

Toxic Substances

---



# **Fourth Report of the Interagency Toxic Substances Data Committee January 1980**



STATUS REPORT #4 ON THE  
CHEMICAL SUBSTANCES INFORMATION NETWORK

PREPARED BY THE  
PUBLIC LIAISON SUBCOMMITTEE OF THE  
INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

U.S. ENVIRONMENTAL PROTECTION AGENCY  
WASHINGTON, D. C.

OCTOBER 1979

# INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

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

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

JAN 28 1980

Dear Colleague:

This is our fourth report on the Interagency Toxic Substances Data Committee (ITSDC), and its efforts to create a Chemical Substances Information Network (CSIN).

The following steps have been taken to ensure that all those concerned with the administration of CSIN development have an understanding of related federal and private sector information system activities:

- . The Toxicology Information Subcommittee (TIS) of the Committee to Coordinate Environmental and Related Programs of the Department of Health, Education and Welfare and the Chemical Substances Information Network Subcommittee of the ITSDC are holding their meetings back to back at the same site on the same day.
- . Maintaining a standing invitation for subcommittee members to attend each others meetings.
- . Routinely exchanging meeting agenda and minutes, and, when appropriate, technical presentations.

When feasible, we hope to coordinate project development across subcommittee lines so as to avoid duplicative efforts.

We hope you will find this information useful in keeping up with our progress and plans. If you have any comments please address them to :

Cathleen Brooks  
EPA PM-218  
401 M Street, S.W.  
Washington, D.C. 20460

Thank You,

A handwritten signature in black ink, appearing to read "Morris".

Morris Yaguda  
Chairman of the Public  
Liaison Subcommittee  
of the ITSDC

## Table of Contents

Agenda and Minutes of July 10, 1979 Meeting	
Agenda and Minutes of August 7, 1979 Meeting	
Agenda and Minutes of September 11, 1979 Meeting	
The Information Response to Chemical Crises Project	
Agenda and Minutes of October 2, 1979 Meeting	
Agenda and Minutes of November 13, 1979 Meeting	
Minutes of December 4, 1979 Meeting	
Agenda of January 8, 1980 Meeting	
CSIN Subcommittee Minutes of June 28, 1979 Meeting	
CSIN Subcommittee Agenda and Minutes of August 23, 1979 Meeting	
CSIN Subcommittee Agenda and Minutes of October 11, 1979 Meeting	
CSIN Subcommittee Minutes of November 28, 1979 Meeting	
CSIN Subcommittee Agenda of January 17, 1980 Meeting	
Toxicology Information Subcommittee Minutes of June 28, 1979 Meeting	
Toxicology Information Subcommittee Agenda of August 23, 1979 Meeting	

## INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

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

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

### AGENDA - July 10, 1979

1. Review of Minutes of June 5, 1979.
2. CSIN Subcommittee Report - Dr. Henry Kissman, NLM.
3. Toxicology Information Subcommittee (TIS) Report - Dr. Henry Kissman, NLM.
4. Report from the CSIN Network Administrator - Dr. Sidney Siegel, EPA.
5. Overview of United States Efforts on Compilation and Evaluation of Physical Science Data - Dr. David R. Lide, Chief, Office of Standard Reference Data, National Bureau of Standards (NBS).
6. Physical and Chemical Properties Data Pertinent to Transport and Persistence of Toxic Substances - Dr. David B. Garvin, Chief, Chemical Thermodynamics Division, NBS.
7. Spectroscopic and Other Data Relevant to the Detection and Monitoring of Specific Compounds - Dr. Stephen A. Rossmassler, Program Manager, Materials Utilization Data, Office of Standard Reference Data, NBS.
  - o Next meeting on August 7, 1979.

## INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

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

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

### MEMORANDUM

SUBJECT: Minutes of July 10, 1979 Meeting

FROM: Nan Fremont, Executive Secretary

TO: Members, Interagency Toxic Substances  
Data Committee

Enclosed are the Minutes of the July 10, 1979 Meeting of the Interagency Toxic Substances Data Committee.

If you have any corrections, additions or business items which should be considered at our next meeting on August 7, 1979, please call me at 202-755-9336.

1. The meeting co-chaired by Ms. Carol Bastian and Dr. Marilyn Bracken, convened in the CEQ Library, 722 Jackson Place, N.W., Washington, D.C.
- I. Review of Minutes of June 5, 1979 Meeting.  
The minutes stood approved as read.
- II. Chemical Substances Information Network (CSIN)  
Subcommittee Report.

Dr. Henry Kissman, NLM, delivered the report. The Subcommittee met on June 28. Various components of the CSIN Chemical Structure Nomenclature System were discussed. Dr. Siegel reported that arrangements for a chaperoned test of the NCI system are being made.

The NLM staff gave a presentation and demonstration of the experimental on-line version of the Chemical Information Resources Director (CIRD) which contains information about some fifteen data bases.

Rita Bergman gave the status of the MITRE contract. Fifty two files have been analyzed and will be compiled into a printed directory of data bases as well as organizations which can disseminate data. This set to become part of the NLM experimental CIRD.



Mr. Charles M. Goldstein NLM, presented a method of linking similar systems, specifically in the area of library catalogues. Using various communicative programming methods, NLM has been experimenting with a user cordial system capable of interfacing the library's catalogue file, Cat-line.

The Computer Corporation of America (CCA) presented a film showing the functioning of a distributed data base management system. It demonstrated all approach to link disparate content systems located across the country to efficiently and effectively coordinate a response to an inquiry.

A discussion followed concerning several ongoing CCA activities: 1) study of system architecture of the CSIN Project. 2) use of intelligent terminals to access a CSIN prototype and 3) and an early version of the prototype.

### III. Report from CSIN Network Administrator

Dr. Sidney Siegel, EPA, delivered the report.

1. Dr. Siegel presented a Federal Register Notice from the Materials Transportation Bureau called "Display of Hazardous Materials Identification Numbers-Improved Emergency Response Capability Proposed Rules." Author George Cushmac, DOT, explained that in the Hazardous Materials Table, the UN number, and not the C.A.S. number, was being used for the transport of dangerous goods. It was apparently felt by the Office of Hazardous Materials

Regulations that the four digit UN Number was more easily recognizable than the C.A.S. number, which may have up to nine digits. Mr. Cushmac invited written comment on the subject to the Office of Hazardous Materials Regulation.

2. Dr. Siegel distributed a report from the CSIN contractor, Computer Corporation of America titled, "Immediate Requirements for Chemical Substance Information Network, Analysis and Recommendation."

3. The CSIN subcommittee chair and the network administrator are currently drafting a letter to agency heads, requesting name commitments to the Interagency Toxic Substances Data Committee.

4. A CSIN presentation was made to the senior Environmental Issues group of DOE during which Dr. Siegel emphasized the importance of CSIN as a tool capable of being used by many agencies and the need for interagency dialogue to better define its functionalities and the importance of interagency support to the project. It was noted that DOE will be making a presentation to the committee, probably in September.

5. Drs. Bracken and Siegel also plan a presentation to CPSC.

6. The issue of information as a national resource will be considered at the upcoming October meeting of the American Society for Information Science (ASIS). Government, industry and academia are expected to be present.

7. The University of Pennsylvania is preparing a matrix of functions of the Chemical Structure Nomenclature Search System (CSNS). A chaperoned test for the National Cancer Institute Chemical Information System is scheduled for July 12. The CSIN subcommittee will formulate a series of recommendations for the development of the CSNS which will be based in part on the University of Pennsylvania's final report on this component of CSIN.

8. The step by step implementation on CSIN was emphasized as a means of insuring the most appropriate expenditure of available funds and thus optimize the use of a much needed system by the user community.

9. The committee discussed user requirements, intelligent terminals and a prototype CSIN. CCA will incorporate these technical comments into a proposal which will be available in the near future. After subcommittee review of the report, recommendations will be forwarded to the parent committee. The importance of written comments was emphasized.

10. A draft of the CSIN five year implemenation plan has been distributed to subcommittee members. The final document will outline the inter-relationships of administrative issues during the development of CSIN. It will also demonstrate the difficulties of planning an interagency supported project.

#### IV. Toxicology Information Subcommittee Report.

Dr. Henry Kissman, NLM delivered the report

1). Report from National Clearing House for Poison Control Centers by Dr. Mark Fow, FDA.

Dr. Fow explained how information about product composition was gathered through direct interaction with manufacturers. He described the newsletter, a directory of poison control centers and also described a case reporting system.

2). A symposium on toxicological information projected for October 1980. The symposium would be co-sponsored by TIS and CSIN, thus involving both biomedical and regulatory segments of the Federal community. Dr. Kissman proposed structuring the symposium along the lines of the steps associated with information transfer from laboratory data to manuscripts to their publication in primary and literature sources.

A task force will be created to begin planning for the symposium.

3). Reports were given concerning progress on the Laboratory Animal Data Bank and two subcommittee publications, Toxicology Research Directory and Tox-Tips.

4). Information Reponse to a Chemical Crisis would be useful to many agencies. The lead group has two objectives (1) this capability be available in the event that an agency faces a crisis situation on a chemical (2) to have this project

function as a quickly available coordination point for data and information gathering once an "event" identifying a chemical crisis is established as being in the domain of an agency(ies). Current member agencies of the project are the Environmental Protection Agency, Department of Agriculture, NOAA, NIEHS, and the parent committee, of the DHEW Committee to Coordinate Environmental and Related Programs.

5). TIS and CSIN subcommittees will meet again on August 23, 1979.

#### V. Overview of United States Efforts on Compilation and Evaluation of Physical Science Data

by Dr. David Lide, NBS.

The main source of numerical data is the scientific literature. To handle the critical evaluation of physical property data, the National Standard Reference Data System was created. It is coordinated by the Office of Standards Reference Data within (OSRD), NBS. Most work for this office is done by experts in Data Centers and in related projects across the country. The Office of Standards Reference Data is thus the center for dissemination of information gathered from the data centers.

Output of the office is still mostly in hard copy form, but several data bases are computer readable files in the EPA/NIH supported Chemical Information System (CIS).

The OSRD interacts with government, professional societies and private sector groups to support the data centers, sponsor projects, and handle inquiries. A strong relationship already exists between ITSDC and NBS through common interests in specific identification of chemicals via the assignment of C.A.S. numbers. And the dissemination of quality physical chemical information.

VI. Physical and Chemical Properties Data Pertinent to  
the Transport and Persistence of Toxic Substances  
by Dr. David B. Garvin, NBS.

Dr. Garvin outlined ways data centers presently are meeting needs for information on chemicals in the environment in two areas, of concern 1) chemical transport, through compartments of the environment and 2) the persistence and/or reactivity of a chemical substance and how these parameters would impact on evaluating the materials toxicologic potential.

Dr. Garvin reported on the type of information available from data centers. In the area of transport, there are projects developing physical and chemical property data such as solubility and vapor pressure of organic compounds and their relationships of these measurements to levels in air, including efforts to standardize information on stratospheric chemistry.

Dr. Garvin emphasized the importance of keeping OSRD abreast of how data will be used so that the data centers and NBS can continue to provide meaningful data to the toxic substances information community.

VII. Spectroscopic and Other Data Relevant to the  
Detection and Monitoring of Specific Compounds  
by Dr. Stephen A. Rossmassler, NBS.

In the ongoing efforts to evaluate reported data, a vocabulary of classes of data has been generated. They are: Class 1 - True and constant facts on nature; Class 2 Data bases generated by spectroscopic instruments rigorously calibrated according to a defined set of procedures with samples of proven purity; Class 3 - Data from evaluations where instrumentation and calibration methods are reported; Class 4 - undefined data. The OSRD also provides recommendations to scientists concerning ways of collecting and presenting data, in an effort to improve and standardize chemical numerical data reporting.

VIII. Miscellaneous

C.E.Q. Memorandum of Understanding/C.A.S. registry numbers.

Last year in the CEQ report to Congress, ITSDC recommended the universal use of C.A.S. numbers by all Federal agencies, not necessarily as an exclusive identifier, but to establish means of referencing across various data bases with a common code. This will likewise be a recommendation in the upcoming Toxic Substances Strategy Committee Report to the President.

To expedite implementation of this policy, a draft of a memorandum of understanding has been prepared for consideration by Federal Agencies. At the next ITSDC meeting comments to the first draft will be considered.

IX. Thanking the guest speakers from NBS, Dr. Bracken announced that the next Data Committee meeting will be on August 7. The Meeting was adjourned at 11:50 a.m.



ATTENDEES-July 10, 1979

FEDERAL

INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

<u>NAME</u>	<u>ORGANIZATION</u>	<u>PHONE</u>
Lottie McClendon	NBS/OEM	
John B. Cox	U.S. Dept. of Commerce	
Bernard Greifer	Commerce/OEA	
Winston, R. deMonsabert	FDA/OC	
Carroll Leslie Bastian	CEQ	
Terri Damstra	NIH/NIEHS	
Sidney Siegel	EPA/OTS/OPII	
Dorothy Drago	CPSC	
Patricia Breslin	OSHA	
Bernard Scharf	CPSC	
Cdr. W. M. Parsons	MSC USN, Naval Medical Research and Development Command, Bethesda	
Cathy Brooks	EPA/MIDSD	
George T. Armstrong	EPA/OTS/ERD	
Rita F. Bergman	Sigma Data Computing Corp.	
Jerry Coffey	OFSPS	
Jerry Calderone	HEW/OASH/PHS	
Roger Connor	EPA	
Richard J. Lewis, Sr.	NIOSH/HEW	
Loren Hall	EPA/OTS/OTE	
Marilyn Bracken	OPII/OTS/EPA	
Henry Kissman	NLM	
Calvin Menzie		

ATTENDEES-July 10, 1979

NON-FEDERAL

INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

<u>Name</u>	<u>ORGANIZATION</u>	<u>PHONE</u>
George R. Hoffmann	NAS	
Ron Grandon	Pesticide & Toxic Chemical New	
Girard Ordway	BNA, Inc.	
Fred A. Tate	CAS	
Ron G. Dunn	CAS	
Twyla Bishop	Mitre Corp.	
Bruce Grogan,	Weyerhauser Co.	
Carie Heller	Sigma Data - Upgrade	
Deborah Mullon	Society of the Plastics	
William H. Krebs	General Motors Corp.	
Adrienne A. Whyte	Biotechnology, Inc.	

## INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

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

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

### AGENDA - AUGUST 7, 1979

1. Review of minutes of July 10, 1979
2. Subcommittee Reports
  - Chemical Substances Information Network (CSIN)
  - Toxicology Information Subcommittee (TIS)
    - Dr. Sid Siegel, EPA
3. C.E.Q. Memorandum of Understanding/C.A.S. registry number
4. Presentations
  - a). UPGRADE - User Prompted Graphic Analysis and Display System
    - Mr. Larry Milask
    - Dr. Carol Graves  
Sigma Data Computing Corporation
  - b). HEEDA - Health Efforts and Environmental Data Analysis Systems
    - Dr. Amy Rispin, Office of Toxic Substances/Testing and Evaluation Program, EPA
  - c). PROPHET
    - Dr. Howard Bilofsky  
Bolt, Beranek and Newman

## INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

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

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

### MEMORANDUM

SUBJECT: Minutes of August 7, 1979 Meeting  
FROM: Nan Fremont, Executive Secretary  
TO: Members, Interagency Toxic Substances  
Data Committee

Enclosed are the Minutes of the August 7, 1979 Meeting of the Interagency Toxic Substances Data Committee.

If you have any corrections, additions or business items which should be considered at our next meeting on September 11, 1979, please call me at 202-755-8050.

1. The meeting, co-chaired by Ms. Carroll Bastian and Dr. Marilyn Bracken, convened at 9:30 a.m. in Room 2010 of the New Executive Office Building 17th and Pennsylvania Avenue, Washington, D. C.

- I. Review of Minutes of the July 10, 1979 meeting.

The minutes stood approved as read.

- II. Subcommittee Reports

- A. Toxicology Information Subcommittee Dr. Sid Siegel reporting in Dr. Henry Kissman's absence.

At the June 28th Meeting Dr. Mark Fow, EPA, described the technical and administrative information activities of the Poison Control Clearinghouse and its network of poison control centers around the country.

Dr. Kissman proposed that a second symposium on handling of toxicology information be held and gave a possible date of early June, 1980. Dr. Siegel suggested co-sponsorship of the symposium by CSIN and TIS.

Mark Southerland, CPSC, spoke about the Chemical Monograph Referral Center and ways of identifying and accessing documents relevant to chemical substances.

Florence Bayard, NLM, spoke about the status of the TD-3 project (Toxicology Document and Data Depository).

Carol Haberman, NLM, discussed the Toxicology Research Projects Directory and TOXTIPS, which blends information on current toxicology testing in industry, government and academia.

Dr. Terry Damstra, NIH/NIEHS, presented the status of the chemical Crisis Information Reponse Project.

Dr. Gary Kielson, NAS, spoke about activities at the National Academy of Sciences as a toxicology advisory and rapid response center.

B. Chemical Substances Information Network (CSIN) Subcommittee.

Dr. Sid Siegel, CSIN Administrator EPA, gave the following report. A letter is being prepared for distribution to agency heads to make "name commitments" to the CSIN and TIS subcommittees.

Dr. Siegel described activities that continue to be ongoing under the contract to develop the Chemical Structure Nomenclature Search System (CSNS) that will be part of the front end of CSIN. A performance test of one chemical structure. substructure search system was carried out at the NCI's Division of Cancer Treatment. The results of that test and of other activities involved in defining the functional requirements and search capabilities of the CSNS were examined and are being integrated into a report through a contract with the University of Pennsylvania. The report will allow the CSIN subcommittee to assess user requirement and functional capabilities presently existing so as to make judgement concerning the structure and functions of phase one of a CSNS.

Bill Caldwell, NLM, demonstrated a computer file which is capable of making use of records generated by the MITRE corporation for the initial phases of establishing the Chemical Data Base Directory. Dr. Siegel suggested renaming the system the Chemical Information Resources Directory, so that the name would more accurately describe the broad spectrum of data and information sets the directory is to be able to identify.

The Computer Corporation of America presented a film which described a distributed data base management system developed by the organization.

Dr. Charles Goldstein, NLM, spoke about a user cordial terminal acting as an interface to on-line card catalogue systems.

Comments were again requested on the five-year development plan. Erika Graf-Webster, author of the plan, is preparing some introductory material for that document.

Dr. Siegel reported on status of development of an intelligent terminal for interfacing the CSIN prototype. A proposal for such capability was submitted for review to the contractor. Subsequently, the contractor prepared a second proposal which detailed four options for intelligent terminal construction. The proposal was distributed to twenty members of the subcommittee and comments were

requested. The choice of option two was based on securing an intelligent terminal that would balance capability with mechanical and software stability for field use. The following are details for the hardware and software specifications for option two:

The hardware configuration for this option consists of an intelligent terminal with two floppy disc drives, auto-dialer and printer.

The floppy disc drives have a combined storage capacity of one million characters. The discs are removable providing additional off-line storage capacity. The storage available in option two supports expansion of function far beyond the capabilities and mode of usage previously envisioned. All programs except a programmable, readable only memory, resident low level loader for terminal start-up will be stored on floppy discs.

Key word lists and captured data may be stored on and retrieved from discs. When data is captured or edited, the user is offered the option of saving it permanently.

Dr. Bracken accepted the general endorsement of the members of the ITSDC to proceed with final negotiations with CCA for the recommended option.



### III. CEQ Memorandum of Understanding on the Use of C.A.S. Numbers.

Ms. Bastian explained that the document is intended to be a brief statement of policy to insure that all the agencies understand and join in the suggested use of CAS numbers in their chemical information systems.

In discussion, it was felt that there was some ambiguity regarding the range of systems affected, as well as implementation guidelines. To address these issues, an ad hoc task force was formed.

### IV. Presentation of the User Prompted Graphic Analysis and Display System (UPGRADE).

This analytic software package was presented by Dr. Carol Graves of Sigma Data Computing Corporation. UPGRADE is being developed under contract with CEQ through interagency agreements with EPA, NCI, DOE and others.

UPGRADE is an English language system designed for use by the analyst or researcher and currently resides in the computer systems of NIH. There are a number of interfaces available through UPGRADE. The general purpose interface allows the user to put in his or her own data. CLIDE, (Country Level Integrated Data base

for Epidemiology) contains data from several agencies among them NIHS, EPA and the census Bureau. It is matched with FIPS to allow the data to be correlated with environmental or demographic data.

Three of UPGRADE's interfaces connect with specific data bases, STORET, a water quality data base, NASQAN, the USGS water quality data base, and SAROAO, the EPA air quality data base. The system has as well various data base management capabilities.

#### V. Presentation of the Health and Environmental Effects Data Analysis System (HEEDA).

Dr. Amy Rispin, EPA, spoke about HEEDA which is being developed in the Office of Toxic Substances and will be available through CSIN. The system, is in part to respond to the need to increase the ability to predict the carcinogenicity, mutagenicity and tetragenicity of chemical substances. At the core of this effort is the development of a base made up of evaluated data and information. The system includes a variety of ecological as well as chemical and biological data bases. A current major effort in the area of data reduction is the gene-tox mutagenicity literature, using this system as the capture mechanism. 3,000 to 5,000 chemicals will eventually be reviewed. With this

development of training sets of toxicological information and having statistical correlation techniques in place, HEEDA will increase the ability of scientists to predict the toxicologic potential of chemicals to biological systems.

#### VI. Presentation of PROPHET

Dr. Howard Bilofsky described the system as an inter-active graphically oriented information handling system. It is funded by the Division of Research Resources of NIH and designed for use by scientists who are not necessarily familiar with computers. PROPHET permits easy access to a wide range of scientific data in the form of tables, graphs and three dimensional schematic illustrations. The system allows an exploration of different data combinations, data extrapolation and manipulation of the graphics by the user. It has specific capabilities for pharmacokinetics, cell kinetics, linear, and regression analysis. It also can manage molecular structures to illustrate molecular binding and building. PROPHET is currently being used in about 20 sites around the country.

Thanking the presenting guests, Dr. Bracken adjourned the meeting at 12:25 p.m.

ATTENDEES - August 7, 1979 - I.T.S.D.C.

<u>NAME</u>	<u>ORGANIZATION</u>	<u>PHONE</u>
Ron Grandon	Pesticide & Toxic Chemical News	783-7472
William H. Krebs	General Motors Corp.	313-556-1597
Adrienne Whyte	Biotechnology, Inc.	573-3700
Howard Bilofsky	BBN	617-491-1850
Judith M. Nushon	MITRE	703-827-6930
John M. Capriccioso	Dow Chemical	517-636-1705
Christine Shine	Mansanta Co.	452-8880
Martin Aronoff	Nat Bur Stds.	921-3491
Doug Sellers	EPA/OTS	755-9336
Carol Graves	Sigma Data Computing Corp.	223-8860
Lewis Gevantan	NBS	395-5763
Marion Suter	EPA/OTS	426-9408
Ron Dunn	CAS	614-421-6940
Fred Tate	CAS	614-421-6940
Patricia Breslin	OSHA/IRLG	523-7177
Jerry Calderone	DHEW/OASH/OPHS	472-5194
Don Hummel	NLM	496-1131
Sidney Siegel	EPA/OTS/OPII	755-8040
Winston R. DeMonsabert	FDA/OC	443-4505
Vera Hudson	NIOSH	443-2100

<u>NAME</u>	<u>ORGANIZATION</u>	<u>PHONE</u>
Bernard Greifer	DOC	377-3078
Kenneth Chu	NCI	496-1152
George E. Cushmac	DOT/MTB	426-2311
Marilyn Bracken	OTS	755-8040
Carroll Leslie Bastian	CEQ	395-4980
Dorothy Drago	CPSC	492-6470
Amy Rispin	EPA/OTS	755-1500
Fred W. Clayton	OSHA/DOL	523-8020
Jerry Coffey	OFSPS	673-7974

# INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

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

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

## AGENDA

September 11, 1979  
Room 2010  
New Executive Office Building  
17th Street and Pennsylvania Avenue, NW  
Washington, DC

1. Review of minutes of August 7, 1979
2. Subcommittee Reports  
    Toxicology Information Subcommittee (TIS)  
    Chemical Substances Information Network Subcommittee (CSIN)  
    - Dr. Henry Kissman, NLM.
3. Highlights of the CCA report, "A Prototype Chemical Substances Information Network"  
    - Rita Bergman, Sigma Data Computing Corporation.
4. General Discussion on the Interagency Memorandum of Understanding on the use of C.A.S. numbers.

# INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

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

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

## MEMORANDUM

SUBJECT    Minutes of September 11 1979, Meeting  
FROM:      Nan Fremont, Executive Secretary  
TO:        Member, Interagency Toxic Substances  
            Data Committee

Enclosed are the minutes of the September 11 1979 meeting  
of the Interagency Toxic Substances Data Committee.

If you have any corrections, additions or business items  
which should be considered at our next meeting on October 2, 1979,  
please call me at 202/755-8050.

Enclosures

The meeting chaired by Dr. Marilyn Bracken convened at 9:40 a.m. Room 2010 of the New Executive Office Building 17th and Pennsylvania Avenue, Washington, D.C.

I. Review of minutes of August 7, 1979, meeting the minutes stood approved as read.

II. Subcommittee Reports.

A. Toxicology Information Subcommittee

- Report by Dr. Henry Kissman, NLM.

The subcommittee met on August 23.

- Activities of the National Referral Center were reviewed by Mr. John Feulner, Library of Congress.

One service offered is a compilation of organizations potential as information resources. This data base lists some 12,000 to 13,000 such groups or organizations (50% of the users of this service are equally divided between government and industry, 25 percent colleges and universities and 25 percent others. They are able to refer 93 percent of their users and of those 85 percent are helped.

NCR published in early 1970 a directory oriented towards toxicology information. The director listed some 1200 organizations with this capability. The book is not currently up to date, but might be made so in the next fiscal year. NCR gets about 125 queries week which are handled by a staff of 16 to 20 people and a budget of about half a million dollars. Information on organizations is updated about once a year.



Mr. Feulner noted that presently the data base was available only to Congress. LC is open to becoming a member of the ITSDC.

- Dr. Terry Damstra, of the NIEHS, gave a report on the Interagency Response to Chemical Crisis Project (IRCC) Work has been completed on the mock crises dealing with "Asbestos in Air". The "Effect of Acid Rain" mock crisis project is still in progress. A mock search on the Health Effects of Formaldehyde-Urea Resins demonstrated that two people can do a quick but useful search in an emergency. The next step will be to examine ways of electronically communicating this information. Collaboration will be explored between this project and the Emergency Response Group of the National Academy of Sciences. Two new non-crisis search projects have been chosen. "The Composition of Toxic Agents in Chemical Dumps including the Oceans", and "Carcinogenicity of Inorganic Compounds in Synthetic Fuels." Currently the IRCC distributes copies of its searches free of charge. At the end of these minutes is a more detailed description of the IRCC project.
- A report on the Laboratory Animal Data Bank was presented by Dr. Warren Hoag. A three year contract has been awarded to The Battelle Columbus Laboratories, (BCL), under which the first task is to establish online public access to the file. BCL will provide user support, training manuals, newsletters, user data, etc. They will also assemble statistics associated with user requirements.

- There was a report on CHEMRIC, a CPSC project under which information on the plans of agencies to prepare monographs is collected and made accessible in an attempt to reduce duplication of effort and encourage the sharing of such efforts. In the light of CPSC's difficulty in continuing support of the project, it was felt that TIS or the Data Committee should aid the effort while leaving it under CPSC Control and direction. It was recommended also that the file be made available in some public on-line computer system.

#### B. CSIN Subcommittee

- There was a report on the pre 1965 C.A.S. registration of compounds project. Negotiations have been started with C.A.S. as to the cost of a feasibility study for this project.
- Rita Bergman reported on the status of contracts supporting CSIN development. An interim report from MITRE was handed out, and a final report on the chemical information resources directly is expected shortly.
- The NLM staff had prepared a draft RFP workscope incorporating information about organizations that collect information about organizations. In gathering information for a Chemical Information Resource Directory, it was felt that part of the task was to define organizations that already do this work, and to investigate the possibility of using those collections for this component of CSIN.

- Rita Bergman reported on the computer corporation of chemicals (CCA) CSIN preprototype. As part of this activity, it is possible that the agencies could set up their own simple or intelligent terminal access stations. Ms. Bergman also outlined CCA's progress in collecting information on CSIN user requirements.
- Dr. Sid Siegel gave the status of the University of Pennsylvania contract on the requirements for the development of a CSNS. A final reprot from the University is due shortly. He said also that the chaperoned test of the NCI system had been completed with results fairly close to those projected by the earlier interim report.
- It was noted that the letter to the agencies requesting to CSIH and TIS committee activities was still in draft form. The draft letter was approved by Dr. Rall and presently awaits EPA approval.
- In discussing the problem of getting reports to be commented on quickly, it was mentioned that there should be a group of three or four stand-by consultants would be available quickly for ad hoc technical reviews especially when an indepth technical response is necessary for further action. Dr. Bracken encouraged those in attendance to suggest to the CSIN subcommittee names of experts in the fields of network architecture, communications, data base design and administration and substructure search system.

### III Highlights of CCA report," a Prototype Chemical Substances Information Network".

- Rita Bergman, CEQ.

In passing out copies of the report, Ms. Bergman emphasized the importance of its careful review as a major way to facilitate efficient and effective CSIN development.

When the prototype can be implemented CSIN will achieve integration of different data bases for the user with a single interface. Thus, there will be a number of components which users will be able to access using one query language. The objective of the prototype is to meet near term requirements for multisystem access.

In this stage of development, analysis of early problems of multisystem interfaces and of data integration will lead to greater efficiency and level of performance as the system matures. Development in part will depend on an evaluation of user experience and feedback.

Four functions are to be included in the system: (1) retrieval of all information on a given chemical by entering its name or its C.A.S. number, (2) defining or storing queries as firmware called script; (3) a uniform query language or query facility which allows access to networked data bases with one language, (4) when required direct use of a CSIN component through each agency-housed terminal station.

In addition, of distributed CCA is reviewing the potential application of data base management systems. None of the existing DBMS are felt to be fully appropriate for CSIN's needs although it is likely that pieces of some of them will at least in part be suitable for application to CSIN. To develop a common query language, an important CCA concept, a global data base query system, will take data from selected CSIN components and try to establish relationships among them. Initially only selected data elements will be mapped and used in development of the query language. Ms. Bergman also noted that a user need not use the common query language, but may access data in any given component by use of SCRIPT, or with assistance from the Chemical Information Resource Directory. In the area of communications, a packet switched network has been recommended because of the flexibility and efficiency it affords. At least two network components will describe some portion of their data as confidential. CICIS and TDMS. Confidentiality will be assured by a separate secure query system and by a physical facility which admits only authorized personnel.

In response to a need for more technical expertise to review some aspects of the prototype design, Dr. Bracken suggested establishing a technical review board comprised of experts who could provide technical evaluation on various aspects of CSIN development.

#### IV. New Business

Dr. Marilyn Bracken, EPA

The Toxic Substances Strategy Committee report to the President included the concept of an early Hazard Warning System whereby relevant information and data bases would be used as resources to alert regulatory agencies and the research community to areas of potential concern. Dr. Bracken asked members to consider ways of organizing and implementing this project. The committee also recommended regarding that a review of Federal legislation be done and recommendations developed that would coordinate or address the whole aspect of exchange of confidential data between the agencies and the ITSDC be responsible for establishing this task.

Dr. Bracken noted that the next meeting would be October 2, 1979 at 9:30

With no further questions or additions to the agenda the meeting was adjourned at 10:50 a.m.

I.T.S.D.C.

ATTENDEES

September 11, 1979

FEDERAL

<u>NAME</u>	<u>ORGANIZATION</u>	<u>TELEPHONE</u>
Barbara Mandula	NAS	389-6932
George R. Hoffman	NAS	389-6351
Martin Aronoff	NBS	921-3491
Josephine L. Walkowicz	NBS	921-3491
Cathleen Brooks	EPA/MIDSD	755-0811
Terri Damstra	NIH/NIEHS	8-629-3471
Henry Kissman	NIH/NLM	496-3147
Vera Hudson	NIOSH	443-2100
George Cushmac	DOT/MTB	426-2311
Marilyn Bracken	EPA/OTS	755-8040
Glenn Simpson	CPSC	492-6962
Bernard Scharf	CPSC	492-6470
Jerry Calderone	DHEW/OASH/PHS	472-5194
Kenneth Chu	NCI/NTP	496-1152

I.T.S.D.C.

ATTENDEES

September 11, 1979

NON-FEDERAL

<u>NAME</u>	<u>ORGANIZATION</u>	<u>TELEPHONE</u>
Deborah Mullon	The Society of the Plastics Industry Inc.	331-0340
John M. Capriccioso	Dow	517-636-1705
Christine Shine	Monsanto	452-8880
Linda Goldstein	Baker & Hostefler	857-1500
Elaine Kaisor	CSMA	872-8110
James A. Hulme	American Cyanamid Co.	737-4800
Jane Rasmussen	Chemical Mfgs. Association	328-4229
Lillian F. Koehler	Johnson & Johnson	201-524-9522
Judith E. John	PM/USA	804-271-3317
Angela Bannon	3M Company	331-5581
Adrienne Whyte	Biotechnology Inc.	573-3700
Jennie Hart	General Motors Co.	313-556-1597
Dale Myers	C.A.S.	614-421-6940
Mary Worobec	BNA	452-4583
Rita Bergman	Sigma Data Computer	223-8860



ITSDC  
September 11, 1979

THE INFORMATION RESPONSE TO CHEMICAL CRISIS PROJECT

Several government organizations have joined in the establishment of an Information Response to Chemical Crises (IRCC) Project.

This project was established in recognition of the fact that various government organizations share a common need for basic information in the event of a "chemical crisis". The primary objective of the IRCC Project is to provide IRCC member organizations with a rapid response literature search service which address the "crisis" chemical(s) or topic(s). A second objective is to prepare bibliographies on substances or topics, which, although non-crisis in nature, are of communal interest to the IRCC project organizations.

IRCC literature searches have been performed on 2,3,7,8-tetrachlorodibenzodioxin; dichloroethylenes; the effects of environmental chemicals on the immune system; asbestos levels in air; and on the toxicity of urea formaldehyde foams. For further information, contact: Dr. Terri Damstra; IRCC Project Coordinating Officer; The National Institute of Environmental Health Sciences; P.O. Box 12233, Research Triangle Park, N.C. 27709.



UNITED STATES ENVIRONMENTAL PROTECTION AGENCY  
WASHINGTON, D.C. 20460

OFFICE OF TOXIC SUBSTANCES

AGENDA

October 2, 1979

Room 2008

New Executive Office Building

17th and Pennsylvania Avenue

Washington, D. C.

1. Subcommittee Reports

Toxicology Information Subcommittee

Dr. Henry M. Kissman, NLM

Chemical Substances Information Network

Dr. Henry M. Kissman, NLM

2. Toxic Substances Strategy Committee

Hazard Warning System

Confidentiality Issues Among Agencies

3. Department of Energy Information Activities

John Wilson

George Kelly

# INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

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

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

## MEMORANDUM

Subject: Minutes of the October 2, 1979 Meeting

From: Nan Fremont, Executive Secretary

To: Members, Interagency Toxic Substances Data Committee

Enclosed are the minutes of the October 2, 1979 meeting of the Interagency Toxic Substances Data Committee.

If you have any additions, corrections or items to be considered at our next meeting on November 13, 1979, please phone me at 202-755-8040.

## I. Introduction

The meeting was held at the New Executive Office Building at 9:30 a.m., and was co-chaired by Dr. Marilyn Bracken, EPA, and Ms. Carroll Leslie Bastian, CEQ. Dr. Bracken introduced Nan Fremont, EPA, as the new Executive Secretary of the ITSDC. After notation of corrections to minutes of the September 11, 1979 meeting, they were approved as corrected.

## II. Old Business

Co-Chair Carroll Bastian reported the signing of two CSIN related contracts at CEQ.

The first is an amendment to the ongoing CCA contract. The amendment provides for an expansion of the use-requirements analysis which will include a broader representation of the pre-prototype system as recommended by the CSIN subcommittee and subsequently accepted by the parent committee, as well as for the development of a design for Version I of the CSIN prototype.

Under the second recently signed contract, Chemical Abstracts Service will conduct a feasibility study for the registration of chemicals from the pre-1965 literature, as recommended in the CEQ 1978 report to Congress.

Dr. Bracken reported on the work done on a suggestion of the parent committee that specialized technical support be made available to the CSIN subcommittee and/or to the parent committee. Methods of identifying and assembling such a group of experts were discussed.

It was agreed that this was an appropriate function of the CSIN subcommittee, and that a suggestion be made to the subcommittee to identify experts who might assist in reviewing technical documents.

Distribution of the CSIN prototype design document by CCA has started the flow of comments and has begun the building of a list of experts to help review the report. A full day open seminar is being planned for November or December to discuss in an open meeting the document contents and its administrative ramifications.

### III. Report from the CSIN Network Administrator

Dr. Sid Siegel, EPA, gave an update of ongoing Network activities.

The final MITRE Corporation report on the Chemical Data Base Directory is due October 5.

The final report from the University of Pennsylvania contract on the Chemical Substances Nomenclature System (CSNS) module will be distributed and a summary presentation of the report is to be made by Dr. David Lefkovitz of the U. of P. to the CSIN subcommittee on October 11. The report will also be distributed to the parent committee for their review. Dr. Lefkovitz will present a summary of the report to this committee on November 13.

A special meeting of the CSIN subcommittee is being considered to discuss and respond to the proposed development schema. This is in

response to concern voiced at the September 2 parent committee meeting that more information and education is needed on some of the technical recommendations before further action can be taken.

#### IV. CEQ Memorandum of Understanding regarding the use of CAS Registry Numbers in Agency Files

Ms. Carroll Bastian, CEQ, commented that issues relevant to the wording and implementation of the CEQ document had been raised and would be turned over to the CSIN subcommittee for further consideration and resolution. The subcommittee is to discuss these issues and develop a final draft of the memorandum for general agency concurrence. Any other comments may be directed, quickly, to Dr. Henry Kissman, NLM, chairman of the subcommittee.

#### V. Toxic Substances Strategy Committee Report (TSSC)

The first item discussed was the recommendation by the TSSC that the Parent Committee look at the feasibility of developing a Chemical Hazard Warning System. Dr. Henry Kissman, NLM, gave some background about past efforts in this area.

Early discussion of this concept has pointed to the need for a chemical intelligence activity, i.e., a review of the literature for reports of events that might have potential implications in the area of toxic substances. It was felt that once such events were

identified they should be further investigated by looking in part at the extent of an involved chemical's use, its biological, physical and chemical properties and activities and properties of chemically related compounds.

Dr. Kissman recommended a feasibility study to examine these and points such as: 1) current availability of this kind of information, or existence of organizations doing this type of work, 2) potential users of such a service, 3) technical guidelines for developing the service and estimating the operating budget, 4) formation of a reporting body and a screening group to evaluate hazard information, 5) actions that might be taken beyond agency alert, 6) the possibility of a retrospective study of hazards to try and ascertain if past incidents could have been avoided or would have been less serious if this kind of analysis service had been available.

Dr. Kissman also mentioned the Information Response Chemical Crisis (IRCC) as a possible monitoring group for this activity, noting its status as a functioning organization. He also said that he would bring this matter up before the Toxicology Information Subcommittee (TIS) and would return to the parent committee with subcommittee comments from members who have insight into this area.

Dr. Edward Bartkus, Dupont, made the following comments: 1) that such a hazard warning system should include consideration of lesser as well as more widely used chemicals, (i.e., and research

chemicals). Early awareness of some hazard could avoid a more complex and widespread hazard later, 2) the committee should take advantage of existing activities like the Environmental Mutagen Information Center, and the Environmental Teratology Information Center, 3) the committee should also consider hazards that do not necessarily fall into the area of toxicology, 4) the effort should be directed toward designing a hazard alert system that could tie in with other domestic and international systems presently under development. In referring to Dupont's own hazard alert efforts, Dr. Bartkus noted that in general, industry would be inclined to contribute to and participate in a national hazard warning system.

Dr. Bracken noted that advantage could be taken of learning more about the NIOSH prepared Intelligence Bulletins, an activity closely related to this one. She suggested having a presentation to the committee on an NSF-supported study done at Clark University on hazard recognition and hazard reduction methodology.

The committee then discussed the suggestion by the TSSC that ITSDC draft recommendations for legislation addressing the issue of confidentiality of trade secrets. Dr. Bracken reported that Mr. Bob Nicholas of CEQ has offered to chair a subcommittee for this task. Mr. Nicholas had led that particular project for the Strategy Committee. Because the exchange of data among the agencies is a critical issue for CSIN development, it was felt that the committee



should be very much involved in the review of comments to the Strategy Committee concerning the trade secrets chapter. It was agreed that a subcommittee should be formed to generate legislative recommendations. Members were asked to go back to their agencies to find appropriate representation to this subcommittee.

#### VI. Presentation of the Information Activities of the Department of Energy (DOE).

Mr. John Wilson began his report with a short history of DOE. This department grew out of the Atomic Energy Commission (AEC). In 1974, the National Regulatory Commission (NRC) component of AEC was set up as a separate regulatory body. In 1977 the AEC, the Federal Energy Administration, the Federal Power Commission, and some departments of the Bureau of Mines were merged to form the Department of Energy. DOE has approximately 20,000 Federal Employees, 100,000 contractor personnel and a budget of \$10 billion. Organizationally, the Department falls into four categories: Federal Energy Regulatory Commission (FERC), the technological arm; Management; and the network of field offices.

Likewise, DOE data can be classified into broad categories. Management data is contained in about 350 data bases and includes information concerning people, staff, payrolls, medical and exposure data systems as well as the whereabouts of some "controlled

substances" The Uniform Contractor Reporting System has information on the contractors associated with DOE.

There are three main regulatory data systems within DOE: the (FERC), formerly the Federal Regulatory Commission, has 40-50 sub-systems. The information included generally has to do with the transportation and sales of natural gas and electricity. It contains statistics of interstate natural gas and pipeline companies, on privately owned US electric utilities, as well as information on power and fuel supplies, electric rates, hydroelectric projects and power surveys.

The second regulatory system is the Economic Regulatory Administration (ERA). It regulates items not under FERC jurisdiction. These are mostly the more public systems, having to do with the allocation of fuel oil, how much oil is to be made into gasoline, etc. Recently, ERA has mapped out a feasibility study for gasoline rationing at a cost of \$1 million per state.

The third regulatory group is the Energy Information Administration, EIA, which projects the supply, demand and consumption of energy. It produces an annual report, drawing information from its 38 data systems.

Other systems are the National Energy Information Center and the Federal Energy Information Locator System, which is reference directory for Energy Information Systems.

DOE also has a number of technical information projects that deal with energy technologies, including engineering as well as applied basic sciences. The major areas are: 1) conservation, 2) solar energy, 3) energy or nuclear Technology, 4) Fossil energy, including coal, 5) Resource Allocation, which moves proven technologies into the commercial sector, 6) Environmental.

Bibliographical information is kept at the Technical Information Center (TIC) at Oak Ridge, which produces Energy Research Abstracts. Energy literature is surveyed and keyed into RECON, an on-line bibliographic system to which DOE as well as some other agencies have access. A contractor appraisal system which monitors a contractor's ability to produce the reports required under the contract for submission to the TIC. There are well over 1000 systems housed there, including management systems and graphic display systems. The largest concentration of these is at the Oak Ridge National Laboratory (ORNL). These are supported by DOE and DHEW (i.e., NLM, NIEHS) as well as other agencies.

In addition, there is considerable interest in analytical and graphic systems which not only store and retrieve data, but also perform some manipulation and calculation, and include statistical packages which can produce tables, bar graphs and charts.

DOE also oversees 25 information centers that are multi-agency supported, including the Berkley Data Center, the Information Center

for Energy Safety, the National Geothermal Information Center. They contain bibliographical as well as hard data.

The real concentration of the technical data and data systems is in the National Laboratories. These energy research centers include specialized biomedical laboratories. They include a facility at Oak Ridge, and the Loveless Laboratory in Albuquerque.

There is varied progress in integrating all of these information resources into the Department of Energy as a whole. Headway is being made in the management areas, while the proprietary nature of much of the information in the regulatory areas raises special problems for the development of such systems. In the technological area things are progressing well. There has been a push for cooperation among laboratories to undertake common projects, necessitating the sharing of information. The Inter-laboratory Working Group on Data Exchange has developed a computerized data exchange system which is widely used throughout the labs.

Currently at the Berkley Laboratory, there is a DOE funded pilot study to develop a distributed computer network. On completion, the software will then be available to all DOE labs, other federal agencies and their contractors. Standards for exchange of data have been set up by the Inter-laboratory Working Group on Data Exchange. Adoption of the standards are voluntary, because most of the labs are contractors, but strong DOE guidelines have encouraged wide

acceptance.

In thanking Mr. Wilson and Dr. George Kelley, also from the Department of Energy, Dr. Bracken encouraged other agencies to think about presenting their own information activities to the committee. Industry is also encouraged to do so.

The next meeting of the ITSDC is scheduled for November 13, 1979 from 9:30 to 12 noon in Room 2010 of the New Executive Office Building, 17th and Pennsylvania Avenue, NW, Washington, D.C. Since there was no further business, Dr. Bracken adjourned the meeting at 10:55 a.m.

ATTENDEES

FEDERAL

Carroll Leslie Bastian	CEQ	395-4980
Rita F. Bergman	CEQ	395-5763
Marilyn Bracken	EPA	755-8040
Jerry Calderone	HEW/OASH/PHS	472-5194
George E. Cushmac	DOT/MTB	426-2311
Bernard Greifer	DOC	377-3078
Vera Hudson	NIOSH	443-2100
George Kelly	DOE	252-4760
Henry Kissman	NLM/NIH	496-3147
Winston R. de Monsabert	FDA/OC	443-4505
Suzanne Rudzinski	EPA	755-5851
Bernard Scharf	CPSC	492-6470
Sid Siegel	EPA	755-8040
Josephine L. Walkowicz	NBS	921-3491
John Wilson	DOE	353-4684
Bruno M. Vasta	EPA	755-5687

ATTENDEES

NON-FEDERAL

Angela Bannon	3M	331-5581
Edward P. Bartkus	Dupont	302/655-0062
Patricia J. Beattie	General Motors	313/556-1597
Sherry Boltz	NPCA	462-6272
Sidney Draggan	National Science Foundation	632-7810
R. G. Dunn	C. A. S.	614/421-6970
Linda Goldstein	Baker & Hootetler	857-1500
Bruce Graham	IIT Research	296-1610
George R. Hoffman	National Academy of Sciences	389-6351
Judith Hushon	Mitre Corp.	827-6930
Lillian Kochler	Johnson & Johnson	201/524-9522
Jane Rosmussen	Chem. Manufactures Assoc.	229-5439
Cheryl Russell	ChemRegister	452-4583
Christine Shine	Monsanto Company	452-8880
Jacqueline Weitzel	P. Q. Corp.	215/293-7352

# INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

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

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

## AGENDA

November 13, 1979

9:30 a.m. - 12:00 noon

New Executive Office Building

Room 2010

- I. Introduction
- II. Subcommittee Reports  
-Dr. Henry Kissman, NLM
- III. Report of the University of Pennsylvania Study  
of System Requirements Analysis for the CSNS  
-Dr. David Lefkowitz, University of Pennsylvania
- IV. Presentation by the Chemical Abstracts Service  
-Mr. Nick Farmer, C.A.S.  
-Mr. Ron Wiggington, C.A.S.



# INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

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

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

## MEMORANDUM

SUBJECT: Minutes of the November 13, 1979 Meeting

FROM: Nan Fremont, Executive Secretary

TO: Members of the Interagency Toxic Substances  
Data Committee

Enclosed are the minutes of the November 13, 1979 meeting of the Interagency Toxic Substances Data Committee. If you have any additions or corrections, please call me at 202/755-8040.

## I. Introduction

The meeting was held at the New Executive Office Building at 9:30 a.m. and was chaired by Ms. Carroll Leslie Bastian, CEQ.

## II. Old Business

### 1. Toxic Substances Strategy Committee (TSSC) Report

Ms. Bastian reported that the chapter of the Toxic substances Strategy Committee (TSSC) report pertaining to the data activities of ITSDC has been distributed and discussed by Data Committee members and has been commented on by various non-federal groups. Generally the comments are supportive of the development of the

Chemical Substances Information Network (CSIN). Other comments include: the need for a Chemical Regulation and Guidelines System; the need for EPA to continue an open ended chemical classification system; a need for coordination among federal groups in their chemical literature search activities including more rapid access to the relevant data and information on chemical substances, especially in emergency situations. Copies of comments from the National Resources Defense Council (RDC) and the DuPont Corporation were distributed. All comments are available through CEQ, and will be distributed to the pertinent agency. Participation by non-federal groups in CSIN development was stressed to insure that the needs of these communities will be integrated into system design.

### 2. Chemical Hazard Alert System (CHA)

In a second item related to the TSSC report to the President, Ms. Bastian invited comments on the Chemical Hazard alert System (CHA) described in the report, to define some points that should be examined in a CHA feasibility study, as recommended by the Toxicology Information Subcommittee (TIS). It was noted that the TSSC report recommendation is intended to "catch the things that fall through the cracks" i.e., it recommends a system for Hazard Alert that spans the scope of literature monitoring being done by all the various agencies, but one that does not duplicate those functioning systems. Important also is the development of a CHA that would bring valid hazards to the attention of the appropriate Federal agencies, while keeping insignificant incidents in the proper focus.

Dr. Bartkus (DuPont Corporation) reiterated that those hazard situations that do not necessarily fall into the area of toxicology should also be included in a CHA. He noted that some incidental chemical-related incident could be a precursor of a more serious environmental one.

Dr. Chu (NCI) suggested the Gene-Tox program at EPA as a program to examine in creating a model for a CHA.

Dr. Damstra stated that while the IRCC does literature searching of a different nature than that described for a CHA, the IRCC will discuss the Hazard Alert issue at their November 27, 1979 meeting. She will return to the ITSDC any suggestions from that group.

### 3. ITSDC Membership

Concerning ITSDC membership, it was noted that the letter to agency heads requesting specific nominations for CSIN membership also requests each agency to confirm and review its membership to ITSDC. The letter presently is being redrafted and tailored to reflect the present relationship of the agency to the ITSDC, CSIN and TIS subcommittees.

## III. Subcommittee Reports

### 1. Chemical Substances Information Network (CSIN) Subcommittee

Dr. Henry Kissman (NLM) first reported on recent activities of the Chemical Substances Information Network (CSIN) Subcommittee. Dr. Bernard Greifer, Department of Commerce (DOC) had been asked at the October 11, 1979 Subcommittee meeting to prepare a paragraph for insertion into the CEQ Memorandum of Understanding to describe the de facto use of the CAS Registry Numbers in many Federal Agencies in their information files. Dr. Greifer's draft of the Memorandum also calls on the ITSDC through DOC to investigate the feasibility of establishing the Registry Numbers as a Federal Information Processing Standard (FIPS). This could allow eventually for reimbursement to agencies for costs of adding the CAS Registry Number to their chemical information files. The draft memorandum was distributed for discussion only, since it had not yet been discussed by the CSIN Subcommittee. After such discussions, the CSIN Subcommittee will make formal recommendations to the ITSDC.

Next, Dr. Kissman presented Dr. Sid Siegel (EPA), CSIN Administrator, who reported on the receipt of the final draft from the University of Pennsylvania (U. of P.) contract. The report titled, "Systems Requirements Analysis for the Chemical Structure and Nomenclature System (CSNS)" was distributed to the Subcommittee in conjunction with the October 11, 1979 presentation of highlights of the study by Dr. David Lefkovitz (U. of P.). Dr. Siegel noted that a subgroup of the CSIN Subcommittee would be meeting with the Chemical Abstracts Service (CAS) in an information gathering follow up. The subgroup of the Subcommittee will meet again on November 20, 1979 to discuss all the facets of CSNS and subsequently draft recommendations to the CSIN Subcommittee as to the developmental direction to be taken for the CSNS.

Receipt of their final report on the Chemical Data Base Directory (CDBD)/Chemical Information Resource Directory (CIRD) marked the end of the MITRE contract. Copies are available through Rita Bergman (CEQ). An RFP is being prepared for continuation of development of the Directory.

In an area related to the CIRD, Dr. Kissman reported as to the possibility of working with the National Referral Center (NRC) of the Library of Congress (LC) to issue a Toxicology related subset of information resources from the NRC data base titled, "Directory of Information Resources in the United States; General Toxicology." An Interagency Agreement is being drafted at NLM and LC, under which the involvement of LC would be defined.

## 2. Toxicology Information Subcommittee

Dr. Kissman then reported on activities of the Toxicology Information Subcommittee (TIS). The Subcommittee discussed the computerization of government monographs including such information sources as the Criteria Documents produced by the National Institute for Occupational Safety and Health (NIOSH) as well as documents published at other agencies. Subcommittee members discussed methods of storing these documents in a system to allow for efficient search and easier reissue of the documents. Dr. Frank Mackeson (NIOSH) presented an overview of the production of the NIOSH Criteria Documents. Subcommittee members were interested in further investigating the possibility of computerizing the Criteria Documents. There was interest also in establishing the feasibility of such a computerization, using a machined Criteria Document in an available system.

The Chemical Hazard Alert System (CHA), a recommendation from the Toxic Substances Strategy Committee report to the President, was also discussed by TIS members. Dr. Kissman stated that the Subcommittee suggested a system to include two groups: one to monitor the literature for potentially hazardous chemical situations either from uses of a new chemical, or from new uses of older, frequently used ones and; a second group that would evaluate reports of the monitoring group, decide on the validity of the indicated potential concern and channel selected information to relevant agencies. The Subcommittee recommends a feasibility study to determine who, if anyone, is currently performing this service, and to try to retrospectively determine if having such an information intelligence service would have helped to avoid or lessened the severity of a past chemical crisis. The scope of resources needed (funds and staff) should be examined as well. The Subcommittee also suggests that the Defense Intelligence Community could constructively advise on the handling of an information analysis system like the CHA. Learning more about activities of the Congressional Research Service might also offer insight into this Hazard Alert Activity.

Dr. Kissman stated that he had written a letter to the ITSDC chairpersons recommending the selection of a small steering group to evaluate development options and to begin work on selecting a group to perform the feasibility study.

Dr. Kissman then reported that the Laboratory Animal Data Bank, a TIS project, is working toward a January, 1980 date for a limited public access to the Data Bank. Administrative plans concerned with user training and fee collection are now in development.

Dr. Terri Damstra (NIEHS) reported that the TIS Interagency Response to Chemical Crisis project, expects to have the bibliography of the non-crisis search, "Effects of Toxic Agents on the Immune System" and "Asbestos in Air" bibliographies available in December. While no distribution mechanism yet exists, copies of bibliographies can be made available. A system to allow the Federation of American Scientists Experimental Biology (FASEB) to publish the IRCC bibliographies is expected to be ready in early 1980. Dr. Damstra also reported that two current searches, "Composition of Toxic Agents in Chemical Dumps," and "Carcinogens in Inorganic Compounds in Synthetic Fuel Technology" are to be completed in March, 1980.

This concluded the Toxicology Information Subcommittee Report.

IV. Overview of the University of Pennsylvania Report on Requirements for a Chemical Structure Nomenclature System (CSNS).

Dr. David Lefkovitz (U. of P.) stated that the report contains both the functional specifications of a CSNS as well as planning recommendations and additional study topics which address increased system flexibility. The Subcommittee was also reminded that any recommendations made by an ad hoc subgroup of technical experts will be reviewed by the Subcommittee to incorporate into their own recommendations to the parent ITSDC.

The U. of P. study began in 1978 and addressed its task in two phases. First, a CSNS users requirements study was performed and a preliminary report issued. The selection of functional specifications of an "ideal" CSNS and a subsequent assessment of some performance characteristics of a number of existing, computerized chemical information systems made up phase two. It was concluded that CSNS development would require the combination of a number of features available in individual systems but that no one existing system would fulfill all of CSNS requirements. The study produced a matrix of about 150 functional requirements for the system of which three are major requirements.

Continuing, Dr. Lefkovitz explained the approach taken to the study. A sample of government agencies were interviewed. Site trips were taken to view actual systems. A substructure search methodology workshop, attended by 20 representatives from government, industry and academia, was held. Some 11 systems were analyzed, including the proposed CAS System. Next, there was in depth direct testing of three "finalist" systems, CHEMLINE, the system which is used by NCI's Division of Cancer Treatment, and the SANSS component of CIS. Finally, there was communication with the Computer Corporation of America (CCA) on interface problems between CSNS and CSIN.

Dr. Lefkovitz first reviewed the conclusions of the workshop/study group that specified the functional requirements of CSNS. Recommended search capabilities included searching by: CAS Registry Number, synonymous names, name fragments; complete or partial molecular formula; substructure; spectra and by algorithmically encoded features (i.e., "screens" or "keys").

The CSNS search must be able to accomodate Undefined, Variable Composition and Substances of Biological Origin (UVCB), i.e., substances that cannot be readily defined by structural formula.

Dr. Lefkovitz then described some operational system requirements which were delineated by the workshop participants. Regarding file size, the group felt that the system should be able to support a minimum of one million compounds, and that ultimately, it might need to be able to search the total CAS Registry File.

The workshop determined that the community of users would be almost universal, including government, industry and research groups. The mode of operation would be interactive. CSNS should be able to serve 200 full structure searcher simultaneously or 50 substructure searchers. Dr. Lefkovitz noted here that the technical approach taken by CAS in creating an interactive search system is different from the one suggested in the U. of P. study.

In the area of more specific system requirements, the CSNS inquiry language should be capable of both great precision and flexibility (variability). (See page 1 of the attachment). Among the most effective such languages at present are the Merck system with its "X-Z" capability, and the Upjohn system. Dr. Lefkovitz noted that there should be informational compatibility with the CAS Registry III Connection Table. Lastly in this area, there should be user-created catalogue keys for performing current awareness searches, or creating profiles of interest.

Dr. Lefkovitz then listed three major development problems. The first is graphic input and inquiry variability. This problem is fairly well solved by approaches developed by Merck and Upjohn, as well as by some methods being used by SANSS, NCI, and others. The second problem revolves around operational issues associated with the interactive search of a multi-million compound file. The third problem pertains to output; there should be high quality structural display for both low and high speed terminals.

Dr. Lefkovitz then spoke about the solutions to these problems. For the input, and more specifically the graphic input problem, the solution is partly a system of keyboard commands, as developed by Dr. Richard Feldman (DCRT, NIH), which has been incorporated into both the NCI and SANSS systems. This approach has two advantages: (1) it can be used on non-graphic, ordinary character terminals; and (2) it can enter a structure rapidly. However, the system lacks the ability to craft certain types of configurations. Three or four bonds from a single origin, or a bridge across the middle of a ring structure are examples of chemical graphics not well processed by the system. In this area, the Rohm and Haas Company has developed a cursor graphic system that handles such structures better.

From an inquiry variability standpoint, the X-Z substitutions of Merck are recommended. The report also cites the ring/nucleus keys of the NCI system, and the atom-centered fragments in both the NCI and SANSS systems, as being potentially useful here. SANSS has a powerful set of keys called RPROBE which could possibly be incorporated into the CSNS. Some upgrading of the RPROBE keys is currently being considered for SANSS.

Dr. Lefkovitz then addressed the output/printing problem. He explained that chemical structures may be represented in two ways in the computer. One is by a vector method, and the second by a matrix. The difficulty is that while the structure legibility is much higher in the vector representations, the quality of current printer terminals is substantially better in the matrix mode. Laser printers with vector capability are starting to be produced, but are not yet well developed. Transferring of chemical structure information into computer form, is a "solved" software problem. The study recommends the use of a program developed by CAS to assign coordinates to the Connection Table. Thus, the solution to the output problem is firstly, the use of the CAS vector display program with existing devices, like Tektronix or Hewlett-Packard terminals. Secondly, while waiting for the development of a quality laser printer in vector mode, the study recommends the development of a conversion program to print on a matrix device, or use the existing matrix display program in SANSS, which would need some upgrading for CSNS use.

To resolve the operational problem, Dr. Lefkovitz described three alternative system configurations (page 2 of the attached graphics). One uses CHEMLINE which is able to accomodate 750,000 compounds in an interactive search mode. A second approach is to develop a new file system based on bit-map technology as in the National Cancer Institute's system which he feels is upgradable to the handling of 5 million compounds. These two options are Inverted List systems.

A third approach is the Chemical Abstracts Search Machine, which Dr. Lefkovitz describes as a serial/parallel approach, i.e., while it processes the bit screens in a serial mode, it simultaneously breaks them up and performs certain parallel operations.

The differences between the Inverted List systems and the CA Search Machine are these: (1) Mode of interaction - the Inverted List screen search is faster, while the Search Machine achieves interaction by interspersing screen and iterative search; (2) Inverted Lists can accomodate arbitrarily large index term (screen) vocabularies without response time degradation - this allows for open ended structure screen vocabularies, the type in NCI and SANSS. This also leads to more effective combining of structural and non-structural search. Some "secondary" differences are that the Search Machine file is easier to maintain, but that multiple file logic is simpler with the Inverted List Systems.

Dr. Lefkovitz then discussed some CSNS development plans for a three year period (see page 3 of attached graphics). The U. of P. study recommends proceeding with the common development of structure input and output, an estimated two year effort. Early during this two years, a decision should be made on whether to upgrade CHEMLINE (a



1.5 year effort) or to develop a new file system, an estimated two year effort. The CHEMLINE option offers lower cost and quicker implementation of a system, but a file size limit of 750,000 compounds. A new file system, while more costly in funds and time, would accomodate a one to five million compound file.

At the two year mark, then, a functioning CSNS will exist. Three ultimate CSNS configurations then emerge (page 4 of graphics). The first addresses the possibility that the CAS system for whatever reason, would be unsuitable. Here, the new file system would be extended to five million compounds. The second proposes a common front end with the CAS system to allow full Registry Search, while maintaining one of the Inverted List systems, CHEMLINE or a new file system, which would accomodate 750,000 to one million compounds. The third possibility is a total CAS system, with the ability to search a five million compound file.

#### V. Technical Development of the Chemical Abstracts Service System

Ms. Bastian introduced Mr. Nick Farmer, Chemical Abstracts Service (CAS) who gave this presentation. Mr. Farmer first explained that CAS is a non-profit division of the American Chemical Society, and has offices in Columbus Ohio. Since 1907 they have abstracted and indexed the chemistry literature and produced various hard copy and machine readable compilations.

Current services are what Mr. Farmer called standard or comprehensive services, including printed Chemical Abstracts (CA) and CA search services as well as more customized services tailored to a user's needs, represented by CA selects and corporate profiles. The CA Registry is a key part of the operation, now containing information on 4.6 million substances, with 7.2 million names (synonyms, common names, trade names, etc.). Future services, the provision of on-line use via networks out of a Columbus based computer, will be based on the Registry.

In designing a system for new services, CAS spoke to 40-50 organizations, their chemists, and information scientists. The ideal system environment includes an integrated system architecture, public and private files, a common user interface, and flexibility for growth. Mr. Farmer used a "Service Delivery Model" (page 5 of attached graphics) illustration. Four major components are: the terminal component, which will exist at the user site, perhaps as an intelligent microprocessor; a data base component, which will build and search the data base; an interactive component, a major part of the program which will interact with the user to do query framing and results

display; a computational component which would require a CPU for functions including iterative search. These components will be clearly separated and connected by a network.

Mr. Farmer then described the parts of the system related to substance search (see graphics, page 6). He pointed out the interactive component which has two sub-parts: the online structure input system which is an interactive graphic system to put structure diagrams into the computer in readable form; and an online name input system; and the full structure identification component which includes the registry substructure file.

The CAS approach to substructure searching is called the Substructure Search Machine, a network of interconnected mini computers. Two major activities are currently underway: (1) the development of a pilot substructure search system which will start with 750,000 compounds, and (2) a private file substructure search system for the National Cancer Institute. Separate now, the goal is to bring these two systems together into a target substructure search system which will have a common interface between the public (pilot) and private files (graphics page 7).

Mr. Farmer then addressed the Target Substructure Search System and listed its four major components: a data base for search and retrieval; a query input capability; a screen search and; a retrieval output component.

The data base component for searching will have a Connection Table component and a screen search component. These screens (approximately 2000) will be as a dictionary type that will allow selection of 1%-2% of the file as an answer set. The screen sets have an efficiency of 77.7% (see graphic page 8).

The CAS Registry III Connection table will be used as a basis for the search system, to do atom-by-atom search to reduce "noise", to eliminate non-relevant answers, and to generate structure diagrams, using the algorithmic structure display program.

The Question Input component allows the user to translate the mental image of a question into search representation. It includes an intellectual process that the chemist uses in interacting with the system, and a mechanical process which entails getting the structure representation into a computer usable form.

The substructure search component is expected to use a Hewlett-Packard Intelligent graphic terminal with a table and stylus, a keyboard and a display service which has two major parts, a structure display area and a menu area. The system is menu-oriented, rather than command oriented.

The menu (an approximate example is seen on page 9 of the graphics) has a create mode and a modify mode. Pointing to the menu with the stylus, a user can choose the create mode and specify a structure (see pages 10-12 of graphics). A phenol group (Ph) for example, can be added by using the "Fetch" command, and bonded to the structure by using the stylus to point out the atom to which the (Ph)

group is to be attached. The system corrects for proper geometry. The system can also accomodate multiple variables in one question (page 13 of graphics) as well as coordinated structural and non-structural search (page 15 of graphics).

Some output devices are a standard character terminal/printer, a modified character terminal/printer, a graphics terminal, a dot matrix printer, and a photo composer.

Mr. Farmer then discussed a search scenario (page 16 of graphics). It would take approximately 35 minutes on the average. The diagram shows, in solid lines, the part of the system that would be in use, interactive components, data base component, the computational component. The screen search and iterative search overlap in time since they are independent parallel processors. In this way, it would take 30 seconds to search a 10 percent sample of the file. The user might then rethink the question, modify the structure, and resubmit for another search. This may be yet again repeated. Once the user feels, from the type of answers returned from the sample 10 percent search, that the question has been framed properly, the system can be directed to search the whole file, overlapping the screen and iterative searches. It might then take one to five minutes to get the final answer.

Mr. Farmer noted that the system offers Full Substance Identification by Registry Number, name, structure and molecular formula. Also featured is substructure search with structure input, efficiency query input, combined structure/non-structure searching and retrospective and current awareness.

Current related work includes work on system integration problems, evaluating some non-substance screens, improvement of Search Machine design to reduce search time, addition of a Xerox laser printer.

Plans for 1980 include installation of the pilot and its operation which will continue through 1982. In the spring of 1981, development priorities will be considered, with input from the market place, customers and the government. Some possible future priorities for development are expanding some features currently in the pilot; addition of applications such as molecular modeling; and expansion of substructure search to the full file search with the new Search Machine architecture.

The next meeting of the ITSDC will be on November 13, 1979 at the New Executive Office Building. As there was no further business, the meeting was adjourned at 12:05 p.m.

ATTENDEES  
FEDERAL

<u>Name</u>	<u>Organization</u>	<u>Phone</u>
Carroll Leslie Bastian	CEQ	395-4980
Rita Bergman	CEQ	395-5763
Linda Billings	EPA/OTS/OPII	755-8040
Joseph D. Bloom	DOD	295-1453
Cathy Brooks	EPA/MIDSD	755-0811
Jerry Calderone	OASH/PHS/DHEW	472-5194
Kenneth Chu	NCI	496-1152
Jerry L. Coffey	OFSPS	673-7974
Ramond E. Corcoran	Bureau of Mines	634-1318
George E. Cushmae	DOT/MTB	426-2311
Terri Damstra	NIEHS/NIH	8-629-3471
Winston de Monsabert	FDA/OC	443-4505
Erika Graf-Webster	EPA	821-2166
Gary Green	NCTR	501-541-4534
Bernard Greifer	DOC	377-3078
Vera Hudson	NIOSH	443-2100
Henry Kissman	NLM/NIH	496-3147
Albert Konvicka	NCTR	501-541-4534
Richard J. Lewis, Sr.	NIOSH/DHEW	8-684-8317
Calvin M. Menzie	FWS/FER	343-6521
Robert Nicholas	CEQ	395-4980
William E. Rhode	NIH/OD	496-9285
Sidney Siegel	EPA	755-8040

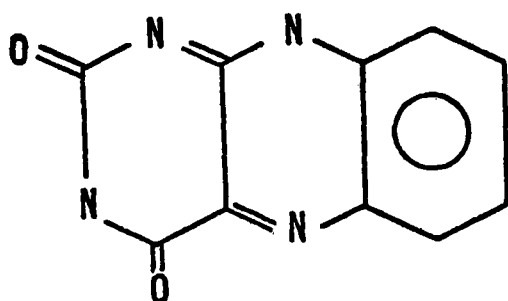
ATTENDEES  
NON-FEDERAL

<u>Name</u>	<u>Organization</u>	<u>Phone</u>
E. P. Bartkus	Du Pont	302-999-4248
George Hoffman	NAS	389-6351
Judith M. Hushon	MITRE	827-6930
N. A. Kippenhan	3M	612-736-1807
Lillian F. Koehler	Johnson & Johnson	201-524-9522
William H. Krebs	GM	313-556-1597
R. E. Maizell	Olin Corp.	203-789-6038
J. H. Prager	3M	612-733-5535
Cheryl Russel	BNA	452-4583
J. N. Weitzel	P. Q. Corp.	215-293-7352
Adrienne Whyte	Biotechnology	703-573-3700
Richard Winter	CCA	617-491-3670

## MORE SPECIFIC SYSTEM REQUIREMENTS

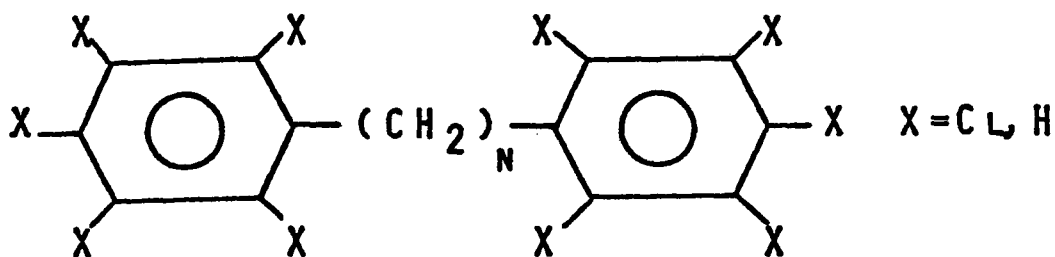
- INQUIRY LANGUAGE CAPABLE OF BOTH PRECISION AND VARIABILITY

### PRECISION



NO TAUTOMERS ALLOWED

### VARIABILITY



X = C, L, H

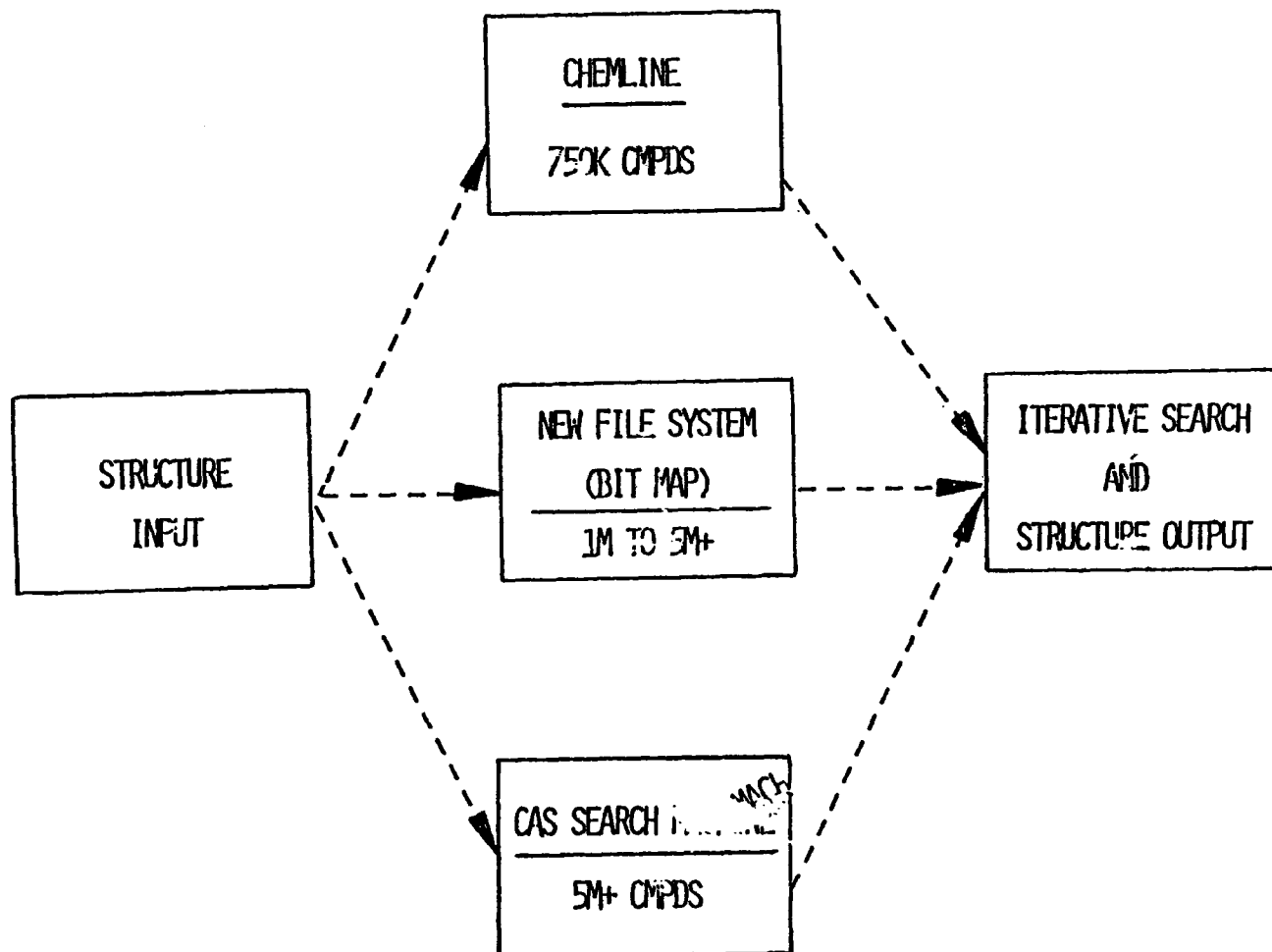
- COMBINED STRUCTURAL AND NON STRUCTURAL INQUIRY PARAMETERS
- INFORMATIONAL COMPATABILITY WITH CAS REGISTRY III CONNECTION TABLE

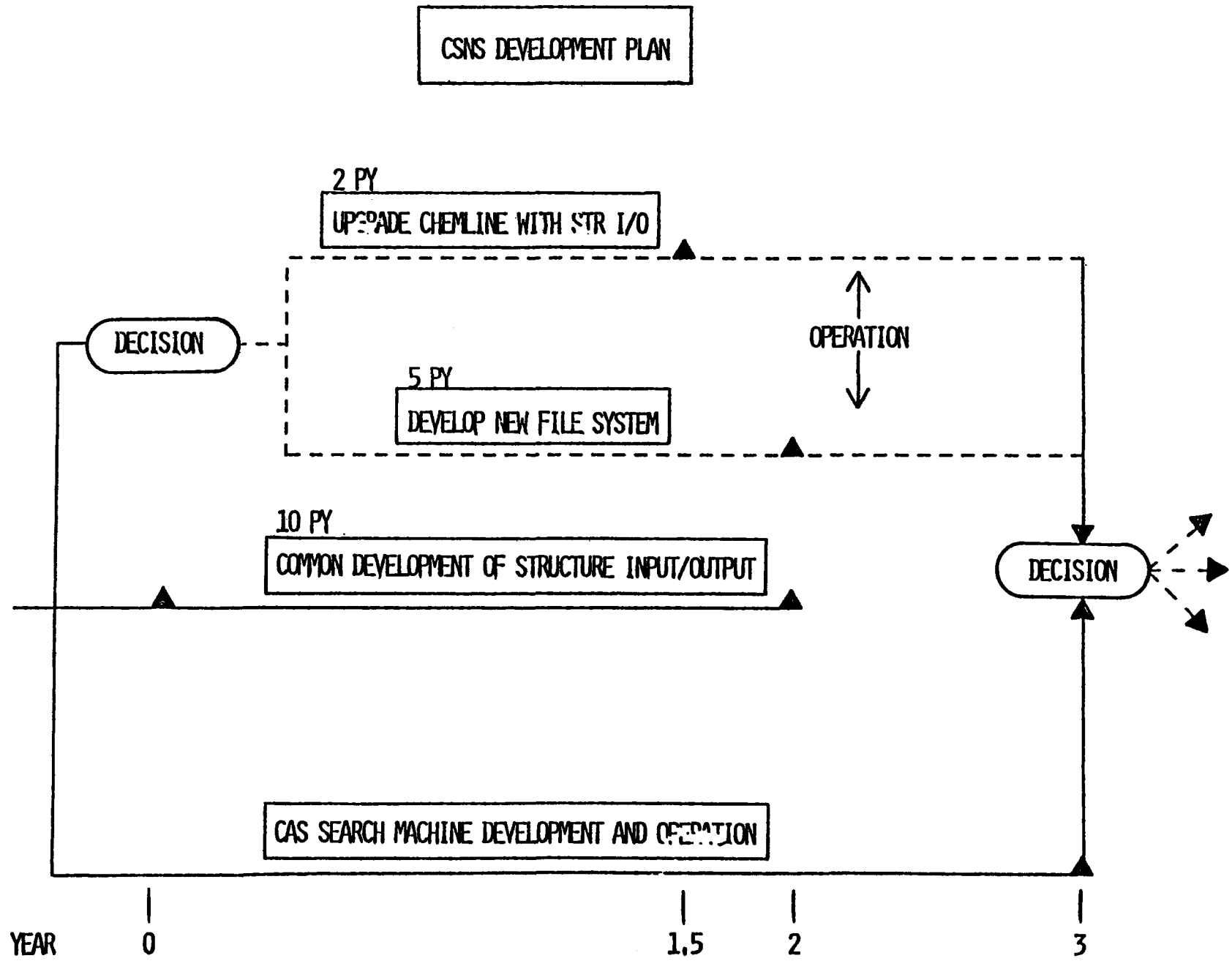
# ALTERNATIVE SYSTEM CONFIGURATIONS

INPUT

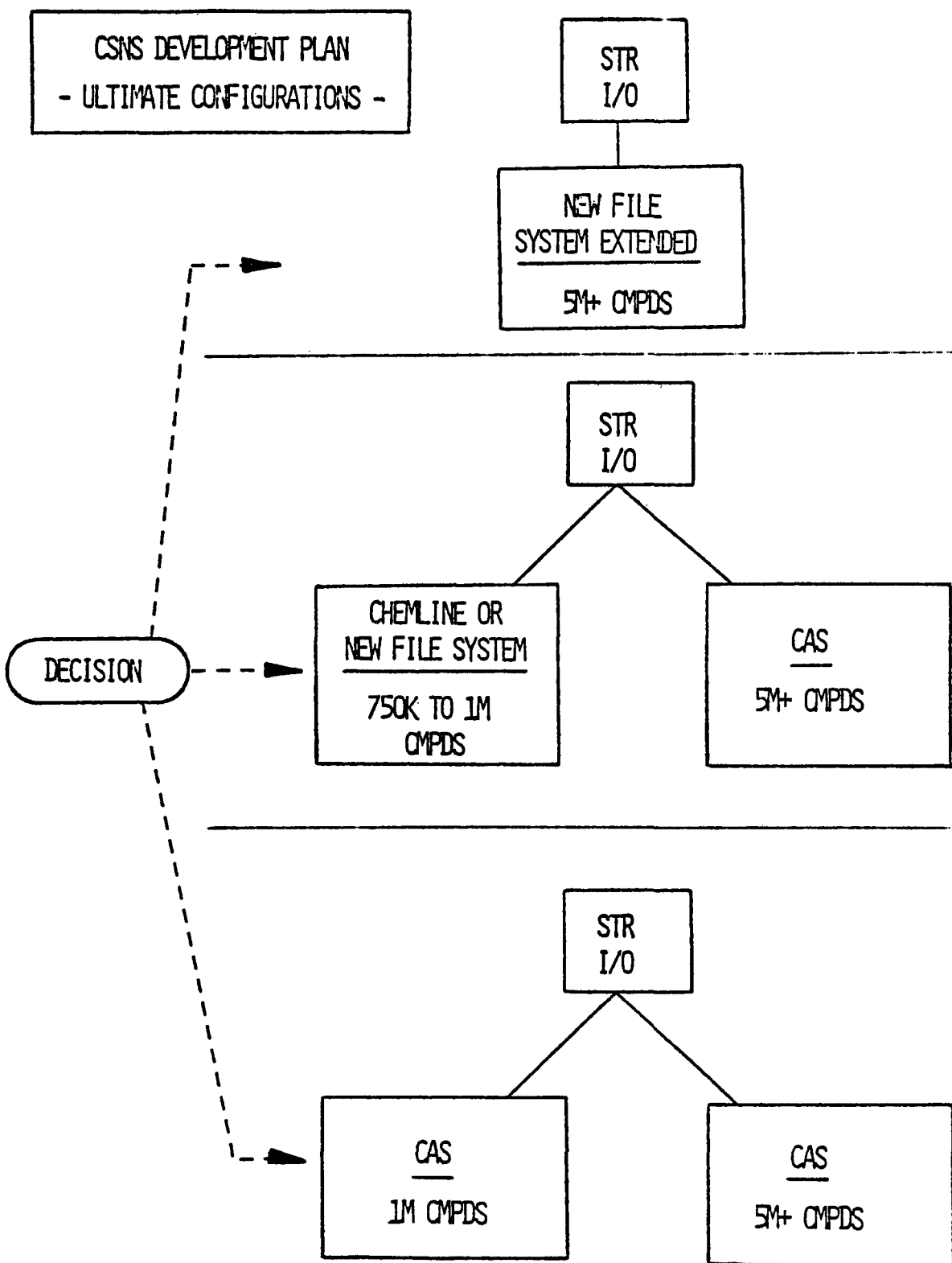
SCREEN SEARCH

OUTPUT

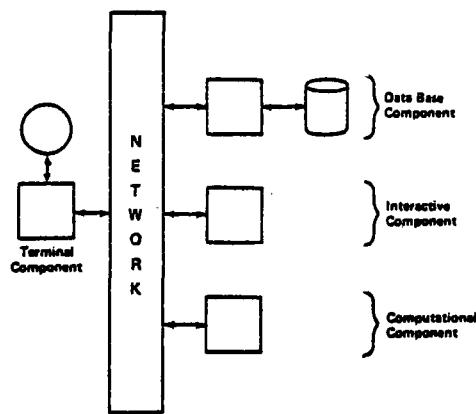




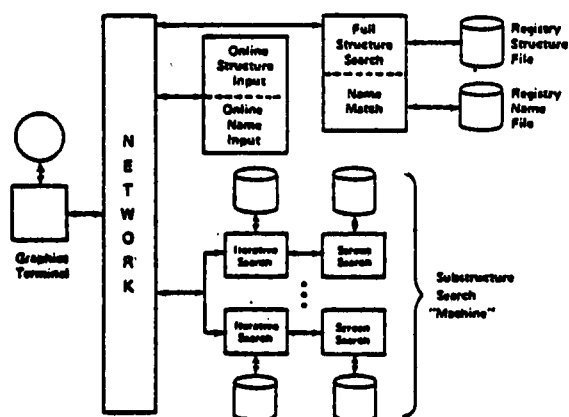




# SERVICE DELIVERY MODEL

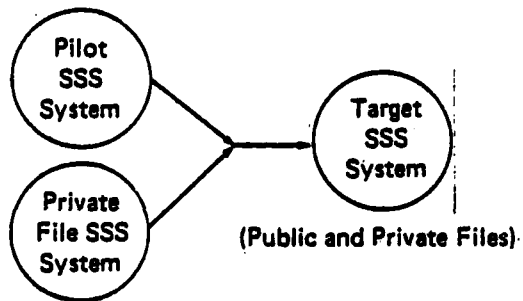


## SUBSTANCE SEARCH RELATED COMPONENTS





### CURRENT SSS ACTIVITIES





### SAMPLE SEARCH QUERIES

	Gross Screenout (%)	Efficiency (%)
NCI	92.9	40 - 50 (est.)
SANSS	98.9	45.0
CAS/BASIC	98.8	77.7

## MENU SUMMARY

THERE ARE TWO MODES OF OPERATION  
'CREATE' AND 'MODIFY'. IN THE CREATE  
MODE, NEW STRUCTURES CAN BE STARTED  
ATOMS OR RINGS CAN BE BONDED TO  
EXISTING STRUCTURES, AND FUSED OR  
SPIRO RING SYSTEMS CAN BE CONSTRUCTED.  
IN THE MODIFY MODE, THE ATOMS OR  
BOND TYPES OF AN EXISTING STRUCTURE  
CAN BE CHANGED.

## CREATE QUERY Menu

MAIN

SEARCH

MOVE

CR MD

( CR mode only )

BOND

SPIRO

R3 R4

R5 R6

R7 R8

( CR or MD modes )

DELETE

C H

N O

P S

F Cl

Br I

Z X

Q Nt

( MD mode only )

BO RCB

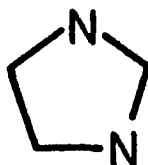
B1 B2

B3 B4

RB CB

B14 B24

20



f ph

# CREATE QUERY Menu

MAIN

SEARCH

MOVE

CR MD

( CR mode only )

BOND

SPIRO

R3 R4

R5 R6

R7 R8

( CR or MD modes )

DELETE

C H

N O

P S

F Cl

Br I

Z X

Q Nt

( MD mode only )

B0 RCB

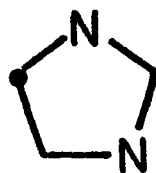
B1 B2

B3 B4

RB CB

B14 B24

Ph



f ph

# CREATE QUERY Menu

MAIN

SEARCH

MOVE

CR MD

( CR mode only )

BOND

SPIRO

R3 R4

R5 R6

R7 R8

( CR or MD modes )

DELETE

C H

N O

P S

F Cl

Br I

Z X

Q Nt

( MD mode only )

B0 RCB

B1 B2

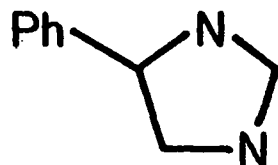
B3 B4

RB CB

B14 B24



Ph



# CREATE QUERY Menu

MAIN

SEARCH

MOVE

CR MD

( CR mode only )

BOND

SPIRO

R3 R4

R5 R6

R7 R8

( CR or MD modes )

DELETE

C H

N O

P S

F Cl

Br I

Z X

Q Nt

( MD mode only )

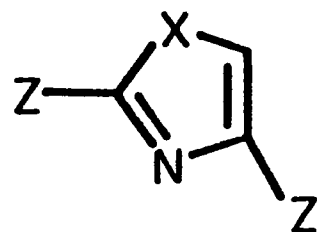
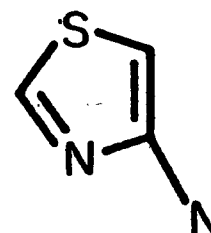
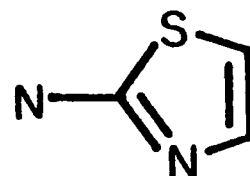
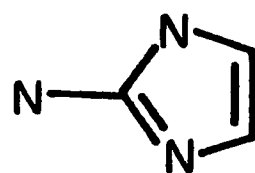
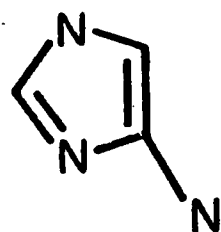
BO RCB

B1 B2

B3 B4

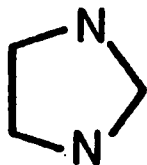
RB CB

B14 B24

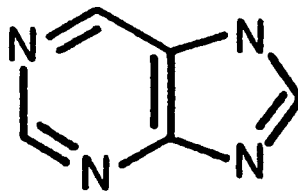


X = N, S  
Z = N, ANY

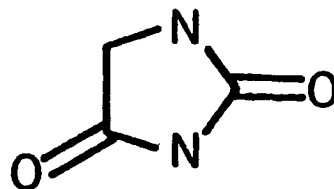
S1



S2



S3



SEARCH Menu

START

STOP

BATCH SEARCH

PAUSE

CONTINUE

DISPLAY

(search control)

KEY LOGIC

STRUCTURE LOGIC

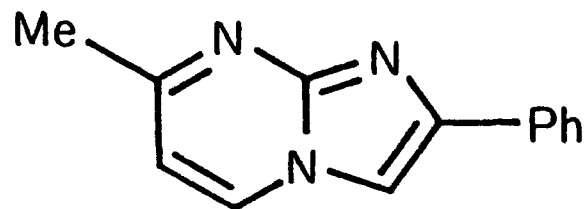
CONTROL PARM

(S1 AND (NOT (S2 OR S3)))

REGISTRY NUMBER 6884-46-4

MOLFORM: C13H11N3

STRUCTURE



CA INDEX NAME (9CI):

Imidazo [1,2-*a*] pyrimidine,

7-methyl-2-phenyl-

SEARCH Menu

START

STOP

BATCH SEARCH

PAUSE

CONTINUE

DISPLAY

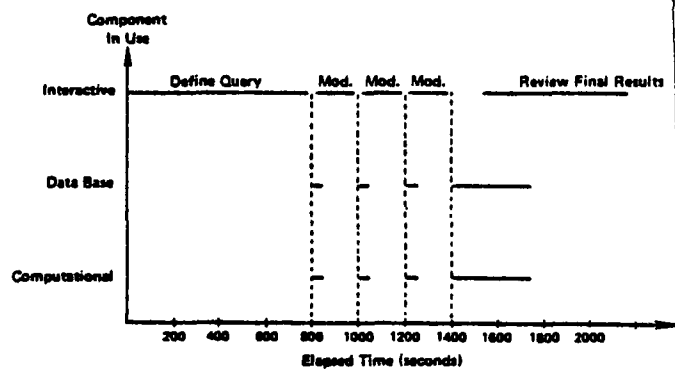
(search control)

KEY LOGIC

STRUCTURE LOGIC

CONTROL PARM

## SAMPLE SCENARIO



# Minutes of the Interagency Toxic Substances Data Committee

December 4, 1979

## I. Introduction

The meeting was held at the New Executive Office Building at 9:30 a.m. and was chaired by Ms. Carroll Leslie Bastian, Council on Environmental Quality (CEQ).

## II. Subcommittee Reports

### 1. Chemical Substances Information Network (CSIN) Subcommittee.

Dr. Henry Kissman, NLM, reported that the Chemical Substances Information Network (CSIN) Subcommittee met November 28, 1979 and discussed plans for the Chemical Data Base Directory (CDBD)/Chemical Information Resource Directory (CIRD). An RFP to further develop this component of CSIN is in preparation at CEQ. A Draft of a Memorandum of Understanding, prepared by Dr. Bernard Greifer, Department of Commerce (DOC), concerning the use of Chemical Abstracts Service (CAS) Registry Numbers in agency files, was distributed and discussed by the Subcommittee. The draft memorandum calls on the ITSDC to investigate the use of DOC mechanisms which could establish the CAS Registry Numbers as a Federal Information Processing Standard (FIPS). Establishing such a FIPS allows the budget process of agencies to seek reimbursement for the cost of adding the CAS Registry Number to their chemical information files. Dr. Kissman added that he subsequently had spoken to a representative of the National Bureau of Standards (NBS) who indicated NBS' interest in the process that would

lead to the FIPS classification of the CAS Registry Numbers. The Subcommittee will present recommendations concerning a final draft of the memorandum to the Parent Committee at the February 5, 1980 meeting.

Rita Bergman, CEQ, at the CSIN Subcommittee meeting, reported on receipt of the first progress report from CAS under their contract to evaluate the feasibility of registering compounds that appear only in the pre-1965 literature. The Subcommittee also was informed of preliminary discussions with the National Academy of Sciences (NAS) on its potential involvement in establishing a peer review of CSIN activities.

## 2. Toxicology Information Subcommittee

Dr. Kissman reported on a presentation made to the Toxicology Information Subcommittee (TIS) concerning the activities of the Information Center Complex at Oak Ridge National Laboratory (ORNL), including efforts of the Toxicology Information Response Center (TIRC). Discussed at the November 28, 1979 meeting was progress in establishing a formal relationship with the National Referral (NRC) of the Library of Congress (LC) to help in the issue of a toxicology related subset of information resources from the NRC data base, entitled "Directory of Information Resources in the United States: General Toxicology." A draft interagency agreement for this activity was presented to the Subcommittee. The agreement allows the NRC to evaluate the possibility of putting together such an information resource, including aspects of indexing and formatting to create and periodically update a publication containing some 1200 relevant information resources.

There was a presentation on the status and background of the publication "Clinical Toxicology of Commercial Products," a sampling of information concerning the toxicology, composition and use of some common commercial products. Status reports were given on the Toxicology Research Directory project, and on Tox-Tips. The Information Response to Chemical Crisis (IRCC) project is finishing two bibliographies, "The Environmental Effects of Acid Rain" and "Composition of Toxic Agents in Chemical Dumps." IRCC is currently discussing the possibility of distribution of its bibliographies by the Federation of American Societies for Experimental Biology (FASEB). The Laboratory Animal Data Bank (LADB) project is expected to be publically available in January, 1980. Access cost will be less than \$20.00 per hour. The LADB will sponsor a workshop on "Generic Nomenclature" on December 6, 1979 at the National Library of Medicine.

Dr. Kissman reported on the start of planning of a second symposium on the handling of toxicological information and invited interested government personnel to join the planning task group.

Both Subcommittees will meet again on January 17, 1980.

### III. Status of the Chemical Substance Information Network (CSIN)

Dr. Sid Siegel, EPA, Network Administrator, reported that through the Preprototype CSIN Task, intelligent terminals will be delivered to the government by February, 1980. Delivery will be to EPA and NLM. One will be kept at the CCA offices in Boston for further developmental work. Dr. Siegel also reported that the CSIN prototype document prepared by CCA had been distributed to the CSIN Subcommittee and to the Parent Committee. A day long meeting to discuss the technical aspects of the document was held



at CEQ on November 27, 1979 with representatives from CCA, the CSIN Subcommittee, and the EPA Office of Program Integration and Information (OPII). Outside experts in attendance were Dr. Shoshani, Senior Computer Scientist, Lawrence Berkeley Laboratories, University of California, Dr. Stanley Su, Professor of Computer Science, University of Florida, Dr. Steve Kimbertone, Manager of Network Development, and his colleague Dr. Pearl Wong of the National Bureau of Standards (NBS). CCA, in presenting in detail their concept of the CSIN prototype, explained user requirements, analysis hardware, software, requirements and issues concerning communication and security. The panel of advisors will each consider all aspects of the prototype including economic feasibility of the project, evaluate the information and give their reports to the CSIN Subcommittee through its prototype subgroup. The experts will later meet with the Subcommittee and CCA for further technical discussion. CCA will then give a one day seminar to the ITSDC in an open meeting. The CSIN Subcommittee will present its formal recommendations to the Parent Committee.

Dr. Siegel reported further activities on development of the Chemical Structure Nomenclature System (CSNS). A CSIN Subcommittee subgroup made up of representatives of NLM, EPA, AND CEQ met with CAS to discuss how the CSIN concept might impact the CA system development activities. A CSNS subgroup met to discuss events of the CAS meeting. A summary of those discussions is in preparation. The CSNS subgroup will make its recommendations for the technical structure of the Phase I CSNS to the CSIN Subcommittee. CEQ and EPA legal counsel are being consulted to help define the Government business posture to be used in USG-CAS talks on the CSNS.

#### IV. An Overview of the Toxic Substances Control Act (TSCA) Section 8.

Mr. Walter Kovalick, EPA, described Section 8 of the Toxic Substances Control Act (TSCA) as a significant tool for information gathering and analysis useful to EPA and other agencies involved in regulatory activities concerning chemical substances. Section 8(a) covers most of the information gathering authority of TSCA. It is further divided into three main levels, each serving a specific purpose toward regulatory development and/or risk assessment. Level A, the Preliminary Assessment Information Rule, can require manufactures to report general use, exposure and production information. It applies to chemical manufacturers and processors. This level will be used to gather basic information on some 2500 chemicals.

Section 8(a) Level B will address a smaller number of chemicals, approximately 100, and require a greater depth of information on them. Section 8(a) Level C would call for even greater depth of information on some 10 to 50 designated chemicals. Levels A and B are expected to be finalized by December, 1980, Level C in March, 1981.

Another rule under development, a follow up to Section 5, can require reports on certain chemical substances that have already been through a premanufacture review process.

An example of a specific application of Section 8(a) is the PPB/TRIS Rule. Proposed on October 2, 1979, it requires any person manufacturing or importing PPB or TRIS, or anyone proposing to do either, to notify EPA. In

another example of specificity, the Section 8(a) Asbestos Rule allows for the gathering of information about asbestos from anyone involved in its mining, manufacture, processing or use in product development. A March, 1980 date is expected for rule finalization. This rule will cover any information needed but not received in response to the Advance Notice of Proposed Rulemaking jointly published by EPA and the Consumer Product Safety Commission (CPSC) in the Federal Register.

Section 8(c) requires the keeping of records and the reporting of allegations of significant adverse reactions to health and the environment by employees and consumers. A Section (c) Rule proposal is expected in January, 1980, and finalization at the end of 1980.

Also under development is a Generic Small Business Rule which would establish standards for the exemption of small business from Section 8(a) rules. Finalization is projected for November, 1980.

Currently EPA and other agencies are working with the help of the Resources Management Act to achieve a gathering of necessary information without duplication of effort between agencies. EPA is currently working on interagency agreements with several other regulatory agencies to allow a free and, where necessary, secure exchange of data.

#### V. New Developments for the Registry of Toxic Effects of Chemical Substances.

Mr. Richard Lewis of the National Institute for Occupational Safety and Health (NIOSH) discussed the Registry of Toxic Effects of Chemical Substances, which is published annually by his organization. The registry is available in book form, microfiche and online through NLM's TOXLINE and

CHEMLINE, and through the NIH/EPA Chemical Information System as one of its tox data files. Updates are available for microfiche and on line versions.

The Registry is a list of toxic substances so far identified in the open scientific literature and those substances' lowest toxic concentrations. 38,000 substances are cross-referenced by 100,000 names. Included is information on molecular weight, molecular formula and structure. A subfile index for manual access to substructure is planned. Mutation data is presently being captured for display. New this year is information on skin and eye irritation properties of some 2,000 chemicals.

Main citations are on toxicity data on chemicals affecting humans and other animals. Specific toxic effects are reported as well as any effects to the Central Nervous System. Animal data concerning carcinogenic potential is reported in three categories: carcinogens, neoplastic, and Equivocal Tumorigenic Agent (ETA). Toxic effects of chemical levels over those administered to control animals is reported. Included also is aquatic toxicity for 400 chemicals. Entries also have reviews of animals carcinogenicity, threshold limit values (taken from the American Conference of Government and Industrial Hygienists, the source of OSHA's standards) and toxicology reviews. Also cited are OSHA standards, the NIOSH criteria documents, and the status of the chemical vis a vis the NCI carcinogenesis bioassay. Recently, the EPA TSCA inventory has been added. NIOSH is currently working to include a "Toxic Effect Code", being developed in cooperation with Dr. Hodge of the University of San Francisco, which relates toxicity effects of a substance to target organ(s) and damage pattern.

## VI. Miscellaneous

Ms. Bastian noted that because of an insufficient agenda from time to time, the ITSDC may not meet every month. Notice of meeting cancellations will be published in the Federal Register.

Ms. Bastian also reported that a Chemical Hazard Alert System subgroup met on November 28, 1979. The system would provide for the review of the scientific literature to identify potential chemical hazards as a result of the production of a new chemical, or the increased production or new use of an older chemical. The subgroup identified some issues to be considered in the framework of a feasibility study and will meet again to further outline such a feasibility study. There is also discussion of a one day meeting to aid in this work. Ms. Bastian invited any comments on this project.

The next ITSDC meeting will be January 8, 1979. The meeting was adjourned at 11:30 a.m.

ATTENDEES

Federal

NAME	ORGANIZATION	PHONE
Martin Aronoff	NBS	921-3491
Carroll Leslie Bastian	CEQ	395-4980
Linda Billings	EPA	755-8040
Kenneth Chu	NCI	496-1152
Jerry Coffey	OFSPS	673-7974
George Cushmac	DOT/MTB	426-2311
Winston R. de Monsabert	FDA	443-4505
Nan Fremont	EPA	755-8040
Bernard Greifer	DOC	377-3234
Vera Hudson	NIOSH	443-2100
Henry Kissman	NLM	496-3147
Walter Kovalick, Jr.	EPA	755-2778
Richard Lewis, Sr.	NIOSH	(913) 684-8317
W. M. Parsons	NMRD	295-1028
Suzanne Rudzinski	EPA	755-5851

ATTENDEES

Non-Federal

NAME	ORGANIZATION	PHONE
Chung-Haeahn	BioTechnology, Inc.	(703) 573-3700
Toni Fedorowski	JRB Associates	(703) 821-4658
Brad Green	Koba Associates	387-8840
Judith Hushon	MITRE	827-6930
Lillian Kochler	Johnson & Johnson	(201) 524-9522
William Krebs	General Motors	(313) 556-1597
Tony Miller	Battelle	(703) 790-1660
Chris Perry	Battelle	(703) 790-8980
Christine Shine	Monsanto Co.	452-8880

# INTERAGENCY TOXIC SUBSTANCES DATA COMMITTEE

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

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

## AGENDA

January 8, 1980

9:30 a.m. - 12:00 noon

New Executive Office Building

Room 2010

- I. Introduction
- II. Status of CSIN
  - Dr. Sidney Siegel, Network Administrator, EPA
- III. DuPont's Environmental Information Programs
  - Dr. Edward Bartkus, Manager of Information Resources, Information Systems Division, DuPont Corporation



# MEMORANDUM

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE  
PUBLIC HEALTH SERVICE  
NATIONAL INSTITUTES OF HEALTH  
NATIONAL LIBRARY OF MEDICINE

TO : Members, CSIN Subcommittee of the  
Interagency Toxic Substances Data Committee

DATE: August 8, 1979

FROM : Chairman, CSIN Subcommittee

SUBJECT: No. 1: Minutes of the June 28, 1979, Meeting  
No. 2: Agenda Items for the August 23, 1979, Meeting

Enclosed are the minutes of the CSIN Subcommittee meeting of June 28, 1979.

The next meeting of the Subcommittee has been scheduled for August 23, 1979, from 1:00 to 4:00 p.m. in Conference Room B, Mezzanine level, National Library of Medicine. The following items are proposed for the agenda:

## I. Report from the Network Administrator

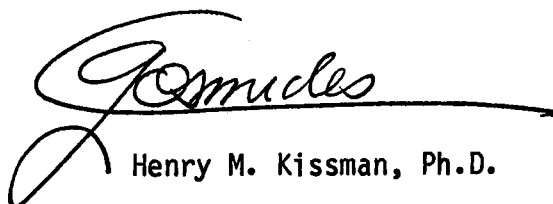
- A. Chemical Data Base Directory (CDBD)
- B. Network Architecture
- C. Chemical Structure and Nomenclature System (CSNS)

## II. Follow-on Contracts

- A. CAS Registration of Historical Information
- B. Proposal for CDBD
- C. Discussion of the Budgetary History for the Chemical Substances Information Network (CSIN)

## III. Miscellaneous

There will be a meeting of the Toxicology Information Subcommittee of the DHEW Committee to Coordinate Environmental and Related Programs on Thursday, August 23, 1979, from 9:00 a.m. to 12:00 in Conference Room B (Mezzanine level) of the National Library of Medicine. The agenda of this meeting includes (1) a discussion of the activities of the National Referral Center, Library of Congress in locating toxicological information; (2) a report from the Environmental Mutagen Information Center (EMIC) and the Environmental Teratology Information Center (ETIC) at Oak Ridge National Laboratory; and (3) status reports on a number of Subcommittee-sponsored projects. Members of the CSIN Subcommittee who wish to attend the Toxicology Information Subcommittee meeting are welcome.

 FOR  
Henry M. Kissman, Ph.D.

Enclosure: Minutes

cc: Dr. Marilyn C. Bracken, EPA  
Mr. Kent A. Smith, NLM

Minutes of the  
Chemical Substances Information Network (CSIN) Subcommittee

June 28, 1979

I.     Introduction

The meeting, held at the National Library of Medicine (NLM), was called to order at 1:04 p.m.

II.    Report from Network Administrator

Dr. Siegel (EPA) reported that a presentation had been made to the Senior Environmental Issues Group at the Department of Energy (DOE) to explain CSIN and relate it to the Group's activities.

Dr. Siegel also listed projects being considered for funding by the Interagency Regulatory Liaison Group (IRLG) in the next fiscal year; they range from the common codes project to establishing a model human health record.

III.   Ongoing CSIN Activities

A.    Chemical Structure and Nomenclature System (CSNS)  
      Developments

Dr. Siegel reported that arrangements have been finalized for a performance characteristic test of the CSNS on July 12, 1979. The chaperoned test is being conducted at the Division of Cancer Therapy (NCI). After the test, the University of Pennsylvania will prepare its final recommendations for a Chemical Structure and Nomenclature System.

B.    Demonstration of the Chemical Data Base Directory (CDBD)  
      Prototype

Mr. William Caldwell (NLM) reported that the Chemical Data Base Directory (CDBD) is a prototype file containing data on 15 chemical information resources.

In demonstrating a search on the CDBD, Mr. Caldwell described the hierarchic vocabulary used to classify the data elements contained in the information resources. The vocabulary consists of nine categories of information, including for example, chemical identification, production, and exposure effects.

C. Network Architecture -- Computer Corporation of America (CCA) Contract

Dr. Siegel (EPA) reported that there are three basic ongoing activities in this area: (1) a study of CSIN architecture and alternative architecture; (2) a report on a CSIN prototype; and (3) an intelligent terminal proposal.

It was suggested that an outline of current proposals be written to help define the technical evaluation of the project and to keep Subcommittee members advised on the order and content of the various CCA documents.

The benefits of intelligent terminals and related software-hardware alternatives were discussed. A revised report on the interim requirements for the CSIN will be available from CCA within the next month.

V. Presentation of the User Cordial Interface for Multiple Online Systems

Mr. Charles Goldstein (NLM) said that the Lister Hill National Center for Biomedical Communications functions as the research and development arm of NLM and is concerned with the technology of information. The User Cordial Interface (UCI) is the result of efforts to give the user better access to the wealth of online information available. The system can access any data base under ELHILL and is multi-level to accommodate all types of users from the student to the trained NLM reference librarian.

VI. Report on CSIN Five-Year Development Plan

Dr. Siegel reported that a draft of the Five-Year Plan has been prepared and distributed. Comments are invited and they will be integrated into the final document. The plan is concerned with long-range plans for CSIN and addresses major administrative issues regarding the network.

VII. Miscellaneous

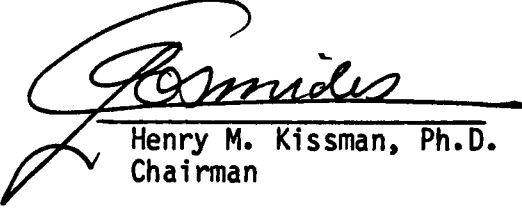
A. Dr. Kissman introduced the idea of a second Symposium on Information Handling in Toxicology, noting that the TIS Subcommittee had approved such a meeting for July, 1980, and was inviting the CSIN to be co-sponsor. He suggested structuring the symposium along the primary, secondary, and tertiary levels of toxicological information. Those present agreed that the CSIN Subcommittee would co-sponsor the proposed symposium.

- B. A memo is being prepared to authorize agency heads, only, to represent agencies involved in CSIN and to clarify CSIN's position in the greater agency structure. Each agency head is being asked to assign a formal agency representative to the Subcommittee.
- C. CCA presented a film about a distributed data-base management system, INQUIRY, that they helped develop for the Department of Defense (DOD). Mr. Richard Winter of CCA noted that the distributive query processing problem addressed in the film is a key point in CSIN development.
- D. The next CSIN Subcommittee meeting will be August 23, 1979, between 1:00 and 4:00 p.m. at the National Library of Medicine.

The meeting was adjourned at 3:57 p.m.

---

Nan Fremont  
Recording Secretary



Henry M. Kissman, Ph.D. *for*  
Chairman

Members Present

Dorothy Drago, CPSC  
John Wilson, DOE  
Sidney Siegel, EPA  
Martin Aronoff, NBS  
L. H. Geventmen, NBS  
Terri Damstra, NIEHS  
Richard J. Lewis, NIOSH  
Henry M. Kissman, NLM

Invited Guests and Observers

Florence Bayard, NLM  
Rita Bergman, CEQ  
Aren Horowitz, CCA  
Tomas Lozano-Perez, CCA  
Richard Winter, CCA  
Leonard Schachter, CPSC  
Bernie Scharf, CPSC  
Roger Connor, EPA  
Erika Graf-Webster, EPA  
Winston R. de Monsabert, FDA  
William Caldwell, NLM  
Charles M. Goldstein, NLM  
Melvin L. Spann, NLM  
R. E. Maizell, Olin Corp.  
William H. Ford, Online Computer Systems  
Helga Gerstner, ORNL

Minutes of the Chemical Substances Information Network (CSIN) Subcommittee of the Interagency Toxic Substances Data Committee, August 23, 1979

I. Introduction

The meeting was held at the National Library of Medicine commencing at 1:20 p.m. Several members of the Toxicology Information Subcommittee, which had met in the morning, stayed over to participate in this session of the CSIN Subcommittee. There were no comments on the minutes of the June 28, 1979 meeting. These minutes therefore stand as written. In the future, minutes of the Toxicology Information Subcommittee and the CSIN Subcommittee will be sent to the membership of both groups.

II. Registration of Compounds in the pre-1965 Chemical Abstracts

This item had been discussed at several meetings of the Interagency Toxic Substances Data Committee (ITSDC) and of the CSIN Subcommittee. A letter requesting Chemical Abstracts Service (CAS) to propose a feasibility study for the registration of compounds that had been indexed in Chemical Abstracts (CA) only prior to 1965 had gone out. The letter was signed by Dr. Kissman as Chairman of the two information committees and Ms. Bastian as co-chairman of the ITSDC. CAS will prepare such a proposal for consideration by the interagency groups in as short a time as possible. Hopefully such a proposal can still be considered for funding in FY '79.

III. Report from the Network Administrator

Dr. Siegel reported that work on a draft interagency agreement with the Council on Environmental Quality (CEQ) concerning funds for CSIN activities had been completed in his office. There is now urgency in placing these funds in contracts so that CSIN work can proceed in the next fiscal year. Dr. Kissman mentioned that an interagency agreement between the National Library of Medicine and CEQ was being finalized. Its purpose is to transfer funds from the DHEW Committee to Coordinate Environmental and Related Programs to the CSIN effort.

IV. Ongoing CSIN Activities

1. Chemical Data Bases Directory (CDBD)

Ms. Rita Bergman reported that the Mitre Corp. is completing work on the descriptive catalogue part of the CDBD.

Mr. John Feulner, National Referral Center (NRC), Library of Congress (LC) had presented a briefing on the activities of the NRC to the TIS in the morning. He had been asked to stay over and present an abridged version of this briefing to the CSIN Subcommittee. (Note: For a report on Mr. Feulner's remarks cf. Toxicology Information Subcommittee minutes of the August 23, 1979 meeting). A discussion of the

NRC and its data base of information resources was of course particularly relevant to the topic of a Chemical Data Bases Directory. When questioned, Mr. Feulner estimated that the NRC data base contained descriptions of some 4,000 information resources in the area of toxicology and chemistry. It seemed clear from the discussion that the NRC had at least the "raw material" for a CDBD. The group wanted to know how this NRC data could be made available to the CSIN effort. Would it be possible to use the data base per se, or perhaps a toxicology subset of it, as a node in CSIN? Could the NRC/LC be funded by the interagency effort to prepare a directory of their information resources for CSIN? Mr. Feulner indicated that he would bring these comments to the attention of LC management and would get back with their reactions to the CSIN group as quickly as possible.

It was mentioned that when the CSIN Subcommittee started its work, it had representation from LC. However, LC participation has not continued. The group thought that it would be useful if NRC could be represented on the CSIN Subcommittee and LC as a whole on the ITSDC. Mr. Feulner agreed also to bring this recommendation to his management.

## 2. Chemical Structure and Nomenclature System (CSNS) Developments

Ms. Bergman reported that a final report from the contractor (i.e. University of Pennsylvania) is expected by the end of August. When received the report will be distributed to the Subcommittee.

Dr. Siegel reported on the "chaperoned" tests comparing three sets of chemical search questions in the chemical information storage and retrieval system of the Division of Cancer Therapy, National Cancer Institute. One question each was submitted by Chemical Information Systems, Chemical Abstracts Service and the University of Pennsylvania. Performance of the NCI system was pretty much as predicted in the preliminary report from the University of Pennsylvania.

## 3. Network Architecture Contract (Computer Corp. of America ) CCA

A report on current activities of this contractor (CCA) was also presented by Ms. Bergman. CCA had been asked to supply a supplemental analysis for their May 29, 1979 report entitled "Immediate Requirements for a Chemical Substances Information Network: Analysis and Recommendation". The supplementary report was presented on July 17, 1979. It listed several options. Option 1 was to proceed with the intelligent terminal system along the lines proposed in the original report. Option 2 would add a disc to the system. Option 3 would also



add word processing capabilities and option 4 would add a general purpose interface capability. Clearly the options represent increasing hardware, and software capabilities with concomitant increases in costs. A telephone poll of "concerned" people, particularly those in EPA, indicated that option 2 should be adopted. It would be easy to move up to options 3 or 4 at a later date if this became necessary. The contractor has been told that option 2 would be accepted and he will act accordingly.

CCA had initiated work on interviewing potential CSIN users. Discussions have been held with NCI, Fish and Wildlife Service (DOI), and Bureau of Foods, FDA. They have interviews underway or planned with OSHA, NIOSH, Bureau of Drugs (FDA), Consumer Products Safety Commission, Department of Energy and Office of Research and Development (EPA). Still to be scheduled and implemented are interviews with Stanford Research Institute and the Environmental Defense Fund.

CCA has also issued a complete report (technical reports, CCA-79-19, August 21, 1979) entitled "A Prototype Chemical Substances Information Network". This is an updated version of an earlier report with the same title. Ms. Bergman asked the Subcommittee members for comments on this report. Dr. Siegel mentioned again that the interagency agreement between EPA and CEQ he had referred to earlier makes sufficient funding available so that work on the CSIN prototype can go forward during the next fiscal year. However CEQ must have comments from the agencies participating in the CSIN activity to make such a decision possible. Dr. Kissman asked whether there is a general workscope for the CCA contract that would allow CEQ to move forward while comments on the recommendations are coming in. Dr. Damstra suggested that there should be expert consultants available to the group for the review of complex systems proposals such as those made by CCA. Ms. Graf-Webster said that at a forthcoming computer meeting in San Francisco, at the end of August, she will try to enlist the help of some experts to form an ad hoc review body. Dr. Siegel pointed out that the EPA-CEQ Interagency Agreement makes it possible to hire such expert consultants.

Ms. Bergman provided some additional details on funding allocation envisioned under the EPA-CEQ agreement. The following areas will be supported.

- (1) secretarial support for three committees, (TIS, ITSDC, CSIN subcommittee).
- (2) technical support for the CSIN network administrator.

- (3) feasibility study of registering compounds in the pre-1965 CA.
- (4) support for the network architecture contract.

There was no specific allocation for continued work on the CDBD or the chemical structure and nomenclature system. However, if some arrangement with the NRC/LC could be worked out in time for funding in FY '79, the former project also could proceed. Mr. Hummel and his staff at NLM had prepared an RFP to identify organizations that collect information about information resources in the general CSIN subject area. If time permits it, one might go forward in this area either with the NRC/LC or some alternate organization.

Ms. Bergman emphasized that if agencies wanted intelligent terminals under the CCA proposal for "immediate requirements for CSIN" this could be done early in 1980. The costs would \$16,000-\$17,000/installation. Agencies should contact her or Ms. Bastion at CEQ.

The discussion reverted to CCA's user study effort. Are the interviews effective? Who in FDA was being contacted? These were some of the questions being asked by the group. Ms. Bergman stated that the agencies had submitted the names of people for such interviews to the network administrator. The people to be interviewed are those who would make use of CSIN to fulfill their daily needs for information. The total list contained the names of some 50 people. CCA made initial contacts and thereby got additional names for subsequent interviews. The methodology of the interviews is as follows: CCA staff conducts a seminar for agency people in which they explain what CSIN is all about. They then go into a question and answer session. They have a questionnaire which they fill out during the interviews. After the session they go back and verify the information they have collected on the questionnaire. Ms. Joan Chase (Carcinogenesis Program, NCI) described her interview session with CCA. She told CCA that she needed information on chemicals that had been selected for long-term bioassay. Ms. Gerstner mentioned that the CCA team had also visited the Oak Ridge National Laboratory. They had sent out some "premeeting" information (slides) and she (Ms. Gerstner) had distributed this material to relevant people at ORNL. Ms. Bergman asked to be informed if the CCA teams were not following up on the interview sessions.

#### 4. Five Year Plan

Both Ms. Bergman and Dr. Siegel emphasized again that comments were needed from the group on the CSIN five year plan that had been distributed. Ms. Graf-Webster stated that she needed comments on: (1) what was left out; and (2) whether the numbers (i.e. funding, milestones, etc.) look unreasonable. In preparing this plan she had input from CCA and the University of Pennsylvania.


#### V. Miscellaneous

Dr. Kissman mentioned that through recent funding action by the Congress, the National Library of Medicine may be given FY '80 funding for implementation of the Chemical Structure Nomenclature System of CSIN. He also asked who was working on the organizational aspects of the interface between the network and nodes designated for the prototype CSIN. Of particular concern here were arrangements that would have to be made if the National Library of Medicine information services (which have been targeted as being in the prototype) are to be linked into the network. Who is planning this work; what sort of discussions with the management of NLM should be initiated, etc.? There was agreement that the network architecture contractor, who would probably be responsible for outlining some of the requirements for such interaction, had not initiated specific work along these lines.

There was concern expressed by Ms. Bergman and Ms. Graf-Webster that the CSIN Subcommittee was too slow a mechanism for obtaining comments on proposals and other action items where a fast turn-around was required. It was suggested that consideration should be given to the formation of a small Steering Group that would be available for immediate response.

The next Subcommittee meeting will be held on October 11, 1979 from 9 a.m. to 12 noon at the National Library of Medicine. It is planned to have CSIN Subcommittee meetings and TIS meetings alternate in using the morning and afternoon time slots for their meetings. The Subcommittee meeting was adjourned at 3:30 p.m.

Signed:

  
Henry M. Kissman, Ph.D.  
Chairman

## CSIN Subcommittee

Members Present

Ms. Claudia Lewis, CDC  
Ms. Rita Bergman, CEQ  
Ms. Dorothy Drago, CPSC  
Mr. Leonard Schachter, CPSC  
Mr. Bernie Scharf, CPSC  
Ms. Erika Graf-Webster, EPA  
Dr. Sidney Siegel, EPA  
Mr. Winston R. deMonsabart, FDA  
Ms. Helga Gerstner, ICC/ORNL  
Dr. George Hoffmann, NAS  
Mr. Martin Aronoff, NBS  
Dr. Terri Damstra, NIEHS  
Dr. Henry Kissman, NLM

Invited Guests and Observers

Mr. John Feulner, NRC/LC

# MEMORANDUM

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE  
PUBLIC HEALTH SERVICE  
NATIONAL INSTITUTES OF HEALTH  
NATIONAL LIBRARY OF MEDICINE

TO : Members, CSIN Subcommittee of the  
Interagency Toxic Substances Data Committee

DATE: October 3, 1979

FROM : Chairman, CSIN Subcommittee

SUBJECT: Agenda Items for the October 11, 1979, Meeting

The next meeting of the CSIN Subcommittee has been scheduled for October 11, 1979, from 9:00 a.m.-12:00 Noon in Conference Room B, National Library of Medicine. The following items are proposed for the agenda:

Memorandum of Understanding Concerning the Use of CAS Registry Numbers in Agency Files

Reports from Other Interagency Committees

Report from the Network Administrator

CSIN Support Activities

Chemical Structure and Nomenclature System (CSNS)

Chemical Data Bases Directory (CDBD)

Network Architecture Contract

Registration of Compounds Indexed in the Pre-1965 Chemical Abstracts

Please note that the meeting of the Toxicology Information Subcommittee is also scheduled for October 11, 1979, at 1:00-4:00 p.m., Conference Room B, National Library of Medicine.

The minutes of the August 23, 1979, meeting are not ready for mailing at this time. They will be distributed at the October 11, 1979, meeting and mailed to those not attending.

  
Henry M. Kissman, Ph.D.

cc: Ms. Erica Graf-Webster, EPA  
Ms. Marion Suter, EPA  
Ms. Rita Bergman, CEQ  
Ms. Nan Fremont, EPA  
Mr. Kent A. Smith  
Ms. Florence M. Bayard

Minutes of the Chemical Substances Information Network (CSIN) Subcommittee  
of the Interagency Toxic Substances Data Committee,  
October 11, 1979.

I. Introduction

The meeting, held at the National Library of Medicine (NLM), was called to order at 9:00 a.m. The minutes of the August 23, 1979 meetings of the CSIN and the Toxicology Information Subcommittees (TIS) were distributed to Subcommittee members.

II. Memorandum of Understanding Concerning the use of Chemical Abstract Registry Numbers in Agency Files

The Chairman reminded the Subcommittee that the Toxic Substances Strategy Committee's Report (in draft) to the President recommended that government agencies adopt CAS Registry Numbers as the identification standard for chemical substances in their files. Consequently, the Council on Environmental Quality (CEQ) prepared the draft of a "Memorandum of Understanding Concerning the Use of Chemical Abstracts Service Registry Numbers in Agency Files" which was distributed to appropriate agencies for comment. Responses to the draft Memorandum resulted in CAS Registry Numbers being recommended, but not required, as the standard chemical identification codes, for both retrospective and current files.

Dr. Bernard Grier, (NBS) added that the Memorandum is essentially an attempt to establish a Federal Information Processing Standard (FIPS) for data bases with chemical information. As the setting of a FIPS is a responsibility of the National Bureau of Standards (NBS), it would be

appropriate for the NBS to explore establishment of the CAS Registry Number as a FIPS. The NBS would first conduct a one to two-year survey to establish the costs and benefits of adopting the CAS Registry Numbers as a standard. During this time, the CAS numbers could still be used on an de facto basis. After the standard is instituted by the NBS, the agencies could be reimbursed for the cost of data conversion. This reimbursement would encourage those agencies whose data conversion costs would make standardization financially prohibitive. Dr. Grier also addressed the effect of instituting CAS Registry Number use on an organization like NTIS which deals primarily with derivative files. If the CAS Registry Number is established as a FIPS, the responsibility for its use would be assumed by agencies submitting information to NTIS, and not by NTIS itself.

Dr. Kissman noted the problems involved in mapping different sets of chemical identification numbers. For example, the ITSDC had heard comments on ways of linking the Department of Transportation (DOT) UN Number, indentifying hazardous chemicals in transport, to the CAS Registry Numbers. As there is no one-to-one relationship between the systems, this is often difficult. Dr. Grier agreed to prepare a paragraph for insertion into the draft CEQ Memorandum of Understanding that identifies the use of the CAS Registry Number as a de facto process so that the agencies can work on conversion problems, while the NBS is invited to study the system with a view to establishing a possible FIPS.

### III. Reports from Other Interagency Committees

Dr. Sid Siegel, EPA, discussed the status of the Information Exchange Group (IEG) of the Interagency Regulatory Liaison Group. The IEG is currently considering the purpose and function of the group, i.e., should IEG simply exchange information between agencies, or should the group, having identified some of the problems and issues, formulate projects to address common needs? Dr. Siegel felt that the group should be concerned with developing projects, and that these projects could be channeled to the CSIN Subcommittee and then be integrated into a larger information activity. During the next few months, the IEG hopes to clarify it's role and to draft a statement of functions.

### IV. CSIN Support Activities

#### 1. Chemical Structure and Nomenclature System (CSNS)

Dr. Siegel, as Network Administrator, reported that while the University of Pennsylvania (U. of P.) feasibility study for CSNS was being completed, the CAS approached the CSIN group and expressed an interest in contributing to CSNS development. (Dr. Kissman noted that CAS is currently creating a substructure search system for the 5 million CAS Registry System compounds, whose capabilities appear to overlap with many of the capabilities described in the U. of P. report in identifying the ideal CSNS.) In response to this CAS interest, Dr. Marilyn Bracken, (EPA), Dr Siegel, Mr. Bruno Vasta (EPA) and Dr. Kissman met in July 1979 with representatives from CAS regarding the U.S. Government's (USG) long and short range



requirements for a CSNS. Following this meeting, the American Chemical Society (ACS) Committee on Chemical Abstracts met in executive session with the aforementioned government group and approved CAS-USG collaboration in CSNS development.

Dr. Chu (NCI), expressed concern that the involvement of CAS at this time might limit our options of reviewing a future route for CSIN development. It was emphasized that no definite commitment had been made to CAS and that discussions with CAS are continuing in an effort to prevent waste of time and resources in overlapping developments of CSNS and the CAS system.

Dr. David Lefkovitz (U. of P.) then gave an overview of the U. of P. report on CSNS requirements. [The report is available for distribution.] He stated that the report contains both the functional specifications of CSNS as well as some planning recommendations and additional study topics that address increased system flexibility. The Subcommittee was also reminded that any recommendations made by an ad hoc subgroup of technical experts currently being identified, will be reviewed by the Subcommittee to incorporate into their own recommendations to the ITSDC.

The U. of P. study began in 1978 and addressed its task in two phases. First, a CSNS users requirements analysis was performed and a preliminary report was issued. The selection of functional specifications of an "ideal" CSNS and a subsequent assessment of a number of existing, computerized chemical information systems made up phase two. It was concluded that CSNS development would require the combination of a number

of features available in individual systems but that no one existing system would fulfill all of CSNS requirements. The study produced a matrix of about 150 functional requirements for the system of which three are major requirements.

Continuing, Dr. Lefkowitz explained the approach taken to the study. (Please see the attached graphics.) A sample of government agencies were interviewed. Site trips were taken to view actual systems. A substructure search methodology workshop, attended by 20 representatives from government, industry and academia, was held. Some 11 systems were analyzed, including the proposed CAS System. Next, there was in depth direct testing of three "finalist" systems, CHEMLINE, the system which is used by NCI's Division of Cancer Treatment, and the SANSS component of CIS. Finally, there was communication with the Computer Corporation of America (CCA) on interface problems between CSNS and CSIN.

Dr. Lefkowitz first reviewed the conclusions of the workshop/study group that specified the functional requirements of CSNS (cf. attached slides). Recommended search capabilities included searching by: CAS Registry Number, synonymous names, name fragments, complete or partial molecular formula, substructure, spectra and by algorithmically encoded features (i.e., "screens" or "keys").

The CSNS search must be able to accomodate Underfined, Variable Composition and Biological substances (UVCB), i.e., substances that cannot be readily defined by structural formula.

Dr. Lefkowitz then described some operational system requirements which were delineated by the workshop participants. Regarding file size,

the group felt that the system should be able to support a minimum of one million compounds, and that ultimately, it should be able to search the content of the total CAS Registry System (i.e., in excess of 5 million compounds).

The workshop determined that the community of users would be almost universal, including government, industry and research groups. The mode of operation would be interactive. CSNS should be able to serve 200 full structure searches or 50 substructure searches users simultaneously. Dr. Lefkovitz noted here that the technical approach taken by CAS in creating an interactive search system is different from the one suggested in the U. of P. study.

In the area of more specific system requirements, the CSNS inquiry language should be capable of both great precision and flexibility. (See page 4 of the attachment). Among the most effective search languages at present are the Merck system with its "X-Z" capability, and the Upjohn system. Dr. Lefkovitz noted that there should be informational compatibility with the CAS Registry III Connection Table. Lastly in this area, there should be user-created catalogue keys for performing current awareness searches, or creating profiles of interest.

Dr. Lefkovitz then listed three major development problems. The first is graphic input and inquiry variability. This problem is fairly well solved by approaches developed by Merck and Upjohn, as well as by some methods being used by SANSS, NCI, and others. The second problem revolves around operational issues associated with the interactive search of a multi-

million compound file. The third problem pertains to output; there should be high quality structural display for both low and high speed terminals.

Dr. Lefkovitz then spoke about the solutions to these problems. For the input, and more specifically the graphic input problem, the solution is partly a system of keyboard commands, as originally developed by Dr. Richard Feldman (DCRT, NIH), which has been incorporated into both the NCI and SANSS systems. This system has two advantages: (1) it can be used on non-graphic, ordinary character terminals; and (2) it can enter structures rapidly. However, the system lacks the ability to craft certain types of configurations. Three or four bonds from a single origin, or a bridge across the middle of a ring structure are examples of chemical graphics not well processed by the system. In this area, the Rohm and Haas Company has developed a cursor graphic system that handles such structures better.

From an inquiry variability standpoint, the X-Z substitutions of Merck are recommended. The report also cites the ring/nucleus keys of the NCI system, and the atom-centered fragments in both the NCI and SANSS systems, as being potentially useful here. SANSS has a powerful set of keys called RPROBE which could possibly be incorporated into the CSNS. Some upgrading of the RPROBE keys is currently being considered by the SANSS development contractor.

There is a two part solution to the large file search problem. First is the key search response time. In the File Structure Feasibility Study, three applicable methodologies were identified. Two of them follow

what is called inverted lists. One is the Bit-Map Inverted List approach used by the NCI. The other is the Pointer List approach, the most efficient example of which is used in the CHEMLINE Software (i.e. ELHILL). The study cites the Bit-Map approach as having far greater extendability for use in multi-million compound files. It is, in fact, a method developed for large files. The third approach is the serial-parallel approach. It involves serially searching a number of screens, but breaking the search up into a number of pieces. This is basically the approach used by CAS. The study recommends either a feasibility study in this area, or simply waiting to observe the degree of effectiveness of the proposed CAS system.

Another part of the "key search response time" solution is a file size limitation, which will depend on the results of a recommended future feasibility study, i.e., can CSNS be extended to five million compounds? Relevant here is the degree of success of the CAS system, how well functional requirements have been met, and possibly a study on the costs and benefits of searching an extremely large file.

Dr. Lefkovitz added that in order to improve the operational response time, some improved search keys are required. Options include:

(1) certain NCI keys, (more efficient form of the atom-centered fragments mentioned earlier); (2) the SANSS RPROBE; and (3) the BASIC (or Swiss) screens used by CAS. [The entire list of improved key option is shown on page 8 of the attached graphics.] An improved key assignment algorithm should also be considered. The partial application of a weighted logic scheme or some other heuristic approach could increase search efficiency. In summarizing this part of the operational solution, Dr. Lefkovitz noted

that he had mentioned methods of increasing efficiency of the system through computer access, through refinement of the keys, and through different key assignment strategies. The second operational problem concerns the iterative search. After the key search is completed, a large number of "hits" may require atom-by-atom, bond-by-bond searching of the Connection Table. The solution is to recode existing iterative search programs to achieve maximum efficiency. A very high speed CPU would have to be used in order to provide sufficiently rapid response times. Alternatively, the use of multiple CPUs can be considered.

Dr. Lefkovitz then addressed the output/printing problem. Chemical structures may be represented in two ways in the computer. One is by a vector method, and the second by a matrix. The difficulty is that while the structure intelligibility is much higher in the vector representations, the quality of current printer terminals is substantially better in the matrix mode. Laser printers with vector capability are starting to be produced, but are not yet well developed. Transferring of chemical structure information into computer form, is a "solved" software problem. The study recommends the use of a very sensitive program developed by CAS to assign coordinates to the Connection Table, an option that works well. Thus, the solution to the output problem is firstly, the use of the CAS vector display program with existing devices, like Tektronix or Hewlett-Packard terminals. Secondly, while waiting for the development of a quality laser printer in vector mode, the study recommends the development of a conversion program to print on a matrix device, or use the existing matrix display program in SANSS, which would need some upgrading for CSNS use.

Dr. Lefkovitz then reviewed the CSNS Design Concept (as shown on page 12 of the attached graphics). Here, the original functional requirements are broken into three broad groups, the spectral systems, name searching, and structural searching. The spectral systems already exists in a highly specialized, fairly well developed form in CIS, and can be made available by incorporation through CSIN. There are also existing name search systems, (e.g., CHEMLINE).

In terms of a CSNS development plan, the study suggests that a CSNS-1 can be developed to handle 1-1.5 million compounds in a two year period. The system would begin operation in the third year, during which a decision would be made regarding expansion to total CAS Registry search. The basis of such a decision would be presumably the success and acceptability of the CAS system. At that point, CSNS will go in one of two directions. Either it will expand, to a CSNS-2, to handle the 5 million compound file, or the CAS system will be used.

Dr. Lefkovitz added that he will present highlights of the U. of P. report to the ITSDC on November 13, 1979.

Following Dr. Lefkovitz's presentation, the Chairman began to assemble a group to conduct an in-depth review of the U. of P. study. The group includes: Dr. Chu, Ms. Graf-Webster, Dr. Kissman, Dr. Siegel, Dr. Spann and tentatively, Dr. Grier.

## 2. Chemical Data Bases Directory

Ms. Rita Bergman, (CEQ) reported on the Chemical Data Bases Directory (CDBD). The MITRE Corporation contract for the CDBD has terminated with the publication of their final report. Volume I of the report delineates

guidelines for developing data base directories. Volume II is an indexed directory of some 50 chemical information resources. The subcommittee agreed that updating and enlarging the scope of the directory would be desirable.

In conjunction with the CDBD Project, Mr. John Feulner (National Referral Center, Library of Congress) said that the NRC/LC could be funded from other agencies for specific projects, and would be interested in participating in the publication of a new edition of the 1969 toxicology information resources directory. He calculated that a directory of approximately 1000 entries would cost around \$75,000. If NRC were not the primary developer of such a directory, Mr. Feulner added that they could provide from their data base a printout of the pertinent information resources that ought to be considered. He estimated that there were 1,500 such organizations.

### 3. Network Architecture Contract

Dr. Siegel reported that because of the complex nature of the CCA proposal for the CSIN prototype, an ad hoc review group of experts is being assembled. Among those being requested to participate on this panel are past project officers of CCA contracts, as well as individuals from NBS, including Dr. Steve Kindelson, who is particularly experienced with distributed data base management systems.

Ms. Bergman reported that the CCA report had been distributed to members. She emphasized the need for written comments by October 31. She also reported the signing of a contract amendment with CCA for the sum of \$540,000 to include the following tasks: (1) the expansion of the original user requirement study to encompass 25 to 30 organizations (this report is



due January 1980); (2) CCA will implement the pre-prototype microprocessor system to meet the immediate needs of the CSIN community as well as to collect use data; (3) CCA will design Version I of the CSIN prototype. Funding to allow CCA to consult with technical experts on various CSIN problems has also been included in the contract amendment.

Ms. Bergman also reported on current CCA progress. Interviews needed to analyze CSIN user requirements are nearly completed at Occupational Safety and Health Administration (OSHA), Office to Toxic Substances (EPA), the Bureau of Foods (FDA), NCI, Stanford Research Institute (SRI) and the Federation of American Societies for Experimental Biology (FASEB). A report entitled "Sample Analysis of Requirements for CSIN" has been completed and is currently being reviewed at the Office of Testing and Evaluation (OTE). The analysis of systems alternatives, which is contingent on results of the requirements analysis study, will be completed in February, 1980.

#### 4. Registration of Compounds Indexed in the Pre- 1965 Chemical Abstracts

Dr. Kissman reminded the Subcommittee that the registration of compounds indexed only in the pre-1965 Chemical Abstracts was proposed in the 1978 CEQ report to Congress.

Ms. Bergman reported that in response to a request from the CSIN Subcommittee to CEQ, CEQ requested that CAS submit a proposal for studying the feasibility of registering chemicals from the pre-1965 CA Indexes. A \$137,590 contract was signed with CAS to conduct that study,

and to select an experimental approach for implementation. Only this phase of the project has been funded. The projected Phase II would apply developed software to portions of the index to determine actual effectiveness. The Phase I final report is due by the end of the fiscal year. In commenting on this, Dr. Kissman cited some of the difficulties in this project. He noted that the compounds are indexed in Chemical Abstracts, subject and molecular indexes. The problem is finding the most efficient and inexpensive method of getting the structure information on the compounds so mentioned into the CAS Registry System. There have been estimates that there are some 2 million compounds involved.

#### V. Other Business.

##### 1. Republication of "A Directory of Information Resources in the U.S.; General Toxicology"

Dr. Kissman said that the publication of this Directory was first sponsored by NLM in 1969. A proposal is currently being considered by NLM for cooperation with NRC/LC to republish the Directory. Such a task would be done in coordination with the CSIN effort. The Subcommittee agreed that such an updating of the Directory would be desirable. Mr. Feulner reiterated that such a republication effort would be feasible (c.f. Mr. Feulner's remarks under "CDBD").

##### 2. Toxicology Information Subcommittee (TIS)

Dr. Kissman reported on two items that might be of particular interest to those CSIN Subcommittee members who could stay for the afternoon meeting of the TIS. One is a discussion of an approach to computerization of what

he called "Government Monographs," including the NIOSH Criteria Documents, and other government reports. The second is the consideration of a proposal of a Chemical Hazard Alert System as generally described by the Toxic Substances Strategy Committee, which will monitor the literature and look for new biological or chemical events that might be of concern to different agencies in either research or regulation areas.

The next meeting of the Subcommittee was scheduled for November 28, 1979, from 9:00 a.m. to 12:00 noon, at the National Library of Medicine. The TIS Subcommittee will meet on the afternoon of that same day. As there was no further business, the meeting was adjourned at 11:45 a.m.



Nan Fremont

Executive Secretary



Henry M. Kissman, Ph.D

Chairman

# ATTENDEES

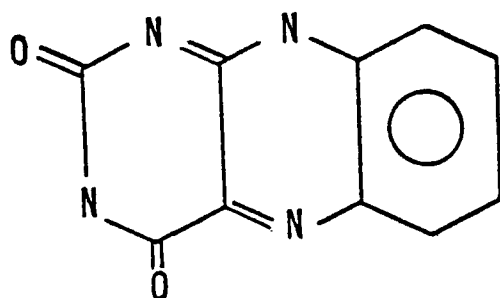
10/11/79

Name	Organization	Phone
Rita Bergman	CEQ	395-5763
Joan Chase	NTP	496-1152
K. Chu	NCI/NIH	496-1152
Paul Craig	NLM-SIS	496-1131
Winston R. deMonsabert	FDA/OC	443-4505
Dorothy Drago	CPSC	492-6990
Jerry Edge	NLM	496-1131
John Feulner	LC/NRC	287-5671
Helga Gerstner	ORNL/ICC	674-7803
Erika Graf-Webster	GA/CEQ	821-2166
G. Green	FDA/NCTR	542-4584
Bernard Greifer	DOC	377-3078
Mike Hazard	NLM	496-1131
H. Kissman	NLM/NIH	496-3147
Claudia Lewis	CDC	FTS-236-3266
Bill Rhode	OD/NIH	496-9285
Leonard Schachter	CPSC	492-6990
Bernard Scharf	CPSC	492-6470
Bob Schultheisz	NLM	496-1131
B. Van Scoyoc	MID/FSQS/USDA	436-8154
S. Siegel	EPA/OTS/OPII	755-8040
Mel Spann	NLM	496-1131
Don Walker	NLM-SIS	496-1131
John Wilson	DOE/EV	353-4684

# MORE SPECIFIC SYSTEM REQUIREMENTS

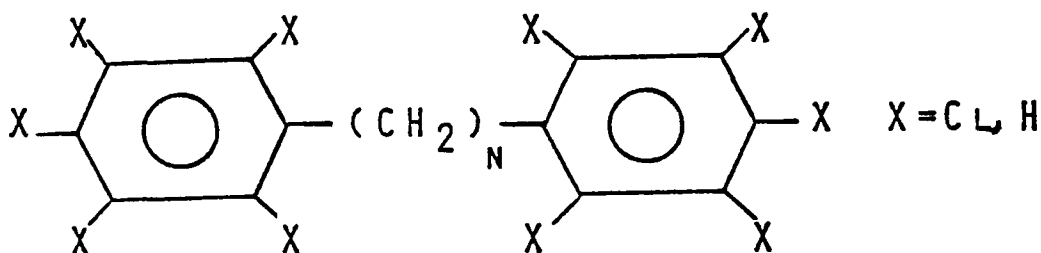
- INQUIRY LANGUAGE CAPABLE OF BOTH PRECISION AND VARIABILITY

## PRECISION



NO TAUTOMERS ALLOWED

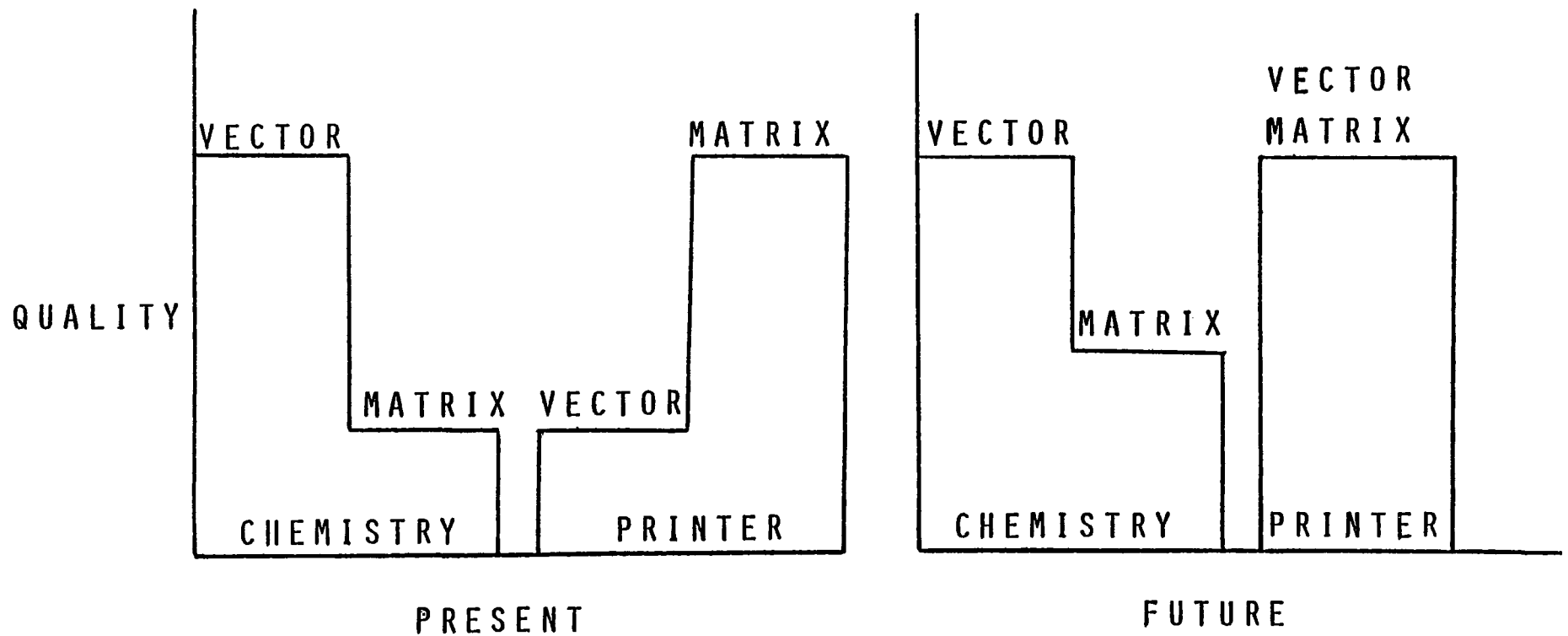
## VARIABILITY



- COMBINED STRUCTURAL AND NON STRUCTURAL INQUIRY PARAMETERS
- INFORMATIONAL COMPATABILITY WITH CAS REGISTRY III CONNECTION TABLE

# THE OUTPUT PROBLEM

## PERFORMANCE OF DISPLAY ALGORITHMS AND HIGH SPEED PRINTERS



SOLUTION TO THE OUTPUT PROBLEM
--------------------------------

1. USE CAS (VECTOR) DISPLAY WITH  
EXISTING DEVICES:

- TEKTRONIX OR HP FOR INTERACTIVE
- ELECTROSTATIC (VERSATEC) FOR  
HIGH SPEED

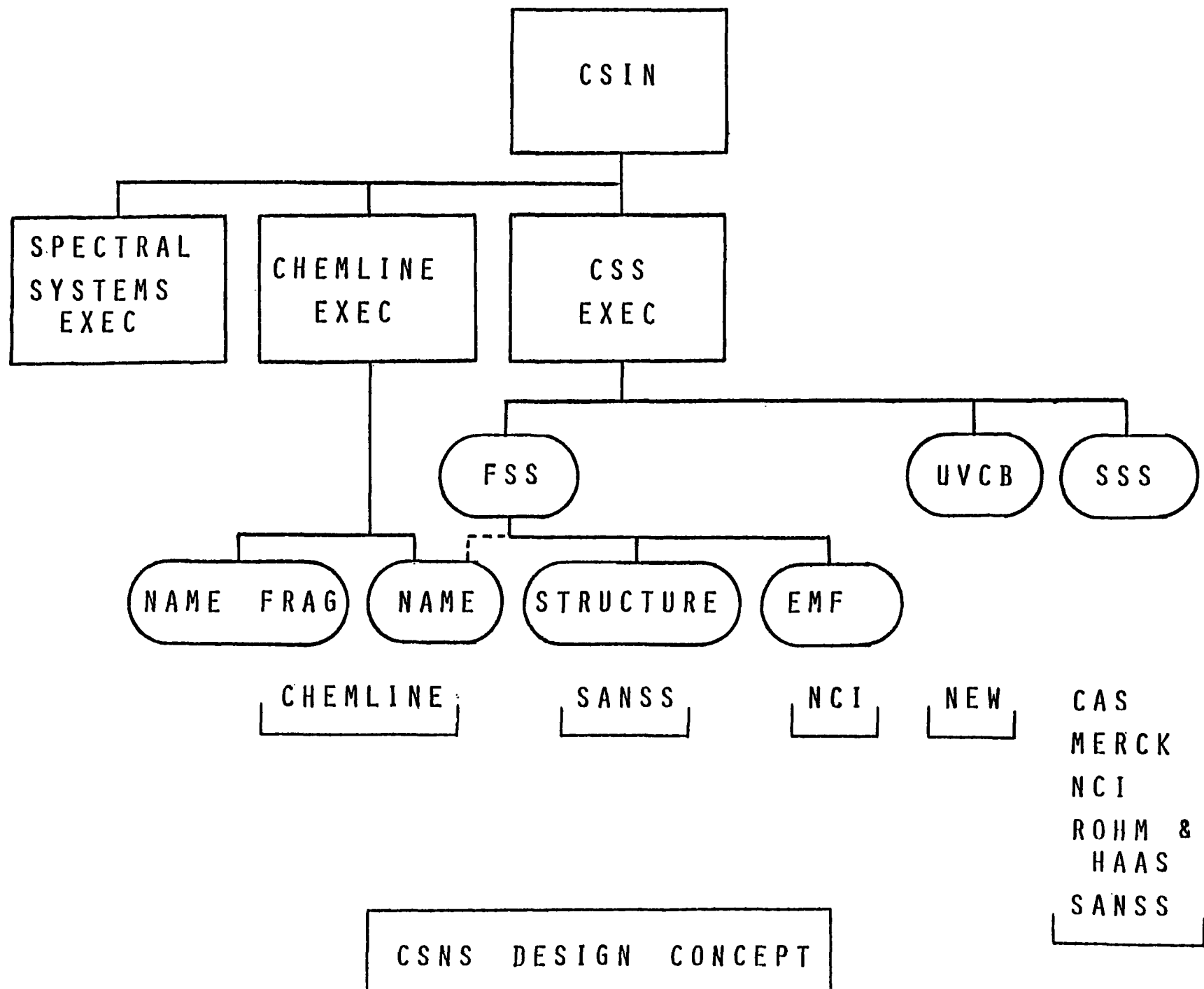
2. EITHER

- DEVELOP A CONVERSION PROGRAM TO  
PRINT ON A MATRIX DEVICE,

OR

- USE THE EXISTING MATRIX DISPLAY  
PROGRAM IN SANSS

3. ANTICIPATE THE AVAILABILITY OF LASER  
PRINTER IN VECTOR MODE.





## CSNS DEVELOPMENT PLAN

- ▶ DEVELOP CSNS-1
  - 1.5M CMPD FILE
- ▶ USE EXISTING SYSTEMS

- ▶ CSNS-1 OPERATION
- ▶ DECIDE ON EXPANSION TO 5M+ CMPD FILE

EXPAND  
CSNSUSE  
CAS

- ▶ DEVELOP CSNS-2
  - 5M+ CMPD FILE

- ▶ CSNS-2
  - INTERACTIVE SEARCH OF 1.5M
  - CAS SEARCH OF 5M+
  - COMMON FRONT END IF POSSIBLE

YEAR

0



2



3



4

*Cathy*

# MEMORANDUM

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE  
PUBLIC HEALTH SERVICE  
NATIONAL INSTITUTES OF HEALTH  
NATIONAL LIBRARY OF MEDICINE

TO : Members, CSIN Subcommittee of the Interagency Toxic Substances Data Committee  
DATE: January 7, 1980

FROM : Chairman, CSIN Subcommittee

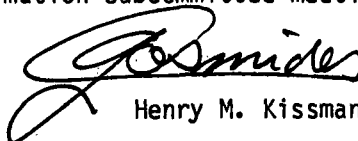
SUBJECT: (1) Minutes of the November 28, 1979, Meeting  
(2) Agenda Items for the January 17, 1980, Meeting

Enclosed are the Minutes of the CSIN Subcommittee Meeting of November 28, 1979.

The next meeting of the Subcommittee has been scheduled for January 17, 1980, from 1:00-4:00 P.M. in the Billings Auditorium, (entrance - "A" level near loading dock), National Library of Medicine. The following items are proposed for the agenda:

Report from the Network Administrator  
Use of CAS Registry Numbers in Agency Files  
Chemical Structure and Nomenclature System  
Network Architecture  
Chemical Data Bases Directory  
Symposium Plans and Organization of Task Group  
Report from Parent Committee

There will also be a meeting of the Toxicology Information Subcommittee, DHEW Committee to Coordinate Environmental and Related Programs on Thursday, January 17, 1980, from 9:00 A.M.-12:00 noon in Conference Room B (Mezzanine level) of the National Library of Medicine. The agenda includes reports from other interagency committees and from subcommittee projects. Members of the CSIN Subcommittee who wish to attend the Toxicology Information Subcommittee meetings are welcomed.

  
Henry M. Kissman, Ph.D. *for*

Enclosure: CSIN Minutes - 11/28/79

cc: Dr. Marilyn C. Bracken, EPA  
Ms. Erica Graf-Webster, EPA  
Ms. Carroll L. Bastian, CEQ  
Ms. Rita Bergman, CEQ  
Dr. George J. Cosmides, NLM  
Mr. Kent A. Smith, NLM

Minutes of the Chemical Substances Information Network (CSIN) Subcommittee  
of the Interagency Toxic Substances Data Committee (ITSDC)

November 28, 1979

I. Introduction

The meeting, held at the National Library of Medicine (NLM), was called to order at 9:00 a.m. The minutes of the October 11, 1979 meeting will be distributed by mail to Subcommittee members.

II. Chemical Substances Information Network (CSIN) Support Activities

1. Chemical Structure Nomenclature System (CSNS)

The Chairman summarized the status of CSNS activities. An overview of the University of Pennsylvania (U. of P.) report concerning requirements for a Chemical Structure and Nomenclature System (CSNS) was presented to the Subcommittee at its October 11, 1979 meeting, and also to the Interagency Toxic Substances Data Committee (ITSDC) at its November 13, 1979 meeting, by Dr. David Lefkovitz (U. of P.). Also at the November ITSDC meeting, Mr. Nick Farmer, Chemical Abstracts Service (CAS), reported on the CAS Target System (i.e. the online substructure search system for the 5 million (plus) compound CAS Registry System ) now being developed. On November 14 and 15, 1979, a subgroup of the CSIN Subcommittee visited CAS in Columbus for an intensive review of plans for the CAS Target System. A site report has been prepared by Ms. Graf-Webster. Other meetings were held with representatives of CAS, on November 19 and 20, 1979. Lastly there was a meeting concerning the CSNS activity on November 27, 1979 with representatives from the CSIN subgroup, CAS, and the Computer Corporation of America (CCA), and the network administrative contractor.

Dr. Siegel, (EPA) reported that the CSIN subgroup, involved in these CAS discussions concerning CSNS, is working to develop a final report within the next 60 days. The subgroup will then make recommendations to the Subcommittee. Counsels from EPA, the Council on Environmental Quality (CEQ) and possibly DHEW also will be asked to the Subcommittee on the legal/business issues involved in a possible collaboration between the US government and CAS on CSNS.

Dr. Kissman then focused on some of the technical aspects of the discussions with CAS. First he reiterated some basic requirements for a CSNS as stated in the University of Pennsylvania study; viz.: full and substructure search capability; minimum file size of one million records expandable to five million plus records; and nomenclature search capability. The system should be interactive, and be

able to handle 50 simultaneous substructure searches and 200 simultaneous full structure searches. It should also have high quality output display and some method for subsetting the files for more efficient searching.

The main issue here is whether the CAS Target System can fill these requirements, and have the technical capability to function as the CSNS for the CSIN project. If the answers to these questions are positive--and it now appears that they are--the Target System probably should be accepted as "the CSNS". The group then must consider the development of an interim system--CSNS-1--to be used while the CAS Target System is in development. In order to be worthwhile, CSNS-1 should be implemented before 1981.

CAS has suggested that the CAS Inquiry System, developed by the National Cancer Institute (NCI), Division of Cancer Treatment, be adapted and used as CSNS-1. The Inquiry System currently provides substructure search capabilities for "private" files (e.g. the NCI/DCT files). Content of the system would be essentially that of CHEMLINE, 450,000 compounds. Initially, the NCI-developed search screens would be used, to be replaced with CAS screens at a later date. The system would have structure input via typewriter, iterative search capability, full-substance identification structure output capability (using a vector mode or by activating the "Feldman Capability") and batch search and output capability. The system would be available to 5-10 simultaneous terminals and be able to accommodate up to 750,000 records. It would be available within 6 to 8 months after the start of the project. There are some weaknesses to an upgraded Inquiry System. The locator capability is not built in and would have to be put in a cross-reference file. The system has no Name or Name Fragment search capability; thus, it would have to be limited to CHEMLINE in the NLM computer through the CSIN query processor. Accomplishing this would require rapid development of the linkage software by CCA. The Subcommittee generally agreed that with the necessary upgrades the Inquiry system linked to CHEMLINE would be a good interim CSNS-1 solution. It was noted that certain other existing chemical information systems in various agencies would continue to be supported and maintained. In this way, capabilities such as the NLM/EPA CIS substructure/spectra search will continue to be available until CSNS is fully functional.

## 2. Chemical Information Resources Directory (CIRD)

The Chairman reported that final copies of the MITRE report on the Chemical Information Resources Directory (CIRD) have been received. Volume I of the report delineates guidelines for developing an Information Resource Directory; Volume II is an indexed directory of some 54 chemical information resources.

In conjunction with the CIRD Dr. Kissman also reported on NLM's discussions with the Library of Congress (LC) National Referral Center (NRC) concerning the reissue of "A Directory of Information Resources in the U.S.: General Toxicology". The volume was first issued

in 1969. An interagency agreement for this activity had been drafted and will be submitted to the NRC for comments. The new directory with its 1000-1200 indexed information resources could function as "raw material" for the continuation of work on the CIRD.

Ms. Rita Bergman (CEQ) said that following finalization of the NLM-LC Interagency Agreement, plans are to draw on the resources identified in the NRC/LC set to form a subset of resources to be used in the CIRD. An RFP to allow continuation of the CIRD project is expected to be finalized at CEQ early in 1980. Tasks will include collection of very specific information about relevant resources in files.

### 3. Network Architecture Contract

Dr. Sidney Siegel (EPA), Network Administrator, reported that through the Preprototype CSIN task, intelligent terminals will be delivered to the government by February, 1980. At first, delivery will be to EPA and NLM (possibly); one terminal will be kept at the CCA offices in Boston for further developmental work.

The CSIN Prototype plan prepared by CCA was reviewed in a technical meeting held on November 27, 1979. CCA, EPA, CEQ and the CSIN Subcommittee were represented. Invited outside experts attending were: Dr. Ari Shoshoni, Senior Computer Scientist, Lawrence Livermore Laboratories, University of California; Dr. Stanley Su, Professor of Computer Information Science, University of Florida; Dr. Steve Kimbelton, Manager of Network Development, National Bureau of Standards (NBS) and his colleague, Dr. Pearl Wong (NBS). In presenting their detailed concept of the CSIN, CCA discussed user requirements analysis, hardware, software, and requirements and issues concerning communication and security. The panel of advisors will each consider all the aspects of the Prototype plan, evaluate the information and give their reports to the CSIN Subcommittee through its Prototype subgroup. The experts will then meet with the Subcommittee and CCA to address any further technical questions. CCA will later present the Prototype plan (revised if necessary) to the Parent Committee in an open seminar.

Ms. Rita Bergman also reported on CCA progress for October and November, 1979. Eighty-five individuals in seven organizations have been interviewed under the Requirements Analysis task. Interviews in the next ten days have been scheduled with another four agencies. The Preprototype design is complete and software development is nearly finished. Most functions are now being tested. The user manual is in draft form. Prototype design continues.

### III. Report from the Network Administrator

Dr. Siegel reported that the Information Exchange Group (IEG) of the Interagency Regulatory Liaison Group (IRLG) is currently considering its purposes and functions, i.e., should the IEG simply exchange information among agencies, or should the group, having identified some of the common

information problems and issues, formulate projects to address these common needs. Further, how would the IEG channel such identified projects to a larger information activity like CSIN. A draft document describing such activity and relationship between the IEG and the CSIN Subcommittee is being prepared.

#### IV. Registration of Compounds in the pre-1965 Chemical Abstracts

Ms. Rita Bergman (CEQ) reported on the status of the CAS feasibility study contract on registration by CAS of compounds only indexed in the pre-1965 Chemical Abstracts (CA). Ways of dividing the pre-1965 volumes into appropriate segments are being examined (e.g. five-year segment divisions or division based on changes in CA indexing policies). Sample pages of the CA indexes are being keyboarded and Optical Character Recognition equipment is being used to obtain data on the computer input costs.

#### V. Use of the CAS Registry Numbers in Agency Files

Dr. Kissman distributed the revised draft of the Memorandum of Understanding prepared by Dr. Bernard Greifer, DOC. This paper proposes the establishment of the CAS Registry Numbers as a Federal Information Processing Standard (FIPS) by the NBS; this would allow (inter alia) reimbursement to the agencies for the cost of linking CAS Registry Numbers to the compound information in their files. Work continues to establish the CAS Registry Number as a FIPS, agencies could continue using it on the current de facto basis. Final recommendations for the Memorandum of Understanding will be presented to the ITSDC at its February 5, 1979 meeting.

#### VI. Peer Review of CSIN Activities

Dr. Kissman reported on discussions with representatives of the National Academy of Sciences (NAS) Board on Toxicology and Environmental Health Hazards (BOTEHH) to establish an advisory group to CSIN as suggested on several occasions by members of the Parent Committee and the Subcommittee. The approach discussed was the formation of a standing advisory group which would prepare periodic reports on aspects of CSIN development. The group would need to have expertise in areas such as computer science and communications as well as toxicology and allied sciences. It would also need to serve as a surrogate for the scientific (i.e. university and industrial research) user community that will be served by CSIN.

The BOTEHH staff is preparing a proposal for this advisory activity. Experts from three areas of the NAS would be asked to take part; BOTEHH the Board on Telecommunications and Computer Sciences, and the Mathematical Sciences group.

Subcommittee members recommended that the NAS proposal when received should be carefully evaluated to insure that issues of interest are properly addressed. Members also noted the importance of other kinds of CSIN review via the

Public Liaison Subcommittee of the ITSDC and public input at ITSDC meetings. Ms. Bastian (CEQ) stressed the importance of representing the needs of Federal and state groups as well as the needs of the academic and scientific communities.

VII. Miscellaneous

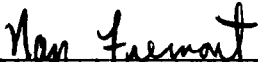
1. EURONET

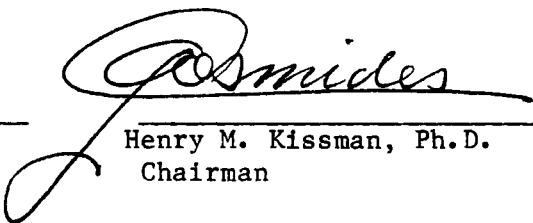
Information about technical and administrative issues of this large distributed data base and communications system, not unlike CSIN will be discussed by EURONET's chief, Dr. Anderlea in Luxemburg and Dr. Marilyn Bracken, EPA.

2. CSIN Presentations

Comments and advice on where and to whom presentations about CSIN should be made are always welcome. Ms. Carroll Bastian (CEQ) will develop a list identifying appropriate people to be contacted in the Toxic Substances area of the Office of Management and Budget (OMB) for such a presentation.

The next meeting of the Subcommittee was scheduled for January 17, 1980 from 1:00 p.m. to 4:00 p.m., at the National Library of Medicine. The Toxicology Information Subcommittee (TIS) will meet that morning at 9:00 a.m. As there was no further business, the meeting was adjourned at 12:00 noon.

  
\_\_\_\_\_  
Nan Fremont  
Recording Secretary

  
\_\_\_\_\_  
Henry M. Kissman, Ph.D.  
Chairman

Attendees

Claudia Lewis, CDC  
Carroll Leslie Bastian, CEQ  
Rita Bergman, CEQ  
Leonard Schachter, CPSC  
Bernie Scharf, CPSC  
John Wilson, DOE  
Sidney Siegel, EPA  
Winston de Monsabert, FDA  
Helga Gerstner, ICC/ORNL  
Barbara Jaffee, NAS  
L. H. Gevantman, NBS  
Kenneth Chu, NCI/NTP  
Terri Damstra, NIEHS  
George Cosmides, NLM  
Paul Craig, NLM/NIEHS  
Mike Hazard, NLM  
Henry Kissman, NLM  
William Rhode, OD/NIH  
Pat Breslin, OSHA  
Florence Bayard, Information Tech, Inc.  
Nan Fremont, Recording Secretary



# MEMORANDUM

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE  
PUBLIC HEALTH SERVICE  
NATIONAL INSTITUTES OF HEALTH  
NATIONAL LIBRARY OF MEDICINE

TO : Members, Toxicology Information Subcommittee of the DHEW Committee to Coordinate Environmental and Related Programs

DATE: August 8, 1979

FROM : Chairman, Toxicology Information Subcommittee

SUBJECT: (1) Minutes of the June 28, 1979, Meeting  
(2) Agenda Items for the August 23, 1979, Meeting

Enclosed are the minutes for the Toxicology Information Subcommittee meeting of June 28, 1979.

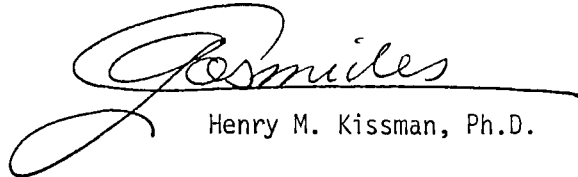
The next meeting of the Subcommittee has been scheduled for Thursday, August 23, 1979, from 9:00 a.m.-12:00 Noon, in Conference Room B, National Library of Medicine. The following items are proposed for the agenda:

- I. The National Referral Center, Library of Congress . . . . . John F. Price and Staffan Rosenberg
- II. Environmental Mutagen Information Center (EMIC) and the Environmental Teratology Information Center (ETIC) . . . . . John S. Wassom
- III. Report from Other Interagency Committees
- IV. Subcommittee Projects
  - A. Information Response to Chemical Crises Project
  - B. Laboratory Animal Data Bank
  - C. Chemical Monograph Referral Center
  - D. Toxicology Research Projects Directory and TOX-TIPS
  - E. Toxicology Data and Document Depository
- V. Other Business

The CSIN Subcommittee will meet in the afternoon of August 23, 1979, between 1:00 and 4:00 p.m. You will find enclosed a set of the slides Dr. Sidney Siegel (EPA) used with his presentation on the Chemical Substances Information Network at the May 16, 1979, meeting of the Toxicology Information Subcommittee.

(more)

There will be a meeting of the LADB Task Group following the Toxicology Information Subcommittee meeting on August 23, 1979, in the Billings Auditorium, National Library of Medicine, from 1:00-3:00 p.m. Dr. Warren G. Hoag has notified Task Group members and will provide the agenda.

  
Henry M. Kissman, Ph.D. FOR

2 Enclosures

cc: Dr. David P. Rall, NIH-NIEHS  
Dr. Raymond E. Shapiro, NIH-NIEHS  
Mr. Kent A. Smith, NIH-NLM  
Mr. John S. Wassom, ORNL  
Mr. John F. Price, LC-NRC (w/o encls.)  
Mr. Staffan Rosenberg, LC-NRC " " "  
Dr. Sidney Siegel, EPA

Minutes of the Toxicology Information Subcommittee of the DHEW  
Committee to Coordinate Environmental and Related Programs

June 28, 1979

I. Introduction

The meeting was held at the National Library of Medicine (NLM), at 9:02 A.M. Dr. Kissman introduced Dr. Mark Fow (FDA), Dr. R. E. Maizell (Olin-Mattheson Corporation), and Dr. Harold Wooster (NLM).

Minutes of the meeting of May 17, 1979 stood.

II. Information Activities of the National Clearinghouse for Poison Control Centers

Dr. Mark Fow reported that the National Clearinghouse for Poison Control Centers which was started in 1957, has been a source of information for approximately 600 Poison Control Centers throughout the country. It provides information concerning the treatment of acute poison emergencies and the probable toxicity of commercial chemical products.

Product composition is requested from manufacturers--either specific ingredients or generic classes of ingredients. From this information, the toxicity of the ingredients is estimated; these toxicity estimates are then disseminated to the poison control centers. Information on about 10,000 products is currently available in a computerized system at the CPSC. A hard copy file, which goes back 20 years, contains information on 25,000 to 50,000 compounds.

Lack of specific information from manufacturers and batch-to-batch variations of the chemical composition of the same product contribute to problems in assigning CAS Registry Numbers.

Another phase of the Clearinghouse information program is the Data Reporting System. Herein, case histories are collected and tabulated. This compilation, going back to 1971, includes specific signs and symptoms of product ingestion. The case history files provide information on the epidemiology of poisons, as well as treatment methodology.

III. Second Symposium on the Handling of Toxicological Information

In proposing a second symposium, Dr. Kissman presented a schematic of a toxicology retrieval system using eight levels of information

handling. They are (1) planning, (2) gathering of data, (3) transfer of data to a file, (4) a primary information channel, (5) secondary information systems, (6) tertiary information channels, (7) a level of aggregation of storage and retrieval, and (8) the user. He asked that the Subcommittee consider structuring the symposium according to these eight levels.

There was discussion regarding potential objectives of the conference, e.g., presenting state-of-the-art information activities in toxicology versus user education. The importance of identifying the user (scientist or information person) was also discussed.

Dr. Siegel (EPA) emphasized the desirability of outlining government agencies' information activities and how they interface with each other to increase the efficiency of information gathering.

Dr. Kissman received the Committee's agreement to arrange a joint TIS-CSIN Committee or task group to proceed with the symposium planning.

#### IV. Report from Other Interagency Committees

Dr. Patricia Breslin (OSHA) reported on recent activities of the Information Section of the Interagency Regulatory Liaison Group (IRLG). Noting the wealth of existing files on toxic substance law, Dr. Breslin said that the IRLG is presently working to formulate an efficient procedure for the coordination and utilization of various files, rather than designing and implementing a new data base on legislative toxicology information.

#### V. Subcommittee Projects

##### 1. The Laboratory Animal Data Bank (LADB)

Dr. Warren Hoag (NCI/NLM) reported on the recent LADB activities including the modification of the Federation of American Societies of Experimental Biology (FASEB) contract to include monthly data evaluation and editorial review by a new Data Acceptance Group. Dr. Hoag discussed the importance of data validation and description to obtain a common language acceptable to the scientific community.

##### 2. The Chemical Monograph Referral Center (CHEMRiC)

Mr. Mark Sutherland (CPSC) reported that approximately half the CHEMRiC records have been transferred to the CPSC Computer system, STAIRS. He emphasized that all of the records (including those not yet transferred to STAIRS) are available for searching.

The Subcommittee discussed methods to insure the reporting of new technical documents to CHEMRiC by the government agencies.

3. Toxicology Document and Data Depository (TD3)

Ms. Florence Bayard (NLM) described a profile that will be matched against the monthly NTIS data input tape to generate toxicology-related documents; these documents will be available for searching in TOXLINE. Specifications are being written for a program that will allow NLM's ELHILL system to accommodate NTIS' TD3 data in TOXLINE. She reported that the results of the most recent draft search profile look promising. The TD3 Task Force will be developing acquisition policy guidelines for the Depository.

There was a discussion concerning possible publication of a report on the refined profile and data mapping procedure.

4. Toxicology Research Projects Directory (TRPD) and TOX-TIPS

Ms. Carol Haberman (NLM) reported that NTIS has distributed publicity packets on TOX-TIPS and the Toxicology Research Projects Directory (TRPD). The index for the June issue of TOX-TIPS has a new format. Subscriptions to the TRPD will cost \$65, as of January, 1980.

The need for input to TOX-TIPS from government agencies, as well as from industry was emphasized.

In explaining the scope of TOX-TIPS, Ms. Haberman described problems in obtaining reports of the initiation of testing and epidemiologic studies to determine toxicity from industrial, governmental and academic organizations. Because the emphasis of the program is to provide a mechanism for avoiding inadvertent duplication of research effort, publication preference is given to long-term studies, although other projects are also reported. Here, Dr. Kissman reported efforts to link this project with the IARC survey.

5. Information Response to Chemical Crises (IRCC)

Dr. Terri Damstra (NIEHS) reported that the most recent mock crisis search topic was the effect of environmental chemicals on the immune system; a follow-up search on this topic will also be conducted. Two non-crisis searches in progress are (1) asbestos in the air and (2) the effects of acid rainfall. The next mock crisis topic will be sensitivity to urea-formaldehyde foams.

Dr. Kissman presented Dr. Gary Keilson of the National Academy of Sciences (NAS) who is also involved with toxicological crises response.

Dr. Keilson reported that the NAS Committee on Toxicology was formed in response to requests for toxicological advice from the military and the Atomic Energy Commission (AEC). Other sponsors include the NASA, the CPSC, the OSHA's, Office of Drinking Water, the EPA's Office of Research and Development and the DOE.

The Committee generally advises sponsors of safe exposure limits of a given chemical within 4-6 months. However, the need to respond to crisis situations arose and an Emergency Response Group was formed. The Group uses the Committee's library which contains an index card catalog organized basically by chemical, but which is also indexed by general biological categories, e.g., effects on the peripheral nervous system, as well as by LD50 and LC50 values and NIOSH and OSHA exposure limits. The library has journals and reprints on file as well.

The Emergency Response Group usually responds to requests verbally within 24-48 hours. Generally, limits on short term exposure are provided, but depending on the compounds involved, information concerning chronic effects may also be provided. A written report usually follows in 30 days. Benefits of some coordination of efforts between NAS and the IRCC were discussed.

#### VI. Other Business

Dr. Kissman distributed a document prepared by Dr. Philip Schambra (NIEHS).

The document, "Review of Current DHEW Research Related to Toxicology" contains two tables of particular interest. The first is a list of chemicals under investigation by DHEW. Dr. Kissman noted that an expansion of the table in order to reflect other agency investigations, would be useful. The second table organizes the chemicals by the types of test. Likewise, expansion of this table to include agencies would increase insight into current government chemical investigation activities.

Referring to a letter from Ms. Claudia Lewis (CDC), Dr. Kissman discussed the need to update the Toxicology Information Resource Directory. Ms. Lewis suggested using looseleaf additions to update this directory. Dr. Siegel brought up the possibility of including a resource directory in CSIN. It would be a logical addition to the Chemical Information Resource Directory, which will perform a locator function for CSIN.

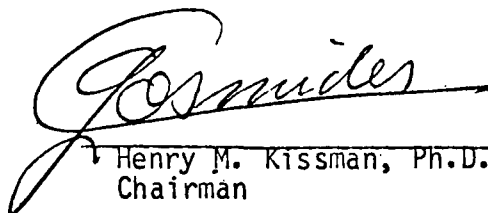
Dr. Siegel invited TIS members to the July 10th Interagency Toxic Substances Data Committee meeting, at 9:00 a.m., 722 Jackson Place, the CEQ Library.

The next TIS meeting was scheduled for August 23, 1979, at 9:00 A.M. at the National Library of Medicine.

As there was no further business, Dr. Kissman adjourned the meeting at 11:52 a.m.

---

Nan Fremont  
Recording Secretary



Henry M. Kissman, Ph.D.  
Chairman

Members Present

Ms. Dorothy Drago, CPSC  
Mr. Mark Sutherland (CPSC)  
Mr. Lester Miller, Dept. of the Army  
Mr. John Wilson, ERDA  
Ms. Anne McCann, FDA  
Ms. Helga Gerstner, ICC/ORNL  
Dr. Ken Chu, NCI  
Dr. Charles Litterst, NCI  
Dr. Terri Damstra, NIEHS  
Mr. Leonard Bahlman, NIOSH  
Dr. George Cosmides, NLM  
Dr. Henry Kissman, NLM  
Ms. Eloise Collins, NTIS  
Dr. Patricia Breslin, OSHA

Invited Guests and Observers

Mr. Bernie Scharf, CPSC  
Mr. Roger Connor, EPA  
Dr. Sidney Siegel, EPA  
Ms. Lois Ann Beaver, FDA  
Dr. Mark Fow, FDA  
Ms. Ann Louis Schick, FDA  
Dr. Gary Keilson, NAS  
Ms. Florence Bayard, NLM  
Ms. Carol Haberman, NLM  
Dr. Warren Hoag, NCI/NLM  
Mr. Harold Wooster, NLM  
Dr. R. E. Maizell, OLIN  
Ms. Nan Fremont, Recording Secretary

# **A Recommendation for CSIN: The Chemical Substances Information Network**



# **CSIN Concept**

## **Definition**

- **A network of coordinated online information systems concerning chemical substances**

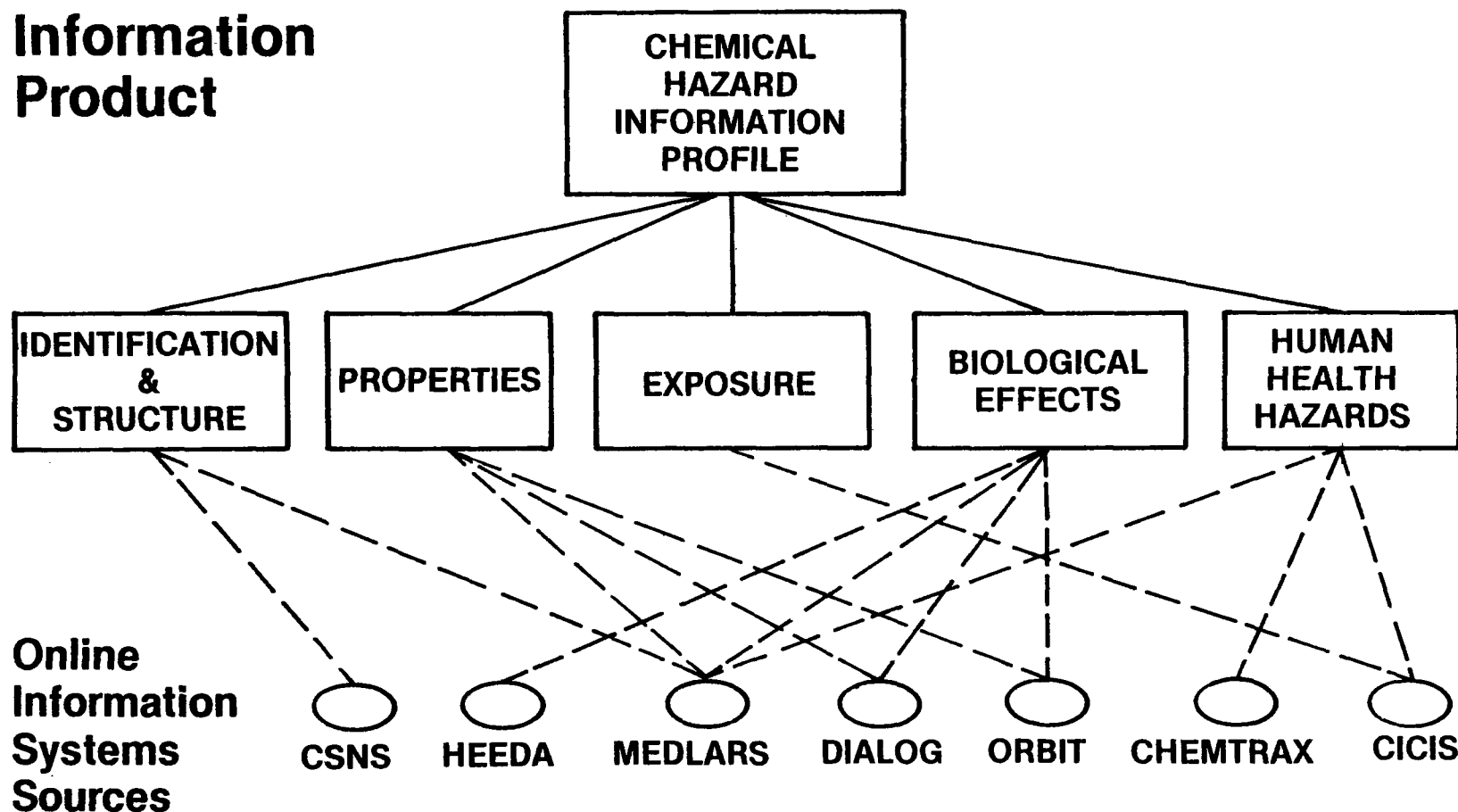
## **Purpose**

- **Satisfy information requirements of toxic substances legislation and a broad spectrum of related activities**

## **Information Content**

- **Provides access to information on**
    - **Nomenclature and Composition**
    - **Properties**
    - **Production and Commerce**
    - **Products and Uses**
    - **Exposure**
    - **Effects**
    - **Studies and Research**
    - **Regulations and Controls**
- of Chemical Substances**

## Information Product

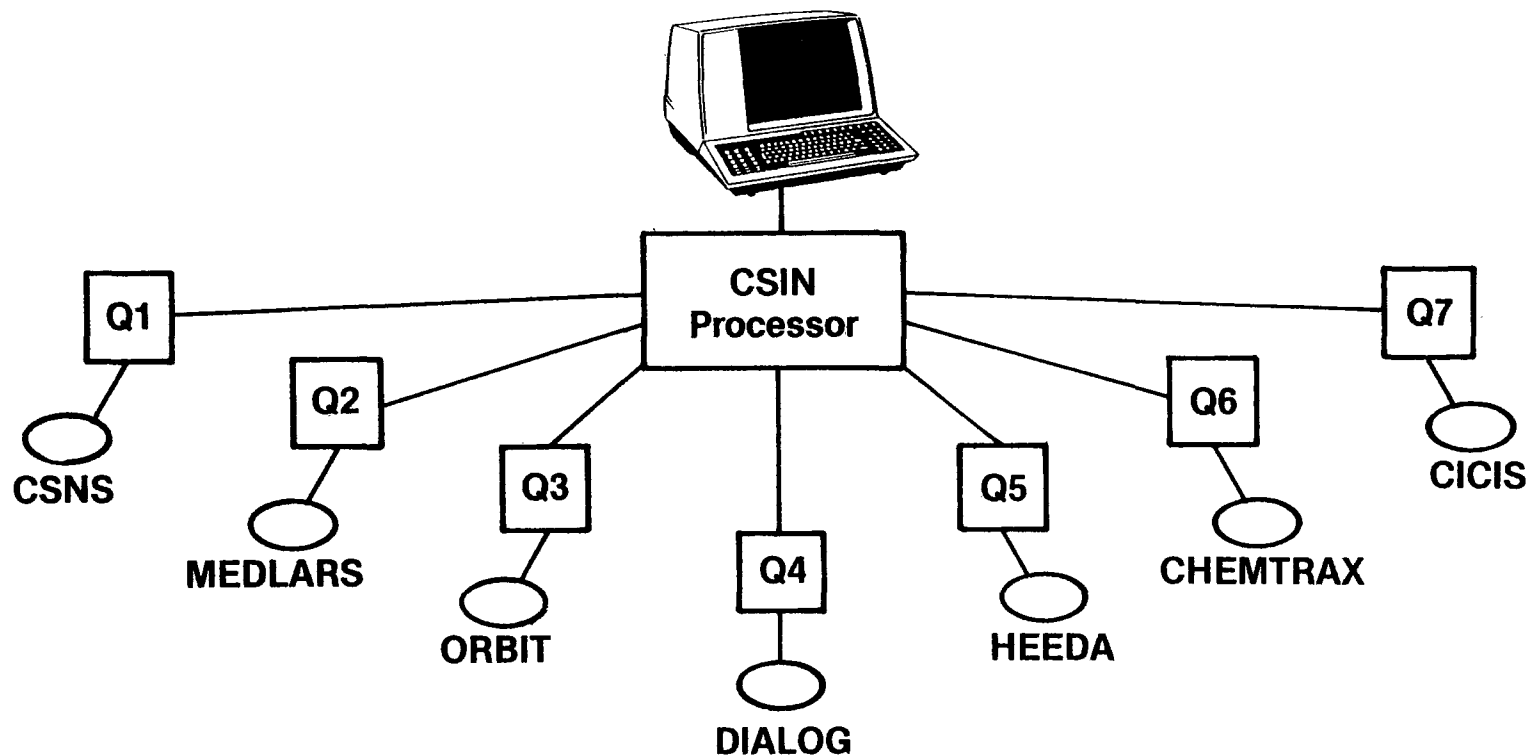


**Typical Information Product Involves Multi-System Access**

## **Multi-System Searches**

- **Common**
- **Required for Essential Information Products**
- **Substantial Professional Effort**
- **Occur Under Time Pressure**
- **Repetition via Related Chemicals**
- **Repetition via Related Systems**
- **Often Sub-Optimal in Selectivity, Exhaustiveness**
- **Will Increase in Frequency**
- **Will Increase in Scale**

# Chemical Hazard Identification



**With CSIN  
User Enters a Single Query**

## **Capabilities Desired in CSIN**

- **Uniform Query Interface**
- **Multi-System Query Procedures**
- **Intersystem Transfer of Query Results**
- **Uniform Data Interface**
- **Integrated Data Description**
- **Information Resource Directory**

# Prototype Program Objectives

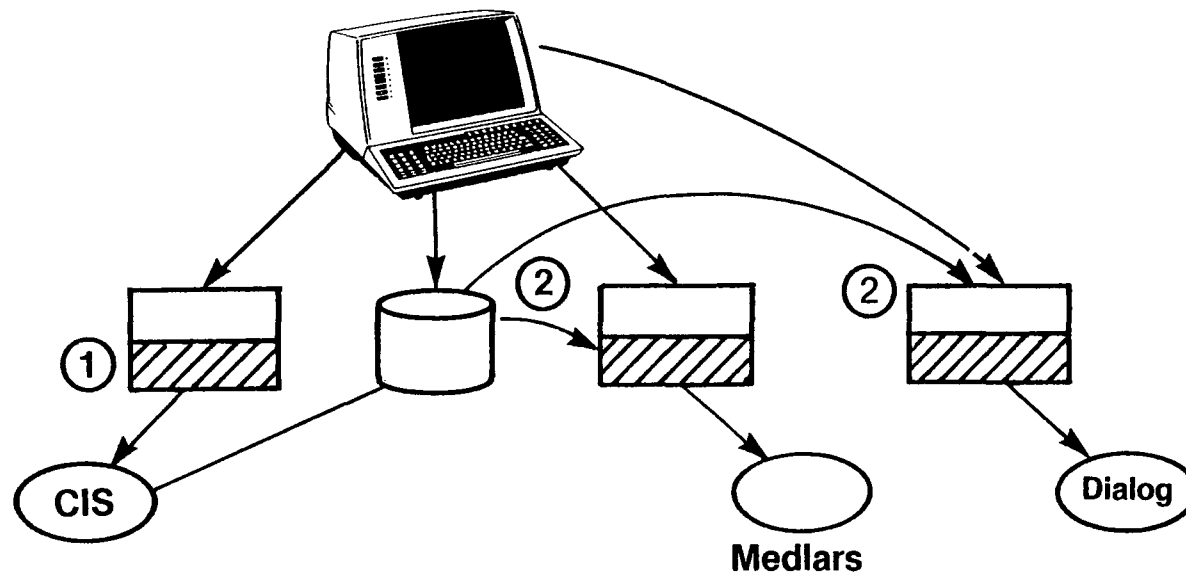
- To satisfy selected near-term requirements for multi-system access to data
- To clarify long-term functional CSIN requirements through user experience with prototype CSIN facilities
- To assure early identification of
  - inter-system interface problems, and
  - data integration problemsthrough early exercise of selected inter-system relationships
- To develop superior user interface design through user experience with prototype interfaces
- To obtain basic cost and performance data for long range design and implementation planning

# **Prototype Program Concept**

- **Use Existing Public Communications Network**
- **Include Only Selected Component Systems**
- **First Augment Direct Component Use**
- **Automate Highest Priority Multi-System Searches**
- **Provide Search Automation Facilities**
- **Develop Integrated Data Description for Selected Data**
- **Provide Uniform Query Language for Selected Data**
- **Prototype Information Resources Directory**

# Intelligent Terminal

**Question: List References to Toxic Effects on Fish and Wildlife for Chemicals Containing a Given Structure.**



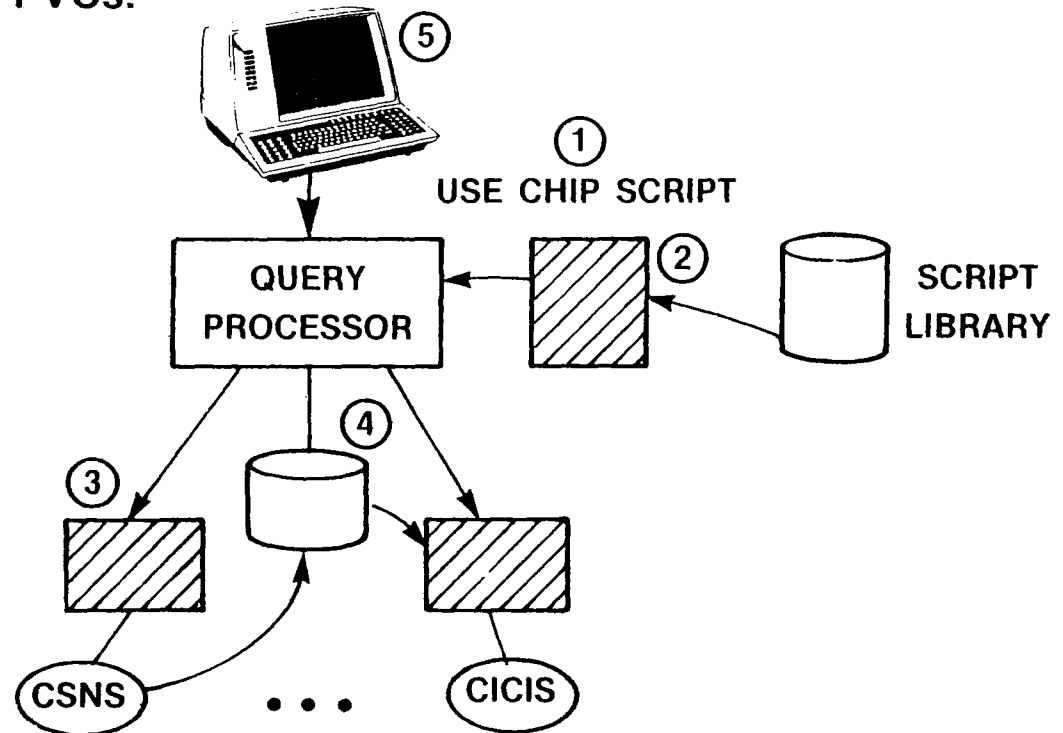
- ① Query Generation is Partially Automated
- ② Data Retrieved From One System can be Transferred Automatically to Others



# Prototype Version I

## Query Scripts

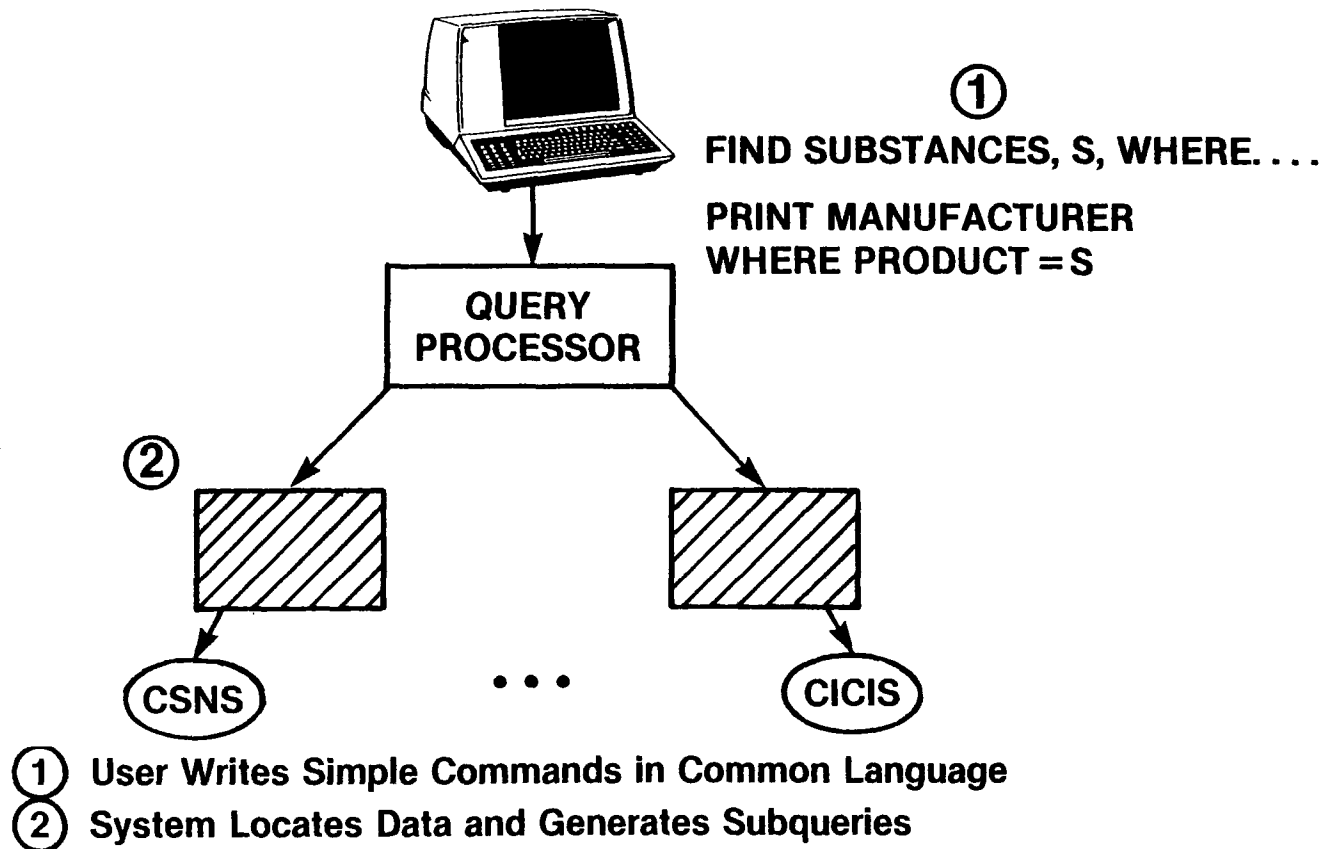
**Question: List Trade Names and Manufacturers and Production Values for Manufacturers of PVCs.**



- (1) User Invokes Script by Name and Responds to Prompts
- (2) Entire Script of Multi-System Search is Pre-Stored
- (3) Typically All Subqueries are Automatically Generated
- (4) Data From one System can be Transferred to Another
- (5) While Process is Automatic, User Must Typically Monitor, Know Components, be Prepared to Handle Exceptions

# Prototype Version 2

## Uniform Query Language



[Not Pictured] Scripts are now More Powerful and Automatic

# Uniform Query Language

- User Oriented Interface
- Clear Simple Presentation of Data to User
- Logically Integrated Description of Data From Multiple Systems
  - Uniform Data Representation
  - Consistent Data Naming
  - Uniform Data Structuring
- Content-Based Data Selection
- Limited Data Domain

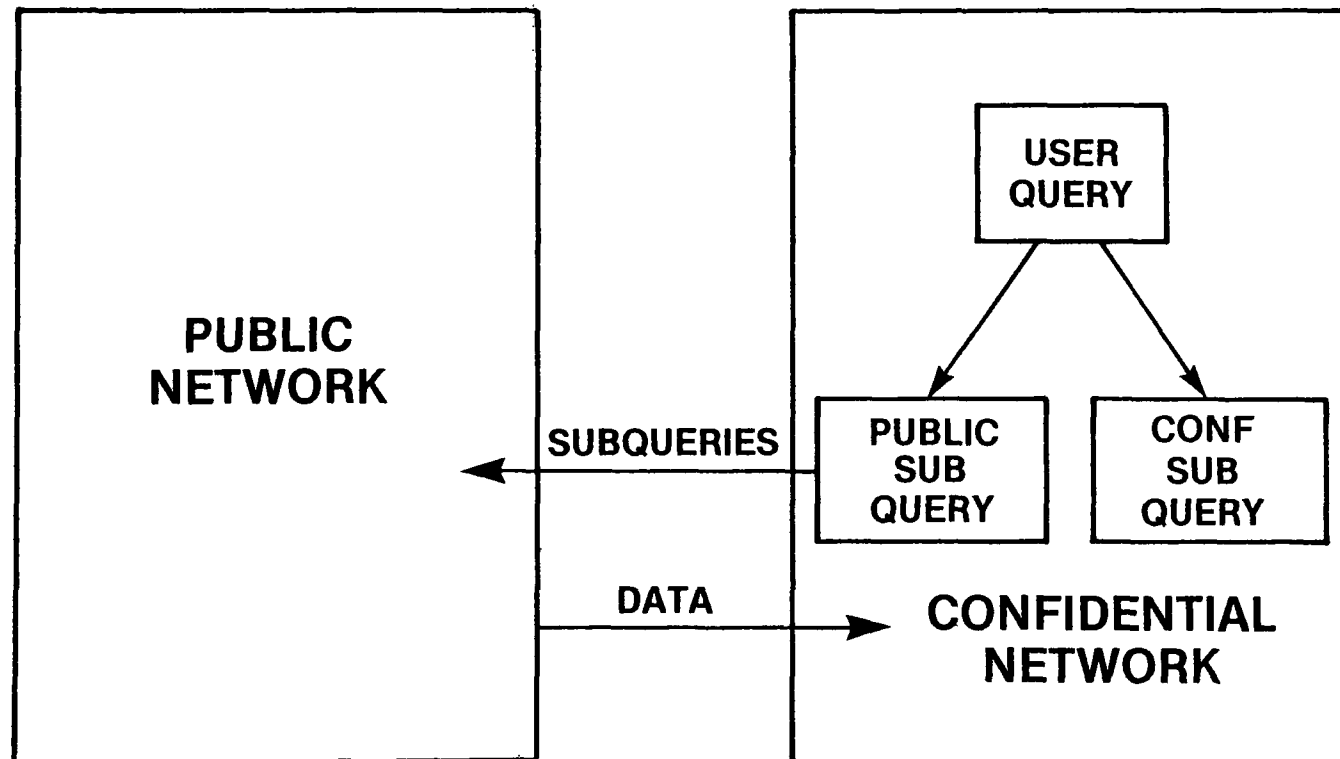
# **Scale of Feasible Prototype CSIN Operations**

	NUMBER OF COMPONENTS		
	LEVEL 1	LEVEL 2	LEVEL 3
Pre-Prototype Phase	2-5		
Prototype Phase I	4-7	2-5	
Prototype Phase II	6-12	4-7	2-5

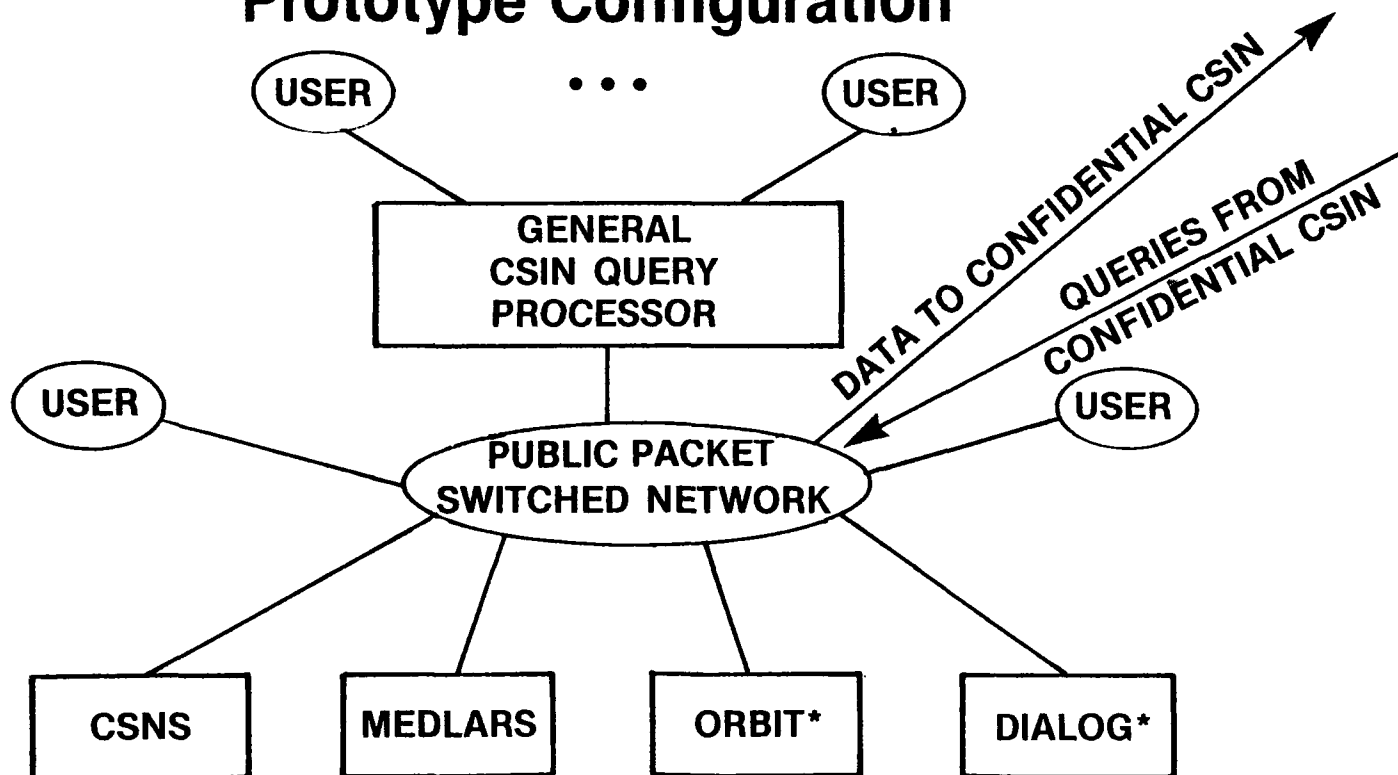
## **Definition of Levels:**

- 1. Communication Support**
- 2. Query Procedure Support**
- 3. End User Query Language Support**

# CSIN Security Scheme

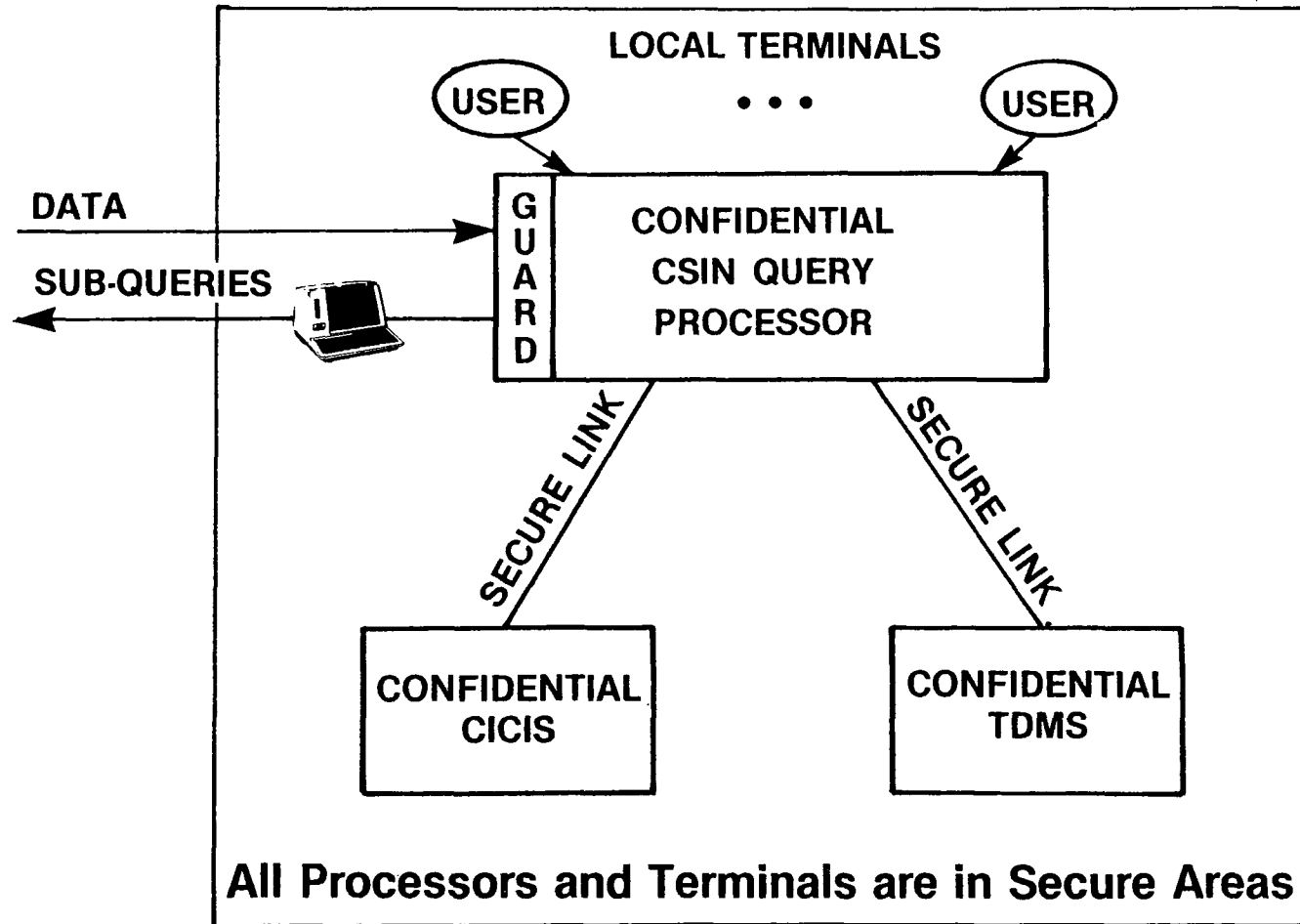


# Public CSIN Prototype Configuration



**\*Provisional Choice of Components; Other Early Candidates Include Public CICIS, CHEMTRAX, TDMS, HEEDA, CRGS and Others.**

# Confidential CSIN Prototype Configuration



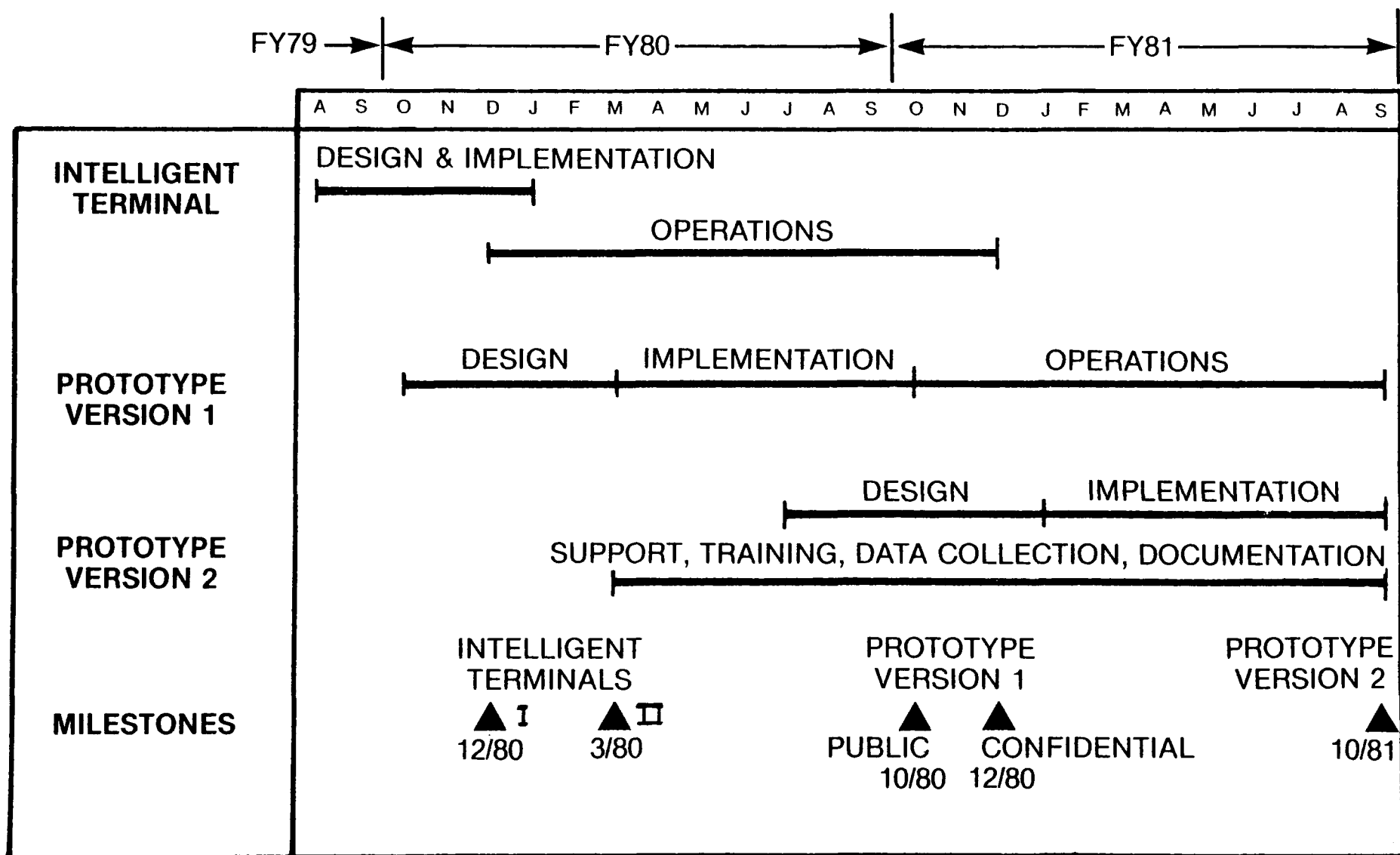
# **Design Principles for CSIN Prototype Configuration**

- 1. Small, Dedicated CSIN Query Processors Will:**
  - Allow Convenient and Economical Scaling**
  - Physically Separate Confidential and Public Access**
  - Simplify Operations and Development**
  - Allow Choice of CSIN Processor to Be Based on Query Processing Requirements Alone**
- 2. Confidential Subnetwork Allows:**
  - Shared Use of Public Data**
  - Good Protection of Confidential Data**
  - Joint Use of Public and Confidential Data Within Confidential Subnetwork**
- 3. Multi-Level Component Support Adds Flexibility**



# CSIN PROTOTYPE PROJECT PLAN

## CSIN PROTOTYPE PROJECT PLAN



# **Benefits of CSIN**

- **Increased Productivity of Professional Staff**
  - Less Effort to Locate and Retrieve Data
  - Routine Searches Handled by Sub-Professional Personnel
  - Reduced Training Requirements
- **Higher Product Quality**
  - At Each Stage of Evaluation, Better Information
  - Standardization of Routine Operations Through Scripts
- **Responsiveness**
  - More Rapid Retrieval of Information in Crises
  - More Adaptable Information Resources for Changes in Requirements
- **Effective Interagency Data Sharing**

## **Benefits of Prototype Approach**

- **Significant utility in the retrieval of information online from a growing set of key component systems**
- **Automation of routinized queries with consequent standardization of operations and saving of professional time**
- **Online query for the casual user on selected data from key component systems**
- **Start of selective global integration of the CSIN database**
- **Early revelation of hidden multi-system inter-system problems through actual implementation and use**
- **Extension, validation, and clarification of long-term CSIN requirements**
- **Progress toward implementation of a full scale CSIN system**