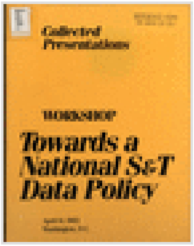


Collected Presentations Presented at the Workshop Towards a National S&T Data Policy (1983)

Numerical Data Advisory Board; National Academy of Sciences; Committee on Science and Technology, U.S. House of Representatives; Congressional Research Service, Library of Congress

Pages
135

Size
8.5 x 10

ISBN
0309327296

Collected Presentations Presented

at the

WORKSHOP

Towards a National S&T Data Policy

April 14, 1983

Whittall Pavilion
Jefferson Building, Library of Congress
First and Independence Avenue, SE
Washington, D.C.

Coordinated by:

Committee on Science and Technology
U.S. House of Representatives

Congressional Research Service
Library of Congress

Numerical Data Advisory Board
National Academy of Sciences

NOTICE: This document reproduces papers by session chairmen and speakers as received. The document has not been subjected to NRC or other review and is intended only to provide reference to points of view expressed by speakers representing their own views.

Limited copies of this unpublished
document are available from
Numerical Data Advisory Board
National Research Council
2101 Constitution Avenue
Washington, D.C. 20418

TABLE OF CONTENTS

Plenary Session Page

Keynote

Needs for an Industrially Based Economy

Workshops

Natural Resources and Energy

Environment

Agriculture and Food

Health

Summaries

Concluding Remarks

IMPROVING R&D PRODUCTIVITY:
THE FEDERAL ROLE

LEWIS M. BRANSCOMB

VICE PRESIDENT AND CHIEF SCIENTIST

INTERNATIONAL BUSINESS MACHINES CORPORATION

ARMONK, NEW YORK

COMMITTEE ON SCIENCE AND TECHNOLOGY
U.S. HOUSE OF REPRESENTATIVES

AND

NUMERICAL DATA ADVISORY BOARD
NATIONAL ACADEMY OF SCIENCES

WORKSHOP

"TOWARDS A NATIONAL S&T DATA POLICY"

WASHINGTON, D.C.

APRIL 14, 1983

In 1955, John von Neumann wrote an article for Fortune called "Can We Survive Technology." He answered his question with a qualified yes. "Yes," he said, "if a long series of small decisions are correctly taken."

That wise and prudent statement was made, I remind you, shortly before the beginning of the golden decade for science in the industrialized world. For 10 years, the scientists and engineers of Europe, the Soviet Union, the United States and Japan enjoyed unprecedented influence, support, even admiration. The race to the moon, the European determination to close the technology gap, the announced-in-advance industrial aspirations of Japan and the cold war were among the driving forces.

Decisions were big ones, and they were not always correctly made. We were too impatient to be cautious, too eager to fulfill the exaggerated expectations that the public was encouraged to entertain. Viewed from the perspective of the technical community, it was an exhilarating period of extraordinary achievement -- while it lasted, but it failed to build a solid infrastructure for science and technology. Then came the morning after, with attendant regrets and a crisis of confidence in the efficacy of science.

In the late 1960s, we began to think through how the world had gotten into the trouble it faced. The technologies we had stressed were largely irrelevant to the world's most crucial unmet needs, although they were all too relevant to the arms race. When the environmental costs and risks were confronted, the reaction was to slow down and take a hard look at the social consequences of new technologies. Technology assessment was born to serve primarily as an early warning system.

The scientific and engineering community was until recently, I think, discouraged about the prospects, but we have reached a turning point in our view of science and technology. Today we see, throughout the Congress and the public at large, a very serious conviction that science and engineering capability, knowledge and, most of all, education are key to economic success. People have a pragmatic view about searching for policies that will build a strong nation and a healthy society.

Fifteen years ago, looking at this issue, I wrote the following: "It is clear, everyone agrees, that a serious problem in information storage and retrieval faces scientists today. But the real problem is that the reader does not know what to believe in the surfeit of so-called 'information' that retrieval systems disgorge. It is just as absurd for the user to tap the total collection of raw material for his data as it would be for a jeweler to order six tons of gold-bearing ore when he wants to make a cuff link. What is the solution? We must increase the density of useful information in the literature. We need better quality control in science."

Today more than ever, now that everyone is focusing on productivity improvement, let's not forget the need for increased R&D productivity, especially in the billions invested by the private sector to use knowledge created by scientists. With the Government spending some $47 billion a year on R&D, it seems to me penny-wise and pound-foolish not also to invest in evaluation, integration and end user packaging of this knowledge produced at Federal expense so that it can and will be put to use by the private sector. More effort in this area would greatly enhance the leverage of economic benefits from the Federal government's investment.

Data and access to it are critical to every R&D project. Recently, in checking our own experience, I discovered that, in one typical IBM lab and plant library, 25 percent of the reference collection is primarily concerned with scientific and technical data. Some 15 percent of the volumes in circulation and over a third of the journals deal heavily with numeric data. This is a very heavily used collection.

When accurate, pertinent data are available, work can proceed. If not, work must stop while a researcher invents a different approach, develops (or redevelops) missing data, or must experimentally verify unevaluated data reported in the literature before daring to commit another period of time and effort on a project that is heading down a critical path.

Progress isn't made by stopping work, or by shifting the goal away from the one you must achieve toward the one you can achieve. That's a good practice in research. It's a terrible way to do business.

There is no way to measure this loss in R&D productivity -- the cost in problems not pursued because they are unpursuable, or in new knowledge delayed -- but the cost of the present policy is obviously high.

As materials science and process technologies mix with piece-parts assembly in high technology manufacturing, we will see another whole new wave of industry requirements for basic scientific data. In this case, the requirements read directly on manufacturing effectiveness and productivity, as well as on the research and development process that most people associate with science and technology data.

Getting a product out the door is no longer a simple linear process, if, indeed, it ever was. Ideas don't necessarily originate in research and flow to development, to manufacturing engineering, to production and to marketing. The relationship we now confront is a triangular one among partners in research, development and manufacturing. It is not unusual today in my company to find large numbers of research scientists -- Ph.D.'s from our corporate laboratory -- working side by side with manufacturing and development engineers. We are often preparing to manufacture with processes that deal with phenomena science has not yet fully mastered. Here, instead of scientific knowledge at risk because of missing data, it might be a whole new product line; it might be several billion dollars in revenue.

None of this is news to many of you. The famous Weinberg Report of the President's Science Advisory Committee that was, in some sense, the starting point for this whole debate -- is now more than 20 years old. The National Bureau of Standards' National Standard Reference Data System is also 20 years old. It is still healthy, I understand, but it has had flat funding for the last decade, and it certainly does not represent a growing segment of our national technical activity.

From the point of view of the dedicated scholars in the data evaluation and review literature field, no doubt progress has been made. They certainly have done yeoman service for the technical community and for the country. But from the point of view of engineers and scientists in a company that employs over 20,000 of them in the United States, the general perception is that one cannot, in general, expect to find existing reliable, evaluated data, even on matters you know have been subjected to scientific research and publication. Handbooks are generally out of date and incomplete; caveat emptor governs the application of results from the primary literature.

That very perception damages progress. I think that, as a nation, we are so far from doing the job properly that there is little demand pressure to get it done, simply because expectations are so low. Another reason, of course, is that the job is not as simple as many people make it out to be. They erroneously focus their attention on access and distribution, rather than on the scholarship required to put data in a form in which users will dare rely upon it.

Institutions end up delegating the job back to individuals and, in most company situations, an individual who confronts a data problem in the middle of a piece of committed work, with a committed schedule, has no choice but to try to get around it empirically. In most cases, there is no time to remedy the situation, and there certainly is no time to alter the knowledge environment. The knowledge environment is assumed by most companies to be something they cannot affect, at least not in first approximation or in the near term.

Government today depends on professional societies to sustain the journals for both the primary and secondary literature. Yet, there is coming into being a commercial capability to provide science and technology data collections through computer networks in a form attractive to users.

I'm sure you're all aware that value-added data communications networks such as Telenet and Tymnet are now rapidly growing in capability. They represent roughly a billon dollar a year market, and explosive growth is predicted.

What a lot of people don't realize is that there is a business four to five times larger in this country, called remote computer services. These companies operate some of the biggest networks in the world, and they are providing access to information and to compute-capability, in a business running close to $5 billion a year.

The fall 1982 Directory of Online Databases -- itself an old fashioned book, I might point out -- describes 1,350 databases, up 40 percent from the year before. The publishers anticipate that their spring 1983 edition will show 20 percent additional growth. Mind you, those are not all science and technology databases; they represent the whole panaply of information for sale via networks and other electronic access methods.

These capabilities can be a vehicle for private sector access to science and technology data, but ensuring reliable, retrievable data is not a function we should leave exclusively to professional societies, the publishing industry or the private sector. The cost of the scientific expertise required to evaluate the data and put it in a form attractive to end users is small compared to the cost of the original research, but it represents a large, high risk front-end investment. I know of no case of a data vendor deciding to invest from scratch in both original research and the means for providing commercial access, at prices adequate to recover the cost of the original knowledge, the cost of added evaluation and the cost of dissemination.

Thus, it seems to me very appropriate for the private sector to provide access to knowledge and as much of the analysis as can be business-justified. Government should accept the premise that funding the review literature and data evaluation is a responsibility integral to funding research activities. Government research agencies should fund both the generation and the evaluation and user-packaging of the knowledge.

A national science and technology data policy would have six elements:

1. Productivity of the research and development process, half of which is sustained with Federal funds, should be a major Federal concern. Agencies sustaining the science and technology infrastructure should be held accountable for increasing R&D productivity.

2. Funding of research to produce generally useful knowledge carries with it a responsibility, not only to ensure publication of results but also their evaluation and preparation in a form suitable for application and the assurance of public access at reasonable cost. Again, all agencies funding research should be held accountable.

3. The private sector should be encouraged to take an increasing, but not exclusive, role in the provision of access to evaluated data, but commercial companies cannot be expected to finance the depth of scholarship required for data analysis and evaluation. That must be viewed as the obligation

of the sponsor of the original research. User fees for the allocated cost of access are appropriate for Government-provided information systems.

4. Unrestricted access to unclassified data, generated at public expense, is a cardinal requirement for a dynamic, high technology economy. The desire to frustrate technical progress by hostile nations must not be allowed to impede the competitiveness of our own economy in its dependence on available scientific and technical data.

5. It is strongly in the U.S. interest to make agreements with other nations, to share the costs and scarce skills for data evaluation and access. We should take the lead in this worldwide cooperative effort as, indeed, we have done in the past. For, as the nation with the most innovative economy, most dependent upon being innovative in the future, we stand to gain the most from global cooperation.

6. The task of evaluating and preparing for application published scientific and technical data of general utility is a joint responsibility of user institutions in the private sector, private information vendors, professional associations and societies and agencies funding, producing and using research. Some national body with Federal and private participation should monitor the adequacy of this science and technology information system and provide policy guidance to the Federal agencies and recommendations to the private sector.

I am sure that other and more specific ideas will emerge from this workshop. I've not given you a recipe for how to bring whatever policy, in the workshop's judgment, is correct into being.

In truth, I'm not even very optimistic. My recommendation number six, for example, was in place in the fifties and sixties when COSATI was in the Office of Science and Technology Policy. The executive coordinating function for science and technology information policy has been moved and downgraded ever since.

We will not make much progress in this area until the Government takes a very pragmatic view of its whole science and technology policy -- not just a data policy. I have long advocated such an overall policy, but I do it with a certain amount of concern.

The science and technology policy I suggest would expend Federal funds on assuring that those solving problems in the interest of our society -- whether they be in the public or private sector -- have the best and most appropriate, most available technical means for doing so.

Now, if that were our Federal science policy, the number one priority would be to make available existing knowledge. Most existing knowledge is very badly underutilized. If users benefitted greatly from information services, there would be an increased demand for new knowledge to fill in the gaps -- because massive gaps there are -- and we would have a driving force for the basic research investment.

Workshop

Towards a National S&T Data Policy

WASHINGTON, D.C. -- April 14, 1983

Dr. Hollis L. Caswell
Laboratory Director
IBM Corp -- E. Fishkill
Rt. 52
Hopewell Jct., NY 12533

INTRODUCTION

WE'VE ALL READ OR HEARD OF THE GREAT ADVANCES WHICH HAVE BEEN MADE IN SILICON TECHNOLOGY. THIS OF COURSE IS THE TECHNOLOGY WHICH HAS LED TO RAPID IMPROVEMENTS IN THE COST PERFORMANCE OF DATA PROCESSING EQUIPMENT, THE AVAILABILITY OF HOME VIDEO GAMES, AND THE CURRENT RUSH TOWARD PERSONAL COMPUTERS. MOST PEOPLE ARE AWARE OF THE PROGRESS. FEWER ARE AWARE OF THE PROBLEMS WHICH HAVE HAD TO BE OVERCOME TO ACHIEVE THIS PROGRESS. IT'S FREQUENTLY WHEN ONE IS IN DIFFICULTY THAT THE AVAILABILITY OF GOOD SOLID SCIENTIFIC DATA ARE PRICELESS IN TERMS OF HELPING TO ARRIVE EXPEDITIOUSLY AT A SOLUTION TO A KEY PROBLEM. YOU MIGHT ASK, "CAN'T A COMPANY THE SIZE OF IBM GENERATE ITS OWN DATA AS REQUIRED?" THE ANSWER IS "YES" BUT THERE IS A PRICE AND THE PRICE IS THE TIME TO GATHER THE SKILLS, SET UP THE REQUIRED APPARATUS, PERFORM THE EXPERIMENTS, AND ANALYZE THE RESULTS. FREQUENTLY, BY THEN, IT'S TOO LATE.

I'D LIKE TO CITE SEVERAL EXAMPLES TO ILLUSTRATE THE VALUE OF GOOD, PUBLISHED INFORMATION IN SOLVING PROBLEMS. THESE ARE EXAMPLES OF WHERE WE AT IBM DID HAVE TO GENERATE KEY DATA WITHIN THE COMPANY BECAUSE IT WAS NOT AVAILABLE IN THE LITERATURE, AND WHERE MISLEADING RESULTS DELAYED PROGRESS IN AN EXCITING ANALYTICAL FIELD.

I WILL START BY GIVING YOU AN EXAMPLE OF WHERE SOME RESEARCH WHICH MOST PEOPLE WOULD HAVE CONSIDERED TO BE VERY ESOTERIC HAD A PROFOUND IMPACT UPON SEMICONDUCTOR DEVICE DESIGN. THE TIME PERIOD WAS THE LATE 1960's AND IBM WAS WORKING TO DEVELOP A LEADING EDGE HIGH PERFORMANCE PROCESSOR. THE PROCESSOR USED A NEW SET OF HIGH PERFORMANCE TRANSISTORS. IT'S NOT UNUSUAL FOR A NEW SYSTEM WITH NEW TECHNOLOGY TO EXPERIENCE MODULE FAILURES DURING THE SYSTEM ENGINEERING PHASE. HOWEVER NORMALLY THE RATE OF THESE FAILURES FALLS OFF AS MORE EXPERIENCE IS GAINED WITH THE TECHNOLOGY AND SUITABLE CORRECTIONS ARE MADE. HOWEVER IN THE CASE OF THIS PROCESSOR THE MODULE FAILURES DID NOT DECREASE AND IT WAS VERY CLEAR THAT WE HAD A PROBLEM. HOWEVER IT WASN'T EQUALLY CLEAR WHAT THE PROBLEM WAS. ALL OF OUR NORMAL FAILURE ANALYSIS ROUTINES INDICATED THE MODULES HAD FAILED BECAUSE OF AN ELECTRICAL OVERLOAD, BUT THERE HAD NOT BEEN ANY RECOGNIZED ELECTRICAL OVERLOADS. ONLY AFTER A DETAILED EXAMINATION OF THE SEMICONDUCTOR CHIPS MOUNTED ON THE MODULE DID WE FIND THAT THERE WAS A VERY FINE HAIRLINE CRACK IN THE ALUMINUM METAL USED TO CARRY CURRENT ON THE CHIP.

IN FACT THE CRACK COULD ONLY BE SEEN UNDER VERY HIGH MAGNIFICATION ON SAMPLES WHICH HAD BEEN SPECIALLY PREPARED AS SHOWN IN THE FIRST SLIDE. A LITERATURE SEARCH WAS INITIATED TO SEE WHAT EFFECTS HAD BEEN OBSERVED IN METALS CARRYING CURRENTS. WE LEARNED FROM THIS SEARCH THAT A PROFESSOR BY THE NAME OF H.B. HUNTINGTON AT RENSSELAER POLYTECHNIC INSTITUTE HAD MADE A NUMBER OF MEASUREMENTS ON WIRES CARRYING HIGH CURRENTS.

HE HAD BEEN STUDYING A PHENOMENON CALLED ELECTROMIGRATION AND BEEN ABLE TO SHOW THAT THE ELECTRONS ASSOCIATED WITH CURRENT FLOW INTERACTED WITH THE METAL ATOMS IN A WIRE AND THESE VERY LIGHT ELECTRONS COULD ACTUALLY MOVE THE METAL ATOMS FROM ONE END OF THE WIRE TO THE OTHER. OF COURSE AS ATOMS PILED UP AT ONE END OF THE WIRE THEY CREATED AN ABSENCE OF ATOMS AT THE OTHER END OF THE WIRE. AS THAT END GOT THINNER, MORE CURRENT HAD TO FLOW THROUGH THE REMAINING MATERIAL AND THE MOVEMENT OF ALUMINUM ATOMS INCREASED. FINALLY THE WIRE FAILED.

THE SIGNIFICANCE OF THE HUNTINGTON WORK WAS THAT THIS WASN'T AN EFFECT WHICH WAS INTUITIVELY OBVIOUS. FOR EXAMPLE, IT TAKES APPROXIMATELY 50,000 ELECTRONS TO EQUAL THE MASS OF ONE ALUMINUM ATOM. THUS TRYING TO MOVE ALUMINUM ATOMS WITH ELECTRONS IS A BIT LIKE ATTEMPTING TO MOVE A BOULDER WITH A PEA SHOOTER. HOWEVER, IN MATERIALS THE INTERACTIONS ARE VERY COMPLEX AND ELECTROMIGRATION IS THEORETICALLY POSSIBLE BECAUSE OF COOPERATIVE FORCES BETWEEN ELECTRONS.

I SHOULD MENTION THAT IN FABRICATING THE DEVICES USED IN THIS HIGH PERFORMANCE PROCESSOR, WE HAD MADE THEM MORE SUSCEPTIBLE TO ELECTROMIGRATION EFFECTS BECAUSE WE HAD MADE THE ALUMINUM VERY THIN. IT WAS MADE THIN BECAUSE WE HAD TO ETCH VERY FINE PATTERNS IN THE ALUMINUM AND THIS WAS EASIER TO DO WITH THIN ALUMINUM. FOR EXAMPLE THE FOUR FINGER STRIPES SHOWN ON THE NEXT SLIDE HAVE A WIDTH LESS THAN ONE TWENTIETH THE THICKNESS OF A HUMAN HAIR. THUS IN ATTEMPTING TO SOLVE THE ETCHING PROBLEM WE HAD UNKNOWINGLY CREATED ANOTHER PROBLEM. EVEN TODAY AFTER THE

DEVELOPMENT OF ALLOYS WHICH ARE MUCH LESS SUSCEPTIBLE TO ELECTROMIGRATION THAN PURE ALUMINUM, ELECTROMIGRATION STILL POSES A FUNDAMENTAL DESIGN CONSTRAINT ON MANY SEMICONDUCTOR DEVICES. WITHOUT THE PIONEERING WORK OF HUNTINGTON AND CO-WORKERS IT WOULD HAVE UNDOUBTEDLY TAKEN FAR MORE TIME AND EFFORT TO HAVE IDENTIFIED THE CAUSE OF THESE FAILURES AND MORE TIME TO ARRIVE AT SUITABLE IMPROVEMENTS FOR HIGH PERFORMANCE COMPONENTS.

LET ME GIVE YOU ANOTHER EXAMPLE DEALING WITH ELECTROMIGRATION; ONE WHERE IBM HAD TO GENERATE FUNDAMENTAL DATA BECAUSE THEY WERE NOT READILY AVAILABLE IN THE LITERATURE. IBM USES A TECHNOLOGY CALLED C4 TO MAKE INTERCONNECTIONS TO THEIR SILICON CHIPS. THE NEXT SLIDE SHOWS A PHOTOGRAPH OF A SILICON CHIP WITH THE LEAD ALLOY SOLDER PADS ON TOP OF THE CHIP. THE CHIP IS SURROUNDED BY GRAINS OF SALT FOR REFERENCE PURPOSES. THIS CHIP IS THEN FLIPPED OVER AND SOLDERED DOWN TO A SUBSTRATE AS SHOWN IN THE NEXT SLIDE. SIGNALS AND POWER ARE BROUGHT TO THE SILICON CHIP FROM THE OUTPUT PINS, THROUGH THE CERAMIC SUBSTRATE TO THE SOLDER COLUMNS WHICH SUPPORT THE CHIP. THE NEXT SLIDE SHOWS AN ACTUAL CROSS SECTION. BECAUSE THE CURRENT DENSITIES ARE RELATIVELY HIGH IN THESE SOLDER COLUMNS THERE'S A CONCERN HERE WITH ELECTROMIGRATION. WHAT ACTUALLY CAN OCCUR IS THAT VOIDS CAN BE FORMED NEXT TO THE SILICON CHIP AND ULTIMATELY LEAD TO FAILURE OF THE C4 PAD.

THE INFORMATION NEEDED IS DATA ON THE RATE AT WHICH METALS DIFFUSE IN THE SOLID STATE THROUGH THESE LEAD ALLOYS. EVEN AT TEMPERATURES AROUND ROOM TEMPERATURE, ATOMS MOVE IN THESE SOLID ALLOYS. WHAT WE NEEDED TO KNOW WAS THE DIFFUSION RATES OF LEAD IN LEAD, OF TIN IN LEAD, AND THEN THE DIFFUSION RATES OF THESE TWO ELEMENTS IN LEAD ALLOYS CONTAINING VARIOUS AMOUNTS OF TIN, INDIUM, GOLD, AND COPPER. THESE DATA ARE SIMPLY NOT AVAILABLE. EXPERIMENTS WERE INITIATED AT THE THOMAS J. WATSON RESEARCH CENTER IN YORKTOWN, NEW YORK TO OBTAIN THESE DATA. WE ALSO CONTRACTED WITH H.B. HUNTINGTON AT RPI TO OBTAIN SIMILAR DATA ON THE FAST DIFFUSERS IN THESE HIGH LEAD ALLOYS. THE WORK CONTINUES TO THIS DAY.

AS THESE DATA BECAME AVAILABLE WE WERE ABLE TO REFINE THE METALLURGICAL COMPOSITIONS OF OUR C4 JOINTS AND OBTAIN INCREASED RESISTANCE TO ELECTROMIGRATION. THIS IS BUT ONE EXAMPLE OF WHERE VERY PRACTICAL CONSIDERATIONS HAVE LEAD US TO SEEK RATHER FUNDAMENTAL MATERIALS ORIENTED DATA.

IN THIS CONNECTION, THE CURRENT PROPOSAL SPONSORED BY THE AMERICAN SOCIETY FOR METALS AND THE NATIONAL BUREAU OF STANDARDS ON ALLOY PHASE DIAGRAMS IS OF INTEREST TO US. PHASE DIAGRAMS TELL WHAT HAPPENS AS ONE MIXES TWO OR MORE ELEMENTS TOGETHER. THEIR PROPOSAL IS TO MAKE AVAILABLE TO REMOTE COMPUTER TERMINALS, METALLURGICAL PHASE DIAGRAMS FOR BINARY, TERNARY, AND HIGH ORDER ALLOY SYSTEMS. THIS IS THE TYPE OF DATA BASE WHICH WILL ASSIST MANY IN THEIR WORK.

LET ME GIVE YOU ANOTHER EXAMPLE. THIS HAS TO DO WITH THE IMPACT OF NUCLEAR PARTICLES ON SEMICONDUCTORS. WHAT HAPPENED SEVERAL YEARS AGO WAS THAT COMPUTER ENGINEERS FOUND THAT THEY WERE OBSERVING VERY STRANGE MEMORY ERRORS IN THEIR COMPUTER SYSTEMS. THE SOURCE OF THESE ERRORS WAS ULTIMATELY TRACED TO MINUTE QUANTITIES OF RADIO ACTIVE MATERIAL CONTAINED IN THE PACKAGES AND MATERIALS USED TO FABRICATE THE SILICON MEMORY DEVICES. THESE RADIO ACTIVE MATERIALS WERE EMITTING ALPHA PARTICLES WHICH ARE CHARGED HELIUM NUCLEI. THESE PARTICLES WHEN THEY STRUCK THE SILICON SURFACE CREATED ERRORS IN THE COMPUTER STORAGE ELEMENTS. THIS EFFECT WAS HIGHLIGHTED IN A PAPER PRESENTED IN 1978 BY MAY & WOODS AT THE 16th RELIABILITY SYMPOSIUM. THIS PUBLICATION SAVED THE COMPUTER INDUSTRY MUCH EFFORT SINCE, ONCE THE PROBLEM WAS IDENTIFIED, SOLUTIONS EVOLVED RAPIDLY. THE KEY HAD BEEN TO RECOGNIZE THE SOURCE OF THE ERRORS AND INFORM THE TECHNICAL COMMUNITY OF THE PROBLEM.

HOWEVER COSMIC RAYS PRESENTED A SIMILAR, BUT POTENTIALLY MORE DIFFICULT PROBLEM SINCE COSMIC RAYS ARE VERY ENERGETIC AND IT IS IMPOSSIBLE TO STOP COSMIC RAYS FROM INTERACTING WITH COMPUTER COMPONENTS. COSMIC RAYS ORIGINATE WHEN VERY HIGH ENERGETIC PARTICLES FROM OUTER SPACE STRIKE ATOMS IN THE EARTH'S UPPER ATMOSPHERE. THESE COLLISIONS RESULT IN SHOWERS OF PARTICLES KNOWN AS COSMIC RAYS WHICH RAIN DOWN ON THE EARTH.

THUS THE UNLIKELY SITUATION DEVELOPED WHERE PEOPLE CONCERNED WITH COMPUTER MEMORIES BECAME INTERESTED IN THE COSMIC RAY DATA CONTAINED IN THE EVALUATED NUCLEAR DATA FILE PUBLISHED BY THE NATIONAL NUCLEAR DATA CENTER AT THE BROOKHAVEN NATIONAL LABORATORY IN UPTON NEW YORK. WITH THESE DATA IT WAS POSSIBLE TO GO THROUGH AND ASSESS POTENTIAL ERROR RATES THAT ONE WOULD ENCOUNTER IN COMPUTER MEMORIES. ACTUALLY TO VERIFY SOME OF THESE RESULTS COMPUTER MEMORY DEVICES WERE FLOWN IN AIRPLANES AT HIGH ALTITUDE AND THE SAME TYPE OF DEVICES WERE OPERATED IN VERY DEEP GOLD MINES IN ATTEMPTS TO ASSESS COSMIC RAY EFFECTS. THUS A VERY UNLIKELY SOURCE OF RATHER FUNDAMENTAL DATA IN A FIELD COMPLETELY FOREIGN TO COMPUTER SCIENCE TECHNOLOGY PROVED TO BE VALUABLE WHEN COMPUTER TECHNOLOGY RAN INTO THE QUESTION OF ASSESSING THE ROLE OF COSMIC RAYS IN MEMORY ERROR RATES. NEEDLESS TO SAY, IF YOU'RE WORKING WITH SPACE ELECTRONICS, THESE COSMIC RAY DATA ARE OF SIGNIFICANCE TO YOU AS YOU ASSESS RELIABILITY CONCERNS.

MY LAST EXAMPLE IS ONE DESIGNED TO ILLUSTATE THE IMPORTANCE OF HAVING GOOD RELIABLE SCIENTIFIC DATA. IT DEALS WITH ONE OF THE MOST POWERFUL ANALYTICAL TOOLS AVAILABLE IN OUR INDUSTRY TODAY; THE SCANNING ELECTRON MICROSCOPE. THIS IS AN EXTREMELY POWERFUL TOOL IN THAT IT PERMITS PICTURES TO BE TAKEN WHICH ARE EXTREMELY CLEAR AND PROVIDE A THREE DIMENSIONAL REPRESENTATION. LET ME SHOW YOU SEVERAL SLIDES TO ILLUSTRATE THE POWER WHICH THE SEM OFFERS. THE FIRST IS A PHOTOGRAPH TAKEN WITH A HIGH-

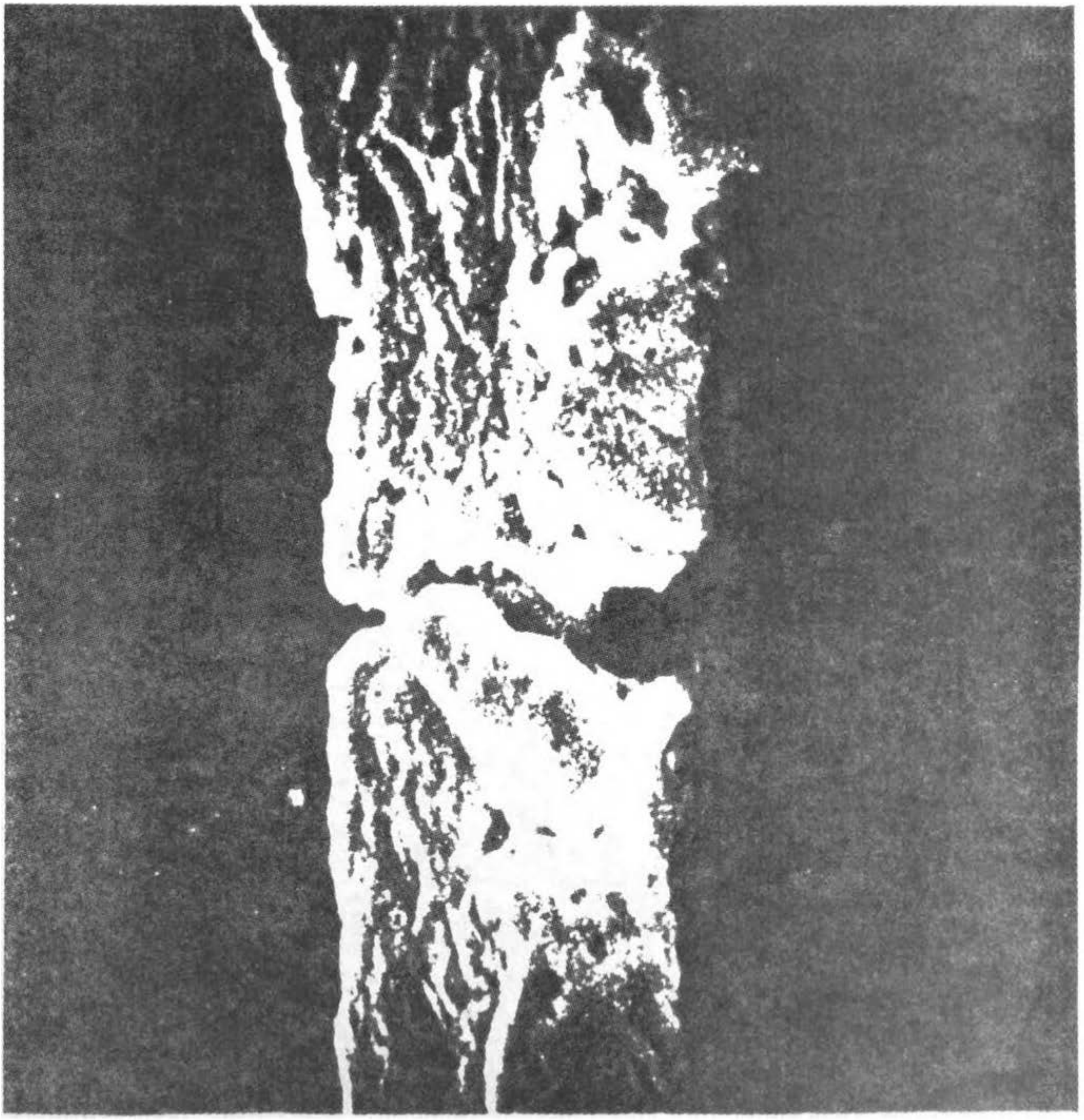
QUALITY OPTICAL MICROSCOPE OF METAL PATTERNS WHICH HAVE BEEN FORMED UPON THE TOP SURFACE OF A SILICON WAFER. YOU'LL NOTE THAT THE RESOLUTION OR THE ABILITY TO SEE DETAIL IS FAIRLY GOOD. HOWEVER WHAT THE SLIDE LACKS IS ANY INFORMATION UPON THE DEPTH IN THE STRUCTURE. IF THE SAME SAMPLE IS EXAMINED WITH A SCANNING ELECTRON MICROSCOPE THEN YOU FIND DEPTH PERCEPTION IS GREATLY ENHANCED AS SHOWN NEXT. FURTHERMORE, THE ELECTRON MICROSCOPE HAS GREATER RESOLUTION CAPABILITY SO THAT WE CAN MAGNIFY STRUCTURES MORE AS SHOWN IN THE NEXT SLIDE TO SEE GREATER DETAIL. ONE CAN LITERALLY FIND CRACKS AND VOIDS IN THESE STRUCTURES WHICH ARE JUST NOT EVIDENT IN AN OPTICAL MICROSCOPE. THE NEXT SLIDE IS ANOTHER ILLUSTRATION OF THE THREE DIMENSIONAL DETAIL THE SEM DISCLOSES.

INTERESTINGLY ENOUGH THERE ARE THOSE IN THE FIELD WHO BELIEVE DEVELOPMENT OF THE SCANNING ELECTRON MICROSCOPE WAS DELAYED BECAUSE OF A TECHNICAL PUBLICATION. IN 1942 A RESPECTED LABORATORY PUBLISHED RESULTS OBTAINED WITH AN EXPERIMENTAL SEM. THESE RESULTS DID NOT LOOK ENCOURAGING AND DISCOURAGED OTHERS FROM ENTERING THE FIELD. IT WASN'T UNTIL 1963 THAT THE SEM CONCEPT WAS SHOWN TO BE PRACTICAL. ONCE THIS WAS DONE PROGRESS WAS RAPID AND THE SEM HAS LONG SINCE BEEN A VALUABLE ANALYTICAL TOOL. HOWEVER, TWENTY YEARS HAD PASSED FROM THAT KEY PUBLICATION IN 1942 TO A RE-AWAKENING OF INTEREST IN 1963.

HOWEVER, IN THE SAME FIELD THERE IS AN EXAMPLE OF WHERE AN EARLY PUBLICATION PROVED OF GREAT VALUE MANY YEARS LATER. ONE OF THE PROBLEMS FACED IN ELECTRON BEAM EQUIPMENT WAS THE TUNGSTEN FILAMENT WHICH WAS USED AS A SOURCE OF ELECTRONS. THIS FILAMENT WAS HEATED VERY HOT AND HAD A LIMITED LIFE ASSOCIATED WITH IT. REPLACING A TUNGSTEN FILAMENT AND REALIGNING IT IN THE ELECTRON BEAM COLUMN WAS A TIME CONSUMING AND TEDIOUS TASK. A SCIENTIST BY THE NAME OF ALEC BROERS WAS LOOKING FOR A BETTER SOURCE OF ELECTRONS. IN SEEKING A MATERIAL FOR THIS APPLICATION HE WENT BACK SOME 16 YEARS TO A PAPER PUBLISHED IN THE JOURNAL OF APPLIED PHYSICS IN 1950 BY J.M. LAFFERTY OF THE GENERAL ELECTRIC RESEARCH LABORATORY IN SCHENECTADY, NEW YORK.

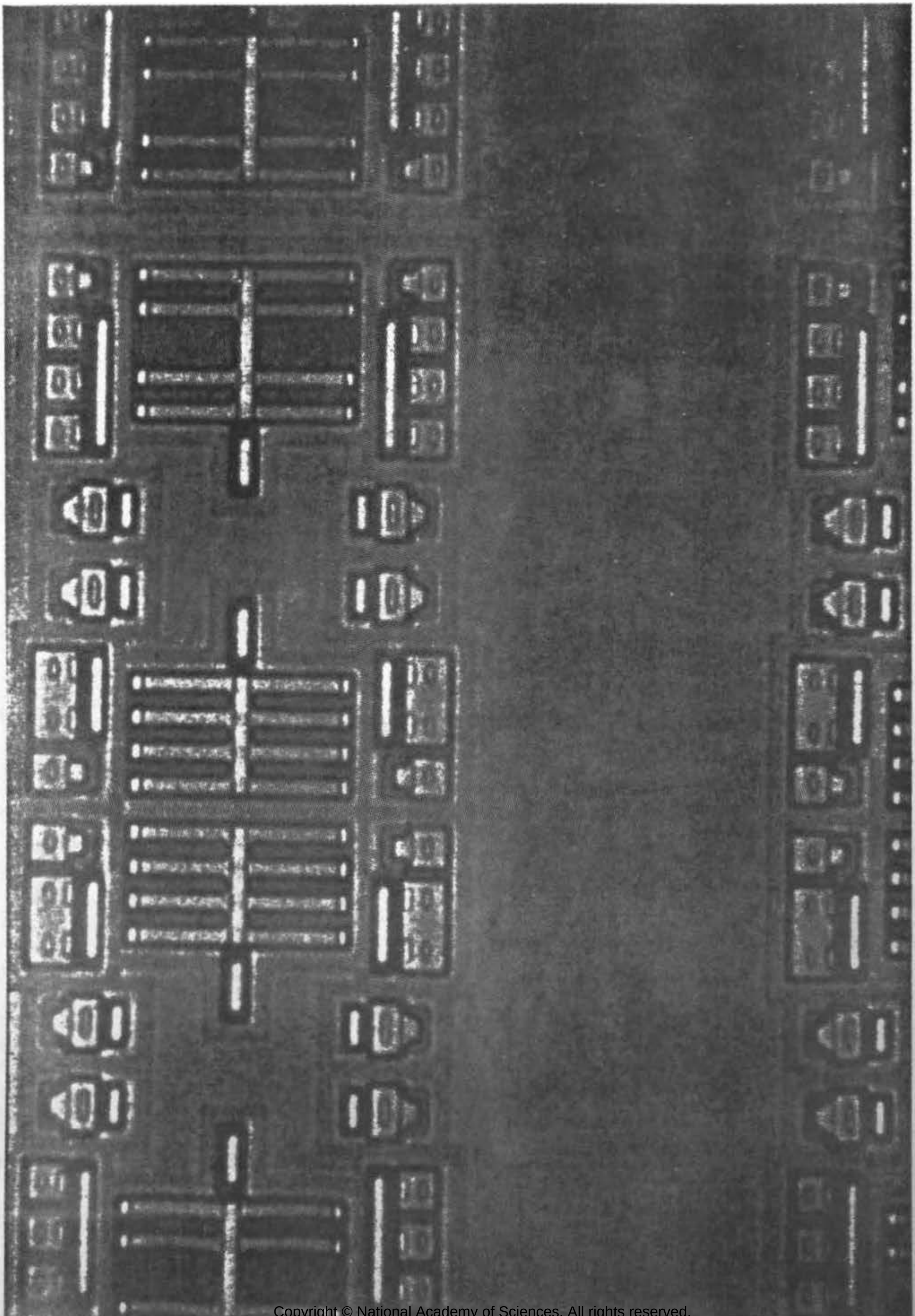
THIS PAPER INDICATED THAT LANTHANUM HEXABORIDE COULD SERVE AS A SOURCE FOR THESE ELECTRONS. THIS LED IN 1966 TO A PUBLICATION IN THE JOURNAL OF APPLIED PHYSICS ENTITLED 'ELECTRON GUN USING LONG-LIFE LANTHANUM HEXABORIDE CATHODE'. WITH THIS DEVELOPMENT THE LIFETIME OF THE CATHODE WAS EXTENDED BY OVER 100 TIMES FROM THAT AVAILABLE WITH A TUNGSTEN HAIRPIN. FURTHER DEVELOPMENTS HAVE LED TO THE AVAILABILITY OF SINGLE CRYSTAL LANTHANUM HEXABORIDE WHICH HAVE EXTENDED THE LIFETIME OF THESE FILAMENTS EVEN MORE.

IN CONCLUSION, I'D LIKE TO CITE SOME LIBRARY STATISTICS FROM IBM SINCE I BELIEVE THESE STATISTICS PERHAPS REFLECT THE VALUE WHICH THE IBM CORPORATION PLACES UPON MAKING AVAILABLE TECHNICAL DATA TO ITS SCIENTISTS AND ENGINEERS. THE IBM CORPORATION HAS LABORATORIES THROUGHOUT THE WORLD AND EACH OF THESE LABORATORIES HAS ITS OWN LIBRARY CONTAINING TECHNICAL INFORMATION APPROPRIATE FOR THAT LABORATORY'S MISSION. HOWEVER, THE RESEARCH DIVISION LABORATORY LOCATED AT YORKTOWN HEIGHTS, NEW YORK IS OUR LARGEST LABORATORY AND I'D LIKE TO CITE A FEW STATISTICS FOR IT. THIS LIBRARY HAS 82,000 BOUND JOURNALS AND BOOKS, RECEIVES 1800 JOURNALS FROM 21 COUNTRIES, HAS 3500 REELS OF MICROFILM AND AN OPERATING BUDGET FOR 1983 OF APPROXIMATELY $1M. I CAN PERSONALLY VOUCH FOR THE FACT THAT THE LIBRARY IS USED EXTENSIVELY AT ALL HOURS OF THE DAY AND NIGHT. THE INFORMATION THE LIBRARY CONTAINS HAS BEEN IMPORTANT TO OUR SUCCESS. IT'S EQUALLY IMPORTANT THAT FREE FLOW OF INFORMATION AND AVAILABILITY OF KEY DATA CONTINUE IN THE FUTURE.



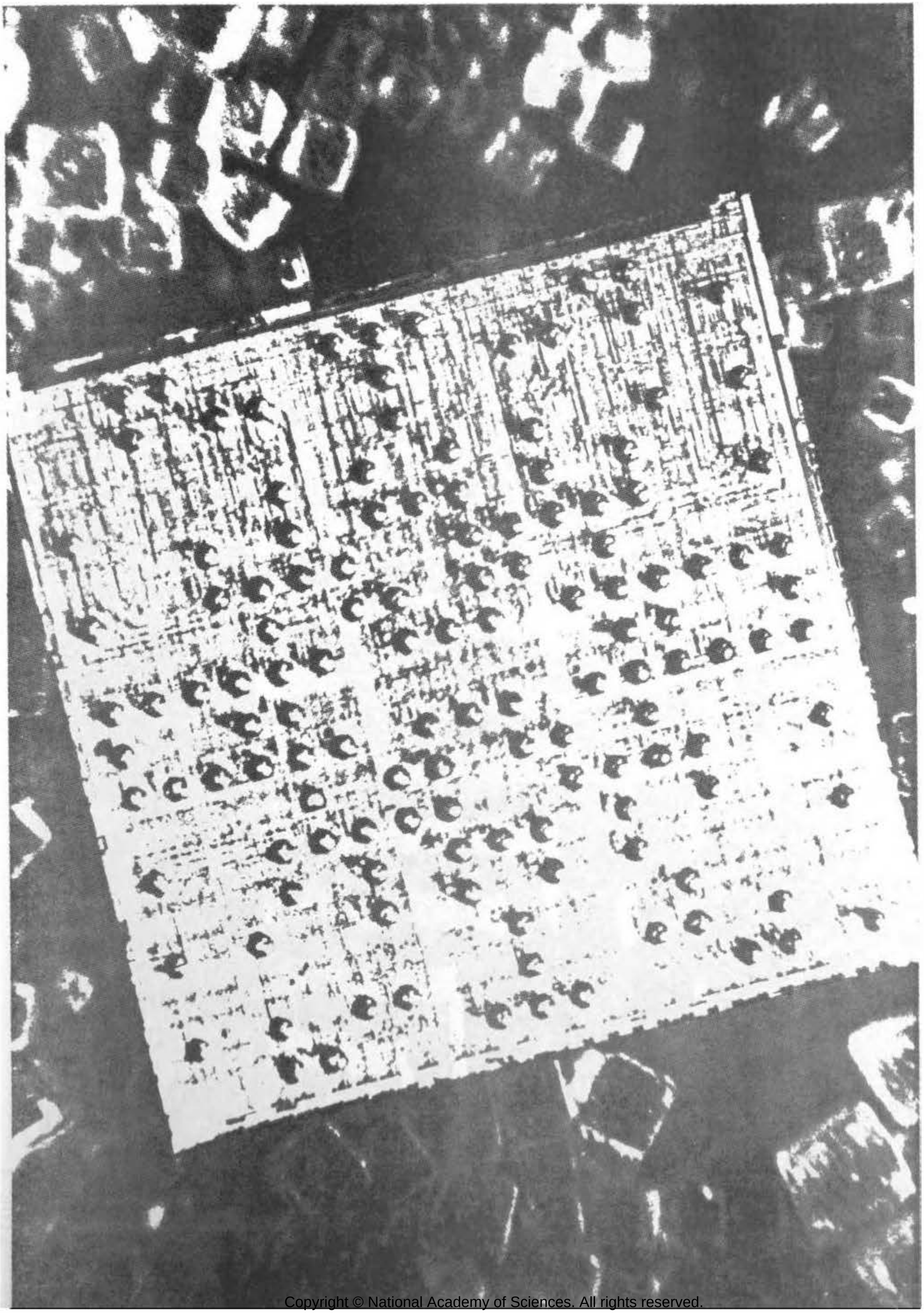
IBM-EF- 47

SLIDE 1



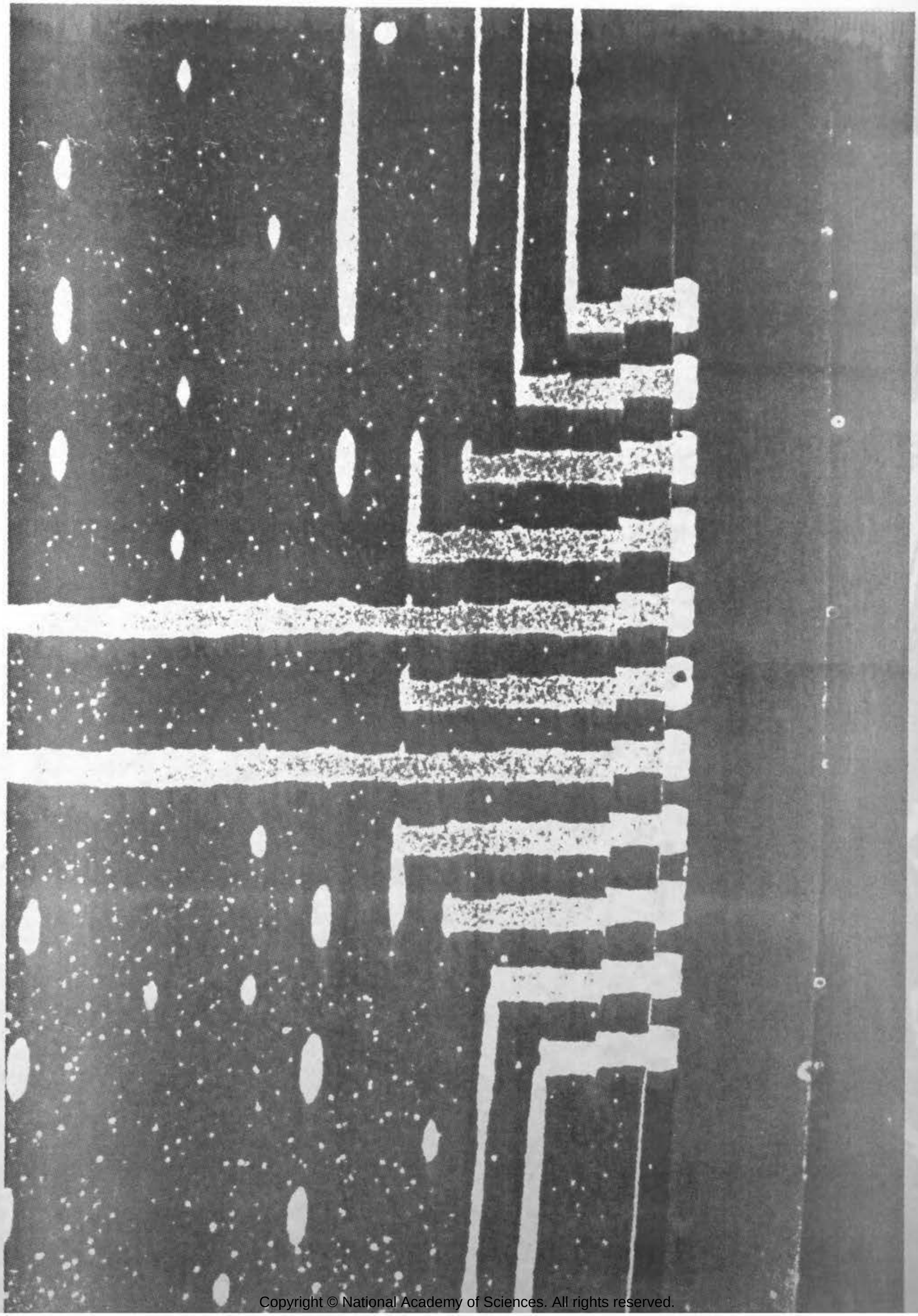
IBM-EF- 47

SLIDE 2



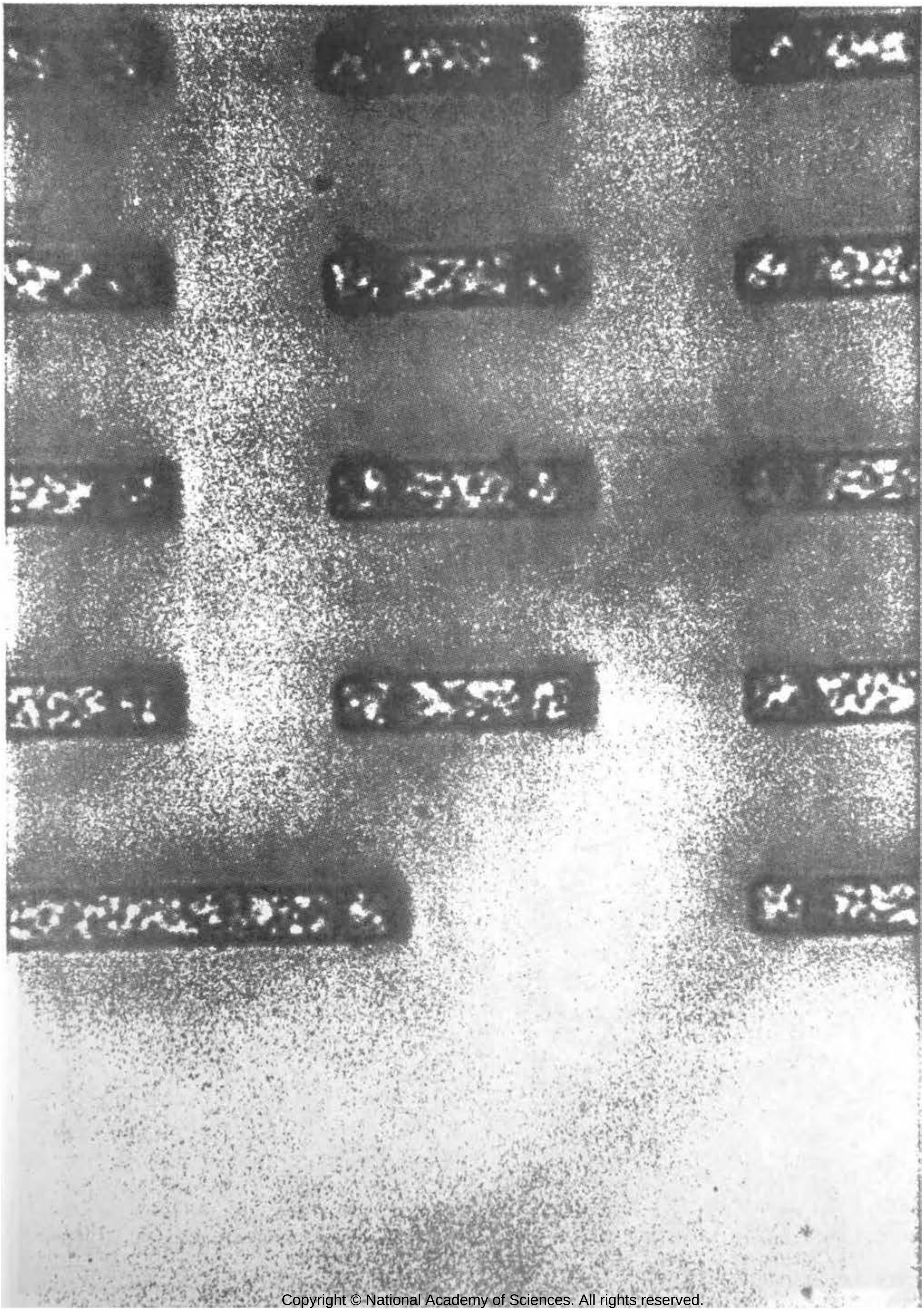
IBM-EF- 47

SLIDE 3



IBM-EF- 47

SLIDE 4

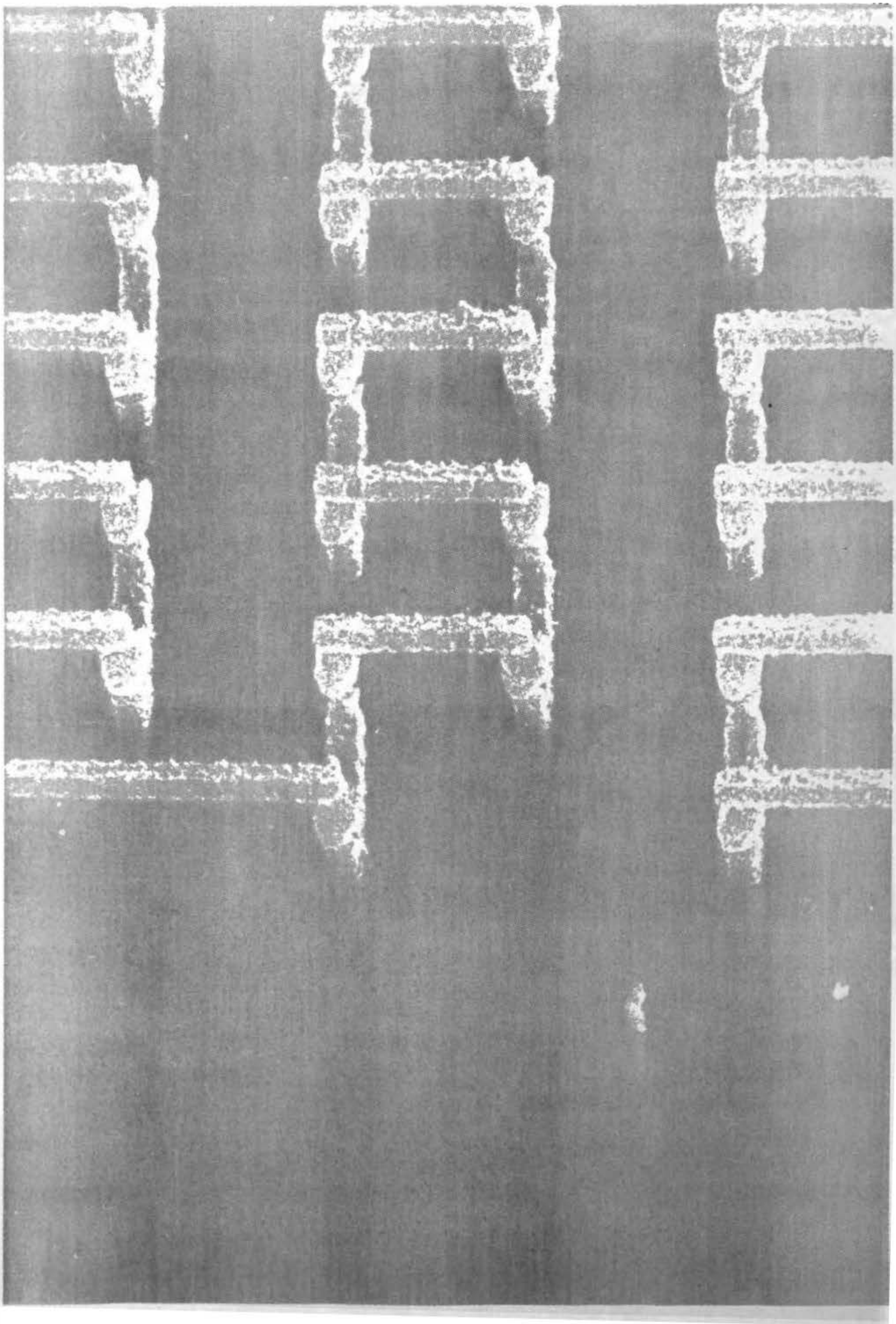


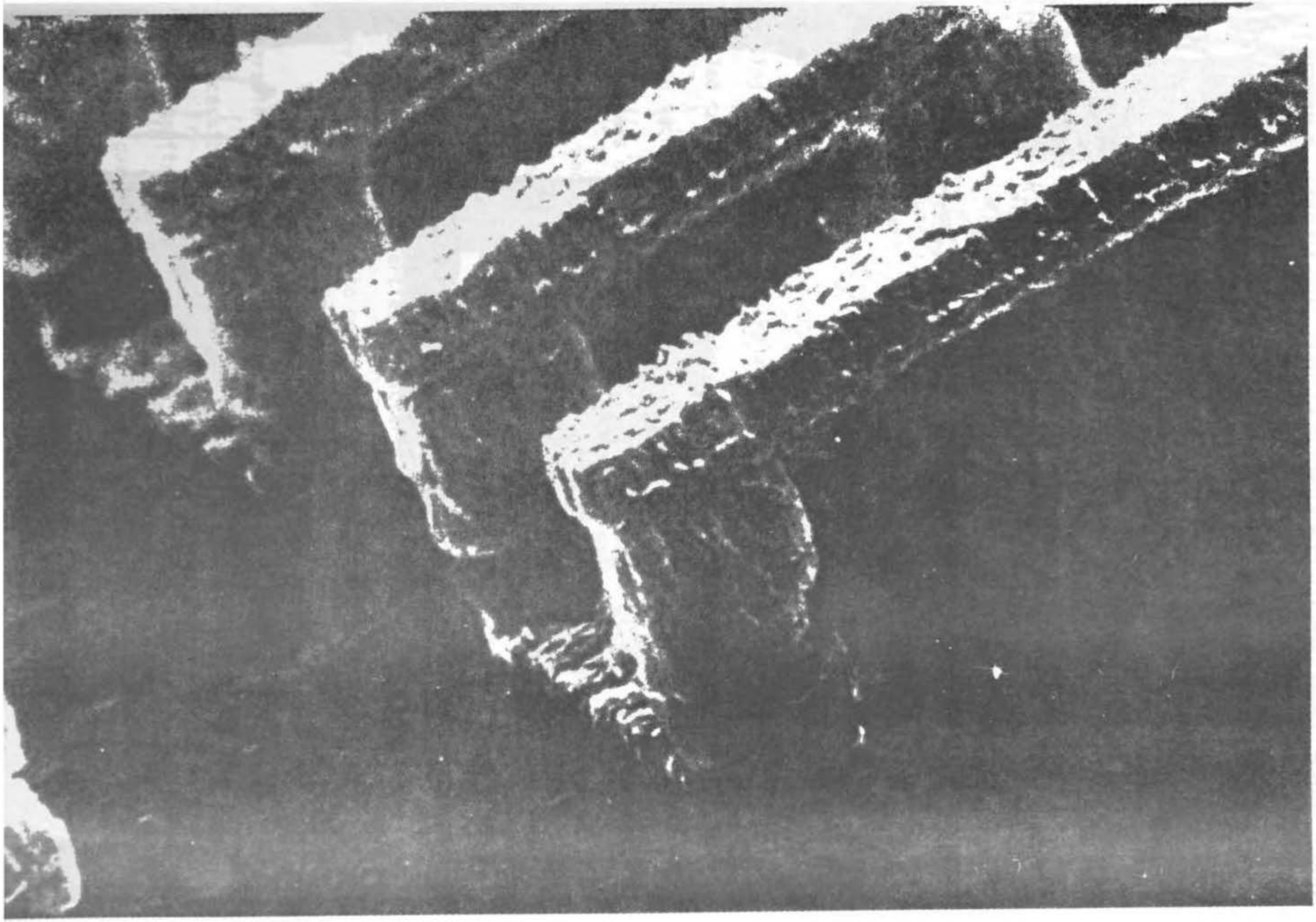
IBM-EF- 47

SLIDE 5

SLIDE 6

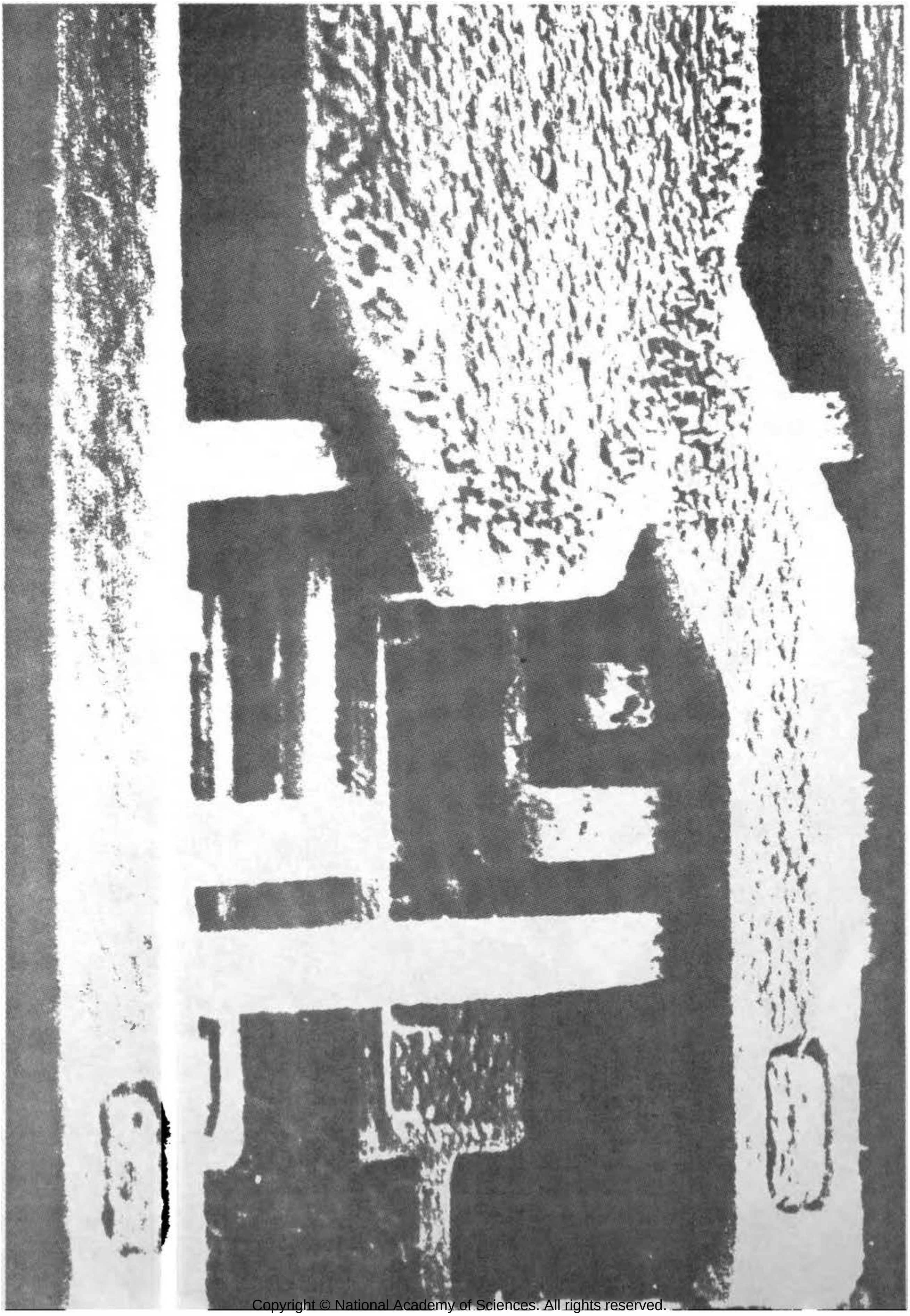
IBM-EF- 47





SLIDE 7

TRM-EF-47



IBM-EF- 47

SLIDE 8

EXAMPLES OF DATA UTILIZATION AND NEEDS IN THE AUTOMOTIVE INDUSTRY

Julius J. Harwood
Research Staff, Ford Motor Company
Dearborn, Michigan

Workshop on "Towards a National S&T Data Policy"
April 14, 1983

It is self evident that the availability of reliable, verifiable and reproducible engineering information is intrinsic to the successful operation of a technologically based industrial enterprise. The needs for a broad and diverse range of data underlie the challenges of high productivity in industrial design, product development and manufacturing operations.

We are coming to understand more and more that in a modern manufacturing organization, a principal activity of people is creating, analyzing, transmitting and managing data and not just the actual transformation of material. Coupled to this are the associated needs for engineering standards and sophisticated techniques and procedures for generating information quickly and with high degrees of accuracy and sensitivity. The spectacular growth in the development and utilization of computer aided technology has underscored the realization that our new strategic resource is information, equal in power to the past effects of capital and physical resources. The power of the computer to integrate and process information has led to the search for better tests and new types of properties characterizations to better simulate real-life service conditions and to more efficient product design methodologies.

This presentation will cite briefly three examples of property data generation and utilization to illustrate the importance of the availability and accessibility of reliable data for automotive product design and development. The first deals with the need for materials property data, particularly cyclic fatigue data, for assuring product reliability and component life prediction capabilities. The second cites the critical need for engineering data characterizing the properties and performance of plastics and fiber reinforced composites to enable their design and utilization as vehicle structural materials. The third illustrates the power of new sophisticated, analytical techniques for rapidly generating data on trace contaminants in exhaust atmospheres and the indispensability of such data for evaluation in emission control procedures and its broad applicability to atmospheric chemistry and gaseous mixtures.

(1) Fatigue Characterization of Engineering Materials

The ability to accurately predict component fatigue performances has become a goal of increasing importance in engineering design practice. Much progress has been made in developing improved fatigue analysis procedures due, in large measure, to detailed studies of the cyclic deformation behavior of materials. Such studies have led to a better understanding of the fatigue process and a much improved quantitative basis for evaluating material fatigue characteristics.

Central to these developments are sets of cyclic properties describing a materials' cyclic deformation behavior and fatigue resistance. The success of modern fatigue prediction techniques has stimulated a number of related and cooperative efforts concerned with specifying standard testing and analysis methods for determining the required materials properties and the establishment of material property data bases to assure the ready availability of such information to designers. Standards also help assure the quality and uniformity of data. In the United States, much of this activity is carried on in the respective fatigue committees of the American Society for Testing and Materials (ASTM) and the Society of Automotive Engineers (SAE). Other societies and government activities also are playing key roles in other technical standard areas.

A past obstacle to the implementation of fatigue analysis procedures has been the lack of an adequate materials property data base. Part of this problem has been overcome in that cyclic properties are now available for a wide range of engineering alloys. Much of this data has been generated by materials users, such as Ford Motor Company and by their materials vendors and suppliers (Fig. 1). We have worked for over 10 years in getting standard procedures established such that vendor companies are now more "comfortable" in providing materials information.

However, no central repository yet exists for this information and it is not as accessible to product designers as we desire. To date, data compilations have generally taken the form of handbooks produced by industrial organizations for their own internal use or by technical societies such as the Annual SAE Handbook. While these have been helpful, this format for large amounts of data can become cumbersome to manage.

Current activities are focussing on the development of computer based data banks and systems (Fig. 2). Together with the Materials Property Council, the National Bureau of Standards and others, we have been promoting the establishment of centralized data banks for mechanical properties. The availability and accessibility of such systems are of tremendous value to users. Properties for particular alloys can be output in the form of data sheets or graphical plots. Materials with specified characteristics can be found through search and sort routines. Materials properties can be accessed by other programs to perform cumulative damage analysis and life predictions. Systems of this type are now under development in a number of organizations. There also is a growing interest in the development of more centralized data banks which would be more extensive in content and readily accessible by the technical community at large.

With the development and availability of a variety of new and powerful design analysis tools has come an increased reliance on analytical design studies for engineering structures (Fig. 3). Mechanical properties of materials play an essential role in such approaches. The coupling of a materials data base with modern structural analysis programs, e.g. finite element analysis, allows design iterations and materials selection decisions to be made early in the design process in a quick and relatively inexpensive way.

Such a computerized fatigue analysis procedure also provides a convenient basis for accounting for the effects on fatigue performance by materials processing effects. The designers can now better anticipate performance problems caused by manufacturing operations, or to enhance fatigue resistance through the judicious choice of materials processing.

With the development of a more sophisticated fatigue analysis capability, there exists an urgent need for better and more extensive cyclic materials property data. Standard testing and analysis procedures for determining cyclic properties appear to be sufficiently well developed to allow for the generation of statistically valid data on a large scale. In the longer term, the more pressing need is for the establishment of automated data bases so that materials information will be readily accessible to engineering design functions. Implementation of these computer aided design/information systems will assist greatly in the development of ever more reliable structures, with high engineering efficiency and cost-effectiveness, and in the optimal selection of materials and processing operations.

We see the need for a coordinated effort involving industrial laboratories, government centers, universities and professional societies for aggressively pushing to establish central data banks appropriately located. It is our understanding that a proposal of this type is being formulated by the Materials Properties Council. Efforts along these lines are moving rapidly in foreign countries and we need to push ahead with our own resources. Somehow the vast amounts of data being generated under government R&D and contracts and in government centers needs to be addressed as to how to better serve these goals.

(2) Characterization of Structural Plastics

Emphasis on higher efficiency, more fuel economical vehicles has focussed attention on the utilization of lightweight materials for vehicle weight reduction. Structural plastics and fiber reinforced composites, as a new engineering class of materials, offer exciting potential for weight reduction opportunities. The Ford graphite fiber experimental car demonstrated that an overall vehicle weight reduction of over 25% was achievable through the use of advanced composite technology, with individual components achieving weight reductions of 50-70%.

However, for these advanced plastics/composites the understanding and data available are in a stage of infancy compared to documentation about metals and alloys. The extensive use of structural plastics in advanced vehicles will occur best with the state of materials characterizations and property data development typical of the information in hand about engineering alloys. The scope of the task cannot be minimized. The wide variety of plastic materials and processing methods in the generic fiber reinforced plastic class of materials makes engineering characterization classification an enormous issue.

Some of the materials characterization information needs are as follows:

(1) Well documented data base about materials properties.

(2) Information on the effect of temperature and environment and time dependent behavior.

(3) Durability and fatigue data.

(4) Energy absorption (crush resistance) data - essential for structural integrity and crash resistance.

(5) Characterization of failure modes.

(6) Control over variability of materials properties.

(7) Statistically reliable data base for efficient design; statistical studies of mechanical property data to assess degree of scatter.

(8) Better correlation between properties of test coupons and behavior of large sheets and components.

(9) Effects of processing variables on mechanical properties.

(10) Effects of composition variations, fiber distribution and fiber orientation on mechanical properties.

(11) Effects of aging on properties and performance.

(12) NDT techniques - quality control procedures for defect identification.

Some of the specific data in these categories can be developed for selected materials by the automotive industry and our associated vendors and suppliers. However, the development of generalized FRP methodologies and the establishment of the desired data base transcends our capabilities and resources. The involvement of university research programs and government laboratories in assisting in transforming these advanced composites into engineering materials classes would be of major benefit. The type of interactions, coordinations and R&D programs which historically underlay the characterization and standardization of engineering alloys and, in more recent times, which made silicon such a thoroughly characterized semi-conductor material, now needs to be applied to the plastic/composite materials classes.

As already indicated in the previous section, the possibility of creating readily accessible computerized data banks, incorporating materials data banks, processing information, and design methodologies would be a major step forward in accelerating the cost-effective, highly efficient design and application of these materials in structural components in future vehicles.

(3) Analytical Data for Emission Control Systems

The challenge of developing reliable and effective auto emissions control systems surfaced the need for a laboratory technique capable of providing detailed compositional analysis of tail pipe emissions. The technique needed to be quantitative over an extended range of concentrations; adequately differentiate among individual molecular species; be sensitive to sub-parts-per-million concentrations, even in the presence of huge amounts of water vapor and carbon dioxide; and be rapid in generating the required chemical spectra. The best available technique, prior to the '70s, was infrared absorption spectroscopy, but this had the unfortunate characteristic of requiring prohibitive times for acquisition of the requisite spectra. Over 25 hours of data recording was needed to obtain a single record.

Fortunately in the mid '70s Fourier Transform Infrared Spectroscopy came about. This new technique afforded a speed enhancement over 1000X greater than conventional spectroscopic methods, computerized the data and yielded spectra of outstanding wavelength precision and excellent reproducibility. The desired infrared spectrum could now be obtained in under a minute's time. At this point there arose the need for "scientific data". The spectra now available could be processed in such a way as to represent the sum of the spectra of the molecular species making up the subject gas, each in proportion to the amount of species present in the sample.

All that was needed to round out the method was a library of reference spectra. This library would consist of spectra of pure compounds diluted by pure synthetic air to concentrations comparable to those encountered in the exhaust analysis. These reference spectra could, in fact, qualify as "standards", in the sense that once recorded they would stand forever. To satisfy the requirements of the exhaust analysis work perhaps 50 such reference spectra would be required.

Over the years, such a library has been established by Ford Research. It services the needs of our basic research atmospheric chemistry groups as well as the emissions analysis work and spans a diverse range of chemical compounds. At present, it includes some 250 spectra of about 225 molecules, each run under standardized conditions and calibrated on average to better than 10%. Those spectra of critical importance are known to better than 1%. Each record is stored on a computer disc and backed up in a magnetic tape archive.

Figure 4 illustrates the nature of the spectral data provided by the FTIR method. The lower trace shows the complete spectrum of a CVS bag sample. Each of the myriad features here distinguishable can be traced to a specific molecule and identified in the corresponding reference spectrum. The middle trace shows on an expanded scale one segment of the total spectrum, with several prominent features labeled by the molecular species causing it - i.e. methane (CH4 - 11 ppm), nitrogen dioxide (NO2 - 47 ppm) and formaldehyde

(CH2O - 2.3 ppm). The upper trace shows another segment with expanded vertical scale, and identifies features due to ethylene (C2H4 - 6.7 ppm), propylene (C3H8 - 3.6 ppm), isobutylene (IC4H8 - 1.5 ppm) and nitrous acid (HONO - 4.9 ppm).

The power of this technique in terms of sensitivity, speed, accuracy and reliability, coupled with the available library of standard spectra have led to its widespread use in Ford Motor Company as an on-line testing system for evaluating the effectiveness and reliability of emissions control systems. With efficacious sampling methods, it has also been used for detecting trace amounts of contamination in many other atmospheric environments. Too, it has proven to be a tool of exceptional value in studies of the chemistry of the atmosphere which have lead to the identification of numerous reaction pathways, intermediate chemical species, and even to the identification of new, stable atmospheric constituents.

(3) Use of Mechanical Properties

Mechanical properties find a major use in the assurance of component and structural reliability. Some of the material considerations facing a design engineer are indicated in Fig. 5. Based on the envisioned mechanical environment, the designer must first consider possible failure modes in his structural design. Mechanical property information then provides an indication of the resistance of various materials to these identified failure modes. Environmental and processing influences are generally assessed through their effect on mechanical properties.

In Fig. 6 are listed the more common materials tests for which standard procedures exist or are being developed. Also shown are typical properties used to describe the elastic, flow and fracture behavior of a material. Large amounts of data from such tests now exist for virtually all classes of engineering alloys. However, this information is often not readily accessible. The designer is thus put in the difficult position of making engineering judgments with inadequate information.

An important factor impacting mechanical property usage is the rapid development of computer-based design and material selection aids. These include routines for structural analysis, mechanical property and fatigue life prediction, and modeling of component processing sequences. Nearly all of these programs require some type of material property input (Fig.7). This serves to point out the desirability of having an on-line data base. In fact, such a data base is essential to the effective utilization of these aids in the design process.

Of greater value to the design engineer is the ability to access mechanical properties through integrated design analysis programs. Such programs are indicative of a growing trend toward the use of analytical design studies in component development. These procedures utilize a finite element model which can be subjected to various service loading conditions in order to assess likely component performance prior to the building of prototypes. Mechanical properties are used both in the structural analyis to determine local stresses and strains and in the fatigue analysis to determine the resistance of a particular material to the imposed loading. After construction of the problem solution components, sensitivity analyses can be quickly performed through systematic variation of geometry, material and loading. Such design procedures hold great potential for developing better engineered structures in a shorter period of time.

Test standardization will take on even greater importance. The goal of such activities must be to more carefully specify test procedures and clearly define the properties derived from test data. This should be done in the context of automated testing with full utilization of the computer's capability to insure conformity to standards. Standard tests will undoubtedly become available on cassettes or discs for direct input into automated systems. The test engineer will then be led through the procedure by responding to keyboard queries and commands. Properties will be determined by the computer in a specified, unbiased way. A great deal of effort will obviously be required to achieve these goals as quickly as possible.

The importance of generating properties that have demonstrated design significance cannot be over emphasized. Skepticism persists in many circles concerning the use of information gained from simple laboratory tests in assessing component performance in service environments. This stems largely from our long term reliance on "easily determined" properties such as yield strength, ultimate strength, elongation, impact energy and endurance limit. These properties were developed in response to problems of another time; as such, they often offer little guidance for the solution of current problems. There is ample evidence from recent efforts in deformation and fracture research to show that properties derived from tests designed to closely simulate service loading and environment can provide valuable design information. The development of better tests and more useful properties must be a prime concern of materials information efforts.

Examples of future trends include the increased use of true stress-strain properties, particularly in structural analysis programs; for fatigue problems, cyclic stress-strain properties will be used. The multiaxial deformation behavior of materials needs to be more thoroughly quantified in support of this application. Increased use will also be made of finite life fatigue properties as well as fracture mechanics properties for handling problems in subcritical crack growth. A continuing need exists to study environmental effects, e.g. temperature and corrosion, on fatigue and fracture behavior and to develop appropriate design procedures.

Material Characterization Sheet

Sheet 1 Of 2

Material SAE 5160

Condition Quenched & Tempered

Hardness 440 HB Other Specs.

Material Source Vendor

Specimen Orientation Longitudinal

Composition 0.61C, 0.86Mn, 0.02P, 0.02S, 0.31Si, 0.76Cr

Reference J. T. Crandall Date 3/77

Microstructure:

Grain Size

Tempered martensite

Comments

Austenitized: 1650°F/30 min.
Agitated oil quench
Tempered: 825°F/1 hr.

Monotonic Properties

Mod. of Elast., E	205	GPa (	29.5	x 10³ ksi)
Yield Strength, 0.2% S_y	1450	MPa (	210	ksi)
Ultimate Strength, S_u	1550	MPa (	225	ksi)
Strength Coeff., K	1895	MPa (	275	ksi)
Strain Hard. Exp., n	0.046			
Red. in Area, % RA	43			
True Frac. Strength, σ_f	1930	MPa (	280	ksi)
True Frac. Ductility, ϵ_f	0.56			

Cyclic Properties

					Correlation Coeff., r
Yield Strength, 0.2% S_y'	1070	MPa (	155	ksi)	
Strength Coeff., K'	2000	MPa (	290	ksi)	0.92
Strain Hard. Exp., n'	0.10				
Fatigue Strength Coeff., σ_f'	2050	MPa (	297	ksi)	-0.94
Fatigue Strength Exp., b	-0.08				
Fatigue Ductility Coeff., ϵ_f'	1.24				-0.95
Fatigue Ductility Exp., c	-0.79				

159-512

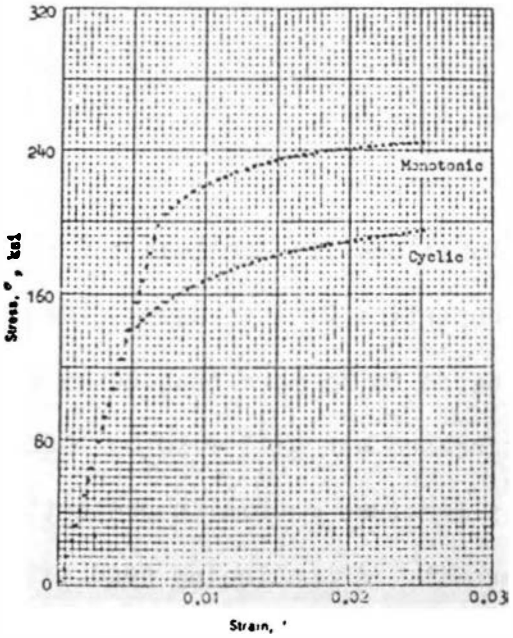
Sheet 2 Of 2

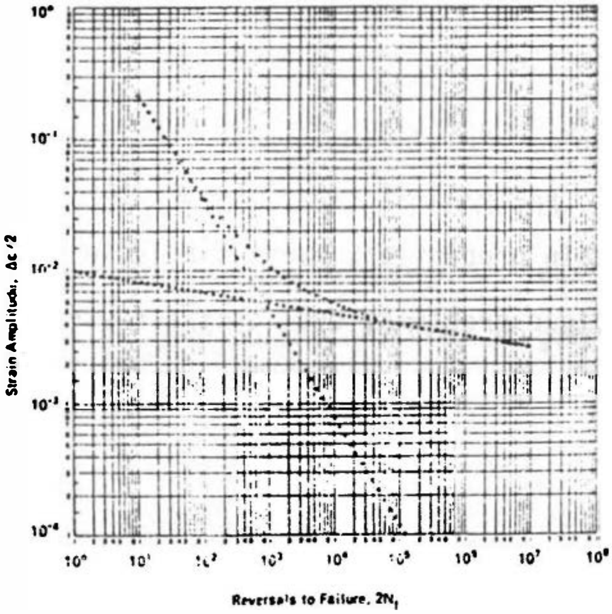
Material SAE 5160

Condition Quenched & Tempered

Hardness 440 HB Other Specs.

Material Source Vendor





159-51201

Fig. 1

COMPUTER-BASED DESIGN/MATERIAL SELECTION AIDS

- Finite Element Analysis
- Service Data Acquisition/Analysis
- Fatigue Life Prediction (Metals)
- Hardenability Prediction (Steels)
- Surface Processing Optimization (Steels)
- Formability Modeling (Sheet metal)
- Mechanical Property Prediction (Composites)

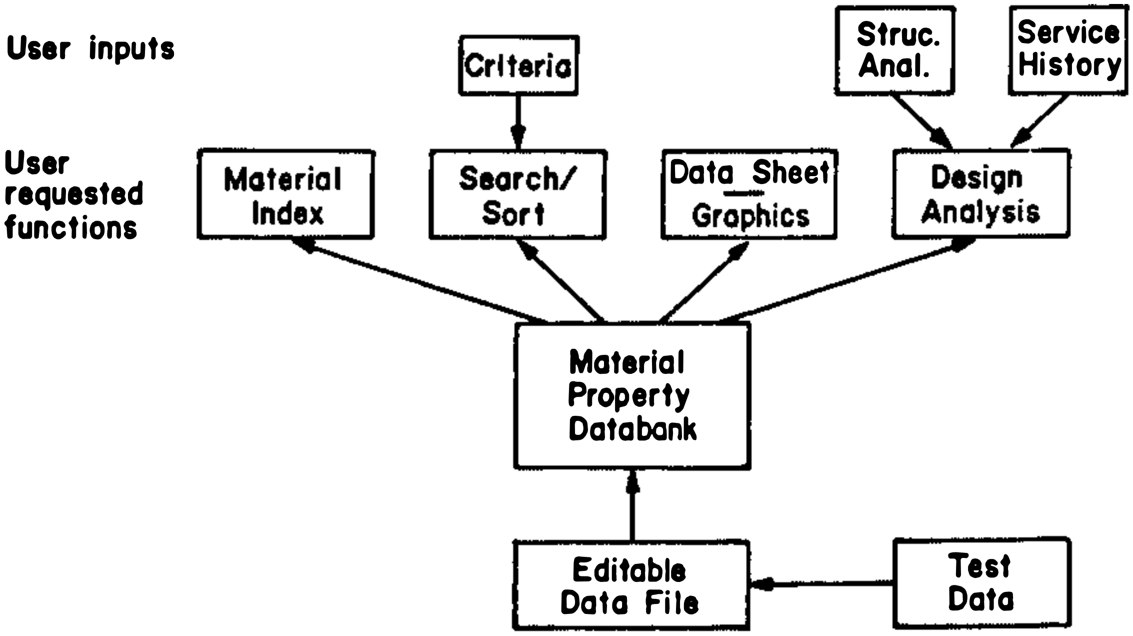


Fig. 2

ANALYTICAL DESIGN

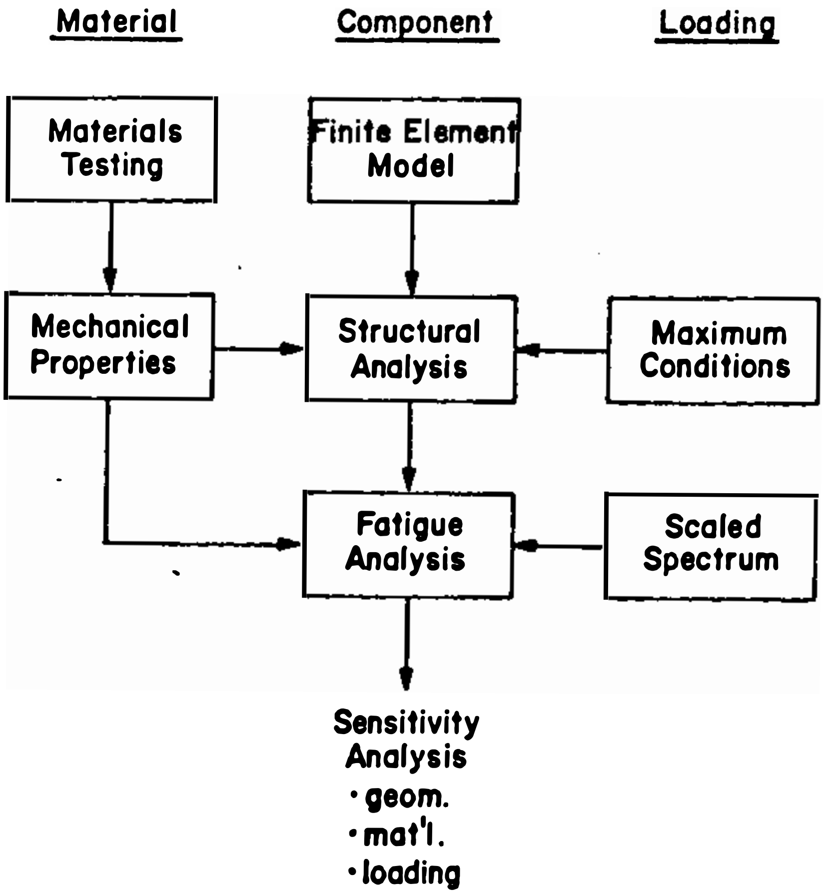


Fig. 3

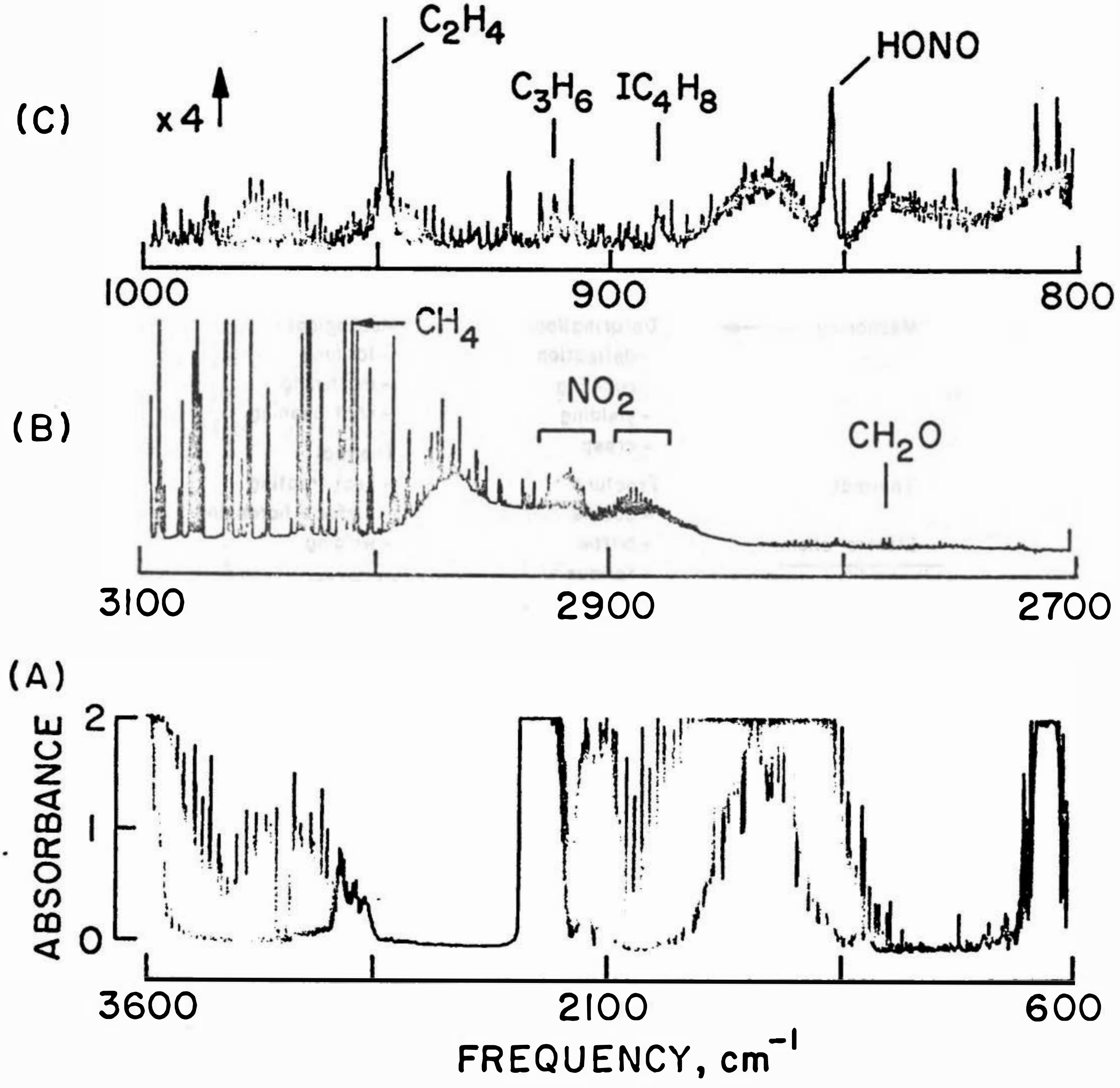

(C)
x4
C2H4
C3H6
IC4H8
HONO
1000
900
800
(B)
CH4
NO2
CH2O
3100
2900
2700
(A)
ABSORBANCE
2
1
0
3600
2100
600
FREQUENCY, cm-1

Fig. 4

MECHANICAL RELIABILITY - MATERIAL CONSIDERATIONS

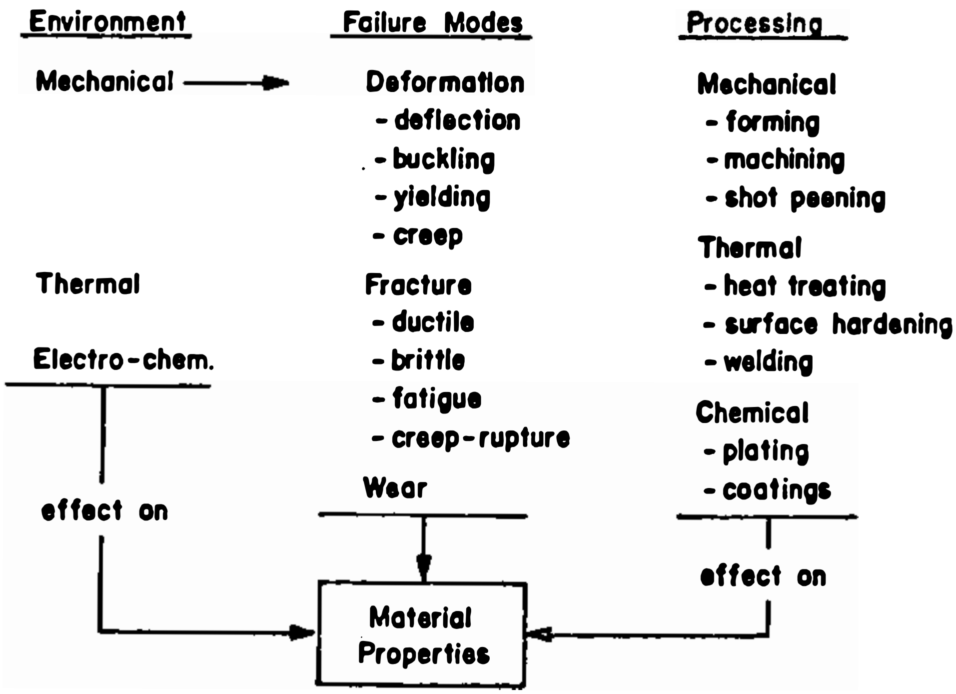


Fig. 5

MATERIALS TESTS

Monotonic

- tension
- compression
- hardness
- impact
- frac. toughness

Static

- creep
- relaxation
- stress-rupture

Cyclic

- damping
- cyclic stress-strain
- fatigue
- crack growth

PROPERTIES

Elastic

- modulus
- Poisson's ratio

Flow

- yield strength
- ultimate strength
- strain hard. rate
- hardness
- creep rate

Fracture

- frac. strength/ductility
- toughness
- strain-life
- crack growth rate

Fig. 6

COMPUTER-BASED FATIGUE ANALYSIS PROCEDURE

Inputs	Operations	Outputs
1. Material Properties - stress-strain - strain-life	1. Local stress-strain analysis	1. Stress-strain response
2. Component Analysis - load-strain relation	2. Damage analysis	2. Damage/event
3. Service History - load (strain) vs time		3. Total damage
		4. Life prediction

Fig. 7

Dr. John P. D. Wilkinson
General Electric Company
Towards a National S&T Data Policy
April 14, 1983

Data Needs of the Electrical Manufacturing Industry

As a diversified, high technology company engaged in fields as varied as medical diagnostics, materials synthesis, and factory automation, General Electric has great need for physical data of virtually every kind. Such information is used in product design, process development, and instrument calibration to name only a few of the functional areas. Some of these data can and must be generated in our own laboratories, but for a considerable amount of what we require, we do turn to existing compilations and data centers. Consequently we are, and have been, concerned with building and maintaining national strengths in these areas. I am pleased to have this opportunity to describe more fully our present data activities and needs and to recommend some actions which we believe would serve our industry and enhance the national capability.

Present General Electric Involvement in Data Programs

General Electric scientists have initiated several data projects which are now a part of national data programs and also have prepared review articles and books in this field. Among these are projects in infrared spectroscopy, heat transfer and fluid flow, properties of superconducting materials, and phase diagrams in metallic systems. Our publications include such titles as "Critical Surveys of Data Sources," "Data Sources for Materials Scientists and Engineers" and "International Metallic Materials Cross-reference." One member of our Corporate Research and Development staff presently serves as Chairman of the U.S. National CODATA Committee and also Chairman of CODATA's Advisory Panel on Data for Industry. Other General Electric scientists and engineers have served on the Numerical Data Advisory Board of the National Academy of Sciences, on the Committee of the National Bureau of Standards/American Society of Metals Binary Metallic Phase Diagram project and on various panels of the NBS Visiting Committee. General Electric was one of the first industrial supporters of the Metal Properties Council. Indeed, the late Dr. Arthur Bueche, our former Senior Vice President, Corporate Technology, recently served as Chairman of its Board of Directors.

Sources of Data

Broadly speaking, General Electric businesses look to three broad categories of data to fulfill their needs:

- Data generated by internal laboratories — for specialized data which could not be found elsewhere and/or which are highly critical to product performance; once in hand these data are considered proprietary, e.g. high temperature mechanical properties of superalloys for aircraft jet engines.

• Data generated from many sources, but centralized by General Electric information centers or services	— for engineering and scientific data of interest to several General Electric businesses; little of this data is considered proprietary but the comprehensiveness, the validation, and the existence of the system itself is felt to be an important business advantage, e.g. electrical conductivity values for copper and aluminum.
• Data from public sources (handbooks, review articles, computerized data-bases, etc.)	— for fundamental technical data, experimental or calculated, of broad interest and applicability which we could not justify measuring or calculating for ourselves, e.g., thermodynamic data, crystal structures, atomic cross-sections.

Perhaps the largest single data activity in the second category above is done by our Materials Information Services component of Corporate Research and Development. A major data service which it provides is called EMPIS, an acronym for Engineering Materials and Properties Information Service. EMPIS was established at the Corporate level in General Electric more than fifty years ago to provide our diverse manufacturing businesses with authoritative information on the characteristics, behavior, identification, availability, and methods of test of a broad range of raw and semifinished materials, commercial hardware (mostly fasteners) and applied finishes. Our General Electric manufacturing operations financially support this service at about the $1M per year level. They are under no compulsion to do so—they find the service to be useful. EMPIS data is contained in 30,000 pages of handbooks covering 12,000 different materials, and we are in the process of placing the information on a computerized data base. Because of its comprehensiveness, EMPIS is proprietary to the General Electric Company. The multifaceted organizational scheme of EMPIS facilitates the usage of the information it contains for materials selection, product design, manufacturing planning, material identification and control, procurement standardization, and communication. We literally could not do business without EMPIS; but rich and reliable as it is, it must be supplemented with data from the other two categories described above.

Examples of the publicly accessible data compilations which we regard as virtually indispensable include:

- the Joint Army, Navy, Air Force (JANAF) tables for thermodynamic data
- the Joint Committee on Powder Diffraction Standards (JCPDS) data base
- the Journal of Chemical and Physical Reference Data
- the various publications of the Committee on Data of the International Council of Scientific Unions (CODATA)

- various publications of the National Bureau of Standards
- the Metals Handbook of the American Society for Metals

Selected General Electric Data Usages—Past and Present

Review of a few selected examples may serve to illustrate a couple of general observations regarding data needs and applications:

- Even a single technical project requires many different types of data to succeed.
- Data, particularly those of a fundamental sort, may have applications quite different than those motivating their original generation.

As examples, I have selected three projects that have been carried out at General Electric's Corporate Research and Development Center. The projects illustrate the need for data at environmental extremes of pressure or temperature.

High Pressure/High Temperature Synthesis of Diamond and Borazon

General Electric entered upon this development because of its own need for industrial diamonds for grinding and polishing cemented carbide, and because of the recognition of the United States dependence on foreign sources of supply. The successful completion of the development in the mid-1950's not only solved the immediate need, and gave General Electric a whole new business, but also led to new complex materials of highly controlled structure and properties, suitable for new applications, and with which the natural diamond is unable to compete. A review of some of the data needs for this area can be structured as follows:

Required Data	Source
Thermodynamic properties of products and reactants as a function of temperature and pressure	JANAF tables
Temperature-pressure calibration standards	NBS publications
Diffusion rates	Various printed compilations of diffusion data
Phase relations and crystal structure stability as functions of temperature and pressure	Various printed compilations of evaluated phase diagrams
Electrical, thermal and mechanical properties of gasketing materials	In-house measurements
Strength and toughness of die materials	In-house measurements

High Pressure Discharge Lamps

Lamps of this kind presently constitute a business of the order of \$200M annual sales for General Electric. Their attraction as a lighting source stems from their high efficiency, their compact size, and their long life. They light many of our nation's cities. Although great progress has already been made with our Lucalox and metal halide lamps, we believe that there is potential for further increases in efficiency, prompting our continued search for improved understanding of the chemistry and physics of lamps. Again, examples of data required in these investigations are as follows:

Required Data	Source
Thermodynamic data	JANAF tables
Oscillator strengths for atomic transitions	NBS-NSRDS publications (DARPA sponsorship)
Stark broadening parameters	Articles in the Journal of Chemical and Physical Reference Data
Electron-atom and atom-atom cross-sections	NASA report on estimated viscosities and thermal conductivities of gases at high temperatures (originally rocket studies, but useful here)
Color calibration data	NBS reports

Superconducting Generator

This year, General Electric conducted a load test of an experimental superconducting electric generator—the first of its type to have been successfully tested anywhere in the world. At full load it generates 20 MVA of electricity—enough for a community of 20,000 people. This is twice as much electricity as could be produced by a conventional generator of comparable size. The rotor is cooled to 452°F below zero and has windings fabricated from filaments of a niobium–titanium alloy in a copper matrix. This is a superconducting material that has near zero electrical resistance, which accounts for the machine's small size. In addition, the low electrical resistance leads to lower losses than the conventional machines, which potentially could result in significantly lower operating costs. The machine is an experimental demonstration of a new technology, and may set the stage for future generations of advanced machines. The types of data required for the design can again be viewed in the three categories as before:

Required Data	Source
Thermal properties of cryogens	NBS
Properties of structural materials at cryogenic temperatures	ARPA/NBS funded Battelle compilation

Required Data (continued)	Source (continued)
Electrical properties of dielectrics	Internally developed
Mechanical properties of structural composites	Internally generated
Magnetic properties of materials	GE materials handbook
Superconducting properties—current density, temperature, etc.	Vendor supplied

Comments on Data Requirements

I've spoken about some representative data in industry. Now I'd like to make some comments on the payoff of having high quality data available in an easily accessible form.

From a national point of view, I believe that high quality, easily accessible data on basic material properties and phenomena will allow us to remain competitive on a worldwide basis, will allow products to be designed, built and tested in a more economical and timely manner, and, in the last analysis, will enhance our national productivity. What is needed in terms of data to do this?

Quality Data for Computer Modeling of Phenomena

Computer models are being used more and more to replace a multitude of physical tests on product prototypes. Since it is less expensive by up to a factor of one hundred, the use of the computer models instead of testing can yield significant cost savings. For the same reason, computer modeling allows a systematic inspection of a great variety of design options which otherwise might not be tested, resulting in a final product that is far closer to optimal. Finally, because it's also faster to run a model on the computer, such modeling cuts down the elapsed time from design to prototype. Only a few key tests are needed to finally verify the effectiveness of the prototype. Computer models depend on both a mathematical description of the phenomena, and accurate physical data to input into the model. Thus, the effective exploitation of this powerful technique requires access to reliable data which has been well evaluated—what we might call quality data.

Computerizing Data

While many data sources themselves have in fact been computerized—that is, placed in such a form as to be accessible by computer—this is by no means true across the data spectrum. Thus, a continued effort needs to be made to not only generate new data, but to place it and existing data on computer-accessible media. This effort will support data users in their computer modeling activities I mentioned earlier.

Data Networks

The computer and computer networking provides a revolutionary asset in handling, evaluating, and using data from many sources. While good examples exist of computerized data networks, it seems well worth while to aim at establishing national data networks over a larger range of data. The networks need to have user-friendly interfaces so that the user at the terminal will have a simple means of accessing data sources.

Specific Recommendations

In view of the importance of establishing a solid data base, I'd like to make two specific recommendations:

- That the National Laboratories and the National Science Foundation and Mission Agencies join together to meet the national need for high quality, easily accessible data, and continue to allocate an appropriate portion of their funding to projects which will compile, evaluate, and efficiently disseminate existing data.

- That the data and standards activities at the National Bureau of Standards be strengthened. In particular, that continued additional funding be made available for:

    - Data compilation and evaluation projects.

    - Cataloging and publicizing sources of data.

    - Initiating national efforts toward cooperation in developing user-friendly computer access to data compilations.

TOWARDS A NATIONAL SCIENCE & TECHNOLOGY DATA POLICY

- The Chemical Industry -

D. W. H. Roth, Jr., Allied Corporation, Morristown, NJ

In outline of my presentation distributed to you regarding the numerical data needs and requirements of the Chemical Industry, I had more broadly defined the data universe involving chemical compounds -- including their manifold derivative products or involvement in energy conversion processes. The thermodynamic and transport properties of chemical substances and their mixtures include:

- Enthalpy and Entropy of Formation, combustion and phase transformation, including heat capacity latent heats.
- Equations of State (PVT) Pressure, Volume & Temperature relationships of compounds, and this includes critical properties, vapor pressures, volumetric properties, and compressibility factors.
- Phase Equilibria, including vapor-liquid equilibria, liquid-liquid equilibria, solid-liquid equilibria.
- Chemical Equilibria, Gibbs free energy.
- Properties of Electrolytes, including mixtures with non-electrolytes.
- Chemical Kinetics, including activation energies.
- Transport Properties, including viscosity, thermal conductivity, and diffusion coefficients.

These properties influence the design of virtually every process in the metallurgical, petroleum, plastic, man-made fiber, pharmaceutical and paper industries - plus industries involving combustion processes such as the aircraft, automotive and power industries, and all processes requiring pollution abatement control systems.

The Chemical Industry has other data needs in order to design, construct, and operate economic, safe, and environmentally sound facilities. These are mechanical and corrosion resistant properties of materials; however, these needs have been disucssed by Julius Harwood or will be addressed by others during this workshop. This is a world and universe in motion. Dynamics are what must be understood in order to design and build any practical operating system. But before we can measure the effectiveness of any practical system, or improve its performance, we must have a sound understanding of the maximum static or equilibrium forces which will ultimately provide the driving forces for our systems in motion. Fortunately, out of a seemingly chaotic microworld of molecular activity - a sound macroworld understanding of molecular activity is found in the thermodynamic laws governing all systems at equilibrium. These laws provide the basis for consistency tests of experimental chemical property measurement, as well as inherent consistency in any empirical correlation equations or methods which have been or will be derived.

As a practitioner of chemical engineering, I have designed or have been in responsible charge of, over $2 billion of chemical, polymer, or fiber plant systems and facilities. It has been my observation that of all of the factors affecting less than optimal performance of these facilities - the most prominent has been the lack of sound chemical compound thermodynamic and transport properties. Or, where not available, the lack of sound predictive methods. The numerical data needs of the Chemical Industry is becoming more acute, not because a small but steady stream of new data and correlation methods continue to flow, but because the computational power at the beck and call of every design engineer allows him to literally examine hundreds of design options, whereas 10 years ago, the engineer might be limited to two or three. Truly optimal systems can now be designed based on simulation methods such as "ASPEN", developed at MIT under DOE contract, but its universal applicability has been restricted by the limitations of the Chemical Industries' data base. We do not refer to sheer volume of data - that fills volumes of books or computer tapes, but critically evaluated data and predictive methods with their limitations clearly delineated. This had led me, and others to promote Industry/Government/Professinal Institute cooperative efforts on a world-wide basis to attempt to move data efforts at the required rate. I was involved with the American Institute of Chemical Engineers along with OSRD of NBS in setting up a Design Institute for Physical Property Data (DIPPR) under AIChE, and am Chairman of that Organization. This organization consists of 50 operating companies, engineering firms, research institutes, and federal agencies, including NBS OSRD - with an annual budget of about $500,000. Other U.S. cooperative data efforts include:

- American Petroleum Institute (API) RP 44
- Gas Processors Association (GPA)
- Fluid Properties Research Institute (FPRI)

In the structural properties of materials area, I am involved with the Metal Properties Council, Inc. - sponsored by ASME, ASTM, ASM, AWS, and the Engineering Foundation - with an annual budget of $3 MM, and nine major subcommittees consisting of over 600 technical experts from just about every major manufacturing company. Other cooperative efforts in the properties of materials include:

- The Materials Technnology Institute of the Chemical Process Industries (MTI).
- ASM/NBS phase diagram program.

Plus others.

Data work is not glamorous, as Steve Berry pointed out, and it does not draw the attention, both in dollar support nor the interest of scientists in making a career in this area. The entire U.S. chemical/petrolum/gas industry is served by a handful, maybe 50-60 industrial thermodynamicists, who many times serve simultaneously on committees in DIPPR, API, AND GPA. These people are good, very good, and they work hard - but just to put their efforts in perspective in the Chemical Industry alone - on sales approximately $150 billion - and new facilities being built or refurbished to the "tune" of about $10 billion/year - and with operating costs of about $125 billion/year - the following figures can be generated.

I would anticipate, based on my own, and others with similar years of experience and responsbility in this field, that a conservative 5% could be saved on capital costs and 5% on operating costs if the data needs of the Chemical Industry were brought to a level of completeness commensurate with the advancement of other design tools.

Thats about $6 billion/year, and I believe the figure is more like 10% - or approaching $10 billion/year! The Codan National Academy of Science/National Research Council (NAS/NRC Committee on data needs) study estimated $7 MM for organized physical & chemical data activities in 1977 - or say about $10 million today - almost insignificant compared to potential benefits.

The U.S. economy could easily justify data efforts in the $100 million range in order to have a first rate economy in the 1990 s. This effort would require a buildup of people and facilities in order to accomplish a ten-fold increase, and would require a great deal of thought and planning on the part of existing industrial/academic/governmental data agencies. Europe and Japan are mounting significant data efforts. We need a firm U.S. policy to stimulate data activity on a national basis and to provide the planning structure for an increased U.S. data effort.

Contribution from the floor on Cost of Failures

William F. Brown, Jr.
Consultant, NASA

$\sigma_p = 1.1\,\sigma_{op}$ $\quad$ $S.F._w = 0.9\,\frac{\sigma_{ys}}{\sigma_{op}} = 1.3$ $\quad$ BURST AT 62% OF σ_{op}

PROOF TEST FAILURE OF THE NASA 260 in. DIAMETER MOTOR CASE

The proof test failure of this large solid propellant rocket motor case resulted in the loss of $17,000,000 and the cancellation of that portion of the program. The motor case was fabricated from a relatively new steel using improper welding methods. The information that these methods were unsuitable was at the time not widely disseminated but available in the Aerospace Structural Metals Handbook sponsored by the Air Force. The cost of preparing the Handbook Chapter for this steel was about $3000.00 or 0.02 percent of the loss.

COST OF SOME FAILURES TRACEABLE TO INADEQUATE DATA BASES

The following list gives some examples of the enormous costs which may be encountered in hardware failures either because the appropriate materials property data were not available or were not readily accessible to the designer who seldom has time within his program schedule to make a comprehensive literature survey.

1) F-111 -- the use of high strength, low toughness, flaw-sensitive steel for the monolithic carry through structure resulted in failures which led to the additional program delays and costs of $150,000,000 to $195,000,000.

2) NASA 260 inch motor -- the use of improper welding processes and inadequate nondestructive test methods resulted in a hydrotest failure of a motor case and a loss of $17,000,000.

3) NASA SPS Tank — the presence of machine tool marks combined with an unexpectedly aggressive influence of methanol on Titanium 6-4 alloy resulted in the destruction of Apollo Service Module No. 17 with a loss of $10,000,000.

4) LM Program Tanks -- undetected, subsurface metallurgical defects in Ti tanks, stress corrosion in an aluminum alloy cover, and inadequate welding techniques and inspection methods resulted in failures which cost NASA $4,700,000 and caused the rejection of the remaining tanks due to reliability concerns.

5) Military Aircraft -- a comprehensive study has shown that in the period of 1962-1970, 1 percent of all noncombat fighter losses were caused by the unexpected failure of a primary structure. Including bombers and tankers, there were 42 wing failures and 16 fuselage/tail failures, some of which resulted in the loss of the entire airplane: Or viewed another way, the USAF has had 16 major structural crises on 15 aircraft types in recent years with catastrophic failures occurring in 10 of them. Twelve of these crises occurred in the last five years. The cost of these failures plus the required correction is beyond estimating abilities, but cannot be below $1,000,000,000.

Source of the above information: Report of Ad Hoc Panel on Fracture Control, NASA Research and Technology Advisory Committee on Materials and Structures, Oct. 1971.

Statement of Dr. James V. Taranik
Dean, Mackay School of Mines
for
Workshop on "Towards a National Science and
Technology Data Policy"

Committee of Science and Technology
U. S. House of Representatives

Numerical Data Advisory Board
National Academy of Sciences

Civil Land Remote Sensing Satellite Data

Background

During the Carter administration a Presidential Directive was issued which transferred the National Land Satellite (LANDSAT) Program from NASA to NOAA. The Department of Commerce established a timetable for eventual transfer of elements of the program to private industry over a 15 year period. Soon after the present administration was established in 1980 the Office of Management and Budget recommended termination of the Government's role in the Land Satellite Program beyond LANDSAT-D' (LS-5), probably in 1988, if not sooner. This recommendation was made to force the commercialization (privatization) to occur more rapidly than planned by the previous administration. In 1982 the Department of Commerce organized government and private sector advisory groups to evaluate mechanisms for the transfer of civil land remote sensing systems to industry. Based on their recommendations in 1982 the Cabinet Council on Commerce and Trade met on December 15, 1982 and evaluated two options:

> Option 1: Transfer to the private sector, by competitive means, the current operational civil remote sensing satellites. Separate bids would be accepted for the land or weather satellites, or a firm could elect to submit a bid for all systems.
>
> Option 2: Continue the current budget policy of bringing the operational land remote sensing systems in the government to a close nominally by 1988 or sooner and retain the civil weather satellites under government control.

On February 28, 1983 the Cabinet Council on Commerce and Trade recommended Option 1 to the President, who concurred in the recommendation on March 8, 1983.

Interagency governmental groups are now preparing a pre-proposal release to inform industry of the Government's intent to solicit proposals. This pre-proposal package should be released mid-May with comments due back from industry by early June. A tentative schedule calls for formal request for proposal to be released in early July and for proposals to be submitted from industry to the Government by early October. Draft legislation would be submitted to Congress in early November and the expectation is that it would be passed before the end of the year. The Government expects to select an industrial entity by January 1984 and complete the transfer by March 1984.

Problem Statement - LANDSAT Data Continuity

LANDSAT-4 was launched in July 1982 and a small amount of data has now been transmitted to earth and processed to images in photographic and computer compatible tape form. The direct data transmission link to the ground is no longer operable and the satellite has a power problem. LANDSAT-4 has a new, advanced sensor called The Thematic Mapper which has produced spectacular data of great value to the natural resource scientific community. In all probablilty LANDSAT D' (LS-5) will have to be launched before the end of 1985 to insure that there is no loss in the opportunity for world-wide coverage with the Thematic Mapper. If LANDSAT-D' is launched early, then the possibility exists that unless Congress appropriates funds in this session for LANDSAT-D" (LS-6), a break in Thematic Mapper data continuity will occur beginning 1987. It seems doubtful that an industry transfer can be arranged in the short time of a year and probably it will take more like two years. Industry is not prepared to fund development of another Thematic Mapper. A similar solid state linear array system cannot be developed until the problem of detectors for the short-wave infrared is solved through basic research. If a Thematic Mapper or similar sensor is to be flown in 1987 or 1988 design and development must begin in FY 1984 to permit satellite integration in 1986 or 1987 and launch a year later.

We have not collected a global seasonal multispectral data set with the Multispectral Scanner (MSS) in 10 years of near continuous operation on LANDSATS 1, 2, and 3. A similar Thematic Mapper data set is needed so optimal seasonal conditions for geological interpretation can be selected. It is also needed to establish an environmental baseline for analysis of changes in the environment which may affect global habitability of planet Earth.

Problem Statement-Costs for LANDSAT Data

Currently scientific investigators face rapidly rising costs for LANDSAT data as the Government prepares to transfer the land system to industry. In 1978 LANDSAT computer data cost $200 per tape set, in 1983 the cost is $2500 per tape set, and by 1985 the Government plans to raise the cost to over $5000 per tape set. Scientific investigators in academia face decreasing Government funds for land remote sensing research, and increasing indirect costs. These trends may significantly reduce important fundamental scientific research nationwide.

Problem Statement-Funding for Basic Research in Land Remote Sensing

Currently funding for basic research with new sensor technology and using space shuttle is declining. NASA will fly the $6.3 million dollar Large Format Camera on STS-17 in 1984 but no funds exist for analysis of the data. Shuttle Imaging Radar-B will be flown on the same mission and over 200 proposers applied for less than $3 million in funding. A Shuttle Imaging Spectrometer is needed for developing new satellite instruments, but only funds for conceptual design studies exist. A combined gravity and magnetic field mapping mission has been unsuccessfully proposed for the past six years. The view has been expressed, by those involved in planning transfer of the land remote sensing satellite program to industry, that basic or fundamental exploratory research should not be conducted in the public domain. Consideration has been given to only conducting such research within the Department of Defense where it can be protected through security procedures or within industry where such research would remain as proprietary. These considerations have arisen as a result of increasing national concern over transfer of space reconnaissance technology to the foreign sector. These trends may significantly reduce data availability and fundamental research by academia on a nationwide basis.

Speech for April 14, 1983 Workshop

Toward a National Scientific and Technical Data Policy

F. J. Feely

It is a privilege for me to be here today as one who has spent his entire career in the Petroleum Industry deeply involved in the fields of Natural Resources and Energy, the subject of our workshop this afternoon.

I'm going to talk about the importance and the wide diversity of scientific and technical data used in the Petroleum Industry and review some of the significant advances which are currently underway in the handling and management of such data using modern computing technology. Finally, I plan to touch on the roles of industry, the academic community, technical societies and government as we look to the future.

During the 40+ years that I spent in Exxon Research and Engineering Company there was a tremendous growth in the quantity and quality of technical data available. Furthermore, today the scientist and engineer can get his hands on the data he needs much more easily thanks to the use of computers to store and provide access to data. These data libraries were developed for use with computer programs to automate the design of practically every aspect of today's oil refinery. Not only has this permitted the engineer to design equipment more accurately, but it has also speeded his job and has permitted studying design options to select the optimum as measured by economics.

Computerized database management systems are being developed today by industry, Government, technical societies, and the academic community to cover a wide variety of scientific and technical fields. Significant progress has been made in establishing scientific and technical databanks but much remains to be done to consolidate and extend present uncoordinated efforts, and to fill in the gaps in the total structure. Along these lines, recently, the Metals Properties Council initiated a project to establish a National Cooperative Materials Property Data Network which would service all industries using metals in our society. The cooperative effort would consolidate the data from a number of autonomous computerized data centers including those of technical societies, private companies, universities, research institutes, government agencies, laboratories and contractors.

A special Task Group on Computerized Data Storage and Evaluation of the Board of MPC will serve as the nucleus of a Board of Governors of the COOPERATIVE. Charter sponsor organizations also will be included. The COOPERATIVE will be organized with an elected Board of Governors and a Chairman representing major industries (aerospace, aluminum, automotive, chemical, petroleum, steel, etc.) technical societies and agencies (e.g., ASM, ASME, ASTM, AWS, DOD, EPRI, NBS, etc.)

Another impressive development currently underway in the Petroleum Industry is the building of integrated data systems which span the collection of raw data from laboratory pilot plants, the work-up of such data along with inspection data from samples of process streams to provide complete material balances for test runs, the validation of pilot plant yield data against process correlations previously developed, and finally the storage of the new runs in the database. Extension of this integration to include commercial plant operating data is also underway. This extension will provide the tie-in between pilot plants and commercial plants, through a common database.

In the past the Petroleum Industry has felt that private enterprise should bear the responsibility and cost for acquiring all of the scientific and technical data required to support its business. Generally the new processes needed in this industry have been developed by individual companies, and these have been widely licensed throughout the industry. However, data which are broadly used such as thermodynamic and physical property data have by and large come from universities and Government laboratories such as NBS. Industry financed research projects sponsored by API have also been of importance with the actual work conducted by Universities or other specially qualified laboratories.

The financing of Synthetic Fuel research data, however, poses a difficult question for industry. There is great uncertainty whether processes for coal gasification, coal liquifaction and shale oil manufacture will be commercialized or if they are, when this will happen. Under these circumstances the Petroleum Industry has sought government participation and has joined with other interested groups such as the EPRI, and overseas groups to spread the risk in this highly uncertain business.

Now let me express some personal views on the roles of Government, the Universities, Technical Societies and Industry, as we look to the future.

- Database technology and software products are in a period of rapid growth. Database software is currently available on all sizes of computers. The field is highly competitive and has great potential for future growth as software products become more oriented to user's needs and easier to use. To date there has been little government intervention, but there have been signs that some foreign governments could implement policies that would hamper the effective use of database systems. This should be avoided.
- Several U.S. Government agencies such as NBS have had a traditional role as data contributors to the scientific community. In this role, they have been partners with academia and industry. This role should continue unchanged.
- Evolution of database technology is best done in a freely competitive environment. The role of government should not go beyond the development of database systems to meet its own needs or to contribute to the overall developments in the field. The natural forces of the marketplace should dominate the evolutionary process.

- U. S. Industry needs to have the freedom to electronically transmit data across state and international boundaries unless it is sensitive from military defense standpoint. Not only will this lead to efficiency improvements, but the free flow of computerized information will help promote the dissemination of technical and scientific information. Some countries have already implemented policies to restrict such flow of data and information. This subject was dealt with in an article from the New York Times of March 13, 1983. Such controls are undesirable and should be avoided.
- Professional Societies and Universities perform a vital function in providing forums for peer review of scientific and technical data to insure that new data are subjected to a continued quality audit.
- Professional Societies can serve a useful function as focal points for the development of standards for computer software. These standards should develop from natural forces and should not be imposed too quickly. Database technology is still changing so rapidly that standards may be difficult to achieve and to implement. As databases application evolve to the point where data is transmitted between companies, then standard data protocols will be needed. Certain industries, such as the banking industry, already have achieved this.

- The development and application of computing technology in the U.S. represents a major strength of our country. In many areas of computing, the U.S. is the world leader. The U.S. Government, through its policies and the support it gives to universities, can help to insure that the U.S. remains in a position of leadership and that the technology is used effectively by the scientific and technical communities. To the extent that foreign governments establish barriers to the free flow of scientific and technical data, our Government needs to take an active role in helping to establish international ground rules which can lead to cooperation and understanding in the crucial field.

April 11, 1983

Eight Points for Consideration with regard to Federal Policy Concerning Acquistion and Preservation Energy and Mineral Resource Data

John W. Harbaugh
Professor of Applied Earth Sciences
Geology Building
Stanford University
Stanford, California 94305
May 31, 1983

This statement deals with eight key points that are suggested for incorporation in the Federal government's policies with regard to acquisition and retention of energy and mineral resource data. The eight points are as follows:

(1) The Federal government should assume general responsibility for the archival preservation of energy and mineral resource data that are relevant to the exploration and exploitation of energy and mineral resrouces _on Federal lands._ The principal energy and mineral resource commodities are petroleum, oil shale, geothermal resources, uranium, metals, and non-metallic deposits.

(2) The archival preservation of information pertaining to the exploration and exploitation of these commodities on Federal lands should be accompanied by a policy that this information will become accessible to the "industrial and scientific public" either upon being placed in a Federal archive, or following some prescribed period (i.e., a "holding period") during which time it is withheld from public access.

(3) The types of data that should be preserved in Federal repositories should be selective and should be confined to those forms of information for which adequate centralized archival repositories do not exist at present. The principal forms of data are as follows:

(a) Exploration seismic data obtained in oil and gas exploration in Federal water on the outer continental shelves (OCS's). The data include "raw" seismic data, seismic data that have computer processed in varying degree, and seismic data that have been geologically interpreted.

(b) Petrophysical logs of wells drilled on Federal OCS leases.

(c) Summaries of wells (e.g. individual "well summaries") containing engineering and geological information for wells drilled on Federal OCS leases.

(d) Logs of boreholes drilled in cool exploration and in advance of cool mining.

(e) Assay of samples and petrophysical logs of bore holes drilled in exploring for uranium and in blocking out uranium ore bodies.

(f) Geochemical prospecting data in reconnaissance exploration for metals and uranium.

(g) Petrophysical logs and summaries of wells drilled for geothermal purposes.

(4) Information stored in archival form should be placed in "machine form" insofar as technologically and economically feasible, as well as maintained in written or graphic form. Mass storage readily accessible in machine form may not be feasible for some forms of information, but the information should be indexed and catalogued so that its retrieval by manual means is feasible. For example, it is neither feasible nor desirable to store raw or processed seismic data in readily retrievable machine form

but the information can be stored on magnetic tape for manual or semi-automated retrieval. Computerized indexes are essential for its retrieval.

(5) The holding period for information, such as proprietary seismic information obtained in Federal waters, probably should not exceed ten years, and may be less. The policies for determining an appropriate holding period may be difficult to devise, and firm recommendations should be forthcoming only after extensive public debate.

(6) Federal agencies engaged in any form of mineral resource exploration should adopt a "multiple commodity" viewpoint, in which data relevant to a variety of mineral commodities are obtained. For example, the Navy (and subsequently the U.S. Geological Survey) should have logged seismic shot holes for coal in exploring for oil and gas in the National Petroleum Reserve in Alaska (NNPRA) from 1975 to 1981. The coal resources of the NPRA probably greatly exceed its petroleum resources in terms of long-term national significance. The costs of running gamma ray logs in the shot holes would have been minute relative to the program's total cost. Furthermore, it is known that coal occurs at shallow depths (as well as substantial depths) within the NPRA. While the shot holes are shallow (60 feet typically), they would have provided some useful information.

(7) The U.S. Geological Survey is probably the agency best suited to maintain archives containing mineral resource information. By assuming this role, the USGS would reemphasize its original role in "surveying" the nations mineral resources.

(8) The recording, encoding, and preparation of information for archival storage should be formally established as an obligation in Federal agencies or in federally funded projects in which geological information relevant to energy and mineral exploration is obtained. The cost of information preparation should be a budget line item for each project.

THE DOW CHEMICAL COMPANY

BARSTOW BUILDING
2020 DOW CENTER
MIDLAND, MICHIGAN 48640

UTILIZATION OF DATA ON THE ENVIRONMENT

by

Etcyl H. Blair, Vice President
Director, Health and Environmental Sciences

1. Introduction

It has been said that a man's judgment is only as good as his information. This implies that in order to make a good judgment a person must have available the information necessary to make that judgment and that the information must be of good quality. Either a lack of information or the use of inadequate information will inevitably lead to a poor decision. In the next few minutes I would like to detail for you what kind of information is needed to make decisions and judgments about the environment; how this information is obtained; the common quality problems associated with such data; and some recommendations for more effective generation and utilization of information.

II. Need For Information About the Environment

In order to make decisions about the environment we need to know first of all what is in or going into the environment. We then need to determine how it gets into the environment. Does it occur naturally or is it the result of natural processes such as decay in forests, waste from animals, emission from volcanoes, etc.? If the material of interest is not generated naturally, we need to know how much and approximately where it is released into the environment.

Another important piece of information is the fate and distribution of the material once it is released into the environment. It is necessary to know if the substance disperses rapidly in the atmosphere, or if it remains confined as a plastic in a landfill. Also, we need to know if it is transformed into another substance or if it is degraded and at what rate these changes occur. A substance which disappears rapidly from the environment either through transformation or degradation is not as great a concern as one that is persistent and remains for an extended period of time.

Presented at the Workshop "Towards A National S&T Data Policy"
Washington, D.C. - April 14, 1983.

In order to define the amount of attention that must be given to a material, we need to know how much is present in the environment and the nature and size of the population (man or beast) which might be affected by exposure to the material. It is important to note that with the exception of the identification of a material, all the other information is quantitative and therefore the accuracy of the data becomes important.

III. Acquisition of Information

There are three ways in which the needed information is obtained. It can be retrieved from a number of sources or compilations if it already exists, it can be generated or it can be predicted from existing information.

A number and variety of sources can be consulted to obtain information about the environment. Through the use of indexing and abstracting services, references to pertinent journal articles and reports can be obtained. Some of these services are computerized which allows rapid retrieval of a large number of references. Once a reference is obtained, the journal article or report must be located and the pertinent information extracted.

In some cases, the data needed is not in a journal or report, but has been compiled into a data bank as it is generated. An example of this type of compilation is the EPA-STORET data bank which contains the results of measurements of amounts of chemicals found in water and air in a great variety of locations. This is a massive compilation and requires a computer to access it. The data in this compilation are simply the results of the measurements without any consideration for accuracy or reproducibility, and in many cases the original data is not available. The data are often derived from a single sample taken at a point in time and may not reflect ambient conditions. The fact that a certain substance was not found in a particular location is not recorded. The meausrements have been done over a period of time, so that the limits of detection have changed, but this is not indicated in the data base.

Other sources are evaluated data collections which were assembled by experts in the field who have read the published literature and reports and have extracted data from these sources and compiled it in the form of tables, handbooks or computerized data bases. In the case of some of these compilations, a scientific judgment has been made of the appropriateness of the data, thus resolving differences between conflicting results and eliminating those data which are obviously wrong or acquired by faulty methods. These evaluated data compilations are most useful because they can be relied upon to a greater extent than other data. However, there is a significant investment required in the extraction, evaluation and selection of data.

If the required data are not available, one can initiate efforts in the laboratory to determine the properties of a substance or to measure its effects on fish or other organisms. Environmental measurements can be made to determine the presence or absence of a substance and, if it does exist, its concentration. This can be done on a one-time basis or a monitoring program can be instituted to obtain these data over a period of time. If one is interested in exposure, a survey can be initiated to determine those persons who are potentially exposed and the levels of that exposure.

The generation of this data is time-consuming and expensive and, in most cases, delays the time when a decision can be made. In an effort to reduce this time lag and the cost associated with the experimental effort involved, techniques have been developed to predict some of these data. In some cases, such as some physical properties, very accurate predictions can be made. However, there are other areas where predictive techniques are not as well developed and the results are at best ball park estimates. Nevertheless, such predictive estimations or guessing can be useful at times.

Earlier we referred to the need for complete data. This does not mean that information has to be available to answer every conceivable question, but that sufficient information must be used to place a question in its proper framework and perspective. To a great extent, the type of question that is being answered or the nature of the decision to be reached will govern the scope of the information needed. However, in each case there is a minimum amount of information which will be required.

The amount of information is determined by an initial (ball park) risk evaluation, consisting of a comparison of estimated exposure concentration with concentrations that produce effects. If the exposure levels are well below effect levels, detailed, highly accurate data are not needed. If they are close, more data can usually reduce the uncertainties.

Some examples may illustrate this decision-making process:

1. Ethylene Oxide

   Let us examine the environmental data available:

   a. Low entry into the environment since it is used primarily as an intermediate.

   b. Enters the environment through air and water.

   c. Moves to the air because it is volatile.

   d. Lifetime. Air: Very short because it is easily degraded by hydroxyl radicals.

      Water: Very short because it is readily hydrolyzed to ethylene glycol.

e. Does not bioconcentrate because it has a low partition coefficient and is non-persistent.

f. Degradation Products: Ethylene glycol is the primary degradation product, and it also degrades rapidly.

g. Fish Toxicity: No adverse effects to fish at 40 mg/l. Ethylene glycol shows no adverse effects on fish at 100 mg/l.

h. Terrestrial Organism Toxicity: Essentially no exposure.

2. DDT

The data on DDT are quite different:

a. High entry into the environment as it is dispersed during use.

b. Low volatility results in it being present in water and sediments.

c. Long lifetime since it does not degrade rapidly.

d. Bioconcentrates because of a high partition coefficient.

e. Toxicity is high. It is a pesticide.

The high persistence and its ability to bioconcentrate requires that much more information be used in assessing the risk that DDT presents to the environment. Data about the effects of DDT on fish, birds and humans were necessary to ultimately decide that it should not be used.

IV. Quality of Data

Once data has been obtained, one has to be concerned about the quality of the data. Let me list some common problems associated with the quality of data used in environmental decisions.

In some cases the assumptions made or the methodology used is such that the results obtained are incorrect. Some years ago, an extensive survey was conducted to determine the extent of occupational exposure to a number of substances. One of the assumptions that was made was that if a certain substance was present in an installation, all the workers in that installation were exposed to that substance. Thus, if, in the course of surveying the exposure in the Capitol Building, a painter in one room of this large building used a paint thinner, then all those working anywhere in the building were considered to be exposed to not only that paint thinner but to any substance which could be classified as a paint thinner. This flawed methodology has resulted in overstating the potential exposure to a substance by as much as 100 to 1,000 fold.

In some cases measurements are not representative of the actual situation. This is a very common problem with environmental measurements. In order to obtain an accurate representation of the level of a substance in a river one has to take samples from the appropriate locations, otherwise distorted figures can result. Not only do the samples need to be taken in the right places, but enough samples must be taken over a period of time. This is not a trivial problem, and a great deal of effort and consideration has to be given to these aspects when setting up experimental or measurement programs.

In addition to these problems associated with data quality, there are some general attitudes which pose problems. One of these attitudes is what we call the computer syndrome. That is the attitude that, if a number comes out of a computer it must be good. Another problem is the situation where one encounters the attitude that any data is better than none, regardless of its quality or appropriateness. Finally, a very real problem is that once some bad data gets used, it seems to be perpetuated and quoted and used forever. It is very difficult to expunge a bad number even if all the experts agree it is not a good one.

V. Use of Environmental Data

Each time that we are called upon to make an environmental decision we must look at all the data we have assembled, and before any decisions can be made we must judge the data we have as to its completeness, relevance and quality. This evaluation effort is necessary if those conclusions reached are to be valid and stand the test of time. Both the gathering and evaluation of data are not trivial tasks. They take time and effort and, in most cases, require a high level of expertise.

Unfortunately, we find ourselves in the situation that everytime we need to make a decision we must go through the data-gathering and evaluation process. This results in a significant amount of duplication of effort which utilizes that very scarce resource - trained manpower.

There are two actions which can be taken to reduce this duplication of effort. One is the systematic gathering, cataloging and indexing the data so as to facilitate access and thus give a greater assurance of completeness. The second is more difficult, but more rewarding. Data can be evaluated by experts once and labeled as to its quality, reliability and limitations. Once this is done, future users need not spend much time evaluating but can rely on the work of these expert evaluators who, in many cases, are much better qualified to make that assessment than the user.

VI. Recommendations

Let me suggest some specific steps which you can take in improving the environmental decisions made in this country:

1. Efforts to generate new data, either in the laboratory or by measurements in the environment, should be provided with the direction and sufficient funding to include the cataloging and indexing of the data generated. This cost is a very small increment of the cost of obtaining the data in the first place.

2. Funding needs to be provided to compile and evaluate existing data in areas of specific interest so that decisions are based on data which have been reviewed and validated.

3. Regulatory agencies should be required to make explicit statements as to the reliability and limitations of the data on which regulations are based.

4. Congress should insist that conclusions and recommendations are based on sound data and the information which Congress receives is complete and reliable, and evidence of this fact should be provided.

Furthermore, in the course of evaluating testimony presented by Congress, staff needs to be sensitive as to whether the evidence is presented objectively and puts the information in perspective or whether it builds a speculative hypothesis upon selected fragments of the data available.

Abstract

by

Arch Park, Consultant
Arnold, Maryland

AGRICULTURAL DATA POLICY ISSUES

- MODELLING -

* THERE IS A NEED FOR A DEDICATED EFFORT TO CREATE FIRST FOR THE UNITED STATES AND THEN FOR THE POTENTIALLY PRODUCTIVE LAND IN WORLD A GEOBASED INFORMATION SYSTEM CONTAINING AT LEAST THE FOLLOWING DATA PLANES --

1. LAND CAPABILITY
2. LAND SUITABILITY (FROM THE PERSPECTIVE OF FARMING)
3. LAND USE
4. CROPPING PATTERNS INCLUDING PHENOLGY
5. AGRICULTURAL CLIMATOLOGY
6. AGRICULTURAL METEOROLGY

* THIS DATA BASE IS NECESSARY FOR THE FOLLOWING MODELS --

1. MONITORING AGRICULTURAL WEATHER IN THE CONTEXT OF EARLY WARNING.
2. YIELD FORECASTING AND YIELD ACCOUNTING
3. AGRICULTURAL DEVELOPMENT
4. LAND REDEVELOPMENT INCLUDING HUMAN RESETTLEMENT
5. ALTERNATIVE CROPPING
6. ECONOMIC FORECASTING

* ALL CURRENT AND HISTORICAL STATISTICAL CROP REPORTS CAN BE ASSESSED AGAINST THIS DATA BASE.

* THIS DATA BASE CAN PROVIDE A GREATLY IMPROVED BASIS FOR INVESTMENT STRATEGIES

* WITH SATELLITE DATA (LANDSAT AND METSAT) THIS PROJECT IS WITHIN THE STATE OF THE ART - THE STAFFING WOULD NOT BE EXCESSIVE - THE BUDGET ACHIEVABLE - THE RESULTS MEASURABLE

OPCON

Operational Economics, Inc.
802 West Alabama / Houston, Texas 77006
(713) 528-3158

TOWARDS A NATIONAL AGRICULTURE AND FOOD SCIENCE AND TECHNOLOGY DATA POLICY

by

Dr. Russell G. Thompson
Chm., Pres., C.E.O. of Operational Economics, Inc.

Prepared for April 14, 1983 Washington, D.C. Workshop Towards a National S & T Data Policy coordinated by Committee on Science and Technology, U.S. House of Representatives; Congressional Research Service, Library of Congress; and Numerical Data Advisory Board, National Academy of Sciences

---

Dr. Thompson is also Professor of Quantitative Management Science, University of Houston, Houston, TX 77004.

TOWARDS A NATIONAL AGRICULTURE AND FOOD SCIENCE AND TECHNOLOGY DATA POLICY

by

Dr. Russell G. Thompson

Because of the growing economic importance of information in the Nation's economy, the Reagan Administration's satellite decision forces the data policy issue to center stage. This is especially the case for agriculture and food where the satellites are a proven technology source for accurate primary data. Joint use of these satellites and traditional reporting procedures has proven its worldwide effectiveness in NASA's Large Area Crop Inventory Experiment (LACIE).

For example, before LACIE, the reported errors in forecasting USSR wheat production in May of each year were several-fold greater in relative terms than they are today. Reportedly, similar gains have been accomplished in reducing the relative range of errors in forecasting the other major crops involved in the LACIE experiment. The economic value of this improved forecasting accuracy must be truly immense when added up across all of the economies of the free world.

The scientific task at hand is to measure the value of the information flowing from the U.S. satellite, aircraft and ground reporting systems. Measuring this value will require completion of the following phases of work:

1. Identification of users and potential users.
2. Delineation of the information role in the decision-making process.
3. Application of a sound criterion for measuring the value of the information.

Satisfactory accomplishment of Phase 1 will require insights into all facets of collecting, analyzing and reporting agriculture and food statistics. Important background information will be the history, objectives and character of the current reporting system. Historical documentation should show the real and perceived needs for the information. The character description should show the informational requirements for timeliness, scope, reliability and validity. A sample survey will probably be necessary to identify users and potential users, which are (or could be) affected by the information.

With knowledge of the users, mathematical modeling of the decision process will encompass the scope of usefulness of the information at each stage in making decisions. Careful structuring of how the information is used will encompass the source-by-source contributors to total value. It will further lend support to the best criterion for measuring the value of the information.

Application of these concepts to the problem at hand will give dollar and cents measures as to what the information is worth. Private sector investors will then be able to make knowledgeable bids for the satellites.

Because the information flows from the satellites will have value beyond trade and commerce (e.g., food stamps, food aid), the U.S. government must keep a hand in the overall operation. Recognizing the benefits of private sector incentives and management argues strongly for a private/public sector joint venture.

In conclusion, the satellite decision is fundamental for accurate agriculture and food statistics. Economically, transfer of this maturing technology to primarily the private sector will enhance the economy's capital base for growth. However, a private/public sector joint venture seems necessary to fulfill well the needs for a growing and informed economy.

Public Data Systems for Agriculture and Food

Kenneth R. Farrell, Director
Food and Agricultural Policy Program
Resources for the Future
Washington, D.C.

The U.S. public data base for agriculture and food has long been regarded as among the best in the world by virtually any standard - comprehensiveness, reliability, timeliness, accessibility, longevity, continuity, etc.

However, it is a system under great stress - a system in need of respecification and redesign to meet future public and private needs; a system requiring greater public support.

These brief comments are in three parts: (1) components of the current system, (2) current and emerging sources of stress in the system, (3) recommendations for improvement of the system.

1. Components of the System

At the Federal level, three Departments provide the bulk of data pertaining to agriculture and food - Agriculture, Commerce, and Labor. In addition, each of the States develops and maintains a data base, in part in collaboration with the Federal government. That Federal-State system is supplemented by data procured through periodic, special-purpose surveys by universities and colleges, a variety of non-profit organizations and increasingly by the private sector. My comments will refer primarily to the public data system, particularly the Federal government component of that system.

Data pertaining to the <u>farm sector</u> - structure, assets, inputs, production, sales, prices, earnings, and income, for example - derive primarily from the Census of Agriculture within the Department of Commerce and from several agencies within the U.S. Department of Agriculture. The Census of

Agriculture, the oldest source of farm related data, provides periodic "benchmarks" by enumerative surveys of those engaged in farming operations. The Statistical Reporting Service of the USDA, the other major source of farm related data, provides frequent, timely information with emphasis on commodity production, prices, inventories, production expenditures, cash receipts from marketing,etc. The SRS data are supplemented by a melange of data from other USDA agencies - economic data from the Economic Research Service; natural resource data from the Soil Conservation, Agricultural Stabilization and Conservation, and Forest Services; foreign trade data from the Foreign Agricultural Service; and other specialized data from such agencies as Farmers Home Administration, Federal Crop Insurance Corporation, Food and Nutrition Service.

Data pertaining to the <u>farm input and farm product marketing</u> sector, the other major components of the food and fiber system are provided by several agencies - the USDA (primarily the ERS and Agricultural Marketing Service) and Commerce being the most important. Major types of data include those pertaining to structure, assets, sales, employment, inventories, and earnings.

<u>Food</u> related data likewise are collected by several agencies - food prices and expenditures from the Bureau of Labor Statistics; marketing margins and costs, food prices, consumption and expenditures from ERS (USDA); food consumption and nutrition data from the Agricultural Research Service (USDA), Food and Nutrition Service (USDA) from periodic survey and special programs.

I also should mention NOAA (Commerce) and NASA, which in collaboration with USDA provide weather and a variety of remote sensing data. The CIA in addition to FAS (USDA) and Commerce, providessupplemental information on foreign agriculture conditions as do several international organizations, notably the Food and Agriculture Organization of the UN and the Organization for Economic Cooperation and Development (OECD).

Information on scientific research in progress in agriculture and food is available from numerous sources, particularly the Current Research Information System (CRIS), a computerized information service maintained by the USDA and the land grant universities. Bibliographic information services are maintained in numerous agencies most notably the National Agricultural Library which has computerized in recent years much of its information.

This very diverse, fragmented system provides an wealth of data (historical, current, and scientific) on agriculture and food and undergirds much of the information available in the private sector.

2. A System Under Stress

As comprehensive and as rich as the data may be in the far-flung collection of data gathering agencies which I refer to loosely as the agricultural and food data system, serious stress and concern relative to its capabilities to meet future needs are emerging. I list six such sources of stress and concern.

1. Changes in the Structure and Organization of the Food and Fiber System:

Major components of the current system were developed upon a model of agriculture which has since undergone vast structural change. Serious gaps exist in the data base, in part, because the conceptual design of the system has not been brought into harmony with the reality of the current and emerging food and agricultural system.

2. Complexity of the Food and Fiber System and its Interdependencies, Nationally and Internationally:

To describe, analyze and understand to-day's food and fiber system, one has to look well beyond the "farm gate" - to the input and product marketing sectors, to international markets, to macroeconomic policies, to the changing patterns of natural resource use, and to demographic and environmental issues. The complexity of the social, economic, institutional, and technological environments in which the food and fiber system functions to-day, creates needs for additional data for both public and private decisionmaking. But our data systems lag seriously in meeting those needs.

3. Fragmentation and Lack of Coordination in Data Collection and Dissemination:

I have illustrated the diversity and decentralized nature of data systems for agriculture and food. There may be strength in diversity. However, the fragmented nature of the system and the lack of coherence and coordination among the component parts also are sources of inefficiency and major concern about the ability of the system to regenerate itself to meet rapidly changing public and private needs.

4. Inadequate Planning to Meet Future Needs:

Related to the above, there is a distinct lack of coordinated longer-range planning among the numerous agencies responsible for agricultural and food data. Planning, such as it is, tends to be ad-hoc and short term in perspective.

5. Inadequate Budgets to Support Improvements in the System:

Real funding levels for data development in the Federal government are declining. Adoption of technology to enhance efficiency in data collection, management, and dissemination are lagging in part, because of inadequate budgets in statistical agencies.

Summing up, public data systems for agriculture and food are under serious stress. A decade ago, a committee of the American Agricultural Economic Association put it this way:[1)]

- "...our data systems are in deep trouble ... demands we make on this system are now outrunning our investment in its continued development. Most significantly, the conceptual foundation of the system is crumbling - and has been for some time."

- "Examination of various Commerce, Census, and USDA measures of income and output suggest not only real conceptual difficulties over what is being measured but grossly incompatible data systems."

- "...we do not have a complete and coordinated body of data that permits description or analysis of the food and fiber industry as a whole."

- "Developing a new theoretical basis for obsolete data systems is an urgent necessity."

- "We must develop a step-by-step strategy in which many varied actors who design, manage, and use data cooperate to retrieve our data systems from the intellectual chaos that now threatens."

Some progress has been made to remedy such problems. Still, it is probably correct to conclude that the system is continuing to obsolesce and still in a condition of intellectual chaos and administrative disarray.

3. Recommendations

There is no simple solution or panacea to the problems I have outlined. However, two initial steps might be considered.

(1) Reestablish a federal statistical coordinating body of some type at the highest levels of the Federal government and in each of the major statistical agencies. Highly authoritative administrative structures may

---

[1] Our Obsolete Data Systems: New Directions and Opportunities, AAEA Committee on Economic Statistics, AJAE, Vol. 54, No. 5, December, 1972, pp. 867-875.

be neither advisable nor feasible. But leadership,and the coordinated planning which can come from effective leadership,are sorely needed.

(2) However it might be accomplished, the most important recommendation pertains to the development of a framework to guide longer-run evolution of the agricultural data system, an endeavor of fundamental importance and one which should involve ultimately a large number of users and data providers.

The fundamental task to be accomplished is that of designing a data system which accords with the current and evolving economic structure of the food and fiber system and rural communities; identifying the types of data required for analysis and decisionmaking, public and private, at key nodes or interfaces within the system; specifying the properties of data (frequency, geographic, commodity, market coverage, etc.) to be collected, and perhaps a generalized approach to data collection. Attention must be given ultimately to where data collection functions might be performed best organizationally, how coordination among parts of the system might best be obtained, and how data can best be managed and delivered for use within the public and private sectors. But it is essential that the initial effort be focused on the conceptual and system design issues, not organizational and process questions.

It is obvious that we must be concerned ultimately with the development of data for related components of what has become a highly interdependent agricultural sector. The development and linkage of international agricultural and trade data with those for domestic agriculture deserves consideration. In a similar vein, the growing need for domestic natural resource related data must be considered. Efforts in the USDA and elsewhere to develop nationally consistent data bases for natural resources have been proceeding largely independent of the system for domestic farm production data. Such an approach seems likely to be not only costly but to leave important data gaps, and inconsistencies. Food consumption and nutrition data should be strengthened. Finally, we must direct more attention to the development of data systems for rural areas, one which encompasses and links not only measures of production activities but measures of the economic and social well-being of rural people, households, and communities.

The design and detailed specification of a data system linking all such components obviously would be a massive undertaking. As a beginning, the broad parameters of the system and the linkages among major components of the system should be designed. Then, progressively and systematically, more detailed specification of the system of the future should be undertaken. There is substantial intellectual capital to undergird the effort we are proposing. The challenge is to get on with the job.

Comments on Data Acquisition During Health Care Claims Processing

James E.D. Gardam
Vice President, Medical Services
Governmental Health Programs Office
The Prudential Insurance Company of America

I am going to speak to you today not as a technical expert in the computerized acquisition, storage, and display of health care data but as a "user", customer, or consumer of such data. In a sense, I am the prototypical friendly user with whom the "computer" technologists has identified in our limited sphere of health care participation. I believe we can state that a satisfying buyer-seller relationship exists. This is an absolute requirement if data is to have meaning. Now let me describe the environment in which this relationship exists.

Prudential's Government Health Programs Office (GHPO), is a totally dedicated, geographically separate, and practically independent claims processing unit of the Company. It administers, under a non-profit, cost only, contract, the reimbursement (plus some other aspects or elements) of five governmental health programs. They are:

> Part B of Medicare in New Jersey, Georgia, and North Carolina.
>
> Part A of Medicare in New Jersey. This program is shared with New Jersey Blue Cross.
>
> The New Jersey Medicaid Program.

As noted, the administration of these activities is separated from our Commercial Health Insurance operations.

Presumably, the distinction between Medicare A and B is known. Part A involves institutions (hospitals, home health care agencies, and skilled nursing facilities). Part B reimburses primarily individual providers (Physicians, Osteopaths, Podiatrists, Chiropractors, and certain other suppliers of health care services). Medicare is a Federal Program; Medicaid is a State administered program with funding shared by the State and Federal Government.

GHPO operates in four locations, communicates through leased lines, and employs 1,570± people. This office provides service to 2¼ million beneficiaries or recipients, 195 institutions (plus 160 CHC's) and about 45,000 individual providers of various types and disciplines involved in the provision of Health and Medical services.

In 1982, GHPO processed slightly less than 20 million claims; disbursed $1.5 billion dollars in benefits at a cost of $45 million dollars. Part A cost per claim was $3.46 compared to almost $5.00 in 1977, or five years previously. The unit cost for a Part B claim was $2.28 compared to $2.77 five years previously despite inflation, postal cost increases, and the changes in medical practice/technology. Pre-coded Medicare claims with mandatory assignment are processed at about $1.30.

It is believed that these costs compare favorably with other processors of Medicare and Medicaid claims and are more than competitive with commercial insurance processing although their mission is often considerably different.

GHPO utilization denials alone have saved many millions and denials or reductions for all causes suggest a savings of more than $5.00 for every $1.00 of administrative costs.

Currently, in use is an IBM 370-168 with attached processor in a central location. It is being supplanted by the third generation in the family of large commercial processors - a 30-81 model with 16 million positions of memory. GHPO maintains its own staff of programmers and system "architects and engineers" for its totally dedicated consumer oriented function. It is my understanding that there is no such thing as "down time"! However, in actuality this statement is simply an indication of the dedication of our EDP people.

With these facts in mind, it becomes obvious that we accumulate a great amount of financial data. However, this data translates into an accumulation of a wealth of health care data. This store can only be characterized, literally, as a mother lode of provider services; patient utilization; institutional care costs; individual, specialty and area professional practices; patient demands; patterns of care; and epidemiological and technological trends. This acquisition of health care financial data simply demands employment of the cliche - it is "boggling".

What are some of the lessions that we have learned from computerized claims processing?

First, data can accumulate rapidly and become enormous in amount. Second, the current and advancing computer technology is nothing less than a national asset. Third, the mental capacity, interest, energy, and education of our people, particularly the young, in computer systems and programming is a second national asset and presages a continually expanding industry for the near and long term. Fourth, the store of accumulated data demands analysis and translation for policy positions. Fifth, the very magnitude of the health care ecomony, now 9/10% of our GNP, demands intensive computer application for the extraction and analysis of data, the ability of computer to computer correspondence or data linkage, and access to data.

It is a principle that language and format control data. The HCFA-AMA, energetic effort to promote universal acceptance of CPT-4 must be carried forward and supplemented by the adoption of ICDA - by everyone: Insurance companies, Academia, Government Agencies. We must promote a common language - not simply a computer language but a technical language. Then any system can talk to any other system with the development of formal line protocol for access to data, preservation of property rights and surety for privacy.

Currently, it is possible to address and resolve technical problems.

It is clearly possible to develop, implement, and maintain policies both for the preservation of individual privacy, the security of data, and the property rights inherent in systems development and the accumulation of data. There is no problem in this area which cannot be addressed and resolved.

The contracts under which we operate reserve all "research" efforts to the government agencies for whom we work. There is no provision in the contracts to fund investigation or research. However, it has not been possible to simply ignore this store of data. Let me simply outline in briefest terms some of our studies. Please note that these are lesser or minor studies, if you will, serving only as examples of the potentials for analysis. There are significant epidemiological and other significant studies of injury, illness, and therapy which are of much greater significance. These studies have been performed, for the most part, within the competitive cost-per-claim constraints cited above. They have been performed in part, as intellectual challenges to enhance the function of "routine" claims processing and challenge or develop our people. They have been rewarding to the participants. When of significance, the results have been communicated to HCFA.

A. We have been able to review the professional services provided to de-institutionalized patients in Boarding Homes (for Region II), and

B. The Medical necessity of routine hospital admission chest films (again for Region II).

C. We have examined the belief that physicians refuse to treat Medicaid patients. A fallacy disproved by a program that indicated that 12,484 or almost all active New Jersey physicians received payment for services rendered to Medicaid recipients.

D. A newly developed electro-magnetic device appears to hasten recovery in patients who have non-union of fractures or failed surgical fusions. We have begun to examine the longitudinal history of knee surgery: the insertion of the total knee prosthesis; failure; removal; failure of the site to unite; surgical fusion; failure of the fusion; and finally, the necessity for employment of the device with an equipment cost of more than $2,500 plus a prolonged disability with loss of possible earnings. Analysis of raw data suggests, first, confirmation of the anatomical fact that the knee joint is distinctly different from the hip - and second, that total knee replacement may be a technique whose time has not yet come.

E. Peer review is an inherent element of hospital practice and it is an accepted "fact" that large costs and high professional fees originate from hospital in-patient services. Our studies suggest that the review of a claim for a single aberrant high fee - the outrageous surgical charge, for example, is not as important as is a pattern of care studies with particular reference to the out-patient setting. In the office setting, we have noted:

   1. Non-pharmacologic use of B12 is a common professional practice.

   2. There is substantial use of injections in place of oral medications.

   3. Interrupted and inappropriate use of antibiotics commonly occurs in the out-patient setting.

   4. Overutilization of tests, services, and procedures in the office setting substantially increases costs.

5. Our state government successfully monitors office prescription abuse to the extent that physician trade in narcotics is disappearing.

6. Returning to utilization of services in the out-patient setting, Please note:

   Two unnecessary ECG's and chest films per day, five days a week <u>in the office</u> represents a potential loss of $25,000 per year.

F. Our current "4G" ($4,000) project is a review of 200 consecutive claims for surgical services with a fee greater than $4,000 - not alone to determine if the fee is without justification but to reliably determine what are the driving forces which result in a claim of this magnitude. One initial impression is that one factor may be $25,000 malpractice insurance premium for Neurosurgery.

G. The patient overutilization project - which was an outstanding success - within designed limits. You are familiar with the axiom - physicians are responsible for 80% of the decisions which result in health care costs. In consideration of actual experience by Medical Staff with patients, it was decided to look at the other 20%. An initial study of the beneficiaries whose claims indicated the largest number of visits per year in all three states, revealed, after personal interviews and medical record documentation, some fascinating patterns: a combination of the demanding patient and the complacent, compliant, older physician with waning skills and who was untrained in the recognition and treatment of functional disorders.

H. In addition to the above studies of service patterns, we are currently implementing utilization screens for hospital stays exceeding 40 days and Podiatric visits for routine care which are in excess of the rate of nail growth; other utilization review screens are being developed for out-patient laboratory tests and comparable services.

I. Finally, our programs for direct physician and hospital claim entry with data exchange regarding eligibility, deductible status, etc., is well underway. Hospital direct imput through other payors, data linkage, in the case of hospital charges and tape to tape exchanges of reimbursement data are actualities.

These are small projects - they have been carefully considered and implemented only after the exercise of slow and painstaking deliberation and within the constraints of very, very limited financial resources. They have represented a challenge; they have brought added satisfaction to what could be a routine claims processing mechanism; and, most important, we believe they point the way toward the recognition of problems. Property rights and privacy can be protected. Access and analysis of data can lead to the solution of health care problems.

THE ROLE OF INFORMATION IN THE PHARMACEUTICAL INDUSTRY

Ralph Hirschmann, PhD
Sr. Vice President, Basic Research
Merck Sharp & Dohme Research Laboratories
Merck & Co., Inc.
Rahway, N.J. 07065

Presented at Workshop
Towards a National S&T Data Policy
April 14, 1983
Whitthall Pavilion
Jefferson Building
Washington, D.C.

My associate, Dr. Horace Brown, and I are pleased to have been invited to present our views to you this afternoon.

We in the pharmaceutical industry, like the others who are speaking today, are vitally concerned that data generated in government, academia and private sector laboratories should be accessible to other scientists, consistent with intellectual property rights. Furthermore, these data must be made available in a manner which permits the user to make scientific judgments about the utility of these data in his own research applications. Thus, we are differentiating between raw data and information, which is derived by an expert's evaluation of data.

I'd like to give you some representative types of information bases which we use at Merck. First, let me emphasize that making a safe and effective drug is a complex process involving more than fifty distinct disciplines. Our major goal is to discover new entities with pharmacologic properties which make a contribution to therapy. Most new entities which we generate go nowhere. The few that do, require a span of ten years from concept to marketplace, and cost 50-100 million dollars to approval for market.

Chemistry is one of the major disciplines involved. The synthesis of new chemical substances requires access to such data bases as Chemical Abstracts and Derwent's Farmdoc. These two services illustrate different methods of funding. Chemical Abstracts is the centerpiece of the world's store of chemistry. It abstracts about 500,000 articles per year. It is operated under the guidance of The American Chemical Society and, by policy, must be self-sustaining, i.e. users must pay for services rendered. However, it is important to note that the current Registry System, established in the 1960's, was heavily supported by grants from The National Science Foundation. "Up-front" dollars for system development can thus be a crucial part of a data base operation.

On the other hand, Derwent Publications Limited is an example of a private enterprise operation. For more than a quarter century Derwent has collected, indexed, abstracted and distributed patent information from major countries of the world. As with the Chemical Abstracts, feedback between users and data base vendors was critical in establishing a useful service.

Now, to return to the discovery phase of drug research. We, at Merck, are keenly interested in a wide variety of diseases related to human and animal health. These include hypertension, infection, inflammation and arthritis, mental health, endocrine diseases, and virus-induced diseases of both children and adults. Thus, we must scan a long list of journal articles and search many information banks to keep abreast of current happenings. I will return to this problem later on.

Let me cite an example from the bacterial infection area. To find a new antibiotic involves the collection and screening of many thousands of cultures. Hence, we frequently must compare specimens of interest with reference material from what is known as The American Type Culture Collection. You will appreciate that it is a complicated task to determine when an antibiotic with interesting properties is truly new. Large volumes of both biological and physico-chemical data must be generated and compared to separate the old from the new.

When an interesting new "lead" is in hand, we must concern ourselves with its potential for side effects or toxicity. This requires yet another data base. Biological Abstracts is a central source for information on the properties of compounds and on methods used to demonstrate therapeutic or pharmacologic actions. Though Biological Abstracts had received support from government grants, it is today a not-for-profit, self-supporting organization.

The National Library of Medicine exemplifies yet another vital source of information for the preclinical and medical areas. Here, Chemline and Toxline serve to give insight into related chemical substances. For example, the search for a chemical class of compounds called imides ought to raise a flag for the chemist re the teratogenic properties of *thalidomide*. Though a compound chemically related to *thalidomide* might not be teratogenic, the investigator would surely wish to be reminded of this possibility.

Following the preliminary selection of a potential product candidate, we must conduct chronic toxicity studies. The volumes of data generated in our laboratories need to be evaluated in the light of extant knowledge related to similar chemical or therapeutic classes. Two data bases from the National Library of Medicine are sometimes helpful - Cancerlit for animal, and Medline for human data.

So far, we have stressed the biological properties of chemicals. We are also concerned with the physico-chemical properties. An interdisciplinary data base supported by National Technical Information Service (formerly NIH/EPA) is quite useful to our physical and biophysical chemists. This data base started in NIH, but was later made available to outside users through government support. Users now pay an annual subscription fee and pay connect charges for access to the files.

Following submission of an IND, careful clinical studies must be carried out. These begin with small numbers of closely monitored subjects in clinical pharmacology studies and then expand to clinical trials in larger patient populations. In some instances, the patients selected for study may be based on epidemiological data obtained from the Communicable Disease Control Center.

From this information source we can get morbidity and mortality data, viral disease incidence for vaccine studies, etc.

Let us now say a word about the building and operation of data bases to which we have referred above. It is a very complex and costly endeavor which nowadays is almost surely computer-based. The computer system must be designed to capture, store and retrieve the data. Communication networks must be chosen for transmission of the retrieved data to users. This whole operation requires the cooperative efforts of experts in all of the disciplines involved. Most important is the selection of the data to be stored. The accumulation of unreliable data is worse than no data. In our pharmaceutical work the interdisciplinary aspect of the information needed adds to the complexity of system design and operation.

Finally, a word about alternative philosophies for the funding, building and operating of data bases. In our view, there is no single optimal solution. Some countries, such as Japan, France and England, regard information bases as a national asset, and consider it a foregone conclusion that such systems should receive support. The British Lending Library and the overnight availability of documents in England are examples of a service we lack in this country. We heartily support the concept of free enterprise, but we note again that some of the successful larger systems for distributing information had their inception in government-supported activities. I also want to stress that the fragmentation of information systems complicates information extraction for the user. For our interdisciplinary needs, it would be desirable to be able to go to a more limited number of sources to get all the information needed. Multiple systems, especially fragmented ones, mean added expense in training users and in rearranging formats for distribution. A more integrated approach and more standardized means of access would be helpful. Another problem that should not be overlooked is what I would call the orphan information problem, i.e. information of only limited marketability.

A good example of a cooperative venture is the National Library of Medicine. Medlars has been a printed source of clinical information for decades. In the 60's, through a cooperative effort involving National Library of Medicine, the Pharmaceutical Manufacturers Association and Systems Development Corporation, this data base was made available on-line. We have used it continuously since that time.

Taxpayer, personal or industrial, and government find themselves being both provider and user of information. We all clearly recognize the role of information as a useful resource. Unlike most other resources, it is not consumed in the process of use.

Knowledge has long been regarded as power. The elite in our information age will thus have power to be used for our common good. We think we require a strategy which assures that our informational needs are met in an optimal manner. In the past, cooperative involvement of government, free enterprise and the end users has often provided reasonably satisfactory results. The way is complex, and paradox is part of the game. Our national endeavors are heavily based in information. We must find a way to maximize its availability and simplify the extraction process as much as possible.

Finally, I should like to mention that we at Merck also contribute to the data bank. For example, we publish some 400 papers in refereed journals a year, and we make many of our compounds available to other scientists, both in industry, in government and in universities. This operation often generates additional data, some of which are then published by the outside investigators.

NATURAL RESOURCES AND ENERGY - GROUP SUMMARY

John C. Davis, Co-Chairman

My report will be somewhat, should we say, imaginative, in that it is very difficult to come up with a concise, coherent summary of what was an extremely heterogeneous discussion. Some of the remarks that I will make reflect in part the various speakers' presentations, and in part my own thoughts which I interject at this time as a personal reward for having served as chairman.

First, I would like to introduce a concept that really didn't come out in the meeting: the concept of the National Patrimony. This means "irreplaceable information of value to the nation."

In the session on natural resources, the data that we were mostly concerned about were data which are time- and space-dependent. This means that they are unique. These data cannot be replaced once they are lost, because they will never occur again (as someone pointed out). Mt. St. Helen's is apt to erupt only once, we hope, in our lifetime.

Also, there is something a little different about the type of data and the type that the gentlemen involved in engineering materials or chemical properties are concerned with. This difference is the vast quantity of data that must be preserved. When processing such data, it is essential that we use computers. The oil companies, for example, maintain some of the largest computers in the world outside the Pentagon, and they process some of the biggest data files that exist anywhere. A basic question--a fundamental question--is how to coordinate and facilitate the exchange and consolidation of the data that are already available. There are basically three choices: Coordination and facilitation can be done by (1) the government, by (2) the industry, perhaps through professional societies, or (3) it can be done by a mixture of the two.

Now, the reasons for preservation of this type of data must extend beyond the narrow concept of their current commercial worth. Economic conditions change. What is uneconomic today may be vitally important tomorrow. As a simple example, we can consider the incredible fluctuations that we have seen in the past years in the price of oil and gas.

If OPEC collapses and the price of crude oil drops significantly, then many of the data archiving and data exchange mechanisms which presently exist in the private petroleum sector will not be economically viable. And if these uneconomic conditions persist for a sufficiently long period of time, the data will be lost, irretrievably lost. For instance, the information in records from dry holes drilled in the exploration for oil and gas represents the only tangible achievement from an expenditure of several millions of dollars for each hole. And its unlikely that the industry or the government or anyone else will go out and drill another hole in the ground in that spot to get that data back once it's gone.

Secondly, our objectives change with time. This was brought out by Dr. Harbaugh in his presentation. The information gathered for one person may someday be extremely valuable for entirely different purposes that were not thought of at the time the data were initially gathered. If that information is lost, it will have to be regathered at great expense. For example, if the nation decides that it needs to investigate the North Slope to determine its coal resources, it's going to have to go back and embark on a major, very expensive program of sampling to obtain this data which could have been gathered at marginal cost at the time petroleum evaluation was underway in this area.

Thirdly, there are important private sectors of the economy that may require information in order to survive, and I can give you a small example from Kansas. The State of Kansas maintains, with industry cooperation, a program for the archiving and distribution of information on oil wells--samples, top cards, driller's logs, and the like. The reason we do this is because the individual members of the petroleum industry that exist in the State now are not financially able to do it by themselves; with the cooperation of the State, they can.

The oil and gas industry in Kansas, although composed of small companies, includes a very large number of such companies, and constitutes the second most important source of income--tax income--in the state. It is also the second largest employer of taxpayers in Kansas. My agency is extremely interested in the health and well-being of this particular segment of the public. I think that equivalent situations exist in the nation as a whole.

In addition, there are certain noncommercial but socially worthwhile uses for temporal and spatial, natural-resource data. For example, trace element analyses have been collected in certain national surveys such as the NURE program of the DOE. Such information may have been gathered for one purpose, perhaps the assessment of uranium resources, but this same data can provide very, very important health information. You can't really place a price on the knowledge of the distribution of selenium in surface samples, for instance, or other elements that may have a very important health-affecting component.

And finally, as we saw in one rather dramatic example this morning, many of the benefits are negative in the sense that the absence of good data has extracted a cost which was unknown or unappreciated at the outset.

The summary of all of this and, I think, a concensus that emerged from the meeting, was that a successful approach to the data problem in the natural resources area requires cooperation between all interested segments of industry and government. The data come from all sources. The government certainly does not have a monopoly on their generation; in fact, it is probably a small generator in the area of earth-science data. There is a concern about funding, about obtaining basic support for the start-up of any exchange mechanism. There is an initial lead-in cost for any mechanism for the dissemination and preservation of such data, and it is difficult to fund such a project from a purely commercial viewpoint.

I think everyone agrees, and I think the industry itself would be the first to say, that there should be a "cost" for data. There would be a tangible support contribution which is made by the users of data. But the general view is that this fee should be tagged to the marginal cost of the data and not to the initial cost of generating the data. The government should recognize (I think this is the most important point) that there is an ongoing cost associated with the generation, maintenance, and distribution of the data that it creates; and that it has an obligation to the National Patrimony-- it must recognize this at the outset of its data-generating activities.

Critical information on certain national resources are not available even though the information exists. Again, this is a problem of coordination. The volume of the data involved makes it a very significant technological and budgetary problem. Also, there is a problem in government/industry data exchange of information from federal lands.

I think this summarizes both my own prejudices and the views expressed in the comments of the committee.

Thank you.

ENVIRONMENT - GROUP SUMMARY

R. Stephen Berry, Co-Chairman

Both our speakers were from the chemical industry, so naturally they shared many of their concerns. One point that came out very early is that for industry, the climate--the pressures on them to deal with environmental matters, to deal with pollution--is more severe than ever before.

The only ways available to deal with these problems demand that we have good environmental data on sources, on ambient conditions, on effects, and on the transformations that occur to pollutants as they go into and dwell in the atmosphere. The points were made that most firms, most emitters of environmental pollutants, behave responsibly and have expertise for measurement and for dealing with controls, but that there are always rogues and there are always new situations arising that prompt regulation. While we have public agencies to gather data, in fact, private firms themselves, for their own interests and for public interests, do gather data, often on many substances that are not under any controls at all, but simply for their own interests and for their own purposes.

Their data is caught, or their data gathering is often caught between a Scylla and a Charybdis. On one hand their data are seen as generated by a private institution having a vested interest and at the same time those data are generated by an institution that has more expertise than any other body. This is a dilemma that we haven't really resolved and it comes in further down the line, when we begin to ask questions about where the responsibility should lie for gathering data, for validating data, for disseminating data.

Firms carry out these activities when they see some gain for themselves in the broadest sense but still gain in terms of the firm. There are public gains that are not necessarily in the firm's interest that justify activities by government agencies. The question then comes as to what are the roles of the agencies and bodies, the academics, the government laboratories in that gathering, validating and disseminating activity. The validation aspect particularly, as Mr. Terry pointed out again and again, requires one being in a kind of policing role. It's inescapable. One cannot expect to have well-validated data unless there are not only inducements for doing it right, but penalties if you don't accept that responsibility. The questions then arise as well, are there institutional ways of attacking the problems of generating and validating and disseminating reliable data?

The way I've discussed the issue so far makes it sound like assembling reliable environmental data can't be done without some kind of confrontational situation. But there is at least one example that was brought out by Dr. Bonczek that--in the case of substances regulated by OSHA--the regulations, standards and regulations for acrylonitrile were decided upon in an amicable, collaborative way. There is one example that we can take as an existence proof that it may be possible to get environmental data and standards without confrontation. One of the questions that the Science and Technolocy Committee will no doubt be facing is what mechanisms could make that one experience general, should the national laboratories, not only the National Bureau of Standards but perhaps other national laboratories, be playing an institutionalized role in data gathering and validation for the environment beyond what they do now.

Should agencies that support research on data, particularly in our context, environmental data, constitute from their constituencies of users advisory boards to deal with problems of generation and validation and dissemination of data--boards that would put together the data generators and data users to find some common ground?

There were several specific recommendations that Dr. Blair proposed, and some others that emerged as well--those Dr. Blair proposed were that data generation activities should have support large enough and specifically indicated for making those data available for cataloging and indexing, and I would generalize that to say validation as well.

Secondly, he comes to the point of validation: funding needs to be provided to compile and evaluate existing data, not only new data, so that those data can be available and validated when decisions have to be made that depend on their validity.

Thirdly, he recommended that the regulatory agencies carry responsibility to examine, state and make public the reliability of the data on which they are basing their decisions. Finally, he dealt with the broader issue, that Congress has a responsibility to insist on the soundness of the data on which it is trying to make its decisions.

There were some other more specific recommendations in terms of data responsibilities that need to be fulfilled, particularly needs for background data and anticipatory research that will tell us how things are changing and with which we can make comparisons in the future of new environmental perceptions or new substances added into the environment.

Thank you.

AGRICULTURE AND FOOD - GROUP SUMMARY

Russell R. Thompson, Co-Chairman

We had two speakers--Ken Farrell and Arch Park. Mr. Farrell, focused on three aspects--the institutional setting for food and agriculture data. He emphasized how it's organized historically and up to the current time. He pointed out that the current system is in great stress for many different reasons, and this stress has been existing now for some time, and there has been a continuing deterioration in the quality and breadth of the data. He made two recommendations for the future and that is there has to be some type of new system for coordinating agricultural and food data, and there must be a framework within which this coordination is planned.

There werea number of questions that came up, one of them being that there seems to be an awful lot of shuffling around, and that shuffling around has existed for some time, but really no one seems to be necessarily in charge. There was some discussion of that and it seems like there is more need for management at a higher level--I didn't discern that anyone had any answers specifically for that. It was pointed out that there is an international dimension and that the international dimension is becoming increasingly important. It was pointed out that there are very definite micro data needs and in the past some of these micro data needs have been fulfilled upon requirements of the Congress for like cost of cotton production data and things like that.

It was also pointed out there are major gaps in the data. One of the major gaps that was singled out was gaps in the case of nutrition.

Mr. Park focused on a global rice model that he has been involved in developing. This was sponsored by the U.S. Government and has had joint review at the international level. He went into how the model was formulated and pointed out the role of the remote sensing technology in both accounting and forecasting. He pointed out the need for aggregating in this model to a relatively large area dimension which was found to be necessary. He pointed out that the data base for this model was necessary and he went into a number of characteristics of the data base. And he said he believes that there is a national interest; that models such as that (and data to support those models) are definitely in the national interest, because the nation must understand not only its own agriculture but worldwide agriculture; and we need to have a sound data base as an early warning system for answering agricultural development questions, for human resettlement questions, proposing alternative cropping patterns, and coming up with investment strategies as to how the investment should be made in the weather satellites versus the remote sensing satellites.

Now, several questions were asked. One set of questions dealt with the privacy issues. Do people feel that their privacy will be invaded by the satellites in collecting data which will be used in very fundamental ways? As I recall he believed that the privacy issue could be handled well.

The ownership/management issue in terms of collection, analysis and use was raised; and there was considerable discussion on that point. Considerable skepticism was raised with regard to private control of a data collection, analysis and reporting system, becase they didn't believe that standards would be maintained with private control.

Some willingness seems to be present to at least consider appropriate private/public sector joint ventures. However, there didn't seem to be any definite insights as to how they might be structured.

It was expressed that there is considerable knowledge as to the users of agricultural data and the potential uses of the information emanating from it, but virtually nothing is known as to the value of this information in use, which is fundamental for economic decision making. For example, what is the likely return-on-investment from private ownerhsip of the remote sensing satellites? As you know, it has been proposed by the Reagan Administration that the weather and remote sensing satellites be transferred to the private sector, and that these satellites are proven technology with regard to agricultural food and fiber data and collection. It is the one area in which the remote sensing technology was proven under the LACIE program.

In conclusion, we had a very dynamic discussion. It lasted the full period, and there was some controversy at the end with regard to the satellite ownership/management issues.

As Spokesman of the Session, I would draw the following conclusions. We are on the threshold of a new era. The nation needs a totally new ownership/management system for data collection, analysis and reporting. This new system must timely and inexpensively fulfill society's needs for accurate information. Incentives and responsive management are necessary, which can only be provided efficiently by the private sector. The private sector must become intricately involved in the decisions associated with the ownership and management of the nation's agriculture and food data collection, analysis and reporting system. But for private sector decisions, some criterion (or criteria) is needed by which to determine the rate-of-return on the investment.

This requires meaningful prices. Unfortunately, nothing is known as to the likely value of the information in use. This value can be measured; this measurement will give the needed prices for decision-making. However, resources must be directed to measurement. To date, virtually none have been.

So, the issue before the scientific community is clear. This community must deisgn the research programs to provide the information values required. It must further become actively involved in designing and planning the new system. Then, the private sector will be able to bid competitively to own and manage maturing technologies developed and owned by the Government.

Thank You.

HEALTH - GROUP SUMMARY

Donald Steinwachs, Co-Chairman

With the Chairman's permission, Wes Clark and I have decided to divide the summary, so I will be very brief and describe issues pertaining to health data and Wes will discuss policy implications.

In the health data area, the group talked about a broad range of types of data, as I think most of you have, ranging from primary data sources to abstract summaries from journal articles that are available through bibliographic services.

I would like to make some general observations that came out of the presentations of the two speakers. First, there is a wealth of data currently in the health arena that is not being fully exploited. Some of the concerns about achieving maximum value from these data include the limited availability of funds for studies to examine the data, issues of data quality and reliability, as well as concerns regarding the content and definition of items of data, which are issues that others of you have raised.

Concerning data that goes into abstracting services, Medlar, Medline and others, the issues focused on questions of compatibility, since the classification systems and definition of terms are not the same. Thus, it is difficult to know whether a literature search and review is comprehensive before arriving at a judgment regarding the effectiveness of health care services.

I'm going to focus on key federal roles for a moment. There was considerable discussion about where the federal role is in health data, particularly since the federal government supports a major part of the health care system, many of the requirements within health care are imposed through legislative mandates, and the government supports health research.

Let me just run through these rather quickly without too much comment and then let Wes talk about some of the policy implications that came up in discussion.

A function that was viewed as a key federal role was in the initiation of new data systems. Many of the existing health data systems, ranging from bibliographic services to primary data sources on health status and utilization, would not exist if it were not for federal support. In some instances this represented initial support that led to support through the private sector based on subscription fees and user fees.

There was not much discussion, nor a concensus, about the extent to which the federal government gets involved in providing guidance as to standards of reliability, content of data bases and the definition of terms. I think there was a perspective emerging that there is an appropriate federal role in developing guidelines pertaining to content, classifications and reliability standards. The need is not so much in the cost and pricing area that was discussed previously, but more so in the definition and classification of clinical variables. Some of the concerns have been sharpened by more recent attempts to use clinical information contained in billing and discharge files to look at the effectiveness of health care.

The federal government enters this arena in terms of establishing data requirements where they pay for services as well as promoting national efforts in the uniform collection of health data.

The federal government is also involved in setting standards for privacy and confidentiality. There was a concern about access to data and the federal government's role. The federal government has paid for the collection of a wide range of data, from billing data to research data. The question is under what circumstances and how broadly should data be made accessible; what's the role of the person who develops the data, what's the role of the user, and what safeguards limiting the share of information are needed.

There was also some discussion of data retention policies. As many of you are well aware, data collected through billing systems are frequently discarded as soon as the bill is paid, and do not become available as a longitudinal data base. Questions were raised whether there should be federal policies encouraging data retention.

Lastly, in terms of issues of support, the federal government has played a major role and we would expect it to continue. This has involved support for both the development and collection of many types of health data as well as the analysis of data. There was considerable feeling that more support is needed for the analysis of data and attention should be given to setting priorities.

HEALTH - GROUP SUMMARY

Westley Clark, Co-Chairman

I couldn't help but be struck by the similarity of comments by each of the chairmen as you discussed natural resources and environment, agriculture, and health. Some of the policy issues that inhere in each analysis seem to be quite the same.

One of the critical elements that we discussed was the issue of commercialization of data bases and of the marketing of programs. The Hill has recently been confronted with policy situation of what should be done with the National Library of Medicine when the private sector complains of unfair competition.

There is no central concept on the Hill of what the role of the federal government should be in terms of defining what information is. There is the issue of role and there is the issue of definition that create major problems for policies. The conservatives seem to feel that if the private sector can do it, then let the private sector do it. Hence you can get administration posture on information which would allow the privatization of weather collecting devices that the government put into operation in the first place, without any kind of assessment of long-term implications.

One of the major themes that came up was the whole concept of orphan information. Orphan information turns out to be a major variable because it is the phenomenon that leads you into asking certain kinds of questions such as the questions that the natural resources panel raised. What happens when the government collects a bit of information that is not very popular at the time and no one else is interested in keeping a comprehensive account of what is developing, them somewhere down the line that information turns out to be very important? Does the government withdraw from that information market, even if that information is critical to a larger unprofitable body of information?

I appreciate being invited here because it is critical that the people who are participating in the information field begin to synthesize an input to the federal government so that people in legislative activities and legislative roles can understand the full magnitude of the problem. We are approaching a time where vested interest individuals color things to fit their vision of the information world. It is understandable and it is a rational thing for them to do. But it may make for poor public policy and therefore we do require the input from all other interested parties.

I think that's about all.

Concluding Remarks

Gilbert Gude
Director, Congressional Research Service
Library of Congress

Thank you for the opportunity of speaking for just a few minutes before your forum. From what I have heard, there has been a good spirited discussion concerning the need for a national scientific and technical data policy.

During the 1970s , economists and futurists alike predicted the coming of the post-industrial age -- an era in which the growth of information industries would begin to outpace traditional manufacturing industries. Today, we are in the midst of this transformation. In the last generation, more than two-thirds of the rise in GNP was contributed by the information and service sectors of our economy; during this time, nine out of every ten jobs created were attributed to these new economic forces. Currently, some experts estimate that the processing of knowledge or facts rather than of physical goods accounts for one-half of all economic activity in the United States, and that proportion is increasing through time.

Information and its associated technologies are the driving force behind this shift from a manufacturing-based economy to one involved in technology, information and services. Indeed there is a growing feeling in the legislative and Executive branches that the key to long-term economic growth lies with the so-called high-technology industries. The diffusion of computer and communications technologies into all segments of the U.S. economy not only can create new opportunities for employment and economic expansion, but also can play a vital role in rejuvenating the manufacturing and agricultural sectors of our society. Here at CRS, we too have become aware of the pervasive nature of these technologies as evidenced by the increasing number of requests coming into our research service.

As we move into a high-technology age, it is essential for both the public and private sectors to manage these information resources effectively to achieve maximum benefits. The success of the United States in the world economy has resulted largely from its preeminence in scientific, engineering and technical achievement. In the past, information from scientific and technical research and access to that data have served as the catalysts for innovation in our Nation's industries as well as increases in productivity. For example, most semiconductor researchers generally agree that it would have taken years longer to shift from vacuum tubes to semiconductor chips without the resources of Bell Labs which at the time were required to license their patents for use by firms in the private sector. In order for the U.S. to maintain this leadership in a high-technology future, it is crucial that information resources be managed even more effectively. Rational, comprehensive, and coordinated mechanisms must be established to guarantee that the volumes of scientific and technical data which are generated are also evaluated, packaged, and disseminated in ways which will best serve societal goals.

It is clear that today, scientific and technical information is increasingly being viewed as a valuable national commodity. The expansion of information technology into all aspects of our lives along with rapid advances in this technology have heightened this awareness. However, what remains unclear is whether the United States needs a national scientific and technical data policy, and if so, whether Government should assume a role to ensure productive utilization of these resources. Opinions concerning the appropriate role of Federal agencies cover a broad spectrum. Some advocate a marketplace approach and claim that information services such as those provided Federal weather and land-observation satellites and the National Technical Information Service are better handled by private sector companies. Critics of this approach maintain that such information is essential to the public interest and economic well-being of the Nation and should be supported by the Government. Other initiatives — including tightened security classification and proposed amendments to the Freedom of Information and Export Administration Acts -- have been undertaken to halt the flow of U.S. technology and technical data abroad. However, such actions frequently conflict with the traditions of public access to public records, open scientific communications and exchanges, and free trade.

These difficulties illustrate the types of public policy issues encompassed in the debate over the need for a national scientific and technical data policy. Some of these issues include the following:

— What is the role of Government, versus other sectors of society, in generating, evaluating, and disseminating scientific and technical data;

— What is the most effective means for acquiring and accessing necessary data for scientific advancement, public policy decisions, and the economic and productive well-being of commerce and society;

— What criteria should be used to determine which data are reliable;

— Who should be responsible for critically evaluating scientific numerical data and who should pay for or benefit from its evaluation and availability;

-- How should data responsibilities among public and private sectors be coordinated to minimize redundancies and reduce barriers to data exchange;

— What is the cost or value to society of having critically evaluated numerical data; and

— Does a national data policy require legislative initiatives or can its goals be accomplished by other means.

Conferences such as this one today can play an important role in finding answers and developing a consensus to these public policy issues. As you know, the work of the Congressional Research Service, like that of private and public enterprises, is highly dependent upon reliable, accessible data. However, unless the public has a clear understanding of the need for critically evaluated scientific data and its contribution to society, such a policy will not be realized.

Scientific and technical information will be vital to the growth of high-technology industries in the future as well as to the revitalization of today's manufacturing and agricultural sectors. Effective management of these information resources has the potential to boost productivity and increase the competitiveness of American business. The ability of policy makers to respond to changes brought by scientific and technical information and to provide guidance for the utilization of these resources will have a direct impact on the economic and social condition of the Nation in the future.

List of Attendees at Workshop, April 14, 1983

Susan Abbassi
Congressional Research Service
Library of Congress
Washington,DC 20540

Theodore Albert
Data Administrator
U.S. Geological Survey
Mail Stop 115
Reston, VA 22092

Harold E. Bamford
Div. of Information Science and Technology
National Science Foundation
Washington, DC 20550

William Bedesen
Committee on Science and Technology
2319 Rayburn HOB
Washington, DC 20515

Louis Blair
Senate Commerce Committee
Russell SOB
Washington, DC 20515

Michael Bloomfield
Office of Senator Riegle
1207 Dirksen SOB
Washington, DC 20515

Jane Bortnick
Congressional Research Service
Library of Congress
Washington, DC 20540

Carl O. Bowin
Woods Hole Oceanographic Institution
Woods Hole, MA 02543

Edward L. Brady
Assoc. Director for International Affairs
National Burean of Standards
Washington, DC 20234

Horace D. Brown
Merck & Co., Inc.
P.O. Box 2000
Rahway, NJ 07065

William F. Brown, Jr.
NASA-Lewis Research Center
21000 Brookpark Road
Mail Stop 49-6
Cleveland, OH 44135

Bonnie C. Carroll
Asst. Manager for Program Development
Dept. of Energy
Technical Information Center
P.O. Box 62
Oak Ridge, TN 37830

Alan Casper
Office of Sen. Metzenbaum
347 Russell SOB
Washington, DC 20515

Nancy Chapman
House Committee on Agriculture
1301 LHOB
Washington, DC 20515

Michael A. Chinnery
Director
National Geophysical Data Center
325 Broadway
Boulder, CO 80303

Mary Corning
Asst. Director for International Programs
National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20209

Marjorie Courain
Deputy Asst. Administrator
for Information Services
NOAA
3300 Whitehaven Street, NW
Washington, DC 20016

Vincent De Santi
General Accounting Office
441 G Street, NW
Washington, DC 20548

Nucgaek Devivian
Office of Space Science & Applications
NASA
Washington, DC 20546

Earl Dowdy
OTA
CIT Program

Greg Farmer
Office of Cong. MacKay
503 Cannon HOB
Washington, DC 20515

Manning Feinleib
Director
National Center for Health Statistics
3700 East-West Highway
Hyattsville, MD 20782

Hillary Frank
Office of Cong. MacKay
503 Cannon HOB
Washington, DC 20515

Joseph F. Fraumeni, Jr.
Associate Director for
  Field Studies and Statistics
National Cancer Institute
National Institutes of Health
Bethesda, MD 20205

William Gahr
Associate Director of Resources
  Community and Economic Development
General Accounting Office
3125 N. 4th Street
Arlington, VA 22201

Alan Gittelsohn
Department of Biostatistics
School of Hygiene and Public Health
The Johns Hopkins University
Baltimore, MD 21205

James A. Graham
Senior Research Associate
Deere and Company Technical Center
3300 River Drive
Moline, IL 61265

Viktor Hampel
Lawrence Livermore Laboratory
P.O. Box 808
Livermore, CA 94550

Joan Harn
Office of Technology Assessment
U.S. Congress
Washington, DC 20510

A. Hellman
Congressional Research Service
Library of Congress
Washington, DC 20540

James Hill
National Bureau of Standards
Bldg. 101, Rm. B113
Gaithersburg, MD

Michael Hoffman
Congressional Research Service
Library of Congress
Washington, DC 20540

David P. Jacobus
37 Cleveland Lane
Princeton, NJ 08540

Don Johnson
National Bureau of Standards
Washington, DC 20234

John Lamb
Program Manager, Chemical Separation and Analysis
Div. of Chemical Sciences
Basic Energy Sciences ER-142 (GTN)
Department of Energy
Washington, DC 20545

Emanuel Landau
American Public Health Association
1015 15th Street, NW
Washington, DC 20005

David R. Lide, Jr.
Chief, Office of Standard Reference Data
National Bureau of Standards
Washington, DC 20234

Walter S. Liggett, Jr.
Stat. Engineering Division
Center for Applied Mathematics
National Bureau of Standards
Washington, DC 20234

Cyrana Longest
Office of Senator Thad Cochran
346 RSOB
Washington, DC 20510

Gail Marcus
Congressional Research Service
Library of Congress
Washington, DC 20540

Buddy MacKay
U.S. Representative
503 CHOB
Washington, DC

Nancy Miller
Congressional Research Service
Library of Congress
Washington, DC 20540

Michael Moran
Chief, International Organization Div.
Inter-American Institute for
Cooperation on Agriculture
Inter-American Cooperation Agency
1889 F Street, NW
Washington, DC 20006

S. Passman
National Science Foundation
Washington, DC 20546

Andrew H. Pettifor
Office of Science and Technology Policy
Executive Office of the President
New Executive Office Bldg.
Washington, DC 20500

Donna Porter
Library of Congress
Washington, DC 20540

David Richman
Office of Energy Research
Department of Energy
Washington, DC

Gloria Ruby
Health Program - OTA
U. S. Congress
Washington, DC 20500

Hubert E. Sauter
Defense Technical Information Center
Cameron Station
Alexandria, VA 22314

John Scheft
Office of Senator Metzenbaum
373 RSOB
Washington, DC 20500

Frank Sobieszczyk
Office of Deputy Under Secretary of
Defense for Research and Engineering
Department of Defense
1400 Wilson Boulevard
Arlington, VA 22209

Don Steinwachs
Health Services R&D Center
Johns Hopkins University
624 N. Broadway
Baltimore, MD 21205

Mark H. Tennant
Office of Senator Thurmond
218 RSOB
Washington, DC 20500

Milton Terry
Consultant
Bell Laboratories MH7F-502
Murray Hill, NJ 07092

Thomas M. Usselman
Geophysics Research Board
National Academy of Sciences
Washington, DC 20418

Anthony Villasenor
Chief, Information Systems Office
Office of Space Science and Applications
NASA Headquarters
Washington, DC 20546

George H. Wakefield
Energy Information Administration (EI-443)
Department of Energy
1000 Independence Avenue
Washington, DC 20585

Hal Wallach
Principal Statistician
General Accounting Office
441 G Street, NW
Washington, DC 20548

Randy Wedin
Committee on Science and Technology
House of Representatives
2256 RHOB
Washington, DC 20515

E. C. Weiss
Director
Div. of Information Science and Technology
National Science Foundation
Washington, DC 20550

Stan Whetstone
Program Manager for
  Low Energy Nuclear Physics
Division of Nuclear Physics ER-23 GTN
Department of Energy
Washington, DC 20545