physical and chemical data for chemicals not included in TDB or CIS.

4.2.3 Composition Data

CPSC, FDA, NIOSH, OSHA, and OPP/EPA require product composition data for chemical formulations that fall under their respective authorities. These agencies utilize the composition data to accomplish first level screening since they are concerned about chemicals in products that are manufactured in large quantities and/or offer potentially high human exposure levels. The regulatory agencies also use the files extensively in hazard analysis and enforcement activities.

Chemical composition of feedstocks (i.e., ingredients) and process intermediates are required in order to set screening priorities for more intensive second level testing. This information is needed on an annual basis but is not required to be automated. The Office of Enforcement within EPA, however, stated a need for chemical composition data on an interactive basis to be responsive to short term or emergency situations, where the identification of all components in a particular substance, as formulated, is important to establish or substantiate violations.

Files which contain product composition data include: Astro-4 Drug Information System, Component Information for Chemical Consumer Products (CPSC), Trade Names Ingredient Clarification File (NIOSH), PARCS, Pesticides and Industrial Chemicals, Research Program of Chemicals That Impact Man (SRI/NCI), Clinical Toxicology of Commercial Products (CTCP), Poison Control On-Line Inquiry, and POISINDEX.

Most of the existing files of composition data respond to a specific Federal mandate and describe end-product formulation rather than providing the detailed chemical composition of all components of a process or product mixture. In addition, product composition files, such as those in FDA, CPSC, and NIOSH, contain a large percentage of data which are confidential and cannot be made available to other agencies. The NCI file maintained by SRI (Research Program of Chemicals that Impact Man) has general composition data but it is limited both in coverage and specificity. POISINDEX, CTCP and the Poison Control file of FDA provide composition data, but it is for products that are typically ingested. Data in these files are general, usually presented as ranges and focus on the active ingredients. POISINDEX has composition data for the broadest coverage of products (160,000 entries).

PARCS provides composition of pesticide products but primarily for active ingredients only. They are looking to OTS to obtain information on the "inerts".

4.2.4 Compound Impurities

Data on compound impurities are required for all analysis functions associated with Category II and Category III type data. There is no requirement for an automated system but there is an increasing requirement for specificity, particularly for research and monitoring functions. Groups conducting extensive testing (NCI, and the Testing Group [OTS]) were particularly concerned about adequate characterization of impurities before compounds enter long term test. This information is generally not available except on a limited basis. PARCS contains limited information on impurities in pesticides. Component Information for Chemical Consumer Products contains formulary information to the 0.1% level for consumer products. The NIOSH Trade Name Ingredient Clarification File has formulary information to the 1.0% level for industrial products. However, the primary intent of these files is to provide product composition data on purposely included chemicals. Firms who provided information to these agencies may not have reported impurities if they were insignificant or not recognized.

Generally, this information is received from the chemical manufacturer on a case-by-case basis. Impurities in technical grade chemicals will vary depending on the purity of the feedstock and on the process. Most agencies stated that they needed knowledge of impurities for proper assessment of risk, and were particularly concerned that when testing results were reported detailed chemical analyses should be provided as to the identification of the chemical and its purity.

4.2.5 Chemical Analysis Techniques

Knowledge of methods for chemical analysis including suitable techniques and standard protocols was cited as a requirement for Category II and Category III with greater specificity required for a manual file updated annually or as changes in methodology occur.

Several sources of this information are available such as handbooks of standard protocols (ASTM, AOAC). The Pesticides and Industrial Chemicals file provides some information, but most is obtained from searching bibliographic Chemical Abstract files. Several agencies, including NCI and EPA, indicated the need for development of a centralized file of analytical techniques for determining impurities in chemicals and methodologies for decontaminating chemicals.

4.3 Production Aspects

The discussion of systems applicable to production has been divided into three subsections. These are: Production Quantity, Plant Location and Manufacturer; Production Process and Control Technology; and By-Products and Impurities.

4.3.1 Production Quantity, Plant Location and Manufacturer

Production information is needed on a site specific basis for all three functional categories with provisions for an annual update. For Category I first screening, range data would be sufficient for site as well as quantity. However, as a chemical proceeds through Categories II and III the need for more exacting information becomes imperative. Hazard identification, hazard analysis and enforcement/compliance have the greatest needs. Because of the volume of data and

the short time frame required for response, an interactive computerized system will be required.

The following are existing data files or systems which may be able to supply information on production: SRI's Directory of Chemical Producers and Chemical Economics Handbook, Predicast Marketing Systems, the Data Base of the U.S. ITC, the IPC Data Base, the Mineral Commodity Survey System, Organic Chemical Producers Data Base, PARCS, the Research Program on Chemicals that Impact Man, the Toxicology Data Base, the Census of Manufacturers, the Annual Survey of Manufacturers, the Current Industrial Reports of the Bureau of the Census, and the Annual Survey of Injury and Illnesses.

No file provides site specific information on all chemicals in commercial production. The Data Base of the U.S. ITC contains quantities of synthetic organic chemicals produced, but the information is confidential when there are less than three manufacturers or production volumes of less than 1000 lbs. per year. They also have manufacturers and plant location information, but they do not have chemical information by plant. The Current Industrial Reports of the Bureau of the Census do contain production quantities by location, but only in terms of SIC code. This information is also proprietary and only summary statistics are released annually. The SRI files have production information by site, but only for a limited number of chemicals and the accuracy of some of their values has been questioned. All of the other data bases contain some pertinent information on production but the coverage is uneven.

Import quantities are also available through the IPC Data Base and others, but again the information is by generic class rather than specific chemical. The information of the Bureau of the Census covers inorganic chemicals production and shipment data on both a monthly and annual basis in their Current Industrial Reports Series and major organic and inorganic product class value of shipment data in their Annual Survey of Manufacturers. These data series do not list separate information on all the chemicals covered by the TSCA inventory.

A large number of agencies are looking to OTS to provide this site specific production information under their TSCA mandate. All would like to have access to an interactive computer file in order to reduce response time and to increase the ease with which the data can be accessed, but they realize that much of the data in the file would be of a proprietary nature. They are currently using these other systems, but the procedure is time consuming, often costly, and may not produce the desired information.

4.3.2 Production Process and Control Technology

Process information is required for the activities in the first functional Category only with respect to the identification of evolving technological changes. Specific process and control technology information, however, are required for Categories II and III activities. This information could exist in a manual form but there was a request to have it regularly updated.

Existing sources of process and accompanying control technology information include: the Kirk-Othmer Encyclopedia of Chemical

Technology, the NEDS subsystem of AEROS, the Organic Chemical Producers Data Base, the EIS and F & S subsystems of the Predicasts Marketing Systems and the Toxicology Data Base. There is much less collected information on available control technology than on production processes. Control technology is often only included in the above sources as it affects the process being discussed.

Probably the most complete existing source of production process information is the Kirk-Othmer Encyclopedia of Chemical Technology. This is, however, manual and somewhat dated. The Organic Chemical Producers Data Base is probably the best mechanized file, but it is limited in scope to 400 chemicals. The Predicast EIS and F & S systems available through Lockheed could supplement the above systems.

In the near future, it is expected that process and control technology information would only be required on a case-by-case basis. In order to update process trends as requested, especially for the Early Warning function, however, it might be necessary to organize a baseline process file. Process information is regarded as highly proprietary by a number of manufacturers, so if it were decided to set up a manual process file, strong industry resistance could be expected. This need for the information will have to be carefully evaluated as the plans for the implementation of TSCA become more firm.

4.3.3 By-Products and Impurities

Information on by-products and impurities is not required for the primary screening under the Category I functions. For the secondary

screening, however, range data are required. For Categories II and III functions, greater specificity is required as to the nature of the by-products and the impurities. In all functional areas a manual file would be sufficient, with provisions for a regular updating.

Two existing data systems could be accessed to provide some of the required information. The Organic Chemical Producers Data Base does contain by-product information on the 400 chemicals which it covers. The Research Program of Chemicals that Impact Man prepared by SRI for NCI also contains this type of information on 3200 compounds, but the file is incomplete in that not all information is included for all compounds.

Several Government agencies, among them NCI, NIEHS and OSHA, would like access to this sort of information were it to be available. The Interagency Testing Committee is also looking to OTS to provide by-products information since no data base exists and OTS has the unique authority to collect this information under section 8(a) of TSCA.

4.4 Marketing

The systems applicable to Marketing can best be discussed if they are divided into two areas. The first area covers Use Information and includes information on uses, users and places of use. The second area includes Economic information and covers sales volumes, costs, and market trend data.

4.4.1 Usage Information

Range use data is required of all chemicals for the initial screening step required to perform Category I functions. By the second screening, information is required on uses including amounts and how much of a chemical is involved per use. This same level of specificity is required for Categories II and III. The majority of the functional areas require this use information to be available in an interactive mode. Updating use information annually will assist in providing indicators of "significant new use."

Existing data systems and files which could supply useful information include: PARCS (for pesticides), the Data Base of the U.S. ITC, the Mineral Commodity Survey System, the Predicasts Marketing Systems, Research Chemicals That Impact Man, the SRI Chemical Economics Handbook and the Kirk-Othmer Encyclopedia of Chemical Technology.

No comprehensive file of uses of all chemicals in commerce currently exists. Those files described above which will probably be most useful in supplying usage information are the SRI Chemical Economics Handbook, Research Chemicals That Impact Man and the Kirk-Othmer Encyclopedia of Chemical Technology. The Chemical Economics Handbook and Kirk-Othmer, however, are not automated at the present time so they could not fulfill the interactive requirement expressed during the interviews. The file of Research Chemicals That Impact Man is automated, but was reported as covering 3200 chemicals. The National Occupational Hazard Survey contains occupationally oriented use information, but it was the result of a one-time plant survey

conducted in 1973 and is thus dated and the duplicability of its results might be questionable.

Some of the composition data discussed in detail in Section 4.2 which might aid in the defining of uses and amounts is contained in machine searchable files by use category. The CTCP, and POISINDEX files have this capability for a number of consumer products. CPSC and NIOSH also have composition data by use code but due to the proprietary nature of their files they are not publicly available.

Some additional use data is contained in the Predicast Marketing Systems and the Data Base of the U.S. ITC, but the uses are generally consolidated into generic categories.

All of the above mentioned files employ different terminologies to denote use. The creation of an interactive file which would determine "significant new uses" would require the existence of a base-line use file and a standardized vocabulary for reporting use. A number of government agencies including OSHA, CPSC, NCI, and DOD, in addition to several consumer action groups, are looking to OTS to provide this base line use information in an easily accessible form for all chemicals in commerce. Some agencies are currently using contractors to supply use data on a compound by compound basis which is both costly and time intensive. TSCA's authority could be effectively utilized to provide a centralized file of usage information.

4.4.2 Economic Information

Marketing Information is required in a non-specific form for the second screening under the Category I functions. Increasingly more

specific data are required for Categories II and III functions. The information can be collected in a manual form, but there is a requirement that it be updated annually.

The following existing files may be helpful in supplying the required informatinon needs: the Data Base of U.S. ITC, the IPC Chemical Data Base, the Mineral Commodity Survey System, Predicasts Marketing Systems, the Reporting of Economic Data for Negotiation of International Transportation Conventions, Dun's Market Identifiers and the SRI Chemical Economics Handbook.

Of the above files, those which are probably most useful are the Data Base of the U.S. ITC and the Chemical Economics Handbook. The Data Base of the U.S. ITC contains data on synthetic organics but it is publicly releasable only for those chemicals produced by more than three manufacturers and in quantities over 1000 lbs. per year. The SRI Chemical Economics Handbook is manual and again is not exhaustive in its coverage. The Predicast Marketing Systems and Dun's Market Identifiers are commercial systems which supply valuable supplementary information. Their coverage is unever, however, since they rely on the release of this type of economic information in journals, government reports, corporate annual reports, etc. Import/export information is available through the IPC Data Base and the Reporting of Economic Data for Negotiation of International Transportation Conventions - the latter of which is concerned with commodities relative to the tariff quotas.

Most of the agencies queried do require this type of information, and are currently obtaining it through the use of contractors. The fact that only a manual file is required for this type of information may suggest that OTS might cooperate with other agencies in defining common information needs for these data, locator files to existing data and designing access patterns.

4.5 Exposure

Information requirements and systems to fulfill these requirements with respect to Exposure can best be discussed in terms of Occupational Exposure, Environmental Exposure and Consumer Exposure. Eventually it would be desirable to be able to also discuss cumulative exposure doses due to a variety of these sources, but these types of data are very difficult to obtain at the present time.

4.5.1 Occupational Exposure

To perform Category I functions, occupational exposure information is required at a moderate degree of specificity. This specificity requirement increases with passage to a Category II or III function. It has been requested that this information be available through an interactive mechanized file.

The following files and data systems may be of use in providing this type of occupational exposure information: AEROS, Dun's Market Identifiers, the National Occupational Hazard Survey, the Research Program of Chemicals That Impact Man and several bibliographic systems such as CANCERLINE, TOXLINE and the OSHA Technical Data Center.

No one comprehensive source of occupational exposure exists. Of the above systems, the most applicable is that associated with the National Occupational Hazard Survey conducted by NIOSH in 1973. This survey covered approximately 5000 of the estimated 5 million workplaces in the United States and was a one-time effort. Information was collected on products, processes, number of workers, exposure, presence of medical exams, protective equipment, etc. The exposure data generated in this survey are mechanized and being used by NIOSH in establishing program priorities. These data are also being used by the Interagency Testing Committee as a primary source of occupational exposure data. The Dun's Market Identifiers and the BLS Annual Survey of Injuries and Illnesses also contain information on the number of workers per workplace as long as it exceeds eleven. In addition, AEROS in its NEDS subsystem contains emissions and resultant work force exposure data.

Several Federal agencies requested occupational exposure information including OSHA, DOD, and NCI. The Interagency Testing Committee was also concerned about occupational exposure. OSHA has a mandate to safeguard worker health and has required industry to maintain health and safety files including work assignments, exposures in excess of the TLVs, adverse reactions, etc., for at least 30 years. OSHA merely retains the ability to request this information from industry on an as needed basis. OTS has a justifiable requirement for occupational exposure information. It may be that OTS's requirements for occupational exposure could be combined with those of OSHA. Certainly a

common request format for information storage and retrieval relative to industry should be considered.

4.5.2 Environmental Exposure

For the first level screening under the Category I function, environmental exposure data are required of a moderate specificity. For Categories II and III functions, more specificity is required. The environmental exposure data available should cover air, water, soil and plants. As with the occupational exposure information, an interactive automated file would be required.

The following existing data systems and files can potentially supply environmental exposure information: AEROS, AGRICOLA, WATERDROP, STORET, the Research Program of Chemicals That Impact Man, CANCERLINE and TOXLINE.

For non-criteria pollutants there is no comprehensive file of environmental exposure. EPA's AEROS system collects air pollution data from a number of state and local agencies as well as from the EPA monitoring network. This includes air quality as well as emissions monitoring information. STORET serves a similar function for water. The Research Program of Chemicals That Impact Man contains available environmental exposure information on selected chemicals. In addition, for water WATERDROP plans to collect monitoring data to determine the presence of organic chemicals in water. AGRICOLA, formerly known as CAIN is owned by the USDA and contains bibliographic references to the effects of various emissions and effluents on crops and livestock.

TOXLINE and CANCERLINE would similarly cite wildlife and plant effects due to environmental exposure to toxic substances.

Environmental exposure information was widely requested, but minimally available from present systems. Various Federal agencies, including DOI, DOD, NCI, EPA, ERDA, in addition to the Interagency Testing Committee, expressed a justifiable need for environmental exposure information. EPA seems to be a logical focal point for the collection of this sort of information and for the synthesis of their existing raw data into a more reliable and usable form.

4.5.3 Consumer Exposure

Generalized consumer exposure information is sufficient to meet Category I activities, but for Categories II and III more explicit information is required. Again an interactive data file capability has been requested for accessing this type of exposure data.

Several existing systems could help to supply some of the required information on consumer exposure. They are: the National Electronic Injury Surveillance System, the Meat and Poultry Inspection Monitoring System, the Research Program of Chemicals That Impact Man and the bibliographic files of CANCERLINE and TOXLINE.

There is no primary information system in the area of consumer exposure. The best existing system is probably the Research Program of Chemicals That Impact Man developed by SRI for NCI. This file is limited to nine categories of information. The Meat and Poultry Inspection Monitoring and a similar fish monitoring system supply

concentrations of a number of pesticides and heavy metals in animals and can be used as an indicator of human exposure due to their ingestion. The National Electronic Injury Surveillance System contains information on accidents associated with consumer products reported to hospital emergency rooms. The information is reported by generic classes of chemicals and is useful for acute type injuries only. Several of the poison information systems such as POISINDEX, and the Poison Control On-Line Inquiry System might also contain some information on adverse consumer reactions, but the bulk of these data relate to ingestion of substances by small children.

NCI, CPSC and ERDA, in addition to the Interagency Testing Committee, voiced a need for this type of information. There is no adequate source to meet this sort of request in either a manual or interactive mode. It therefore remains that a new system will be required to fulfill this justifiable need.

4.6 Epidemiology

Epidemiology studies are concerned with identification of populations exposed to toxic substances and their resulting adverse reactions. These studies generally deal either with an occupational population or with an identified section of the general population.

Epidemiological information is required for Category I secondary screening functions though it need not be highly precise. In order to complete Categories II and III activities greater specificity is required. A manual system would be sufficient to meet the expressed user information needs.

A number of existing files contain epidemiological information. These include: BLS's Annual Survey of Injuries and Illness and Supplementary Data System, the Atlas of Cancer Mortality, the International Cancer Epidemiology Clearinghouse, the National Center for Health Statistics, the National Electronic Injury Surveillance System (NEISS), the National Occupational Hazard Survey, Poison Control On-Line Inquiry, the Population Studies System, the Standards Completion Program, the Toxicology Data Bank and a number of bibliographic systems including BIOSIS, CANCERLINE, NIOSHTIC, the Technical Data Center, and TOXLINE.

With regard to supplying occupational epidemiology information, there are several important systems. BLS's Annual Survey of Illnesses and Injuries and Supplementary Data System together provide biannual information on all work-related adverse effects requiring workman's compensation by facility in addition to supplying total workforce numbers for comparison. These data are proprietary in nature which means that only summarized statistics by state and SIC code are publicly available. In addition, the illness codes are not very specific. For example, it is impossible to differentiate a liver cancer from an ulcer. The National Occupational Hazard Survey provides data on 5000 establishments collected during a single survey conducted in 1973. Worker health data were collected as well as information on exposure levels to various potentially toxic substances. It is extremely difficult to obtain chronic occupational effects data

and most of the above collected data are of an acute nature. In general, only where individual industries have been surveyed in detail for a long number of years can valid conclusions be drawn.

Epidemiology information on the general population is also available, but again it is usually based on acute toxic reactions. Information on consumer-related adverse reactions to toxic substances is available form the National Electronic Injury Surveillance System which collects hospital emergency room data relative to injuries associated with consumer products. Similar information is available from the poison control adverse report systems. The International Cancer Epidemiology Clearinghouse and the Atlas of Cancer Mortality produced by NCI provide geographic as well as body site, sex, and race information on cancer development and death.

The National Center for Health Statistics has collected U.S. prevalence data on a number of conditions based on a nationwide survey system. They are also one of the best sources of mortality information. Their disease prevalences can be used to provide a baseline against which alternative incidences can be measured. The number of conditions on which they have collected data, however, are limited and often generic in nature.

If baseline demographic data were required for comparison of exposed and unexposed populations, this could be obtained from the Census Bureau by all geographic divisions down to census tract.

Need for epidemiological information was expressed by OSHA, DOI, DOD, CPSC, NIEHS, NIOSH and NCI as well as the Interagency Testing Committee.

OTS has the authority to request health and safety information from industry under section 8(d) of TSCA. TSCA also has a provision which requires industry to maintain this sort of data in an accessible form. OSHA may also solicit health and safety information, and industry has begun to design computer systems to provide this information to OSHA in an approved format. Care must be taken in designing OTS's industry request format to ensure that it is compatible with OSHA's.

4.7 Biological Effects

The systems that provide relevant data concerning biological effects are discussed in terms of systems which provide data on Acute Toxicity, Chronic Toxicity, and Metabolism. Both acute and chronic toxicity data are necessary for second level screening in Category I, however the data are not required to be precise. Data are required to be available on an interactive basis and the information should be as current as possible. NCI, OSHA, and various offices within EPA preferred the data to be available in a centralized file for easy accessibility. Category II functions required access to toxicity data on an interactive basis with a little more specificity. Category III functions, however, required greater specificity to substantiate the need for regulations. Since more time is available for preparation of information for selected high priority chemicals, interactive access to the data was not justified for this function.

4.7.1 Acute Toxicity

There are a number of systems which contain Acute Toxicity information. Systems which contain relevant information are the Advisory Center on Toxicology file, CTCP, POISINDEX, Poison Control On-Line Inquiry, the Fish Pesticide Research System, the Mammal Toxicity and Repellency Data Base, the Military Entomology Information System, OHM-TADS, the Organic Chemical Producers Data Base, the Supplemental Data System, the Research Program of Chemicals That Impact Man, the Registry of Toxic Effects and the Toxicology Data Bank. Bibliographic files which are frequently the source of acute toxicity data include BIOSIS, NTIS, MEDLINE, TOXLINE and the Toxicology Research Projects Directory.

CTCP, POISINDEX and the Poison Control On-Line Inquiry System provide acute toxicity data collected from the published literature, which have been evaluated before entry into the system. In addition, these systems provide antidotal information for treatment of poisonings involving the referenced chemicals. These systems are limited in size (although POISINDEX now has 160,000 entries) and have been designated primarily to assist physicians in the treatment of poisoning cases.

The Registry of Toxic Effects is the largest mechanized file of acute toxicity data. Data are extracted from the published literature and the sources are cited. There has been no evaluation made of the data before entry into the system, but as stated, the reference is cited so it is possible to obtain the original source for evaluation.

This file will be available on-line in the near future through the National Library of Medicine.

The TDB is a smaller file, now providing data on 1000 chemicals. This file provides acute toxicity data for all data which are evaluated before entry into the file. Also, the Research Program of Chemicals That Impact Man developed by SRI for NCI provides data on acute toxicity. It is used primarily by NCI to assist in selection of chemicals for entry into the Carcinogenesis Bioassay Program.

OHM-TADS provides acute toxicity data for 1000 chemicals frequently transported and therefore, subject to spills, fires, etc. In addition, PARCS provides acute toxicity data for all registered pesticides.

The Registry of Toxic Effects and TDB provide immediate sources of relevant information pertinent to the needs of agencies involved in controlling toxic substances.

4.7.2 Chronic Toxicity

Systems which provide data relevant to chronic toxicity are CANCERLINE, CANCERPROJ, the Carcinogenesis Bioassay Data System, the Information Bulletin of the Survey of Chemicals Being Tested for Carcinogenicity, Information Storage and Referral Section, the International Cancer Epidemiology Clearinghouse, the Laboratory Animal Data Base, the NCTR Experiment Information System, the Organic Chemical Producers Data Base, the Registry of Toxic Effects of Chemical Products, the Research Program of Chemicals That Impact Man, the Survey of

Compounds that Have been Tested for Carcinogenicity, the Toxicology Research Project Directory, EMIC, ETIC, TOX-TIPS and TOXLINE.

The number of systems mentioned provides evidence that there is no one file which provides information on an interactive basis which would be responsive to the stated need for a coordinated file. The concern was expressed by many offices and agencies that a coordinated file is critical to discover what testing is being conducted in order to reduce duplication. The coordinated file should contain as a minimum, the type of test, the test method utilized, the investigator's name and association, the species utilized and the results. The file would serve as a validation both of testing methodology and of the effectiveness of *in vitro* tests as a predictor of *in vivo* test results.

TOX-TIPS and the IARC Information Bulletin of Chemicals Tested for Carcinogeneis provide the basis for a coordinated file of Carcinogenesis Testing Information. EMIC is the focal point for mutagenic testing data. However, with the widespread use of mutagenic screening tests, much of the data that were formerly being published in the journal literature and collected by EMIC are not now being published, but are remaining part of company or governmental files. Testing data submitted in response to testing regulations and as part of a pre-manufacturing notifications will be entered into the EPA/OTS Reports Management System. These data will need to be analyzed and made publically available. In addition, EPA can require health and safety data to be submitted under 8(d) of TSCA and, again, there would be the requirement to make these data publically available. It is clear that

to be responsive to the user requirement for an interactive file of chronic toxicity information, considerable planning is necessary to link existing files of data with files to be generated as a result of TSCA regulations.

Testing regulations for submission of chronic toxicity data will, by necessity, have to include standard formats for reporting of the data. These formats should be consistent with data submissions required by other government agencies. Industry is most agreeable to working out standardized reporting systems for both acute and chronic toxicity test reporting, particularly with so many of the larger firms considering the development of in-house information systems. Systems such as CBDS and the NCTR Experiment Information System will be examined for possible utilization by EPA for storage and retrieval of long term testing data.

ETIC provides the best source of teratology data, even though it is a very new system. TDB has assembled information on carcinogenicity, mutagenicity and teratogenicity in one file, but only for selected chemicals.

4.7.3 Metabolism

There is no cited requirement to have data with respect to metabolism in a computerized file. TDB does, however, provide an on-line source for metabolism data for approximately 1000 chemicals. Other sources of this type of data are CANCERLINE, BIOSIS, Fish Pesticide Research, Information Storage and Referral Section, MEDLINE, NTIS,

TOXLINE, and the Toxicology Research Project Directory. Most of these files are bibliographic and provide references to journal articles. Very little effort has been spent in terms of putting this type of information into a machine file, although the value of access to this data has become increasingly important.

4.8 Environmental Effects

Environmental Effects data can be divided into three main classes: Degradation, Transport and Fate, and Disposal Procedures; Ecological Effects and Bioaccumulation; and Weather and Materials Effects. Of the above, the information required for first screening Category I functions is degradation, transport and fate and bioaccumulation data, with weather and materials effects also being required for the second screening. For Categories II and III functions data in all areas are needed with increasing degrees of specificity. A manual system would be sufficient to supply these needs.

Most of the existing applicable files are bibliographic or referral in nature. These include AGRICOLA, APTIC, BIOSIS, the Defense Documentation Center, The Federal Inventory of Environment and Safety Research, NTIS, OASIS, SSIE, SWIRS, Toxicology Research Projects Directory and TOXLINE. The only identified non-bibliographic source of information on degradation, transport and fate and disposal procedures is the Research Program of Chemicals That Impact Man which has collected information, when it is available, on 3200 chemicals in commerce. Some information on bioaccumulation levels of heavy metals and pesticides in several animal species can be gained from the Fish Control Laboratory

Data Bank, the Fish Pesticide Research, the Environmental Contaminant Monitoring Program and the Military Entomology Information Service. These data are usually collected only in a few species and testing is done on only a few prescribed chemicals. Its general usefulness to the broad consideration of the effects of toxic substances on the environment is questionable. In the area of weather and materials effects, no useful non-bibliographic systems have been uncovered.

From the above discussion of existing data sources, it becomes apparent that no primary file exists which adequately addresses the area of environmental effects. This information was requested, however, during almost all of the Federal interviews and those with the Interagency Testing Committee.

EPA has had a history of being the focus for the collection of environmental effects-type data especially regarding selected criteria pollutants. It seems reasonable to consider that they would provide the central point for the collection of the types of data required in the area of environmental effects for toxic substances regulation and decision making. Only ranges of bioaccumulation data were required for all chemicals, and the other types of environmental effects information were only required for a narrowed list. A manual file could therefore exist which would collect this sort of information using the applicable bibliographic files to make it more readily accessible.

4.9 Standards and Regulations

Information concerning relevant Federal, state, local and international standards and regulations would be required to perform the

second level of screening under Category I and for all subsequent Categories II and III activities. For all purposes, a manual file could be sufficient.

Several existing interactive files contain Federal regulations such as CRECORD and the Congressional Information Service, Inc., and are searchable by the chemical names under which the regulations were promulgated. The NIOSH Registry of Toxic Effects of Chemical Substances contains all OSHA occupational exposure standards in both an automated and a manual form. Other sources of Federal standards information include the Standards Completion Program and the Technical Data Center. Specific pesticides information is contained in the Pesticide Reporting System and the Pesticides Enforcement Management System. Import/Export information is also available in the Data Base of the U.S. ITC and the IPC Data Base and the Multilateral Trade Negotiations Data Base.

Only the Advisory Center on Toxicology and the Pesticide Reporting System have been identified as containing state regulations, on a national basis. The files of the Advisory Center on Toxicology are manual, making them not readily available for reference. Likewise, international regulations and standards only exist in a bibliographic form in EPA's Environmental Reports Summaries.

There does not seem to be a good centralized source for obtaining all relevant standards and regulations concerning a given chemical. All government agencies and the Interagency Testing Committee contacted

during the course of this project expressed a desire to have access to this sort of a file, though a manual file would be adequate.

4.10 Summary and Conclusions

The following discussion summarizes the conclusions drawn in Section 4 regarding those systems most capable of providing information in a given subject area. In Sections 5 and 6 of this report those data bases selected for inclusion in the core systems will be more closely analyzed. Other data bases, including many of those mentioned in this summary section, should be peripherally available but a need does not exist for them to be directly linked.

Substance identification is required for all chemicals. The Chemical Information System and CHEMLINE, both of which will have incorporated the candidate inventory list, are potentially able to satisfy this user requirement. Information on chemical and physical properties, composition and impurities is required for only a selected subset of chemicals and access to it need not be mechanized. The Chemical Information System and the Toxicology Data Bank do provide selected data concerning physical and chemical properties, but more attention is required in this area. With regard to composition, several agencies have responsibility within their mandated areas of concern to collect composition information. A major distinction must be made, however, between chemicals and products. When areas of responsibility overlap, plans for cooperation and file linkage need to be developed. An additional difficulty associated with composition

information regardless of the ownership, is that it is generally confidential in nature. No comprehensive file exists of impurities in commercial chemicals Similarly regarding chemical analysis techniques, no data base exists which can adequately satisfy the identified user requirements.

There is a justifiable need for information on production, quantity, plant location and identity of the manufacturers for all chemicals in commerce. This production quantity information is not currently available for all chemicals. The best available sources being the Data Base of the U.S. ITC which only covers synthetic organics and the Census of Manufacturers whose data are collected solely by SIC code. In addition, much of the data in the above two data bases are confidential, with only summary statistics available for public release. The SRI Directory of Chemical Producers has only selected site specific production information.

Information has been requested on changes in processes and control technology. The Kirk-Othmer Encyclopedia of Chemical Technology is probably the best source covering the largest number of chemicals, but it is manual and somewhat dated. The Organic Chemical Producers Data Base constructed by Radian for EPA is probably the best automated file of this type of data.

Range information on chemical by-products and impurities is required for a large number of chemicals for prioritization and hazard identification. The Organic Chemical Producers Data Base and the SRI

files are the only real sources of by-products and the number of chemicals covered is limited. No comprehensive source of information on impurities exists.

In order to perform hazard identification and early warning functions, there is a justifiable requirement for usage information on all chemicals in commerce. No file contains comprehensive use information for all chemicals, but the SRI files, The National Occupational Hazard Survey files and those of the CPSC and the poison control centers such as POISINDEX and the Poison Control On-Line Inquiry System could provide some relevant information. One of the greatest problems associated with the utility of these various files is that they all have unique use terminologies. There is a critical need to adopt a common use terminology to permit multiple file access.

Economic data are typically required on a base-by-case basis to assess the impact of a proposed regulation. No justifiable need exists to have a comprehensive file. In general, highly specific data are required only for a particular chemical or chemical group and include market share by use, the availability of substitute chemicals, etc. What is required, however, is an awareness of the existence of such data in other agencies where it has been collected for their mandated purposes.

The need for summarized data with respect to occupational, environmental and consumer exposure is justified for all chemicals for hazard identification and early warning. There is no comprehensive file of

occupational exposure. The Occupational Hazard Survey is useful though limited to data collected during a one-time walk-through of 5000 workplaces and extrapolated statistically to cover all workplaces. The SRI Research Program of Chemicals That Impact Man provide exposure data for select categories of chemicals. Monitoring files can be used to derive some exposure information, but they are generally structured on a priority and criteria pollutant basis. There is no general consumer exposure file, although the CPSC file provides some range data.

There is no expressed requirement for epidemiological data for all chemicals in commerce. Studies of this type are usually required to substantiate regulatory activities and for that purpose are performed on a case-by-case basis. There is, however, a need to know what previously conducted studies are available and their results. There is also a need to collect baseline data for comparison with observed results in order to perform early warning and hazard identification functions. The National Center for Health Statistics collects information only for the presence and progression of certain diseases. NCI's Atlas of Cancer Mortality is also highly specific.

There is a justifiable requirement for a comprehensive index of all types of acute and chronic toxicity testing for the purposes of (1) identification of those chemicals which have been tested, (2) the validity of the test methods, and (3) the results. However, much of this type of information is unevaluated and would require a reference

to the original source for verification. Several existing files such as the NIOSH Registry of Toxic Effects and the Toxicology Data Base could be used to build a comprehensive toxicity index. TOX-TIPS and the IARC file of Chemicals Being Tested for Carcinogenicity could prove useful in identifying compounds under test. Currently, EMIC provides the only centralized collection point for results of mutagenic testing, but with the wide use of bacterial screening tests, much of this information will probably go unreported and hence uncollected. ETIC serves a similar function for teratological testing.

Information on bioaccumulation, degradation and transport and fate are required for all chemicals in order to prioritize them and for Early Warning and evaluation of pre-manufacturing notices. Environmental effects information is not available in a coordinated fashion for all chemicals, though a number of relevant bibliographic files do exist. Several files contain bioaccumulation data for pesticides or heavy metals in a selected list of species but both the chemicals and species are very limited in number. A baseline file of normal accumulation, transport, degradation, etc., levels is required on a large number of chemicals to provide a basis for comparison of values submitted by industry as a part of a pre-manufacturing notification data package. Such data do not presently exist in a collected form.

Several Management Systems have been identified in the user requirements study for purposes of assisting OTS and other offices in EPA to more efficiently manage activities associated with TSCA. They

are primarily tracking systems for decision packages, petition and substantial risk notifications, and correspondence. In addition, a compliance and monitoring management system is required. The requirements for automation are not defined yet because the volume of transactions is not clearly identified at this time.

Several areas have been identified above where there are not adequate files to meet justifiable user requirements. Other areas have been identified where there are existing files of information which satisfy all or some of the total requirements for specific types of information.

In the next sections of this report, METREK will present suggestions for the creation of new files and the agency which should have lead development responsibilities. In addition, recommendations will be made for linking existing and proposed files with various systems development options.

5.0 DEVELOPMENT OF AN INTEGRATED RETRIEVAL SYSTEM

5.1 Background

Section 10(b) of TSCA provides the EPA Administrator with the authority to establish a system within EPA to collect, use and disseminate data submitted to the Administrator under this Act. Section 10(b)(2)(A) authorizes the Administrator, with the cooperation of the Secretary of HEW and other heads of appropriate agencies, to develop an efficient and effective system for the retrieval of toxicological and other scientific data necessary to carrying out the purposes of this Act. The Act also explicitly states that systematized retrieval shall be developed for use by all Federal agencies with responsibilities in the area of regulation or study of chemicals and their effects on health or the environment.

The legislative intent is clear in calling for EPA to establish a system which will collect, store and disseminate data received in response to regulations promulgated under TSCA. It is also clear that EPA is to use the information gathering authorities of the Act not only to assist other agencies in carrying out their respective responsibilities under TSCA, but also to apply this information to the regulation of chemicals under various other legislative mandates.

The implementation of TSCA provides a unique opportunity for EPA to design and build an information system capable of being responsive to the needs of decision-makers in all government agencies. The Act provides extensive authorities to collect information necessary to

assess the environmental aspects of industrial chemicals. It provides the authority to fill in the "information gaps" that exist in the authorities of such Acts as the Federal Insecticide, Fungicide and Rodenticide Act and the Food, Drug and Cosmetic Act which focus on regulation of chemicals for specific uses. Table 2-2 in Section 2.3 demonstrates the overlapping authorities of existing legislation, but also shows the impact that TSCA will have on the "universe" of chemicals.

EPA has stated in its "strategy document" published in February 1977, that it intends to utilize TSCA as an "important tool for developing the information base which will undergird many major decisions of the future." It is further stated, that the explicit provisions of TSCA underscore the clear intent of Congress that this legislation serve the interests of many organizations in a variety of ways, particularly with regard to acquisition and dissemination of data.

Furthermore, EPA recognizes that just as a coordinated approach to the assessment and control of toxic substances is necessary, a coordinated approach to data systems development is also necessary. A critical first step was the assessment of user requirements of EPA, other Federal agencies and private groups for information concerning chemical substances, with particular attention being given to common information requirements that could best be satisfied through TSCA. It was also clearly recognized by EPA and the designers of the legislation that a multiplicity of data activities presently exist which collect, store, and disseminate data relevant to toxic substances

regulation. Coordination within the government is critical to: (1) assess the information requirements; (2) assess the existing systems which satisfy these requirements; (3) identify the gaps in information needed for regulatory purposes; (4) limit the total reporting burden on industry; and (5) identify ways to make the information acquired under TSCA available as widely and as promptly as possible.

Section 2 of this report presents the results of the user requirements study. These users are looking to EPA to exercise TSCA's information collection authorities and provide a comprehensive system capability that permits access to these data. Furthermore, they are looking for a capability to perform data correlations to assist in the assessment of health and environmental effects. EPA and other agencies plan to use the information system to support decision-making activities such as early warning, and selection of chemicals for test, risk assessment, etc. Furthermore, they are expecting to use this information to assist in the prediction of health and environmental effects, in establishment of priorities for long-term testing and in development of regulations.

For the desired systems capability to be responsive to user requirements, it must be a comprehensive, integrated system capable of providing a variety of data on a large number of chemicals. The system must permit public access to the non-proprietary information obtained under TSCA, but still provide full confidential protection

to the data that are "trade secret." TSCA specifically excludes from claims of confidentiality health and safety studies on chemicals offered for commercial distribution and on chemicals subject to pre-market notification or testing requirements. Other data such as process information, may be considered confidential and will require protection from disclosure. Systems development options responsive to user requirements and consistent with EPA strategy to implement TSCA are presented in the following sections.

5.2 Approach to Defining Systems Development Options

The results presented in Sections 2 and 3 identified the user requirements for data and inventoried the currently available data bases and systems which are potentially able to satisfy those needs. Certain information gaps were found as were some apparent duplications of effort. The objective of this section is to define system design concepts which will satisfy the user requests.

At the direction of the project officers, the METREK analysis of systems development options was confined to examining feasible concepts of systems integration. Consideration was given at the outset to an on-line retrieval system that would link a series of computerized information files and direct an on-line user to other external information files (which may or may not have on-line access). Recommendations were to be formulated with the ultimate goal of achieving a system usable by the "end-user" rather than being limited to information specialists or librarians. The priorities and policies

of EPA/OTS were to be considered in terms of their impact on information needs, data acquisition and system implementation. Scheduling considerations from the user point of view were to be emphasized in recommending systems development options.

A complementary effort by another independent contractor addresses the development of a program to implement the systems development recommendations presented in this report. Included in the complementary effort are an analysis of software and hardware characteristics and requirements, system maintenance requirements, detailed systems specifications, and the costs associated with implementation of the recommended plan.

A first step in defining systems development options is to formulate the long-range goals and objectives for a comprehensive integrated system that is responsive to satisfying information requests. Once the long-term capability has been formulated, alternative approaches for achieving that capability by utilizing or modifying currently existing systems can be developed and evaluated. Alternative approaches are dependent, however, upon EPA policies and priorities for exercising the data gathering authorities granted by TSCA. The extent to which EPA issues regulations requiring the submission of various categories of data greatly affects the nature of any data base or system capability at a given point in time.

To limit the number of possible alternatives which could be considered and to provide a framework for recommending systems

development options, three scenarios have been developed. The first is based on EPA information gathering policies as stated in their strategy for implementing TSCA and current plans for section 8 rule-making. The second is based on an incrementally increased information gathering policy of EPA in terms of sections 8(a) and 8(b). The third is based on an EPA policy to fully implement all data gathering authorities listed in 8(a)(2).

Within each of these scenarios, specific system development plans are presented. The designs are presented in terms of definitions of component files, system linkages, file ownership and accessibility. The relationships with other Federal files are defined in a way consistent with their functional responsibilities. In developing the system plans, a number of considerations are addressed. These include: the current stage of development of data systems and bases; the degree to which information requirements can be fully satisfied; the systems ability to facilitate analysis of potential hazards and to disseminate information to a large community of users while simultaneously providing for protecting confidential information; and the impact on the users of the time frame within which implementation of enhancements is possible.

5.3 Long-Range Objective of a Comprehensive Chemical Substance Information System

5.3.1 Requirement for Integrated Computer Network

When examining the information requirements integrated across functions for all users (Table 2-6), it can be seen that there is a justifiable requirement for an interactive system containing substance identification data, production data by plant location, use data, exposure data and biological effects data for Category I chemicals. After analysis of existing systems, it is clear that this information, for Category I chemicals, is not presently available in existing data bases with the one exception of substance identification data (that is, molecular formula, CAS registry number, CAS name, synonyms, and chemical structure) for chemicals that are presently on the "candidate list."*

It is also clear upon careful review of the legislation concerning regulation of chemical substances, that the authority for obtaining such information (production, use and exposure) resides in EPA. EPA, utilizing the industrial reporting and recordkeeping provisions of TSCA, is in the position to build a comprehensive data facility required by EPA and other Federal agencies with responsibilities in the area of regulation or study of chemical substances and mixtures in commerce and their effects on health or the environment. EPA,

---

* The assumption is made that the inventory of chemical substances authorized under section 8(b) is an adequate definition of Category I chemicals for most users.

therefore, has a major role in the creation of such a comprehensive, integrated system that provides data on all chemical substances.

When one examines the long-term need for an integrated system to support EPA and other Federal agencies and one which permits rapid access to information on chemical substances for purposes of making risk assessments, predicting toxicity, selecting chemicals for testing, approving chemicals for pre-manufacturing, etc., certain components appear to be critical. These components vary in size and detail. Extensive, detailed information similar to that outlined in Table 2-1 may eventually be collected by EPA for all chemicals in the inventory (50,000 to 100,000) for regulatory purposes. On the other hand, selective data such as substance identification and structure information may be obtained on as many as 500,000 chemicals by various agencies involved in research or regulation under other legislative mandates.

MITRE recommends the following:

(a) That the information required to support TSCA activities be implemented in a set of function specific on-line data bases;

(b) That all data bases which are of primary importance to TSCA activities and which are likely to be accessed as part of a coordinate search, should have compatible data structures and should utilize common, standardized nomenclature. These primary data bases are identified on Page 5-15, and are referred to as "core components;"

(c) That a network of data bases called the Chemical Substances Information Network (CSIN) be developed. This network system shall have the capability to greatly facilitate access to core component systems, and to direct users to other, non-core component data bases which contain useful information, but which are not part of the network system.

A diagram of the proposed CSIN is shown in Figure 5-1. CSIN has as its primary objective the service of those Federal agencies involved in the study and regulation of chemical substances. A secondary goal must be for CSIN to become a fundamental new information tool for R&D activities in the biomedical community. This systems network as shown in Figure 5-1 builds on existing systems, where appropriate, and provides those additional analytical capabilities necessary to support the decision-making activities and other governmental functions previously described in Section 2.4.1.

In developing the network concept illustrated by Figure 5-1, recognition has been given to the fact that the legislative responsibilities of Federal agencies vary. In some instances, the agencies are concerned with different types of chemicals (e.g., food, drugs, or pesticides) although they may require similar categories of data. In other instances, the agencies may be concerned with different aspects of regulating the same chemicals and hence could use a common data facility. This circumstance is perhaps better illustrated by considering Figure 5-2 and the following example. In support of TSCA-related regulatory activities, there is a requirement for general information for Category I chemicals (i.e., all chemicals subject to regulation under TSCA) in a large number of data categories. There is also a requirement for more detailed information within all categories of data but for fewer chemicals (i.e., Categories II and III). To regulate pesticides, there is a similar requirement for data for chemicals used as pesticides. Simultaneously, NCI is concerned with

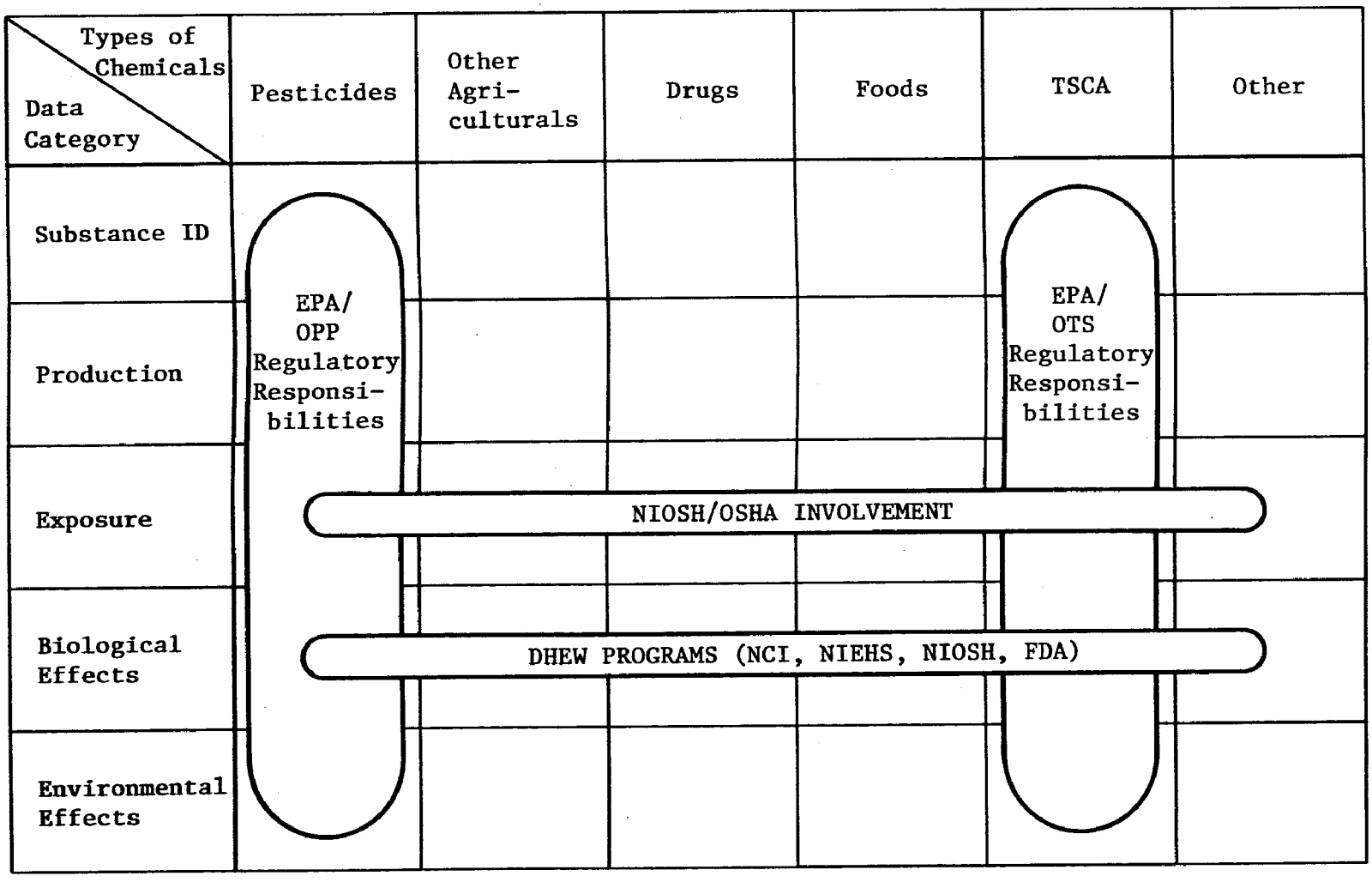

FIGURE 5-2
DATA INVOLVEMENT OF SELECTED REGULATORY AGENCIES

carcinogenic effects of chemicals, such as pesticides, food additives, drugs, other agriculturals, "TSCA," or other types of chemicals, while NIOSH/OSHA are concerned with occupational exposure and associated health effects of chemicals regardless of their type.

The system design implications of this are that (1) no single data base can fully satisfy all user requirements, and (2) multiple data bases must be designed and developed in a manner that facilitates cross-exchanges of data and retrieval of particular data by chemical substance identifiers. A further implication is that the ultimate direction in which various information systems will evolve is dependent upon the requirements to be placed upon them by many and varying users, each with his own unique data requirements. Although it is beyond the scope of this effort to conduct a rigorous and comprehensive analysis of alternative network systems responsive to these varying requirements (since the primary intent of this effort has been directed at TSCA requirements), a general awareness of their implications has been incorporated into the analysis used in developing the concept expressed in Figure 5-1.

The concept of the CSIN proposed herein consists of a set of core component data bases which are distributed over a network, and a set of non-core component files, which are known and referenced in the network, but physically do not reside within it. Several options are available for linkage of the core-component systems. At the most sophisticated level, data bases are directly linked, so that they

appear to the user as a single, coherent system. The user, in essence, deals with a data resource executive (a piece of computer software), which in turn deals with the component data bases and their data base management systems. This is relatively easy to implement if all the core component systems reside at a single computer facility and under control of the same data base management system (DBMS). It becomes more difficult if the data bases are on different computers, and it is prohibitively complex if the data bases are under the control of different data base management systems. It is probably necessary to require that all directly linked core component systems be implemented under the same data base management system. (It should be noted that several commercial DBMS's are implemented on a variety of hardware systems). Direct linkage appears to be appropriate for some component systems, but is probably not required for all. For those systems which are not directly linked, a user would access the network directory, which would inform him of the location of the data of interest, and would transfer him to the site of the data base. From that point on, however, the user would be interacting directly with the target data base. In order to access another data base, the user would have to access the directory again, and be redirected. The precise method of linkage used and the important issues of file backup and security must be addressed during the CSIN design phase.

Another alternative might involve component systems utilizing a variety of software packages linked to a minicomputer that provides

a common "macro language" and query capability which makes accessing a variety of different systems transparent to the user. Detailed decisions regarding the type of hardware and software are beyond the scope of this report, and will be considered in the subsequent analysis. Those data bases which are not core components will be referenced in the network directory, and their location and method of access will be given. The actual access will be left to the user.

5.3.2 Individual Components of the Chemical Substance Information Network

Long-term user requirements for chemical substances information can best be satisfied by the development, in an evolutionary manner, of a distributed network of data bases and systems. Within the recommended network, certain files and systems are of primary importance and are core components. For these core components, user requirements are best satisfied if these components are structured using common, standardized nomenclature for data elements and categories. The data bases of the core components should be maintained by a single data base management system (DBMS) to facilitate cross exchanges of data and retrievals of particular data. This, however, does not imply that a single repository is required. In fact, user requirements are probably best satisfied by maintaining the core components in different computer facilities which provide time-shared systems capable of supporting large numbers of terminals of various degrees of sophistication. Actual location of core components is not critical as long as their access is equable and widely available to the public.

The core components of the recommended distributed network are:

- Chemical Data Bases Directory
- Chemical Structure/Nomenclature System
- TSCA Chemical Data System (Proprietary)
- TSCA Chemical Data System (Public)
- TSCA Reports Management System
- Toxicology Data System
- Chronic Testing Support System
- Bibliographic Literature Scanning System
- Laboratory Animal Data System
- Regulated Chemicals Standards System

Other data bases and systems must also exist to provide access to information on additional chemical substances and other categories of data. Access to these non-core components is by referral with coordination provided by the Chemical Data Bases Directory. Compatibility between data formats, nomenclature, data base management systems, and overall system capabilities is less critical for these non-core systems. In some cases, these non-core components are repositories for categories of data similar to those contained in the core components, but the set of chemical substances for which the data are maintained is specific to certain legislative mandates or research responsibilities. The specific contents of each of the "core" components of the network is discussed with what follows.

5.3.2.1 Chemical Data Bases Directory. Within the core components, the Chemical Data Bases Directory (CDBD) is the pivotal file in that it is a "help file" and provides detailed information on the nature of the data bases/systems in the network. It directs the user to data systems which will satisfy his requirements for information. It includes component file identifiers, data element identifiers, and a general discussion of the types of compounds for which there is data coverage. It does not identify specific chemicals for which there is coverage. It indicates the specific mode of access, including file names, telephone numbers, file ownership, file location, system characteristics, size of file, update frequency, searching capability, and output media. The CDBD provides standardized data element terminology for all data elements in the core component systems. The Directory also includes references to non-core files that may maintain other data element names, with the Directory indicating the necessary cross-reference terminology.

The Directory file must be widely available to the general public, and structured for easy access. Maintenance responsibilities will be shared by individual file owners, but the data resource administrator of the network will have full responsibility for updating and maintenance of file integrity. Section 5.5 contains a further discussion of the data base administrator and network management.

5.3.2.2 Chemical Structure/Nomenclature System. The Chemical Structure/Nomenclature System is the second critical element of the

comprehensive Chemical Substance Information Network. This system provides chemical identification data for approximately 500,000 chemical substances. It provides a sub-structure searching capability and a locator designator which points to other files in the system containing information on that particular chemical substance. The size of the file is important, because this file must serve all agencies concerned with the study and/or regulation of chemicals. It must contain chemicals that are used as drugs, pesticides, industrial chemicals or those of research interest. The file must be searchable by CAS number, CAS preferred name, synonyms, structure, structure fragment, nucleus probe, molecular weight, etc. System output must include display of the structure. The system must also be usable without extensive knowledge of chemistry. The locator designator (referencing all relevant files which contain the chemical) is clearly feasible and should be an integral part of this system. Updating such a system is clearly a sizeable responsibility since this system contains the critical data linkage elements (CAS number, synonyms, other identification codes and structure).

A continual interface with Chemical Abstract Services will be necessary to allow for updates to the file as CAS numbers change, and new chemicals with their respective registry numbers and structures are added to the file. Industry files and other government files will also require updates when there are changes in the Chemical Structure/ Nomenclature System.

The Chemical Structure/Nomenclature System must be publicly and widely available. As defined above, this system builds on and includes features of both the present CHEMLINE and CIS/SSS systems.

5.3.2.3 TSCA Chemical Data Systems. The TSCA Chemical Data Systems are also major components of the network and essentially provide much of the critical data associated with chemical compounds in commerce. The systems use a hierarchical file structure with the chemical compound being the key data element. It is envisioned that the record hierarchical structure would be similar to the general scheme illustrated in Figure 5-3. The systems contain varying amounts of data on the approximately 50,000-100,000 chemical compounds in commerce and those chemicals that are subject to pre-manufacturing review.* The systems are constructed primarily from data submitted as a result of regulations promulgated by EPA under TSCA and may contain extensive amounts of confidential information. The TSCA Chemical Data Systems contain both unevaluated and evaluated data, (e.g., reviewed testing data). They are the source of, and home for, chemical information necessary for environmental and health hazard analyses (i.e., as defined in Table 2-1). Beyond providing a structured data base, the TSCA Chemical Data Systems must provide,

*Although data bases similar to the TSCA Chemical Data Systems are required to contain similar data for other types of chemicals (e.g., pesticides), it is not recommended that these other data bases be considered core components of the network. Cross-reference linkage to these systems provided by the Directory, results in their inclusion in the Chemical Substances Information Network.

an analytical data manipulation capability to permit the system user to identify correlations and interactions between various categories of data by allowing the creation of specific temporary subsets of the data file.

Because of the state-of-the-art of the technology associated with protecting confidential data and the potential for inadvertent disclosure, METREK recommends creating two systems: one that is a proprietary system and one that is a public system. Direct access to the proprietary system is limited to those persons in EPA and other government agencies who are "approved users." EPA is responsible for releasing non-proprietary data immediately into a second file for repackaging to make it publicly available to a large number of users simultaneously. EPA is also responsible for making decisions concerning repackaging of the proprietary data. Industry representatives have expressed considerable concern about the release of confidential data, but agree that summarized or tabulated data or data aggregated in ranges are acceptable. Consumer groups and environmentalists are seeking release of as much data as possible in order to have that information available for scientific review and assessment. A balanced publicly available data base is the long-range goal to provide protection to the data claimed as confidential on the one hand, and on the other hand, make much of the data publicly available.

5.3.2.4 Reports Management System. The Reports Management System, while not representing a major system from the point of view

of providing either data or an analytical capability to support risk assessment-related activities, is critical to EPA's TSCA activities since it provides a record of individual corporate submissions and references to stored industrial health and safety studies and other reports. Its primary function is to provide a reports locator and tracking capability. The file is organized on a corporation basis and contains corporation identification information (name, address, etc.), plant identification and location data, and references to reports both requested and submitted on individual chemical substances.

5.3.2.5 Toxicology Data System. A Toxicology Data System is another critical element of a comprehensive chemical substance information network. The purpose of this system is to provide a structured and consolidated source of biological effects data (e.g., acute toxicity, carcinogenicity, mutagenicity, teratogenicity, and other chronic toxicity data). The system makes available test results and biological effects data for all types of chemicals or those intended for general research. It serves as a source for research data from government, industry, academics, and other international sources. It permits a user to examine chemicals analyzed using mutagenic screening tests and compare the results with in vivo carcinogenic testing. Verification of methodologies across laboratories and/or species will be facilitated. The system contains the type of study, methodology, race/age/sex, species/strain, route, site, effects, investigator, length of test, degree of evaluation of the data, and a reference. The system will evolve by combining, restructuring and enhancing capabilities currently

available in TDB, EMIC, ETIC, TOX-TIPS, the IARC Bulletin of Chemicals Tested for Carcinogenicity, the Survey of Chemicals which have been Tested for Carcinogenicity (PHS-149), the Registry of Toxic Effects of Chemical Substances, the Fish Control Laboratory Data Base and the Fish-Pesticide Research. Access to this system is through on-line terminals with the files directly linked to the TSCA Chemical Data System (Public).

5.3.2.6 Chronic Testing Support System. The Chronic Testing Support System provides a software capability and storage and retrieval module for the results of long term chronic toxicity monitoring studies. The system may be used by government agencies in the conduct of long term carcinogenesis bioassays (e.g., NCTR, NCI), by EPA in carrying out its testing responsibilities under TSCA or its other Acts, or by industry when required to conduct chronic tests in response to government regulation. The system incorporates the requirements of the Carcinogenesis Bioassay Program of the National Cancer Institute and the integrated laboratory support capability required by the National Center for Toxicological Research. It is designed to support private, independent agency or industry files with access and update privileges limited to "approved users." The primary intent of the system is to provide a computer utility for collection, monitoring, evaluating, and reporting of bioassay information. The system permits collection of data on chemicals and chemical preparations, the experimental procedures and test environment, the observation data and

complete pathology reports on individual animals. The system interfaces with various statistical application programs and a report generator. Use of such a system by government agencies and industry encourages standardization of testing protocols, forces standardization of reporting, and incorporates concepts of good laboratory practice. Summary results from bioassays should be structured for entry into the Toxicology Data System.

5.3.2.7 Bibliographic Literature Scanning System. Another major component of the distributed network is a Bibliographic Literature Scanning System containing references to toxicological and biomedical journals. It is designed to assist researchers and other health professionals in ascertaining what has been published on any specific biomedical subject, including results of human and animal toxicity studies, effects of environmental chemicals and pollutants, cancer research, and analytical methodologies. The system is searchable by CAS number, chemical name, and citation (title, author, journal, etc.). Text searching of the abstract is also permitted. This component is structured around existing systems including TOXLINE, MEDLINE, CANCERLINE and CHEMRiC.

5.3.2.8 Laboratory Animal Data System. The Laboratory Animal Data System is also recommended for inclusion in the network. This system contains information on control animals including species, strain, colony and observed terminal pathology collected from numerous government and private sources. It provides baseline information on

control animals and is useful in designing test systems and selecting appropriate species. For increased compatibility with the other components of the network, the Laboratory Animal Data System should be transferred to the data base management system selected for the network where it will be widely accessible to the public.

5.3.2.9 Regulated Chemical Standards System. Also incorporated in the network is the Regulated Chemical Standards System which provides the user with information on standards or regulations which have been proposed or promulgated concerning individual chemical substances or classes of chemicals. The system incorporates occupational standards, transportation, packaging, and labeling requirements, threshold levels, and various procedural regulations which impose industrial reporting requirements with respect to individual chemical substances or classes of substances. State, Federal and international standards are all included. The system is implemented on a data base management system that is publicy available, thereby providing information to manufacturers and processors as to their respective responsibilities under various legislative authorities. Government agencies and international organizations require this system to maintain awareness of proposed and promulgated standards in order to minimize the development of conflicting standards.

5.4 Supporting Rationale for the Recommended Network Design

The network as defined in Figure 5-1 responds to the user requirements for an integrated, comprehensive data network that can be used

for hazard identification and hazard assessment in the control of chemicals affecting health and the environment. The network is designed to coordinate collection and storage of like kinds of data and to make as much of the data available to the public as possible. It permits comparison of diverse elements of information, provides easily updated systems and on-line interactive access.

The network provides a system for OTS to maintain information collected under TSCA and make available the health and safety data in a manner consistent with the requirements stated in the EPA/OTS RFP No. WA 77-D072. It also facilitates access to a sub-structure and chemical nomenclature system for a large number of chemicals. The use of a common data base management system for all applicable component members of the network permits efficient storage of the data, eliminates redundancy of data items in separate data files, and promotes more efficient processing and accessing of information. It also enables a user to integrate information across many files providing a much broader analytical capability. The network design shows direct linkage of the TSCA Public Chemical Data System and the Toxicology Data System since much of the data residing in these systems will be needed simultaneously to respond to the type of queries where correlation among varying types of data is needed. For example, a query might require the system to identify high volume, high exposure chemicals correlated with chronic toxicity data.

There is no direct linkage of the Chemical Structure/Nomenclature System with the TSCA Chemical Data System and the Toxicology Data

System since sequential searches are acceptable to most users. However, if the Chemical Structure/Nomenclature System utilized a common DBMS facility and did not require unique software, direct linkage would be an automatic by-product available at no additional cost other than that of converting the nomenclature system to the common DBMS.

Direct linkage of systems is thus recommended only for the Chemical Data Base Directory, the TSCA Public Chemical Data Systems, and the Toxicology Data System. The Chemical Structure/Nomenclature System can be directly linked to these files if it resides in the same data base management system at the same computer facility or if the selected DBMS permits distributed data base management at different computer facilities. Direct linkage of the other files is not necessary since sequential accessing is adequate.

The systems selected as core components are included primarily because 1) the data contained therein are critical to the study and regulation of chemicals or 2) the data system's software is critically needed to store and retrieve necessary data. The core component systems potentially provide the data necessary for hazard identification, hazard analysis, and support for regulations regarding commercial chemicals as well as for enforcement activities. Coverage of large numbers of chemicals in the Structure/Nomenclature Systems and in the Toxicology Data System fulfills requirements of research groups to look at structural relationships regardless of the use of a chemical.

The Chronic Testing Support System is included to provide a sophisticated data handling capability for groups involved in long term testing resulting in large amounts of monitoring data on many individual animals.

A data system of environmental monitoring data consolidated across all media was frequently mentioned as being "desirable" by several groups interviewed. This type of system was not included as a core component of the network since the feasibility of creating such a system appears difficult and the requirement does not exist at this time. Monitoring data for select chemicals are contained in the TSCA Chemical Data System, and other existing systems are referenced by the Directory. The UPGRADE System, developed by CEQ, provides an analytical capability to retrieve data from such files as SAROAD and STORET and may answer the requirement to link environmental monitoring data. As these systems include coverage of larger numbers of chemicals, consideration may be given to consolidating summary data into a core component system.

Private files of agencies which contain large amounts of proprietary data are not included as core components but are referenced by the Directory. Such files as the product composition files of CPSC and NIOSH, the pesticide registration files of EPA and the drug application systems of FDA are examples. However, consideration should be given by these agencies to "spinning off" publicly accessible files similar to that suggested for the TSCA Chemical Data System. Agencies

with proprietary data have a responsibility to protect such data, but they have an additional responsibility to make non-proprietary data available if its release contributes to science.

Bibliographic files sponsored by Federal agencies other than those in the NLM System are not included as core components. Files such as NIOSHTIC and SWIRS should be made available on-line through a time-shared network. If the public usage is too limited to support the system in this manner, then the file should be dropped unless the respective agencies find them critical to their operations.

Inclusion of core system components in the network and actual development of the systems themselves will result from a dynamic decision process. Not only do policy decisions within EPA and other Federal agencies dictate program planning, they also impact on network development. It is important to recognize the impact of these policy decisions so that subsequent adjustments to the network design can be made as required.

New data bases, responsive to particular requirements, must continue to be developed. They do not, however, always have to be developed under the umbrella of the network or as part of an existing system (even though cognizance should be given to inclusion of standardized nomenclature and compatible file structures, etc.)

5.5 Data Base Administration Responsibilities

Management of the network is best provided by an independent organization having a mandate to apply its resources to the advancement

of science by collecting, storing and disseminating chemical and toxicological information to investigators, educators, government regulatory agencies and the public at large. Responsibility for overall development and maintenance of the comprehensive network as defined in this report should be placed in an organization where crises, emergencies and program activities will not take priority over the information dissemination function. A regulatory agency typically must respond to situations somewhat beyond their control (e.g., citizen's petitions, court decisions, emergency situations) which cause continual shifts in program activities. Historically, information activities in regulatory agencies have been neglected and resources cut back or reprogrammed in times of crisis.

In the case of CSIN, EPA will have the responsibility for maintenance of the proprietary data collected under TSCA. Furthermore, EPA will have the responsibility to separate the publicly releasable information from the proprietary data. However, the maintenance of the resulting public file does not need to be an EPA function. It can physically reside in a government-owned or a privately-owned computer accessed through a time-shared network.

The interagency committee authorized in section 10(b) or the Council on Environmental Quality as designated by section 25(b) can provide advice concerning which office should have the designated responsibility for the network, and can continue to serve in an advisory capacity as the network develops. A data resource administrator should be selected who is responsible for the design,

development, operation and maintenance of the system. Cognizance should be given by the data base administrator to the relationship between the implementation of the reporting provisions of the Toxic Substances Control Act and its impact on the network development. In addition, development of other component systems of the network must be appropriately scheduled in a manner consistent with the user requirements.

During the development of the network, considerable attention must be given by the network management to the creation of publicly available data bases and packaging of data to serve the diverse community of users. Consequently the data resources administrator should possess sufficient knowledge of user applications to perform a satisfactory trade-off among user demands.

The evolution of a standardized nomenclature is a requirement for the continued maintenance of the directory and locator designators. Sensitivity to the problem of unevaluated data versus evaluated data must be recognized and handled. Where possible, references and sources must be tagged. Where there are no citable references, greater detail must be provided in the systems record to allow user evaluation of the data. Maintenance of data integrity and data currency of the core component systems are additional responsibilities of the data resources administrator.

6.0 RECOMMENDED SYSTEMS DEVELOPMENT OPTIONS

In Section 5.0 a comprehensive information network to satisfy user requirements for information on chemical substances is defined. It was noted that the specific data gathering policies and plans of EPA will have a direct impact on recommendations concerning development options for data systems. To provide a framework within which specific and detailed recommendations could be formulated, three data collection scenarios are defined. In this section, systems development recommendations are made in response to those scenarios.

6.1 Clarification of Scenarios and Their Systems Development Implications

Prior to discussing the specific system development recommendations, the implications of the various data collection scenarios must be considered. Each scenario must be analyzed both with respect to its specific data base and system options and to the setting of priorities for systems development.

The first scenario assumes that EPA will collect site specific production information as a part of the inventory reporting under section 8(b) of TSCA and that it can be processed within the next three years. It is further assumed that a regulation under TSCA section 8(a)(2) will require submission of information on amounts produced by each category of use, descriptions of by-products resulting from production, uses, environmental and health effects, exposure information and the methods used for disposal for approximately 1,000-2,000 chemical substances of particular interest to EPA.

At the end of the initial three-year time period, EPA will reach a decision point. At this time, a choice will have to be made between continuing the limited data gathering activities described under Scenario I or initiating an increased data collection policy, i.e., the second scenario. Should EPA decide not to increase their data acquisition activities, no changes in their systems development plans would be required. If, however, the decision favors adoption of the second scenario with its increased data collection requirements, changes in systems development activities must occur.

Under the second scenario, it is assumed that EPA will initiate a policy requiring submission of information on use, users and exposure in addition to the information already being collected. Further, it is assumed that the list of chemical substances for which reporting under TSCA section 8(a)(2) is required will be extended to include a total of between 7,000 to 10,000 chemical substances. External information files, previously developed for other purposes, are used under Scenario I to provide some use, user and exposure data. Since Scenario II provides for these data to exist in the TSCA Chemical Data Systems, access to external files containing use/exposure data would no longer be critical. Continued maintenance of these files would have to be predicated on a justification other than TSCA.

After approximately five years, EPA is assumed to make a second policy decision. This decision involves a choice between continuing the Scenario II data collection activities or initiating a policy to fully implement all data collection activities authorized under TSCA

section 8(a)(2). This would add to the TSCA Chemical Data Systems information on by-products, environmental and health effects and disposal methods for all chemical substances on the inventory. It is further assumed that under this third scenario, EPA will implement regulations requiring reporting of new uses of chemicals already on the inventory in accordance with section 5(a) of TSCA.

Initiation of the third scenario will result in a major expansion of the volume of data held for all chemicals on the inventory in the TSCA Chemical Data Systems. This expansion will permit the satisfaction of user requirements for all inventory chemicals which were previously only satisfied for selected chemicals. It will also facilitate the data searching activities required to access the information, for under Scenario III, most information previously only available from external files will be available in a single system. However, EPA and other regulatory agencies will always have to rely on outside sources such as the scientific literature, reports from the research agencies, epidemiological studies, etc., for science-based decision-making.

The implications of the three scenarios with regard to information required from industry, and external files needed to supplement this information are integral to the developmental systems recommendations presented in Section 5.3. A greater emphasis is placed on the consequences of interactions between the first two scenarios since adoption of these by EPA is considered most likely. It should be noted that

full development of the Chemical Substances Information Network is the ultimate goal of all recommendations regardless of the specific data gathering scenario.

It is apparent that as EPA collects data under the TSCA reporting provisions, including sections 4, 5, and 8, the TSCA Chemical Data Systems will increase in size and dependence on other systems which only partially satisfy their information needs will diminish. Data bases, however, will continue to be developed which respond to specific Federal responsibilities with respect to the TSCA chemicals and others not covered under this Act (e.g., drugs and pesticides). Network components which satisfy multiple user requirements and which contribute to the satisfaction of EPA stated goals for implementing TSCA, are given priority in the design of the network. In the long term, network development will be accomplished by increased data collection and enhancement of the TSCA Chemical Data System. Simultaneously, concurrent enhancement or development of other core components must occur in a manner consistent with 1) TSCA implementation plans, 2) network user requirements, 3) available funding, and 4) a willingness on the part of concerned Federal agencies to cooperate in data acquisition and data base development.

6.2 Scenario I Systems Options

As noted above, the TSCA implementation strategy, which impacts the design of the first scenario, includes the collection of site specific production data for all chemicals on the inventory during the

next year. Therefore, the user requirements for this information would be satisfied by the TSCA Chemical Data Systems to be developed by the Office of Toxic Substances. Under the assumptions of the first scenario, EPA obtains information on use, exposure, and biological effects data for only about 1,000 - 2,000 chemicals on the inventory under 8(b) or 8(a)(2). This, then places greater near-term reliance on existing data bases to satisfy the identified user needs expressed in Section 2.5.1.

In Figure 6-1 existing data systems and systems to be developed in the future are displayed. The relationships of existing data bases to the planned data bases are presented in such a way as to illustrate the modular development of the network. Systems that are required, during the interim, to at least partially satisfy user requirements are discussed below in the following text as well as those data bases which are recommended as integral components of the planned network.

In terms of responding to identified user requirements, under Scenario I, within the next two or three years site specific production information will be contained in the TSCA Chemical Data System with a public file being developed. Chemical substances identification data are available from CHEMLINE or from the Chemical Information System/Substructure Search System. Marketing and use data, exposure data, biological and epidemiological data, and environmental effects data for chemicals of concern to the interviewed community of users will be only partially available from a variety of systems.

6.2.1 Directory Development Recommendations

The Chemical Data Base Directory is of great importance since it will be the focal point for information on data bases and reference sources. Construction of the Directory should be given the overall highest priority and should begin as soon as possible. Although responsibility for construction can be decided by the TSCA section 10 interagency committee or the section 25(b) CEQ Committee, responsibility for administration and maintenance of the Directory is logically within the National Library of Medicine since their system currently provides terminal access to a large number of potential users of such a directory. They have already initiated preliminary design work for a Directory. The Chemical Information System, through its time-shared network, could also serve as a temporary residence of the Directory -- the only differences being that the existing users of the CIS tend to be a limited subset of the potential user community, and that the system utilizes a private computer network rather than a Federally-funded computer network.

6.2.2 Nomenclature and Structure Development Recommendations

As the next highest priority for the network, it is recommended that CHEMLINE and CIS/SSS be enhanced along the lines of the present planning for these files. Beyond those plans, it is recommended that CHEMLINE include a locator designator for all files identified in Figure 6-1 with primary attention being given to those files which become merged or contribute to "core component" files. Improvements to the CHEMLINE structure searching capability should also continue.

For CIS/SSS, it is recommended, beyond the current plans to increase the chemical substance coverage, that a nomenclature search capability and a locator designator be provided. Enhancements to substructure searching features of CIS/SSS are also necessary at this time since the desirable state-of-the-art has not been reached.

Substructure searching is also inherent in the Army's CIDS system. It is recommended that a coordinated activity in terms of funding and development of a unified Chemical Structure/Nomenclature System be initiated in the near future with the specific objective of planning for development of the more comprehensive system described in Section 5.3.1. An indication of the advantages and disadvantages of these systems with respect to nomenclature and structure searching is presented in Table 6-1. A more definitive evaluation of these systems with respect to their structure search capabilities is desirable and is recommended.

Moreover, for all existing systems, emphasis must be placed on chemical substance identification, since these data elements become the critical linkages or connections between existing data bases. Chemical Abstract Service preferred names (the widely accepted standardized nomenclature for chemical files) are used in a number of files, but the majority of files have not been name-matched and provided with CAS numbers and names. Clearly, use of a CAS number provides a universally acceptable standardized nomenclature and its use should be encouraged. EPA, through the Chemical Information

TABLE 6-1

SELECTIVE COMPARISON OF STRUCTURE SEARCHING APPROACHES

ASPECT	CHEMLINE	CIS/SSS FILE	CIDS

ADVANTAGES	• LARGE NUMBER OF CHEMICALS
	• LOCATOR FILE
	• SEARCHABLE BY
	  • CAS NUMBER
	  • CAS NAME
	  • SYNONYM
	  • WLN
	  • MOLECULAR FORMULA/WEIGHT
	  • RING CHARACTERISTICS
	  • NAME FRAGMENTS
	• PUBLICLY AVAILABLE
		• SEARCHABLE BY
		  • CAS NUMBER
		  • SUBSTRUCTURE COMPONENT
		  • CIDS KEYS
		  • NUCLEUS PROBE
		  • ATOM BY ATOM APPROACH
		  • MOLECULAR FORMULA
		  • MOLECULAR WEIGHT
		• COMPOUNDS FROM MANY FILES INCLUDED
		• PUBLICALLY AVAILABLE
		• SYSTEM BASED ON CAS CONNECTION TABLES
			• SEARCHABLE BY
			  • MOLECULAR FORMULA
			  • STRUCTURAL FRAGMENTS
			• SHORT STRUCTURE SEARCH LEARNING TIME
			• GOOD STRUCTURAL DISPLAY CAPABILITY

DISADVANTAGES	• NON-CYCLIC STRUCTURES SEARCHABLE ONLY BY NOMENCLATURE • FEW FILES INCLUDED IN LOCATOR • NO STRUCTURE DIAGRAM ENTRY AND RETRIEVAL CAPABILITY .	• NO NOMENCLATURE SEARCHING CAPABILITY • MUST SEARCH EACH FILE INDEPENDENTLY • STRUCTURE SEARCH METHODOLOGY DIFFICULT TO LEARN AND REQUIRES MORE ADVANCED CHEMICAL KNOWLEDGE • STRUCTURAL DISPLAY NEEDS IMPROVEMENT	• LIMITED KEYS • LIMITED TYPE OF CHEMICALS INCLUDED • SPECIAL HARDWARE REQUIRED FOR PRINTOUTS • NOT WIDELY AVAILABLE TO PUBLIC

System, has registered a large number of files and made the CAS number, name and structure available through CIS/SSS. This has been extremely useful in terms of standardizing nomenclature and making structure information for these chemicals available. It is important that owners of the file follow the registration of the CAS name and number with incorporation of this information into the file. Priority must be given to name matching files which will be needed in the interim and which are identified in Figure 6-1.

Data elements such as the CAS number, CAS name, or Wiswesser line notation code (WLN), when present in more than one file, can provide a linkage between those and other files, also containing these data.* Figure 6-2 examines the substance identification data elements included in a number of relevant files as reported in the CEQ Survey and identifies the common data elements which would permit file interconnections.** It can be clearly seen that the common link is the non-standardized chemical name or synonym.

---

* A data base mapping model and search scheme was developed at the University of Illinois under National Science Foundation support in order to test the feasibility of data element linkage among various chemical files. Results demonstrated that use of a consistent scheme for classification of data bases by subject and common data elements greatly increases the potential for accessing data bases.

** In developing Figure 6-2 no indication of data items other than CAS No. were cited unless it was definitely known that those additional items had been incorporated into the file after it was name matched.

6.2.3 Toxicology Data Systems Development Recommendations

Biological effects data for selected chemicals are available during the interim from TDB, EMIC, ETIC, TOX-TIPS, the Registry for Toxic Effects, the IARC Bulletin of the Survey of Chemicals Being Tested for Carcinogenicity (PHS-149), the Fish Control Laboratory Data, and Fish-Pesticide Research. Physical/Chemical property data are available from TDB, CIS, and the Thermophysical Properties Research Center. Since biological effects data were cited in the user requirements study as being necessary, immediate consideration should be given to the feasibility of developing an interactive system containing data on a wide variety of chemicals. The 1,000 - 2,000 chemicals for which OTS is considering requesting 8(a)(2) data in this calendar year are prime candidates for inclusion in TDB. TDB should continue to be enhanced during the interim period, with major attention being devoted to the chemical and biological effects data, and minimal effort made to include production data since this will be available in the TSCA Chemical Data System.

6.2.4 Exposure/Use Systems Development Recommendations

Other critical data categories identified in the user requirements survey include use and exposure data. Systems which provide some of these data include the NCI/SRI Research Chemicals That Impact Man, the U.S. International Trade Commission Data Base, the Mineral Commodity Survey, the Chemical Economics Handbook, Dun's Market Identifiers, and the National Occupational Hazard Survey. Decisions

to incorporate these systems into the network and enhance them by extending their coverage are predicated on EPA's strategy relative to data collection activities. For example, the NCI/SRI system provides exposure profile data for approximately 3,200 compounds (some of which are pesticides, cosmetics and drugs). The system provides the best attempt to date to model the uptake of chemicals by biological systems as a result of use and exposure data which SRI collects from various sources. It would provide the network with a source for limited amounts of these data. Consideration must also be given to the economics of enhancing this data base to assist in satisfying user requirements and the recommended coverage of chemicals.

The use and exposure data collected under section 8(a)(2) will need to be supplemented with body uptake information. The NCI/SRI data base, since it now includes a methodology for generating these update data, would be a logical candidate for federal support. If this were decided to be the case, the NCI/SRI data base should focus on generating uptake information on those chemicals selected by EPA for section 8 reporting, excluding from their operation the obtaining of the use and exposure data (these would be supplied to them by EPA). An Alternative to support of the NCI/SRI data base would be the development of an update algorithm through interagency R&D funding by those agencies requiring this information (e.g., EPA, NIOSH, NCI, FDA, CPSC) with lead responsibility assigned to one agency. The value and cost of generating these uptake data through NCI/SRI or through an interagency

agreement will have to be considered in light of EPA's expected decision to require use and exposure data.

If EPA defers the decision to collect these data for all chemicals in 1980 and continues in a first scenario data collection mode, then enhancement use and exposure as well as uptake components of the NCI/SRI file as part of the network would be a viable alternative to reaching the long-term objective. It is thus recommended that the NCI/SRI file be referenced as a relevant file for Scenario I. In addition, consideration should be given to carefully enhancing its coverage in such a way that it supports and does not overlap with EPA's industrial reporting plans. As decisions are made in EPA regarding data collection of production and use data, additional adjustments can be made concerning further enhancement of the NCI/SRI file or the initiation of a new interagency research effort.

Furthermore, the Directory should provide pointers to the files and reference tools mentioned above as being potential sources of production and use data during the interim stages in the development of the Chemical Information System Network.

6.2.5 Development Recommendations for Other Systems

Other data requirements for physical/chemical property data, environmental effects data, epidemiological data and additional economic data can be partially satisfied by results published in the open literature and by current research studies. The literature scanning activities of the NLM and the other agencies, professional

societies, etc., which are made publicly available through time-shared government and private networks are a vital part of the interim system. Relevant bibliographic files such as TOXLINE, MEDLINE, CANCERLINE, SWIRS, NIOSHTIC, and those available through SCD, LOCKHEED and BRS, are to be referenced by the Directory.

The two files developed for EPA by Radian and PEDCO provide process data for selected chemical industries. As noted in Sections 4.5.2 and 4.8, environmental effects data and environmental monitoring data are not readily available for a wide range of chemicals. As additional data are collected, efforts should be made by EPA to incorporate them into AEROS, STORET or other appropriate systems for wider dissemination.

During this interim period, product composition data can be obtained in varying degrees of specificities from the CPSC System, the NIOSH System, CTCP, POISINDEX and the Poison Control On-line Inquiry System. EPA users felt that these data bases were particularly valuable for obtaining use information, but as specific use information became available to them through the TSCA Chemical Data Systems, they would not use these systems as extensively.

Existing sources of some relevant epidemiology data include NEISS, National Center for Health Statistics, the Atlas of Cancer Morality and other systems identified in Section 4.6. None of these sources specifically respond to user requirements for epidemiological or adverse

effects data. These needs would only be met if EPA implements the section 8 reporting and recordkeeping requirements.

The Directory and ultimately the locator in the Chemical Structure/Nomenclature System will provide references to manual files as well as automated files. The Merck Index, the Chemical Economics Handbook, etc., continue to be needed during this interim period to provide varying types of data.

6.2.6 Limitation on Recommendations

The additional components typically included in a system development plan (such as cost considerations, personnel requirements, specific recommendations for both software and hardware capabilities, and required storage capacities) are not addressed here since they are not in the scope of this effort. These components are the subject of a second, concurrent effort by an independent analysis team and will be published separately.

6.3 Scenario II and III Systems Options

6.3.1 Scenario II Systems Implications

A Scenario II assumption is that EPA will initiate a policy of requiring submission of information on use, users and exposure for all chemicals in the inventory in addition to site specific production information. It is further assumed that the list of chemical substances for which 8(a)(2) reporting is required will be extended to include a total of 5 to 10,000 chemicals. With the increased data being collected by EPA, the content of the TSCA Chemical Data System will be greatly

increased. Less reliance will therefore be placed on external systems capable of providing limited interim use and exposure data for those chemicals that fall under the jurisdiction of TSCA. Data bases of lesser concern include the NCI/SRI, the U.S. ITC Data Base, Dun's Market Identifiers and The Mineral Commodity Survey. Reference tools which diminish in need include the Chemical Economics Handbook and the Kirk-Othmer Encyclopedia of Chemical Technology.

The core components described as part of the long range objective and illustrated in Figure 5-1 are still required to satisfy user requirements to provide a more comprehensive system that will be useful in carrying out the purposes of TSCA. The priorities for implementation of the core components do not change from those identified in Scenario I. The major difference between the first and second scenarios is that under the second scenario, use and exposure data requirements for commercial chemicals will be more adequately satisfied by the increased reporting requirements and less dependence on the other systems is necessary. That, in essence, is the only significant change in the Scenario II systems development option.

6.3.2 Scenario III Systems Implications

In Scenario III it is assumed that EPA exercises its full 8(a)(2) reporting requirements for all chemicals in the inventory. In addition, it is assumed EPA implements significant new use reporting under section 5(a). Implications of these assumptions are that both the proprietary and public TSCA Chemical Data Systems will be

greatly expanded as far as the number of chemicals included. In addition, stated user requirements for data previously available only for selected chemicals (e.g., by-products data) will be available on a large number of chemicals. Scenario III assumptions do not impact on the systems included in the network or referenced by the Directory. Reliance on external systems to partially satisfy data needs is required to the same extent as in Scenario II. The Scenario III increased reporting causes previously unmet data needs to be satisfied more fully. Development of the core components of the network is just as critical under Scenario III assumptions as under those of I and II, and furthermore the priorities for implementation remain unchanged.

6.4 Other Considerations of Systems Development Options

6.4.1 Systems Options, Their Compatibility and Development

Comparing of Figure 5-1 with Figure 6-1, one can see the similarity in basic design structure from a user point of view between a network which has the potential to be responsive to user requirements and the currently existing systems which are partially responsive to some data requirements and unresponsive to others. Previous discussion has emphasized that satisfaction of user requirements is predicated on the ability to obtain access to varying types of information necessary to make assessments concerning the hazards of chemicals and their impact on man and the environment. Although much of this information

can be obtained by EPA using the industrial reporting provisions of TSCA, much of this information must be generated from additional testing and research.

As new data become available, they must be collected, structured and made available in systems for easy retrieval. The Chemical Substances Information Network provides the potential structure for these systems. It potentially satisfies needs for substance identification data, production, marketing, and exposure data. It will also provide a centralized source of existing epidemiological data, biological effects data and environmental effects data for commercial chemicals. Information on standards and regulations with respect to chemical substances that have been promulgated by international, Federal, state and local governments will be available.

Scenario I systems development options satisfy user requirements for systems identification data (that is, chemical nomenclature and a structure search capability) and for site specific production data for chemicals in the Inventory. It does not satisfy requirements for use, exposure and biological effects data nor does it provide adequate data on epidemiology or environmental effects. It provides for development of a Directory file which points to existing systems where useful data can be found. However, coverage of these data bases is very weak with respect to some categories of data (e.g., environmental effects data) and not well coordinated for others (e.g., biological effects data).

Scenario II systems development option satisfies the user requirements for substance identification data, site specific production, use and exposure data. These data fulfill some user's specific requirements associated with the hazard identification function. Other user requirements are still unmet by Scenario II system options (e.g., biological effects data, epidemiology data and environmental effects data for all chemicals on the Inventory.)

Scenario III satisfies previously unmet requirements and provides for collection of all 8(a)(2) data for all chemicals on the Inventory. However, these data, available in a structured data base, do not in themselves respond to all user requirements. Linkage with other component systems of the network is critical to coordination of the entire spectrum of data which must be considered when making hazard evaluations on chemicals or establishing regulations affecting their control or release into the environment.

Network development is evolutionary and is dependent on EPA decisions to implement TSCA. The approach most likely to be taken by EPA toward implementation of section 8 rulemaking, will probably most closely resemble the data collection activities described in Scenario I. If this is the case, development of the network must proceed by building on existing systems capabilities. EPA may make a decision to increase section 8 data collection activities some time in the future and it could well choose an incremental approach such is suggested by Scenario II. The actions of EPA significantly impact on the design

of the Proprietary and Public TSCA Chemical Data Systems. In addition, EPA's actions affect the design of the network in terms of decisions to enhance existing data bases or to build new data bases to obtain information that might, otherwise, be collected under section 8 of TSCA*.

EPA actions do not, however, impact on the design of the other core component systems described in Section 5.3.1. Development of these systems must be concurrent with development of the TSCA Chemical Data System no matter what data collection scenario is in place.

6.4.2 Time-phase Implementation of the Core Component Systems

The general events associated with the concurrent development of core component systems and associated time frames are presented in Figure 6-3 which assumes Scenario I data collection option as the initial starting point. The figure also includes the events associated with Scenario II and III systems development recommendations. In some cases, lead agency responsibilities for systems development are identified.

The figure presents a definition of existing systems on the left-hand side and the network component objective on the right. The figure indicates the point in time at which the existing systems are consolidated, restructured, or enhanced. This is illustrated by

---

* These decisions are further complicated by the fact that EPA has the unique authority to collect these data. Any data bases developed or enhanced would require extensive contractor support with no Federal authority to obtain such data from industry.

the merging of horizontal systems lines or the positioning of vertical lines indicating initiation or termination of specific events. Events relating to more than one system are indicated by a box overlaid onto all affected systems.

The time phased implementation for the core components shows development of an interim directory after one and one-half years. This capability is augmented with the ability to access CHEMLINE and CIS/SSS for chemical identification and structural data on specific chemicals. The TSCA Reports Management System is to be operational in 1978 to be responsive to the assumed initial submissions of 8(a)(2) data, inventory data, pre-manufacturing and testing data. The requirements for a system, as expressed in the RFP No. WA77-D072, are not inconsistent with the recommendations made in this report. The RFP's work statement specifies a system which can provide for storage and retrieval of data submitted as a result of regulations promulgated under TSCA. This system would encompass the Reports Management System and the TSCA Proprietary Chemical Data System as described in this report. Recommendations for a subsequent public system are not explicitly stated in the scope of work of the envisioned contract.

The recommended major events associated with the consolidation of existing systems containing biological effects data are also presented in Figure 6-3. A feasibility study to consolidate existing systems with biological effects data into the Toxicology Data Bank is recommended for initiation within year one. Subsequently, software

modifications to TDB are required. It is recommended that EPA and NIEHS take the lead responsibility in coordinating mutagenic data, and NCI in consolidating existing carcinogenic data. NIEHS would be the appropriate agency to coordinate teratology data, NLM and NIOSH would take the responsibility for structuring acute toxicity data using TDB and the Registry of Toxic Effects. Overall responsibility to assure the development of the Toxicology Data System in a timely manner is the responsibility of the network management. It is proposed that the Registry of Toxic Effects be an integral part of the Toxicology Data System. The yearly publication of the Registry (as mandated by Occupational Safety and Health Act of 1970) in the long term will be a product of the Toxicology Data System. EMIC and ETIC, as analysis centers, may still be critically needed operations in the long term with most of the efforts being devoted to evaluation and review of data.

Development of the Toxicology Data System is dependent on the availability of resources. Figure 6-3 indicates mutagenic and carcinogenic data are input into the system within three years. Acute toxicology data and teratology data are loaded into the system in year four. Metabolism data are entered in the system during year five. It is conceivable that all of these data could be entered into the system concurrently if funds are available, but the assumption is made that resources for development of this system are limited. Consequently, priority was given to mutagenic and carcinogenic data since the

existing data are not well coordinated and such a priority best satisfies user needs.

A study to examine the feasibility of developing a Chronic Testing System capable of being responsive to the requirements of NCI, NCTR, EPA and industry is recommended and NCI should take the lead responsibilities. The system will incorporate the best features of the Carcinogenesis Bioassay Data System developed by NCI and the National Center for Toxicological Research Integrated Laboratory Support System.

EPA is recommended as being the lead agency for the development of the Regulated Chemicals Standards System. Priority for inclusion in the data base is given to Federal standards associated with commercial chemicals with subsequent attention be given to state, local and international standards. Standards affecting pesticides should be entered next into the system.

6.4.3 Compatibility of Component Systems

It is recommended that most of the core systems be connected by a common data base management system. This serves to facilitate cross exchange of information, direct linkage of files when necessary and retrieval using a common command language. Similarly, it is recommended that standardized nomenclature or data element terminology be utilized by all core systems in the network. Standardization of nomenclature is never easy to implement and many systems are usually affected. The difficulties usually result from the inability of all system participants to agree on a standardized vocabulary. In the case of CSIN,

arriving at standardized nomenclature may be somewhat easier since most of the components are new or are being developed from existing files. This difficulty must be addressed in the development of the Directory and the subsequent development of other core components.

An alternative to achieving complete conversion of existing nomenclature is the use of minicomputers. These would function essentially as "black boxes" which provide conversion routines to first translate user preferred terminology to the standardized terminology employed by the systems in the network and second, translate into the corresponding terminology employed by the individual data systems to be accessed. This conversion is transparent to the user. Use of a minicomputer increases the flexibility of the users of the system by not requiring their learning the standardized network terminology. Conversion routines can be written to update core system components for feeder systems which employ different nomenclature. As systems are designed for the network, standardized terminology would be used.

The same approach can be utilized for systems that require unique software and do not convert to the common DBMS. A "front end" or "black box" can be employed which permits interrogation of the system through a "macro query language" which in essence connects the specialized software into what appears to be the DBMS.

6.5 Network Development and Management

Development of the Chemical Substances Information Network is operationally feasible and clearly within the state-of-the-art of computer technology. Success of such a network as far as the users are concerned is predicated on their ability to obtain data necessary to carrying out their functional responsibilities of hazard identification, hazard analysis, research, regulation, development and compliance and enforcement. The authority to obtain much of the data, which previously had not been available, now exists. Difficulties of protecting confidentiality of such data exist, but are not insurmountable with proper data handling procedures. Difficulties are also encountered in packaging proprietary data to make it publicly available. Some of these have been handled before to the satisfaction of concerned parties, but this is clearly an area where more innovation is required. Clear delineation of how data are to be used will assist in data aggregations and data packaging.

Success of CSIN is also dependent on the management of the network development and financial support provided. This critical area, discussed briefly in Section 5.5, involves the designation of an agency with the responsibility for data base administration. A decision regarding network development and management responsibilities should be made as soon as possible. CEQ, under its section 25(b) responsibilities, or EPA through the section 10(b) Interagency Committee might make recommendations as to the appropriate agency to undertake this responsibility.

EPA has an explicit responsibility to develop a system for the data submitted under TSCA. However, there is no requirement that the public system developed from these data or the other components of the network needs to reside in EPA. Arguments are presented in Section 5.5 as to the merits of having the network management placed in an independent organization which is not subject to frequent shifts in program priorities.

The decision concerning the physical location and selection of the executive computer and its backup capability is also important. Implementation of the network could be modelled after the National Library of Medicine's system which uses a Federally funded computer and provides for its own program and systems support, or it could be modelled after CIS which utilizes a private contractor who is responsible for the system's support and marketing.

There are pros and cons to both approaches. One could argue that an internally supported system would result in increased control over network development and, consequently, the assurance of adequate systems maintenance. On the other hand, the private contractor would be responsive to the market demands, and would provide for continued enhancements to the components of the network that prove to be self-supporting or profitable. The final decision should depend on which one is more economical in satisfying the performance standards.