化学情報システム構築への提言

—NIH/EPA Chemical Information Systemに基づいて—

~~~~~ シュテファン R. ヘラー*¹

CIS の概要

このプロジェクトは、主に米国政府が財政的に負担し、組織的にも力を入れて作ったものであり、この種のプロジェクトとしては、おそらく世界最初のものだと思われる。これは、米国政府ほどクレージーなところはないからかもしれない。

いまや脱工業化社会が到来したと言われている。現代を支えるのは、エレクトロニクス、テレコミュニケーション、コンピュータの三本柱だとされており、これらのほとんどの分野で、日本が米国よりも進んでいるようである。これらの技術(テクノロジー)の応用(アプリケーション)の分野にお

いて、米国で最初に行われたものが一つ二つある。

コンピュータ利用のコストの推移というものはしばしば認識されずに終わっているが、イスラエルの友人からもらった資料によれば、コンピュータ利用が始まってからの、ある特定の科学上の産物のコストの変化を見ると、1940年から1980年までの推移の中で、1964年ごろに最も著しい変化が起きている。1964年ごろ、この産物の作り手がすべての組織をコンピュータ化した結果、それまでよりもコストの上がり方が一段と激しくなっている。

例えば'Chemical Abstracts'はコンピュータ化を積極的に進めた代表的な機関であるが、そのプ

* 1 Stephen R. Heller 米国環境保護庁(U.S.Environmental Protection Agency)

* 2 この講演は、昭和58年4月20日科学技術庁振興局情報室の主催により、サイエンスビル7階ホールで行われたものを収録し、それに基づきその後Dr.Hellerより送付された資料によって編集したものである。講演内容を補うために、編集者の責任において若干の脚注を付けた。(編集責任者 中井 浩)

ログクツの価格は毎年急激に上昇してきた。

これはコンピュータ化が期待していた状態を実現しなかったケースである。

技術革新を導入するだけでは、必ずしもコストエフェクティブだとは限らず、コストアップを招くこともあり得るわけである。技術を発展させる経済的な側面でのインセンティブは、軍事、ビジネス、娯楽、科学の四つである。

米国では、軍用用のものの調達には、金に糸目をつけない風潮があるが、ビジネスにおいても、売れる限り、ある程度高くてもがまんする傾向がある。娯楽の面では、価格が安いほど売れ行きがいい。例えばソニーの録音機、あるいは自動車などの例で分かるように、大衆消費者、大衆市場にとって価格は極めて大切なポイントである。日本でも米国でも、科学の分野は、ビジネス、消費者、軍事の分野に比べて、科学者の数がそんなに多くなく、この分野の製品は補助金・奨励金がつかない限り、コストが高くならざるを得ない。

オンラインコンピュータシステムを論ずるにあたっては、ハードウェア、ソフトウェア、コンピュータネットワーク、インフォメーションという四つの重要な項目が存在する。

一般的に、ビジネス用、軍用、科学用の分野ではコンピュータの使用が増加しているので、一番手に入れやすいのはハードウェアだと考えられる。

ソフトウェアは問題である。科学用の分野では、科学的な計算や作業をするための特殊なコンピュータプログラムが必要である。というのは、ビジネス用、軍用のプログラムでは役に立たないからである。

コンピュータネットワークは、ハードウェア同様、だれでも利用できる。

インフォメーションは、科学においては、ビジネス、娯楽、軍事の分野のものとは違い、独特のものがあるので、ソフトウェア同様の難しさを持っている。

したがって、CIS (Chemical Information System) の場合においても、あるいは、日本で現在計画さ

れているプロジェクトが実施された場合においても、ハードウェア、コンピュータネットワークの部分での問題の解決は容易であろうが、ソフトウェア、インフォメーションの部分が難しいわけである。

CIS プロジェクトは、1970年に NIH (National Institute of Health) によって開始された。それ以降、他の官庁、団体などが遅々としてはいたが、確実な歩みをもってこのシステムに参加するようになり、現在では、極めて広範なインフォメーションとソフトウェアを擁するようになった。

ここに来て真っ先に、CIS を構築するにあたっての最大の問題点とそれの解決策について質問されたが、これに答えるのはそんなにやさしくない。われわれが遭遇した多くの問題は、政治的・経済的なものであり、これらをインテリジェントな方法で解決する適切な手だてがないことは残念である。

しかし、この12年間を振り返ってみると、当初、ハッキリと確立されたプランがなかったことが、むしろ開発を成功させたのではないかという気がする。というのは、人間に協力してもらうのはなかなか難しく、とりわけ、最初に「これこれをしてくれ」と言ったら、人間というのは、なかなかそのとおりに動いてくれるものではないからである。実際には、まず特定のアイデアを提出し、それをやってみてうまくいったら採用し、うまくいかなかったら採用しない、というやり方を採った。

現在、CIS は情報のデータベースを毒物学 (toxicology)、環境 (environment)、法規制 (regulatory)、分光分析 (spectroscopy)、化学的・物理的な特性 (chemical & physical properties)、生物学 (biology)、分析・モデル化 (analysis & modeling)、書誌的な連関 (bibliographic link)、電子メール (electronic mail) の各分野において持っており、これらを一体化したシステムとして構築しているのが SANSS (Structure & nomenclature search system) と呼ばれるものである (第1図)。

CIS は、化学および法規制のデータベースと、

データ・アナリシス・コンピュータプログラムとを総合したシステムであり、書誌的情報だけではなく、毒物学、環境、法規制、分光分析、化学的・物理的な特性、生物学の各分野の数値的な情報を持ち、このシステムを使って、ストラクチャー、サブストラクチャーの全体および一部分のネームサーチングをすることによって、22万5,000以上の化学物質の探索を行うこ

とができ、さらに、ユーザ間のコミュニケーションの手段として電子メールのシステムを採用している。

CIS の構成

CIS の構成要素の一つ一つを簡単に説明しよう。

まず、中央には中核的な存在として SANSS がある。この中には23万以上の化学物質および、それに関連した65万以上の名前が入っている（第2図）。

このシステムの最初のソフトウェアは、NIH のフェルドマン氏によって開発され、この10年間、改善され続けてきた。23万に及ぶ化学物質のデータベースは、政府が関心を持つ化学物質のリスト

第1図 CISのシステム構成図

を作り、それを化学物質のマスターストとマージして作ったものである。

したがって、このデータベースには、政府が規制を行っている種々の化学物質や、政府自身が持っているデータベースの内容が含まれていることになる。

このデータベースは、文献の中に出てくる化学物質のリストやデータベースとは違うものである。現在、ビブリオグラフィック・サイテーションの中に出てくる化学物質は600万を超えている。この CIS データバンクとビブリオグラフィック用のデータベースとを明確に区別することが重要であり、性質が全く違うものである。

600万というのは多過ぎるし、商業的な価値を持つ化学物質は10万ぐらいであるから、そのぐらいのデータベースでもいい、という説を言う人には、23万でも多過ぎることになるが、商業的な価値を持つ化学物質に限定されたデータベースでも十分ではない。

CIS のデータベースは、どのような内容のものにするかが重要である。日本において特別な法規制の対象となっている化学物質、さらに、政策決定にかかわり合いをもつ化学物質を含めることも大切であるが、使われているものをすべて網羅しようということも賢明ではない。なぜならば、そ

STRUCTURE SEARCHING

STRUCTURE AND NOMENCLATURE SEARCH
SYSTEM (SANSS)

~ 254,000 CHEMICALS WITH
~ 804,000 CHEMICAL NAMES

第2図 SANSS

の大部分は、あまり用途がないものだからである。

実際に毒物学的なテスト、分光検査が行われている化学物質は、化学の分野では数千に過ぎない。化学用に質の高いデータベースをつくらうとするのであれば、対象は極めて狭い範囲になると言えよう。

現在出ている商業的な化学物質を網羅したデータバンクをつくらうとすれば、いまの科学力をしても10年はかかるであろう。本当に使いやすく良質な化学用のデータベースあるいはインフォメーションシステムをつくらうとするのであれば、対象は多少限定されても、質の高いインフォメーションを得ることが大切であり、それを使えば、効率的な意思決定や法の実施が可能であり、ひいては健康・環境の管理・保全にも役立つであろう。

きょうの話で最も大切なポイントは、質と量と、双方の比較・分析であり、10万とか100万とかいった数のものについての必要な情報を集めることは、ほとんど不可能であり、そういった情報の要求もしばしば寄せられるであろうが、それらのすべてに答えることは不可能であり、あまりそういうことには耳を傾けないほうがいいというアドバイスである。

それぞれの分野ごとに次のようなデータバンクがある。

(1) 毒物学 (第3図)

・RTECS：化学物質の毒性効果の登録簿であり、サイテーション (引用) が7万、化学物質が4万4,000入っている。毎年5,000ないし1万ずつ増えていくが、数はまだ少ない。しかし、これで急性の毒性のテストが可能なのは、明らかに短期的な毒性を示す化学物質の数が少ないからである。これについては、与えられた情報の質は問わない。

・CTCP：米国において商品として売られている化学物質の臨床的な毒性。対象となっている製品が2万。しかし、それを構成している化学物質は3,000に過ぎない。つまり、製品の数は膨大だが、ごくわずかの化学物質がそれを構成するに過ぎないことを、ここでも銘記すべきであ

TOXICOLOGY

REGISTRY OF TOXIC EFFECTS OF CHEMICAL SUBSTANCES (RTECS)
~85,000 CITATIONS FOR
~55,000 CHEMICALS

SCIENTIFIC PARAMETERS IN HEALTH AND THE ENVIRONMENT (SPHERE)
DERMAL & AQUATIC TOXICITY FOR
~3,000 CHEMICALS

CLINICAL TOXICOLOGY OF COMMERCIAL PRODUCTS (CTCP)
~20,000 PRODUCTS FOR
~3,000 CHEMICALS

CHEMICAL EVALUATION AND RETRIEVAL SYSTEM (CESARS)
~150 CHEMICALS

CHEMICAL HAZARD RESPONSE INFORMATION SYSTEM (CHRIS)
~900 CHEMICALS

CHEMICAL CARCINOGENESIS RESEARCH INFORMATION SYSTEM (CCRIS)
~900 CHEMICALS

第3図 毒性データに関するデータベースシステム

る。

・CESARS：400の化学物質が対象になっている。この数は増やしていきたいと思っている。ミシガン州の五大湖の汚染を調べるためのものである。やはり化学物質の数は極めて限られており、ミシガン州はこの5年間ほど、数百の化学物質を対象にして分析を行ってきている。

・SPHERE：魚類のとりわけ皮膚の毒性。やはり対象となっている化学物質の数が少なく、1,000から2,000の間である。

・CHRIS：危険物反応情報システム。米国沿岸警備隊が、海上に核物質が流れた場合の危険度を調べるためのもので、900の化学物質が含まれている。

(2) 環境 (第4図)

・OHMTADS：石油および危険物質の技術的援助のデータ・システム。鉄道貨車やトラックから石油および危険物質がこぼれた場合のもので、1,100の化学物質が対象となっている。

・TSCAPP：毒性物質規制法に基づくプラントおよびその製品に関するノン・コンフィデンシャルな化学物質5万5,000を対象に、8,500の報

ENVIRONMENT

OIL & HAZARDOUS MATERIALS TECHNICAL
ASSISTANCE DATA SYSTEMS (OHMTADS)
~1,200 CHEMICALS

TSCA PUBLIC PLANT & PRODUCTION
DATA SYSTEM (TSCAPP)
~90,000 REPORTS FOR
~60,000 CHEMICALS

TSCA UNDEFINED & VARIABLE COMPOSITION
DATA SYSTEM (UVCB)
~1,800 UVCB PRODUCTS

CLEARING HOUSE OF EPA DATABASES &
MODELING SYSTEMS (CLEAR)

第4図 環境情報に関するデータベースシステム

告書という形で持っている。どういう種類の化学物質が、だれによってどれだけ作られているかを調べるものである。

・WDROP：財政的な困難性によって解消されたので、説明を省略する。

・UVCD：やはり TSCA（毒性物質規制法）の規制下にあり、ディファインされていない重要な化学合成製品に関するもので、約1,800の製品が対象となっている。

・CLEAR：最近できたもので、EPA（環境庁）のデータベースとモデリングシステムの情報センターのようなものであり行政用に作られたものである。

(3) 法規制

・FRSS：1977年以降出た14万以上のサイテーション（引用文献）が網羅され、法規制の対象になっている化学物質とその所在、関係する法案などが分かる。一般に化学物質の規制というのは、まず、規制が提案された物質に対して最終的な審査を行った結果、パスすると、行政コードに登録される。米国では、行政コードに登録されたものと、規制が提案された段階の物質と、両方が分かるようになっている。

・CHEMLAW

・CRGS

(4) 分光分析（第5図）

分光分析の分野は CIS において最も古く、ある

意味では、最も多くデータを持っている。

・MSSS：約4万のエレクトロニック・インパクト・スペクトルが入っている。EPAはこのデータを使って、大気中の汚染物を探索している。もちろん、それらの化学物質に毒性があるかどうかを調べるのが大切である。

・CRYST：X線結晶データバンクは英国のもので、2万5,000以上の結晶構造が入っている。

・XTAL：NBSのシングル・クリスタルのファイルには6万ぐらい入っている。

・C-13NMR：オランダにあり、8,000のデータが入っている。

・IRSS：現在4,000ぐらいになっている。

なお、前述の毒物学の分野で、毒性の情報を持っている化学物質は4万から5万の間であった。また、分光分析の分野では、その数が4万ぐらいである。しかしながら、この二つのデータバンクをオーバーラップさせても、双方に共通なものは5,000しかない。というのは、毒性を研究している人は、質量分析計にかけようとするはずがないし、化学を研究している人は、毒物学の勉強をしようとはしないだろう。

しかし、極めてうまくオーバーラップしたデータバンクを持っているところがある。それは民間企業である。民間企業は、自分が作っていたり、かわりを持っていたりする化学物質についての情報は、徹底的に収集したいからであるが、これは企業機密として公表されていない。政府がスボ

SPECTROSCOPY

MASS SPECTRAL SEARCH SYSTEM (MSSS)
~40,000 EI SPECTRA
~1,500 CI SPECTRA

CARBON-13 NUCLEAR MAGNETIC RESONANCE
SEARCH SYSTEM (CNMR)
~6,000 SPECTRA

INFRARED SEARCH SYSTEM (IRSS)
~4,000 SPECTRA

第5図 スペクトルデータに関するデータベースシステム

CHEMICAL & PHYSICAL PROPERTIES

THERMODYNAMICS SEARCH SYSTEM (THERMO)
~ 14,000 INORGANIC CHEMICALS

X-RAY CRYSTALLOGRAPHIC SEARCH SYSTEM (CRYST)
~ 29,000 STRUCTURES
~ 36,000 CITATIONS

NBS CRYSTAL DATA IDENTIFICATION FILE (XTAL)
~ 57,000 STRUCTURES

POWDER DIFFRACTION SEARCH MATCH (PDSM)*
~ 30,000 STRUCTURES

PARTITION COEFFICIENTS (LOG P)*
~ 8,000 MEASUREMENTS
~ 6,000 CHEMICALS

*AVAILABLE ONLY TO ORGANIZATIONS THAT LEASE THE DATABASE

第 6 図 化学・物理学的特性データベースシステム

BIOLOGY

NUCLEOTIDE SEQUENCE SEARCH SYSTEM (NUCSEQ)

第 7 図 生物学データベースシステム

ANALYSIS & MODELING

CHEMICAL MODELING LAB (CHEMLAB)

MATHEMATICAL MODELING WITH CLUSTER ANALYSIS (MLAB)

FACTOR ANALYSIS & PATTERN RECOGNITION (ARTHUR)

COMBINED HANDLING OF ELUCIDATION METHODS
FOR INTERPRETABLE CHEMICAL STRUCTURES (CHEMICS)

第 8 図 分析モデリングデータベースシステム

BIBLIOGRAPHIC LINK

PROVIDES CAS REGISTRY NUMBER LINKS TO

LOCKHEED (DIALOG®)
SDC (ORBIT®)
NLM (MEDLARS)

第 9 図 文献情報システムとのリンク

ンサーとなっている公共的な情報を扱う場合、不備があることはやむを得ない。

(5) 化学および物理学的な特性 (第 6 図)

・THERMO: この NBS の熱力学のデータベースには 1 万 4,000 の無機化学物質が入っている。前述の質量分析および毒性分析に関するデータベースは有機化学物質が対象であったが、このほうは無機化学物質なので、オーバーラップはさらに少ない。データベースを別なソースから持ってきて作り、しかも使い方も違うという場合は、どうしてもオーバーラップする部分が少なくなるのはやむを得ない。例えばオーバーラップする部分があっても、それは最初から意図して行ったものではなくて、偶然の結果である。

(6) 生物学 (第 7 図)

・NUCREO: 最近、遺伝子工学などで脚光を浴びる分野として、重要性を増すであろう。

(7) 分析およびモデリング (第 8 図)

このシステムでもう一つ大事なのが分析およびモデリングを行うシステムであり、現在、四つが出来上がっている。

・CHEMLAB: エネルギー・コンファメーションおよびそれに関連した用途のものである。

・MLAB: 数学のモデリングのほか、クラスター、解析、カーブ・フィッティングなども行う。NIH が開発したもの。

・ARTHUR: ワシントン大学のコワルスキー教授が開発したファクト・アナリシス・アンド・パターン・レコグニションのためのものである。

・CHEMICS: 豊橋技術科学大学の佐々木教授が開発したもので、特定の化学物質がデータベースに入っていない場合、その分光分析によって性質を決定するシステムである。

(8) 書誌的な連関 (第 9 図)

CIS には、アブストラクト・レジストリーナンバーはあるが、書誌的な連関 (リンク) が欠けているので、データをトランスファーするために、次の三つを使っている。

・DIALOG: ロッキード

CIS COLLABORATORS

COUNTRY	ORGANIZATION
USA	NIH—OD, NCI, NIAMDD, DCRT, NLM EPA—OA—MDSO, OOHs OPTS—OTS, OPP ORD—Cincinnati, Athens, Research Triangle Park, Las Vegas OSWER—OSMCD, MDSD Great Lakes Program NBS—OSRD FDA NIOSH CPSC NTIS DOT—Coast Guard Arizona State University Case Western Reserve University Chemical Abstracts Service CRC Systems, Inc. Dartmouth College Fein-Marquart Associates, Inc. Infomatrix, Inc. Lederle Labs Merck & Co. Midwest Research Institute Nicolet Instruments Corporation SRI International State University of New York at Stony Brook State of Michigan—Dept. of Natural Resources Technology Service Corporation Thermodynamics Research Center, Texas A&M University Tracor-Jitco Inc. University of California, Riverside University of California, San Francisco University of Rochester University of Washington Finland University of Helsinki Technical Research Center, Espoo France University of Paris Germany Deutsche Krebsforschungszentrum, Heidelberg Holland Netherlands Information Combine Hungary Hungarian Academy of Sciences Israel Hebrew University, Jerusalem Japan Japan Association for International Chemical Information Kinokuniya Company Ltd. Tokyo Institute of Medical Sciences Toyohashi Institute of Technology University of Tsukuba Sweden University of Goteborg Switzerland University of Bern UK Cambridge University Crystal Data Centre Department of Industry Fraser-Williams (Scientific Systems), Ltd. Mass Spectrometry Data Centre SRC-Daresbury University of Manchester USSR USSR Academy of Sciences, Siberian Division Yugoslavia Boris Kidric Institute, Ljubljana

・ORBIT : SDC

・MEDLARS : MLN

このようなインフォメーションシステムを作る際に最も大切なのは、できるだけ多くの協力者を募ることであり、米国においては、政府、大学、民間企業などが全国的なレベルで協力している。ここで言う「協力」には、めいめいが持っているデータベースの利用やソフトウェアの提供を受けることも含まれる。また、国際的には、ヨーロッパ諸国、日本、ソ連などが協力者のリストに含まれている(第10図)。

CIS の組織

C I S の組織構造は、1982年10月以来、次のようになっている(第11図)。

まず、中央にある「CIS アドミニストレーター」は、一方では、前述のような協力者との連絡を取り、他方では政府の「EPA・関係省庁間マネジメント・ボード」との連絡をとっている。さらに、新しい試みとして、政府以外のメンバーから成る「サイエンス・アドバイザリー・ボード」を置き、われわれのやっていることに対する批判・忠告を行ってもらっている。

CIS アドミニストレーターは、三つの仕事を遂行している。第一に、ユーザに対するサポート。例えば訓練を行う。日本においては、紀伊国屋の業務となっている。第二に、研究開発。コンピュータプログラムおよびデータバンクに関する研究開発を行う。第三が、メイ

第 10 図 米国内・外のCISへの協力機関

ORGANIZATIONAL STRUCTURE

第11図 CISの機構図

メンテナンスである。コンピュータファシリティおよびコンピュータネットワークのメンテナンスを行う。第三の業務が最も簡単である。市販の標準的なコンピュータシステムを使っているからである。第二の業務が最も難しい。しかし、真に最も重要なのが第一の業務である。

現在、22か国に1,500ほどのユーザがいる。団体数で600ぐらいである。ユーザに対するサポートが重要なのは、ユーザが多ければ、政府のお偉方や監督機関も、「こんなにユーザがいるくらいだから、さぞかし有用なシステムなのだろう」ということで、財政援助を打ち切るようなことをしないだろうからである。

われわれの CIS は、民間企業なども積極的に参画して、自然発生的に出来上がったシステムであり、政府が法律をもって設立を命じたものではないので、ユーザの数ができるだけ多いということが、極めて大事なのである。

CIS は法律の根拠を持っていないので、この10年間くらい、政府から予算を獲得することが極めて困難であった。1982年は、ユーザサポートのための予算が全くなかったもので、ユーザの数が増えなかった。1983年は300万ドルの予算がついたが、1984年の予算がどのくらいつか、全く見当がつかない。どこが出すかすら分かっていないからである。したがって、各方面からの協力が必要であり、できるだけ腰を低くして協力を仰がなければならないわけである。

今年は、管理的な分野の新しいシステムを一つ

Option? NPROBE

Fragment or whole name search (F/W) (F)?W

Specify name (CR to exit): MERPHALAN

File 4, 1 compounds having name: MERPHALAN
Specify name (CR to exit):

Option? SSHOW 4

Structure 1 CAS RN 531-76-0
CIS Sources of Information

NIOSH/CIS, RTECS: AY36000,SQ7328000

Non-CIS References

Merck Index: 5746
CPSC, CHEMRIC
IARC, Monographs (Carcinogenicity Reviews)
NLM CHEMLINE: TOXLINE

C13H18C12N2O2

DL-Phenylalanine, 4-[bis(2-chloroethylamino)]- (9CI)
Alanine, 3-[p-[bis(2-chloroethylamino)phenyl]], DL-(8CI)
DL-Phenylalanine mustard
DL-Sarcosine
DL-Sarcosine
Merphalan
.....
Option?

第12図 SANSSによる名称検索の例

加えることになった。というのは、ユーザから費用を徴収するのに商務省の手を借りているからである。いまのところ、毎月6万ドルほどの収入を得ているが、ユーザの75~80%が米国にあり、20~25%がヨーロッパおよび日本にいる。

ここで、実際にシステムを運用する実例を示そう。第12図はSANSSを用いて化学物質の検索をしている例である。まず、SANSSを呼び出し、「命名検索」(NPROB)を指定する。そして「全体名による検索」(W)を選ぶ。そして化合物名「MERPH-A LAN」を入力すると、それに関するデータがFile 4にストアされる。それを出力させる(SSHOW)^(脚注1)。

次に第13図は質量スペクトルシステム(MSSS)における検索例である。まずピークによる検索(PEAK)を指定する。そしてピークの位置(M/Z) 85, 最低強度30, 最高強度100を入力すると該当1008物質の集合がつけられる。その中に、さらに、ピークの位置(M/Z) 125, 最低強度5 最高強度60を入力するとその集合は50物質に減少する。それを続けて最終的に6物質にまで縮めて、結果を出力させる。

OPTION? PEAK

TYPE PEAK, MIN INT, MAX INT
CR TO EXIT, 1 FOR REGN, QI, MW, MF AND NAME

USER: 85, 30, 100

• REFS M/Z PEAKS

1008 85

NEXT REQUEST: 128, 5, 60

• REFS M/Z PEAKS

50 85 128

NEXT REQUEST: 29, 10, 90

• REFS M/Z PEAKS

19 85 128 29

NEXT REQUEST: 69, 10, 100

• REFS M/Z PEAKS

6 85 128 29 69

NEXT REQUEST: 1

REGN	QI	MW	MF
15780-56-0	563	212	C12H20O3
730-46-1	603	254	C16H30O2
502-26-1	688	282	C18H34O2
2420-38-4	728	314	C19H38O3
52262-76-7	610	338	C22H42O2
26933-79-9	704	580	C37H72O4

NEXT REQUEST:

NAME
ACETOACETIC ACID, 1-ETHYL-CYCLOHEXYL ESTER (8CI)
2(3H)-FURANONE, 5-DO-DECYLDIHYDRO- (9CI)
HEXADECANOIC ACID, 4-HYDROXY-, GAMMA-LACTONE (8CI)
GAMMA-PALMITOLACTONE 2(3H)-FURANONE, DIHYDRO-5-TETRADECYL-(8CI)
GAMMA-STEAROLACTONE OCTADECANOIC ACID, 4-HYDROXY-, GAMMA-LACTONE 4-OCTADECANOLIDE
OCTADECANOIC ACID, 4-HYDROXY-, METHYL ESTER (8CI)
7,9-DOCOSANEDIONE (9CI)
HEXADECANOIC ACID, 1,5-PENTANEDIYL ESTER (9CI)
PALMITIC ACID, PENTAMETHYLENE ESTER (8CI)
1,5-PENTANEDIOL DIPALMITATE

第13図 質量スペクトルシステム(MSSS)の検索例

(脚注1)

SANSSにおいては、名称、分子式、構造式、分子量、原子数、リング数と員数によって検索が可能になる。

名称検索

コマンド

NPROB

指定パラメータ

Whode

全体名

Flagment

部分名

トランケーション指定:

前方一致 AMINOPHN:

中間一致 AMINO:OL

後方一致 :CHLORO

分子式検索

コマンド

指定パラメータ

MF

Complete

全分子式

Total

完全一致

Addend

付加物も許容

Partial

部分分子式

Ranged

元素数の範囲指定

分子量検索

コマンド MW

原子数検索

コマンド ACOUNT

リンクと員数

コマンド RCOUNT

Option? SPEC

CAS RN: 54115

Pyridine, 3-(1-methyl-2-pyrrolidinyl)-, (S)-

WLN: T6NJ C- BT5NTJ A

CAS RN = 54-11-5 CMR# 394 MW = 162.12

C10 H14 N2

L.F. JOHNSON, W.C. JANKOWSKI, C-13 NMR SPECTRA, JOHN WILEY & SONS, NY, (1972).

Solvent: DIOXANE

SHIFT	MULT	INTENS	ASSIGN
149.6	D	76	9
148.6	D	89	12
138.9	S	99	6
134.2	D	82	8
123.2	D	92	10
68.7	D	89	3
56.8	T	82	4
40.0	Q	79	1
35.6	T	82	5
22.8	T	82	7

出力項目は

CAS レジストリ番号 REGN
 クォリティ・インダストリ QI
 分子量 MW
 分子式 MF
 名称 NAME

である。

第14図は、C13核磁気共鳴データのシステム (CNMR) における検索例である。CAS のレジストリ番号 (54-11-5) を入力すると、その物質の構造式、名称、ワイスワッサ線型表示等の情報とともに、核磁気共鳴のデータが出力される。

第15図は、SANSS による部分構造検索と質量スペクトルデータ (MSSS) による検索を組み合わせた例である。まず SANSS によって、求める物質の構造を画く。"RING 5" によって 5 員環の骨格を作る (脚注2)。その中の①のノードに二つの物質を含む側鎖をつける。また②のノードに一つの物質から成る側鎖を二つつける。次いでノード③と新しくつけ

第 14 図 C 13 核磁気共鳴 (CNMR) システムの検索例

(脚注 2)

コマンド "RING 5" によってワークエリアに左のような 5 員環が造られる。"ABRAN n ATp" というコマンドは「ノード P に炭素数が n 個の側鎖をつける」を意味している。

"ABRAN 2 AT 1, 1 AT 2, 1 AT 2" で上の 5 員環は左のようになる。"ABOND P₁, P₂" は「ノード P₁ と P₂ を結合せよ」というコマンドである。

"ABOND 3, 7"

"ABRAN 1 AT 3, 1 AT 4" によって 5 員環は下のようにになる。

なお、ABRAN コマンドで造られる側鎖の各ノードは、まず炭素 C であることが仮定されている。しかしこのままで検索すると、必ずしも C ではなく、置換基

を想定した検索が行われる。もし置換基を明示したい時は "S-ATOM P₁, P₂....." のコマンドを用い、P₁, P₂..... に該当する物質を指定すればよい。

Combination Peak and Structure Search

Structure and Nomenclature Search System
(Version 3.11/4.3) March, 1980 (\$60/hr)

Latest news for SANSS . . .
24 April 80; SUBNMR Option Again Fully Operative

Complete data base selected.
Option? RING 5

Option? ABRAN 2 AT 1,1 AT 2,1 AT 2

Option? ABOND 3 7

Option? ABRAN 1 AT 3,1 AT 4

Option? RPROB

?	?	
?	?	
C????C??		
??C ?	Conditions of Search	
? ?C ?	Characteristics to be matched	Type of match
??C ?C	Type of ring or nucleus	EXACT
? ?	No heteroatoms	EXACT
C	Substituents at 2 2 3 4	EXACT
	This ring/nucleus occurs in 135 compounds	

File = 1, 135 compounds contain this ring/nucleus

Option? GO MSSS
NIH/EPA/MSDC Mass Spectral Search System (4.1/1.4; Feb 80) (\$36/hr)

Latest news for MSSS . . .
30 Apr 80; MSSS Version 4.1 Incorporates Many New Features

Option? PEAK

Type peak,min int,max int
CR to exit, 1 for REGN,QI,MW,MF and Name

User: 152,40,60

File 2 contains 113 references to M/Z 152

Next request:

Option? INTER 1 2

File = 3 resulting references = 1
Option? MSHOW 3

Print or Plot? PRINT

File 3 member 1

REGN	QI	MW	Formula, Names
76-22-2	676	152	C10H16O
			Bicyclo[2.2.1]heptan-2-one, 1,7,7-trimethyl-
			(9CI)
			Camphor (8CI)

Instrument: ; Inlet: ; Source Temp.: O C; eV: 70
Contributor:ORMSBG,5 1238 1A
There are 46 peaks in this spectrum. List (Y/N)? Y

User: the min & max m/z are: 0,999

User: type all m/z's with intensity > 5

m/z	INT	m/z	INT	m/z	INT	m/z	INT	m/z	INT
38	18	41	35	43	5	53	10	55	22
67	14	68	14	69	28	77	5	79	5
80	14	81	66	82	11	83	31	93	8
95	100	96	10	97	5	108	45	109	30
110	16	152	50	153	5				

た側鎖とを結び、ノード③と④に側鎖をつける。
このような構造を持つ物質を検索する(RPROB)。
このときの検索の条件は次に示される。

- 環の型の一致は正確であること
- 環は炭素以外を含んでいないこと
- ノード②の二つの側鎖、ノード③と④の側鎖には置換基があってもよい。

この結果、該当する135物質が検索され File 1 にストアされる^(脚注3)。

ここで質量分析データによる検索に移る(GO-MSSS)。そしてピーク検索を指示する(PEAK)

ピーク位置 (M/Z)	152
最低強度	40
最高強度	60

と入力すると、該当物質が113見つかり、それを

File 2 にストアする。

ファイル1とファイル2の積集合を造る(Inter 1, 2)と1物質のみとなる。その物質に関するデータを出力させる。

CIS の将来

最後に、CIS の将来について触れよう。

われわれの目標は、現在の組織でできるだけ効率よく機能を果たすことであり、'83年は組織再編成、効率向上の年である。ユーザ数や利用回数を増やして、将来の財政的な基盤を築いていきたいと思っている。

さらに、新しいデータベースを入れたいと思っているが、最近では、新しいものを捜すのが非常に

(脚注3)

部分構造検索において、次のようなコマンドがある。

(1) 構造を画くコマンド

* 骨格	炭素数 n の直鎖	CHAIN n
	n 員環	RING n
	m, n 縮合環	NUC mm
* 付加	炭素数 n の側鎖	ABR N n AT p
	官能基	AGR COOH AT p
	結合の生成	ABOND p1 p2
	p1 と p2 に炭素一つを挿入	AATOM p1 p2
	p1 と p2 に n 個の炭素を挿入	ALINK p1 p2 n
	p に n 員環をつくる	CRING p n
	p1 と p2 に n 員環をつくる	ARING p1 p2 n
	水素原子の表示	PUTH
	" 除去	ELIMH
* 結合	非炭素元素との差換え	SATOM p1 p2
	X : F, Cl, Br, I	
	M : 金属	
	EL : C, H, D, T 以外の元素	
	結合の種類の指定	SBOND p1 p2
	リングオールタネイトの指定	ALTBD p1 p2

* 変更	原子の消去	DATOM P
	結合の解除	DBOND p1 p2
	構造式の消去	CLEAR
	CAS のローカント番号につけかえ	MORGA
* 表示	直前の構造式	D
	" と分子式	DMF
	一つ前の構造式	REST
	結合表	C

(2) 全構造検索のコマンド

IDENT

(3) 部分構造検索のコマンド

構造入力

● FPROBE

中心元素と結合種の一致

XTAUT : トートマー結合を局在化した結合に限定

TAUT : 上記の解除

SUBSS : 複数部分構造と質問構造式の一致

TERMA : 最高結合原子数の指定

● RPROBE

ヘテロ環を持つ部分構造に有効

コードによる入力

● SPROBE

難しくなってきた。新しくつくるのも非常に費用がかさむようになってきた。これは非常に深刻な問題だが、現在の経済不況の中ではいい解決策は見いだせないようである。

しかし、データバンクあるいはインフォメーションシステムについて一般的に言えることとしては、数年前に比べてずっと多くの人が関心を示し始めたことである。しかし、これにはマイナスもある。たくさんの人が同じようなことをしようとすると結果、共倒れになりかねないからである。と言って、一企業、一機関でやろうとしても、コストがかかり過ぎる。米国政府ですら手に負えないほどになっている。したがって、われわれとしては、協力者の数をできるだけ増やさなければならぬと考えている。

このようなシステムをつくるのは、協力者や共同者がなければ、世界のどこでも成功しないであろう。しかし、これを理解してもらうのは極めて難しく、子供を育てるのと同じで、親の言うことをなかなか聞かないものだし、聞くようになったときにはもはや手おくれになっている。

米国においても、コンピュータシステムの構築によって多くの問題が解決できることが期待されながら、実際には失敗を重ねてきているが、その失敗の歴史をたどることからも教訓が得られるだろう。われわれの経験では、プロジェクトの企画、計画の段階から関係者が一生懸命力を合わせ、しかも、リーズナブルなところでとどめておくことが大切である。つまり、だれか一人が監督者、総括責任者になるという形ではなくて、民主主義的なやり方をすべきである。しかし、これは、言うは易く行は難いことであろう。

要するに、われわれは、この世界を生き延びさせようとするのであれば、みんなが協力しなければならないことをその経験から学んだ。化学情報システムやコンピュータの分野に限らず、どの分野でも必要なことは協力である。

〔質疑応答〕

Q：CISとしては、データベースの拡張と利用システムの拡張と、どちらに重点を置いておられるのか。

A：この2～3年間はアプリケーションの拡張をしていきたいと考えている。CISは一応のレベルにまで到達した感があり、さらに多くのデータを収集するのは、経済的・技術的に大きな困難を伴うであろう。したがって、ここしばらくはその方向の努力を中断し、ユーザのほうを固め、将来の飛躍に備えようと思っている。

Q：現在、シンケム2はどういう状態にあるか。

A：シンケム2(脚注4)のプログラムはIBMのコンピュータを使って行ってきたが、最近、ボックス(VAX)に変えた。10年間以上開発を続けてきたが、現在はあまり使用されていない。というのは、システムそのものの持つ制約とコストという大きな問題があるからである。コンピュータを1回走らせるのに500ドルから1,000ドルかかるので、実験用としてははなはだしく高くつく。したがって、予期したほどの伸びは見えていない。これは、最初のほうで述べた、コンピュータ導入が招いたコストアップの一例かもしれない。原則として、コンピュータは便利な道具だが、完全なものではない。

Q：データの数を増やすことよりも、よく使われるデータを持つほうが大事だという意味では、材料関係のデータが欲しい。合金、セラミックスなどのデータについては、どう考えておられるか。

A：材料のデータベースについては、NBSのサポートによってつくろうという機運はあるが、われわれの組織に協力したがないところがあって、はかどっていない。

私はコンピュータはいいものだと思っているが、世の中には、私の母親のように、コンピュータはあまりよくないものだと思っている人も多い。年配者に限らず、若者の中にもコンピュータを信用し

(脚注4) Synchem II。化学反応によって物質が合成されていくすべての可能な経路をファイルしておき、それに対話型でアクセスできるシステム。

なかったり、怖がったりしている人が見受けられる。まずコンピュータというものを認めさせるという、このシステムとは別個の問題が残っている。

このシステムを拡大する前に、できるだけ多くの人にコンピュータの扱いになれてもらうことが大切である。しかし、われわれのマネジメント・ボードのメンバーにも、ターミナルをいじったことがない人がいるくらいで、問題はそれほど簡単には解決しないだろう。

いかにわれわれが欲しいと思っても、必ずしもデータバンクを提供してくれるとは限らないことを十分認識しておくことが大事である。つまり、独裁者のように強制する態度は禁物であり、このシステムを使いたがっている人がいるし、使って大きな恩恵を受ける人がいるんだ、ということをジックリと時間をかけて説得すべきであろうと思う。さらに、ユーザの立場に立って考えることも大事である。自分の言い分がいつも正しいと思い込みがちであるが、ものの見方は人によってマチマチである。われわれが正しいと思っていることを、ユーザは必ずしも正しいと思っていないかもしれないので、その辺の意思疎通に努力することが大事である。いかにいいシステムをつくろうと、いかに多額の政府の予算がつこうと、だれも使ってくれないものであったらまったく価値がない。そんなことが民間企業で起こったら倒産するに決まっている。使ってもらえるシステムをつくることが大事である。

Q：ユーザを増やしていくための本質的な条件は何か。コンピュータになれた人を増やす以外にどういうことを考えておられるのか。

ユーザが増えることによって、こういうサービスが国から援助なしに採算が合うようになるという展望をお持ちか、それとも、将来ともこういうサービスには国家支出が不可欠だとお考えか。

関連して、CISが民間から自然発生的に生まれたときに存在したニーズと、今後ユーザを増やす場合に存在するニーズと、本質的に同じだとして考えるのか、それとも違ってくるものとして対処するのか。

A：ユーザを増やすためには、純粹に営業的に考え、販売促進をしなければならない。広告をし、ユーザのトレーニングをし、展示会などで積極的にデモを行い、適当なマニュアルを英語や日本語でつくらなければならない。われわれは実際にそれを行っている。

コンピュータになれた人を増やすのは時間がかかるだろう。現在、若者や大学生たちがようやくコンピュータを学び始め、アプリケーションをやり始めたところである。おそらく次の世代にならなければ、満足した状態にはならないであろう。とにかくすべての人にコンピュータになれてもらうのには、少なくとも10年や20年はかかるであろう。したがって、いま一番大切なのは、コンピュータ教育をいかに行うかということであり、これはわれわれの分野ではなくて、教育制度の問題である。

米国では、多くの学校で子供たちにコンピュータを教えているが、教えている先生のほうがよく知らないの、あまり効果が上がっていない。コンピュータ導入したが、2～3回使ったきりでホコリをかぶっているというところのほうが多いのではなかろうか。

実話をご紹介します。米国の下院議員の一人は「小学生用のコンピュータのテキストを買ったが、読むのがこわい。たぶんむずかしくて理解できないと恐れている」と言う。私は、そんなことがないことを強調し、勇気づけなければならなかった。ワシントンで大きな影響力を持っている人物ですらこのありさまなので、こういうことを克服するのには、ずいぶん時間がかかりそうである。しかし、それが克服されないかぎり、ユーザが飛躍的に増えることも期待できないだろう。

次に、このようなシステムが商業ベースに乗って独立採算がとれるかどうか。いまのところ、今後5年ないし10年は、ノーだと答えなければならない。われわれが持っている書誌のデータであるロッキード、SDC、BRSなどは、欧米や日本でも広く利用され、ユーザの数も多く、情報の量も多いが、これでもまだ赤字である。CISは、これらに比べてずっと小さく、化学物質も500万も600万も

入っているわけではなくて、わずかに2万とか3万とかいった程度のものが多いから、こういった科学的な数値データを出すシステムが商業ベースに乗ることは、もし可能であるとしても、ずいぶん先のことであろう。

次に、十数年前に比べてユーザのニーズが変わったかどうか。本質的には変わっていない。科学というものは、そんなに早く変わるものではなく、科学のためのツールが変わり、科学者にもっと便利なツールを提供することはできるようになったけれども、求められている情報そのものは変わっていない。

Q：国際連合の環境専門機関で進められている有害化学物質の登録管理などのプロジェクトと、CISや米国EPAとはどのような関連を持っているのか、具体的な例でお話しいただきたい。

A：国連のそういう委員会との共同作業については、私は知識を持っていない。

Q：われわれが知っている米国のデータセンターは、NBSのもとにあるのがたくさんあるが、そういうものとの関係はどうなっているのか。例えばパーデューのPPRCなどとの調整がうまく行われているのかどうか。

データを採用するにあたって、それがいいデータであるということをどういう手続で決めておられるのか。評価のやり方をお聞きたい。

A：NBSのデータバンクとの協力体制については、実際に一緒にやっており、そのデータバンクをわれわれのプログラムの中に入れているが、NBSというのは保守的な組織なので、われわれが望むほどに早いペースで協力体制にはいつてきているとはいいがたい状態である。NBSの関係者に比べてわれわれのほうがはるかに若いので、われわれのペースが早過ぎるのかもしれない。NBSのほうは、間違っただけをしていないかどうか、ということで、常に慎重な態度をとっているのが現状である。

データの品質に関しては、それぞれのデータの関連の専門家に判断してもらっている。

Q：データメンテナンスについて、特に重要

な特性データの数値は、どのような頻度・方法でメンテナンスしておられるのか。また、それを絶えずユーザにアナウンスする必要があると思うが、それをどのようにやっておられるのか。

分光分析データとか、いろんな種類のデータがあるが、オーバーラップしていない部分を補てんする意向があるのかどうか。

二次元から三次元に座標を変換する場合、完全に正しいパーセンテージで出てくるのか。X線解析などでやったものとピットリ一致するのかどうか。

A：データメンテナンスについては、われわれはできるだけデータをアップデートし、間違いを訂正していきたいと思っているが、それがどれだけ頻繁にできるかというのは、どれだけお金があるかということにかかってくる。データメンテナンスというのは、なかなか費用がかかる。いままでの経験ではミスが発見される頻度が最も高いのは、データを実際に使ってもらったときであるが、これはよしあしである。

データのオーバーラップについては、できるだけ増やそうと思っではいるが、われわれのところに来る情報のソースがあまりにも多岐にわたっており、情報源における調整が至難である。いまのところ、年間25万ドルぐらい使って、商業的に市場に出ている化学物質と質量解析データとの間のオーバーラップを増やそうとしているが、現在市場に出ている化学物質が6万にものぼっているので、毎年1,000ぐらいずつやっていくのでは、長年月を要することになる。

二次元から三次元に座標を転換するプログラムは、一般的に言ってもうまくいっていると言えよう。ボンド（結合）の長さ、角度については全く問題がない。しかし、ボンドの構造、ねじれぐあいのほうは難しい。このプログラムでは、ミニマム・エネルギーが計算できることになっているけれども、難点が全くないわけではない。いずれにしても、一般的に言ってもうまくいっていると言えるし、計算もそれほど難しくない。難しいのは、物理的特性の計算、例えばパーティション係数、ダイポール・モーメントなどのところである。（終）