

ated the operational procedures needed for a data indexing scheme. It was established that a tagging mechanism could be incorporated into the routine production operations of document analysis. Important elements to be considered in such a data-tagging technique are a combination of the number of types of numerical data to be tagged as well as their frequency of occurrence in the source documents.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under Grant DSI75-03491 and was conducted in cooperation with the Department of Energy/Oak Ridge National Laboratory.

REFERENCES AND NOTES

- (1) "Data Needs for Energy", CODATA Bulletin No. 31, International Council of Scientific Unions, March 1979.
- (2) Committee on Data Needs, Numerical Data Advisory Board, Assembly of Mathematical and Physical Sciences, "National Needs for Critically Evaluated Physical and Chemical Data", National Academy of Sciences, National Research Council, May 1978.
- (3) "Evaluating Quantitative Data Information for Research and Development", Numerical Data Advisory Board (NDAB) pamphlet (undated).
- (4) "Flagging and Tagging Data", CODATA Bulletin No. 19, International Council of Scientific Unions, June 1976.
- (5) Murdock, John W. "Current Knowledge on Numerical Data Indexing and Possible Future Developments", Informatics, Inc., Rockville, MD; report 1978; order No. BP279924, 58 pp. (English) Available from NTIS.
- (6) "Introduction", *Chem. Abstr.* 1980, 93(1), ix.

Chemical Information Resources Directory: An Integrating Component of the Chemical Substances Information Network¹

RONALD A. RADER* and SHASHI P. SOOD

Environmental, Information, and Safety Systems Department, MITRE Corporation, Metrek Division, McLean, Virginia 22102

Received November 3, 1980

The Chemical Information Resources Directory (CIRD) serves as an integrating mechanism for the diverse information resources to be included in the Chemical Substances Information Network. The present CIRD, available as a published handbook with its two parts, the Subject Catalog and the Descriptive Catalog, uniformly catalogs and indexes the functions, content, access, and other attributes of 53 chemical-related information resources. The CIRD will soon be available as part of the prototype Chemical Substances Information Network (CSIN).

Implementation of the Toxic Substances Control Act (TSCA) requires that the most reliable information and data be available to regulate the risks associated with the manufacture and use of chemical substances. Section 10b of TSCA mandates the design and establishment of an effective system for managing and disseminating data submitted under the act and other information which could be useful in its implementation.

Subsequent to the passage of TSCA, the MITRE corporation was contracted jointly by the Council on Environmental Quality, the Environmental Protection Agency, and the National Library of Medicine to conduct a study to identify the information needs of the various institutions implementing and affected by TSCA and investigate approaches required to store, integrate, and provide access to information as mandated by the act.² Development of a Chemical Substances Information Network (CSIN) was recommended.

The CSIN concept includes a collection of core component information resources integrated in a computer network, so that they would be accessed automatically and appear as a single resource to the user, and a set of noncore (noncomputer network) component information resources which would be referenced through the network, but not directly accessible through it. CSIN, it is expected, would provide users with convenient means to identify and access diverse chemical-related information and data systems. To accomplish this, two major components of CSIN were envisioned—the Chemical Structure/Nomenclature System (CSNS) and the Chemical Data Bases Directory, subsequently termed the Chemical Information Resources Directory (CIRD).

The CIRD and CSNS would function as integrating mechanisms for the diverse information resources to be incorporated in the CSIN. These data bases would serve as locaters and support such advanced computer network capabilities as automatic user query analysis, automatic selection and querying of relevant data bases, and the integration and coordination of data retrieved from multiple computerized data bases. These locator systems provide two different tools to access CSIN information resources. The CSNS would help the user identify relevant information resources in CSIN on the basis of chemical structure, substructure, and nomenclature of substances referenced by component information resources. The CSNS would incorporate many of the present capabilities of CHEMLINE of the National Library of Medicine's online network and the Structure and Nomenclature Search System (SANSS) of the NIH-EPA Chemical Information System.

The CIRD, on the other hand, would provide the user with a tool which would identify information resources on the basis of their characteristics and subject content and provide descriptions of each resource. With the help of the CIRD, the user could identify and prioritize information resources to be accessed for information and data relevant to his needs. Presently, both locator systems are being developed independently for incorporation in a prototype CSIN computer network. The discussion below will confine itself to CIRD development.

The CIRD in its current published handbook version presents detailed characterizations of 53 information resources (listed in Table I).³ An information resource is defined from the user's point of view—it is the interface between the user

Table 1. Information Resources Covered by the CIRD

APTIC (Lockheed)
 Carbon-13 Nuclear Magnetic Resonance Search System
 Carcinogenesis Bioassay Data System
 CAS77/CAS72-76 (Lockheed)
 Chemical Dictionary On-Line (CHEMLINE)
 Chemical Economics Handbook
 Chemical Formulations of Consumer Products
 Chemical Monograph Referral Center
 Chemical Regulation Reporter
 Chemical Regulations and Guidelines System
 Chemical Transportation Emergency Center
 Chemicals in Commerce Information System
 CHEMNAME
 Clinical Toxicology of Commercial Products, 4th Edition
 Clinical Toxicology of Commercial Products Trade Name Data Base
 Directory of Chemical Producers: United States of America
 Distribution Register of Organic Pollutants in Water
 Drug Registration and Listing System
 Environmental Chemicals Data and Information Network (ECDIN)
 Environmental Mutagen Information Center
 Environmental Mutagen Information Data Base (RECON)
 Environmental Teratology Information Center
 Environmental Teratology Information Data Base (RECON)
 Information Bulletin on the Survey of Chemicals Being Tested for
 Carcinogenicity
 Kirk-Othmer Encyclopedia of Chemical Technology
 Laboratory Animal Data Base
 Library Services Office, Environmental Protection Agency (Research
 Triangle, NC)
 Mass Spectral Search System
 Merck Index: An Encyclopedia of Chemicals and Drugs
 National Center for Health Statistics
 National Institute for Occupational Safety and Health Technical
 Information Center
 Organic Chemical Producers Data Base
 Pesticide Product Information on Microfiche
 Pesticide Product Information System
 The Pharmacopeia of the United States of America, Nineteenth Edition
 POISINDEX
 Poison Control Data Base
 Pollution (Lockheed)
 Pollution (SDC)
 Pollution Abstracts/Data Courier
 Predicasts Terminal System
 Registry of the Toxic Effects of Chemical Substances/NIH-EPA Chemical
 Information System
 Registry of the Toxic Effects of Chemical Substances/Government
 Printing Office
 Registry of the Toxic Effects of Chemical Substances/National Library
 of Medicine
 SAFETY (SDC)
 Safety Science Abstracts Journal
 Structure and Nomenclature Search System
 Survey of Compounds Which Have Been Tested for Carcinogenic Activity
 Synthetic Organic Chemicals: United States Production and Sales
 Toxicology Data Bank
 Toxicology Information On-Line (TOXLINE)
 Toxicology Research Projects Directory
 TOX-TIPS

and the information contents, technologies, and societal factors which govern information/data access and retrieval. Accessibility/dissemination and retrieval/processing capabilities are as important as content and media/format to the definition and utility of an information resource. Examination of the resources covered by the CIRD illustrates this point. Among the 53 information resources are a variety of types of information resources: publicly networked and in-house only online

data bases, referral centers, reference texts, information centers, bibliographic, nonbibliographic, and numeric data bases, regulatory related management information systems, and others. These are accessible to the user using a variety of organizations, mechanisms, and technologies. Some of these resources originate from the same source or contain the same data but are organized or accessed differently.

The CIRD is composed of two main parts: a Descriptive

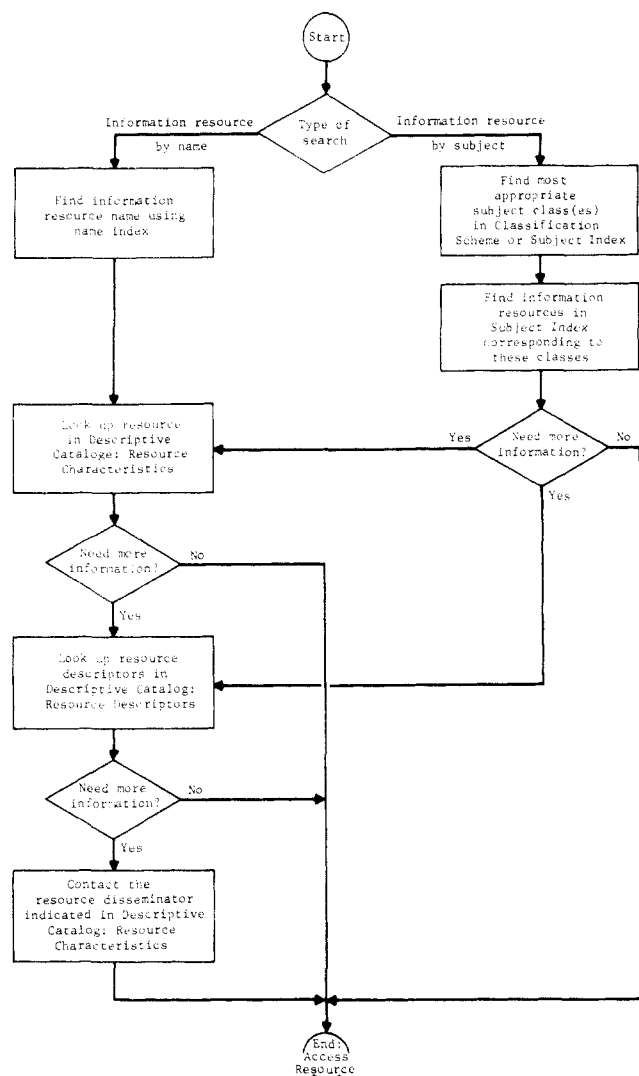


Figure 1. Schematic of CIRD usage.

Catalog containing descriptions and characterizations of the content attributes, access attributes, and organizational support for each resource and a Subject Catalog providing a hierarchical subject classification scheme and a classified index to the subject descriptors or informational units within each resource. A flowchart presenting CIRD usage and the functional relationships among its components is shown in Figure 1.

DESCRIPTIVE CATALOG

The Descriptive Catalog consists of two parts. The first portion of the Descriptive Catalog, the "Resource Characteristics" section, contains formatted descriptions of the 53 information resources. Included are information and data concerning resource identification and description, concerned organizations, access, content, and a bibliography.

The second portion of the Descriptive Catalog, the "Resource Descriptors" section, contains characterizations of each information resource's descriptors designed to uniformly format and briefly describe the useful information available through each resource and its descriptors. Descriptors are defined as the informational units, often analogous to data elements or fields, which are used consistently within a resource to structure information, provide retrieval capabilities, and/or present information or data for the user. Wherever possible, each descriptor is briefly defined, and information is provided regarding its format and permissible values, labels and abbreviations, occurrence and usage in the file, relationships with

1. Chemical Identification

- 1.1 Chemical Name
 - 1.1.1 Systematic Name
 - 1.1.1.1 CAS Name
 - 1.1.1.2 IUPAC Name
 - 1.1.2 Nonsystematic Name (common, trade, trivial names)
- 1.2 Chemical Codes
 - 1.2.1 CAS Registry Number
 - 1.2.2 Wiswesser Line Notation
- 1.3 Chemical Structure
 - 1.3.1 Molecular Weight
 - 1.3.2 Molecular Formula
 - 1.3.2.1 Molecular Formula Fragment
 - 1.3.3 Ring Structure
 - 1.3.4 Atom Centered Fragment
- 1.4 Chemical Properties
 - 1.4.1 pH
 - 1.4.2 Shelf Life
 - 1.4.3 Reactivities
 - 1.4.4 Persistence (includes environmental and biological persistence)
 - 1.4.5 Accumulation
 - 1.4.6 Degradation
- 1.5 Physical Properties
 - 1.5.1 State of a Chemical
 - 1.5.2 Color
 - 1.5.3 Density
 - 1.5.4 Boiling Point
 - 1.5.5 Freezing Point
 - 1.5.6 Melting Point
 - 1.5.7 Flash Point (includes fire and explosive potential)
 - 1.5.8 Volatility
 - 1.5.9 Solubility
 - 1.5.9.1 Partition Coefficient
- 1.6 Spectral Properties
 - 1.6.1 NMR Spectra
 - 1.6.2 X-Ray Diffraction
 - 1.6.3 Mass Spectra
 - 1.6.4 Infrared Spectra
 - 1.6.5 Fluorescence Spectra
 - 1.6.6 Ultraviolet Spectra
 - 1.6.7 Atomic Absorption
- 1.7 Chemical Class

Figure 2. Sample portion of the classification scheme.

other descriptors, retrieval characteristics, and display characteristics.

SUBJECT CATALOG

The Subject Catalog is also presented in two parts: the "Classification Scheme and System Vocabulary" and the "Subject Index to Descriptors". The Classification Scheme presents the index terms used to classify and make retrievable the subject-related content of each information resource and its descriptors. It employs four levels of hierarchy, has a unique decimal notation assigned to each term, and includes required scope notes. The CIRD presently lacks a thesaurus to provide required lead-in vocabulary and information regarding relationships among terms (e.g., broader, narrower, and related term indications).

On the basis of extrapolated user requirements and information resource analyses, nine major classes of chemical-related information were identified and used to structure the Classification Scheme and the Subject Index. These are

- Chemical Identification
- Chemical Production and Manufacture
- Trade and Commerce
- Marketable Products
- Product Application/Use
- Exposure to Chemicals/Chemical Products
- Effects/Treatment of Exposure
- Studies and Research
- Societal Controls

Another category, Bibliographic Items, was included to handle bibliographic-related descriptor information. A sample portion of the Classification Scheme is shown in Figure 2.

The "Subject Index to Information Resources", a subject index to all identifiable descriptors, will probably be the most commonly accessed portion of the CIRD. The Subject Index

1	Chemical Identification	CICIS*	-
		DROP*	-
		ECDIN	Analytical Methods
		Kirk-Othmer	-
1.1	Chemical Name	CRGS	Name
		DROP*	-
		ECDIN	Compound
1.1.1	Systematic Name	CBDS	Preferred Name
		CFCP	Ingredient/Component Names
		CHEMNAME	Descriptor; Substituent; Name Modification
		ECDIN	Other Systematic Names
		Merck Index	Derivative of Title Compound
		RTECS (GPO)	Definition
		RTECS (NLM)	Chemical Name
			Fragments; Definition
1.1.1.1	CAS Name	CBDS	CAS Name
		CHEMLINE	Chemical Abstracts Service Type-1 Name; Name Fragments
		CHEMNAME	Heading Parent; Name Match
		ECDIN	Preferred Systematic Name
		Merck Index	CA Name (8th Collective Index); CA Name (9th Collective index)
		MSSS	CAS Name
		RTECS (CIS)	Chemical Name
		RTECS (GPO)	Chemical Substance Prime Name
		RTECS (NLM)	Chemical Substance prime Name; Source of Identification
		SANSS	CAS Name
		USP XIX	Chemical Names
1.1.1.2	IUPAC Name		-
1.1.2	Nonsystematic Name	CBDS	Chemical Name; Chemotherapy Name; Synonym
		CHEMLINE	Synonyms, Name Fragments

Figure 3. Sample portion of the subject index.

uses the terms from the Classification Scheme to index and facilitate retrieval of relevant information resources' descriptors. This hierarchically arranged index presents information resource abbreviations (acronyms) and descriptor names for each Subject Index term and its numerical notation. Descriptors are assigned as many index terms as are relevant and, thus, may appear in multiple entries in the Subject Index. A sample portion of the Subject Index is shown in Figure 3.

The CIRD project had two major complementary goals: first, to conceptually design an information resources directory for maximum user and CSIN network utility and second, to develop and implement a functioning chemical information resources directory for dissemination in handbook and machine-readable form. In order to achieve these goals, the CIRD concentrates on information resources at the interface where information exchange with the user takes place and concentrates on resource descriptors as its means for presenting and indexing the subject content of information resources.

CIRD development is continuing. An online implementation has been demonstrated by the National Library of Medicine. A version of the CIRD will be available in the prototype CSIN in spring, 1981. A contract has recently been awarded for continued CIRD development and expansion through coverage of an additional 100 information resources. Eventually the CIRD will probably be available online through one or more publicly accessible computer networks.

REFERENCES AND NOTES

- (1) The work upon which this paper is based was performed under contract to Council on Environmental Quality, Contract EQ7AC010.
- (2) Bracken, M.; Dorigan, J.; Hushon, J.; Overbey, II, J. "Chemical Substances Information Network", MTR-7558; MITRE Corp.: McLean, VA, June 1977 (NTIS: PB-274-638/8GI and PB-284-638/6GI).
- (3) Levitan, K.; Bishop, T.; Kuch, T.; Rader, R.; Sood, S. "Chemical Information Resources Directory: Volume I, Development and Uses of the Chemical Information Resources Directory, and Volume II, Descriptive and Subject Catalogs", MTR-79W00128; MITRE Corp.: McLean, VA, Aug 1979 (NTIS: PB80-105190 and PB80-105208).