The United States Government and Chemical Information*

ANDREW A. AINES
National Science Foundation
1800 G Street, Washington, D. C. 20550

Received September 14, 1971

It is always a pleasure to talk information to chemists. I know of no other profession in science and technology that measures up to the accomplishments and the sophistication of your community in the field of information. This statement is based on a decade of interaction with men and women of chemistry striving to attain progress not only for chemistry, but across the board in information arts and sciences. I salute these good allies.

Early in my salad years in the information business, I wondered why chemists took this no nonsense approach, eventually concluding that it had much to do with the importance of quality information and data for the individual chemist in his daily work and the result of some process of selection of profession, anything but random, both of which combine to make for a unique *homo chemico-informibus.*

I must admit to a long-time interest, but not necessarily expertise, in the chemical information business. People like Pep Vlannes, Paul Olejar, Dave Jacobus, Al Feldman, Gelberg, George Hager, and others should be held responsible for my conversion. I cut, then perhaps dulled, my teeth on the battle of notations systems that raged about a decade ago; it was then that I learned about Dyson, Wiswesser, structure-searching by means of typewriters, and other—what then appeared to me—exotica. I began to meet many chemists in the U. S. Government who were searching for advanced chemical information systems. This acquaintanceship spread to people in the American Chemical Society and the Chemical Abstracts Service: Milton Harris, Bob Cairns, Byron Riegel, Dale Baker, Fred Tate, Dick Kenyon, and many more. To keep acquainted with the trends in the chemical information area, I have made occasional trips to the mecca of chemical information at Columbus, Ohio. Special quarterly meetings were arranged for top government officials to interchange information and views with the ACS Board of Directors.

While wearing my recently discarded OST hat, EOP backup was given to the National Science Foundation in its trailblazing support of CAS in establishing the registry system. We all fervently hope that the millions of taxpayers' dollars that have been provided for R&D at Columbus will bring great dividends in the future. Support was also given to the NSF program that sought to improve the market for chemical data by assisting university centers to create programs to interact cooperatively with CAS in dissemination of data products.

Another bit of pioneering with ACS was undertaken, when we invited it to work through the 23-nation Organization of Economic Cooperation and Development in encouraging actions, to lead to better international sharing of chemical information and data. ACS made extraordinary progress in this program, including provision of a training program for interns at CAS, which has received NSF support and much applause.

While my assigned topic today is to talk about U. S. Government and Chemical Information, those of you who know me could have easily predicted that it would be impossible to keep me down only on *that* farm. The least inspiring thing I could do for you would be to give you an atom-by-atom cook's tour, an annotated directory of what is happening only in the Government wing of the chemical information business, important as this may be. To understand what the Government is up to in the chemical information field, it is also important to get a barometer reading on what the Government men sense is happening in the larger information area, what trends and forces they are watching, and those to which they react. It is my intention to escort you on a tour along the trail that interconnects a number of peaks in the information range. I hope that I select the most appropriate peaks and avoid leading you into cul-de-sacs and treacherous terrain. I also hope I do not traverse too much ground covered by earlier speakers.

If I were to choose the highest peaks that tower over the others, I think my first selection would be the extraordinary proliferation of information and data. You see it in chemistry and in a number of other fields; each annual accretion forming a mountain in itself. Unfortunately, the higher the mountain grows, the more it becomes an amorphous blob, and the harder it is to extract the specific information one wants or needs. Information systems now being built to handle the current data store, rather than the volume that even conservative growth curves can predict for 1975 and 1980, could be an exercise in futility, if they are based only on simple linear extrapolations. Handling a bushel *vs.* a ton of apples is not the same kind of an operation.

The second peak that looms high in my view is that of information-processing technology and techniques—the extraordinary, emerging tools that offer us some hope that we can rapidly and successfully search through the mountain of information and data electronically. Here, too, we find irony. These tools do not come cheap. Their mastery is still a long way off. Few of us want to give up the printed word systems that are our heritage, our precious security blankets. Mechanized information systems take a lot of capital. To get the resources, there is almost a Faustian recompense few of us want to pay. Employing them as our valued servants, we can find all too ruefully that they are eating us out of resources we need for other purposes. We obviously cannot trust them sufficiently to switch completely away from inkprint information systems, so that the traditional and the nonconventional systems must operate in parallel. Just as we find that money needed to resurrect cities, overcome poverty, and restore the quality of the environment is in short supply—so is it hard in the competition to obtain the funds to establish and maintain computerized information systems. But while electronic data processing may not be the best show in town, it is the only one that offers us hope to scale the slopes of the first mountain.

The third peak in our range involves cultural and institutional change. More specifically, I mean that differential ability of men and their institutions to take advantage of the opportunities presented by the new technologies. The world famous brain researcher, the late Warren McCulloch of MIT, pondering on the subject of the communication revolution and if the world can survive it, observed that we have about twenty years to find out. My old friend, Edward Safford, reflecting on McCulloch's comments,

*Address presented at the Division of Chemical Literature Luncheon, 162nd Meeting, ACS, Washington, D. C., September 14, 1971.

pointed out that he really was talking about the sociological and psychological results of a new environment which establishes completely new patterns of relationships among people, and a completely new relationship between knowledge and man. Safford, a well-known publisher and communications expert, believes that the impact of technological change is not that of technology on man, but rather that of man-upon-man through technology. The technological advances and devices have merely extended man's capability of influencing man, and it is *this* phenomenon that really composes the communication revolution.

I do not want to dwell on this peak too long, but it deserves a long residency in my judgment. As the world goes through a cosmic process of rewiring its communications, the power and challenge of McCulloch's words will be better understood. I have long ago come to the conclusion that the problem of the improvement of scientific and technical communication and information-handling, including that of chemistry, is a subset of the larger problem and cannot be solved in isolation.

But let us move on to another mountain. In just a couple of decades, we have been propelled from an era of information scarcity, when we battled with recalcitrant Mother Nature for a few nuggets of new knowledge which she begrudgingly gave up each year, into an era of information superabundance. With something akin to frenzy, we have almost lemming-like continued to beget new knowledge, too often oblivious of the quality. In a sense, we have mutated into a fast-breeding information production machine. The cry: "Publish or Perish" is seemingly changing into: "Publish, then Prepare to Perish in the Flood." Whoever invented the advertising slogan, maybe it was a chemist, "Won't somebody turn off the bubble machine?" must have been thinking of the dilemma of the confraternity charged with keeping the literature flow within bounds. Physical scientists, including chemists, wear white hats comparatively in this one.

This issue reminds me of a trip I made to a Government laboratory a few years back as a staff member of the Office of Defense Research and Engineering and long before I became involved in information and communication. It is a true story. Thumbing through a sheaf of technical reports prepared at the lab as I made my tour, I noticed one that described a study of four commercial can openers. In deadpan fashion, so dear to Consumers Research and Consumers Union, it recounted how each can opener was lovingly tested and ranked one with another. Not being able to contain myself with the thrill of this astonishing research revelation, I asked the officer in charge what prompted him to disseminate a report on such a trivial subject. His reply has haunted me ever since; it was so monstrously reasonable. He pointed out that his laboratory turned out hundreds of technical reports annually, many of them of top quality, even for unpublished literature, but there were times when report production was low. Because he had a corps of civil service editors, artists, printers, writers, and others in the operation whose work showed up on performance charts with the usual ascending curves, as happened on occasion, a few marginal reports made their appearance. That is the inevitable consequence of what happens when you inflict a production schedule more appropriate for Detroit on a research and development facility.

You can anticipate my next comment. I have been obsessed by the fear that unless we take great care when we mechanize, nationalize, and internationalize our great information systems of the future, the machines will produce materials as a function of their being, and a large part of the production will nightmarishly turn out to be reports of tests of can openers. Earlier I pointed out that the third peak dealt with cultural and institutional change. Perhaps I can sum it up by simply stating that machines start out simple and become more sophisticated, while human attitudes, customs, conventions, and institutions come equipped with a built-in lag mechanism. Part of the resistance to change is absolutely mandatory, makes a lot of sense; but on the other hand, too much makes it difficult to take advantage of the new tools we can use to solve other problems. Nevertheless, the one thing none of us wants is to change the productive open-endedness of science by mortgaging our souls to data networks that may tend to inhibit the creativity of scientists.

For the past two years, I have been listening to statements made by scientists, engineers, industrialists, government officials, and others who plead for funds to obtain data they need badly to accomplish their tasks, their missions, and their research. For example, diverse communities, like the Council of Economic Advisers, scientists who specialize in seismology, and members of Congress are joined in a plea for better data. Urbanologists, poverty workers, and ecologists issue report after report about the harmful lack of required knowledge. Social scientists call for the systematic collection of social indicators; environmentalists make the welkin ring with prescriptions for a global monitoring system. A recent report on this subject calls for the creation of a program that will result in a time series of 112 EQ indicators and indices. An earlier report for the Marine Council called for a multiyear program that would cost anywhere from a half-billion to a billion dollars to make better data available for this community. A congressman has a bill before Congress seeking the establishment of a national environmental data system. Another congressman wants to have a research and development project data center, under the control of the General Accounting Office, to make available data pertaining to Government research and development programs to all who want these data in and out of the Government. Report after report on health problems lament the lack of key information and data. Ralph Nader and other consumer champions complain about the need of data that would permit the consumer to make better decisions about their purchases. Officials in the National Oceanic and Atmospheric Agency candidly admit that they are unable to mine the immense bank of data they obtain because of the lack of resources and people, making a persuasive case that their data are attractively ready for further exploitation. I have visited information analysis centers specializing in certain kinds of physical data that are years behind in the preparation, announcement, and dissemination. I could go on and on. This is a fourth mountain, a mountain of expressed need for essential data, a mountain of half-processed, unprocessed, or exploitable data. I distinguish this mountain from the first, which has a sign describing it largely as *documents and literature,* rather than data. I associate this mountain closely with the second, the advent of the new information-communications technology, largely because the supplicants for better data vibrate to the promise of the technology and clamor for its employment.

Those of us who have been seeking ways and means to harness the new information-processing technologies are painfully aware of the difficulties and hazzards that stand in the way of their full employment; but like our data-hungry associates, we, too, are compelled to seek progress as rationalists who recognize the opportunities for mankind in what I like to refer to as cyboelectronics.

While we Americans relish the advantages inherent in our pluralistic society, we recognize how difficult it is to construct a national information system or systems that will combine the private and public programs. While many scientists and information specialists have come to the conclusion that only the Government can accomplish this objective, it is not that simple, I can assure you.

Earlier, I mentioned an activity in the Government which is leading towards an information system for environmentalists. It is one of many, actually. In the Federal Government alone, I estimate that there are upwards of 50 programs scattered through various agencies that directly or indirectly involve the environment. More than that, programs are springing up in international groups like the Organization for Economic Cooperation and Development and the Council for Europe. The mass of documentation that involves the many facets of environmental quality is as explosive as any field I am aware of today. Although the Government is seeking to comprehend and control the literature in a bibliographic sense, I have a proposal on my desk that shows the interest of an unexpected candidate like the Woodrow Wilson International Center for Scholars ready to take on the task of creating a selective, annotated environmental bibliographic service. The only way that the Government can influence a nongovernment group like the Wilson International Center, indeed, if it still plans to go ahead, is if that group seeks or requires Governmental funds, otherwise it can organize the program as it sees fit. The talent and the resources needed to build a national system, or a large component of one, are scarce commodities. I hope that the Council on Environmental Quality and the Environmental Protection Agency will be able to build up enough steam to obviate the need for competitive systems.

I wish I could see it possible for the private sector, both profit and nonprofit, to take on a chore of this magnitude; but I see little hope for this, especially if the revenue to maintain the information system must come primarily from the sales of services. On the other hand, by means of contracts of grants, some components of such a system can be effectively handled outside of the Government, assuming that we can continue to make headway in the kind of team effort exemplified in the venture of the National Science Foundation and Chemical Abstracts Service.

Having mentioned environmental quality information activities, it might be helpful to discuss for a moment a dilemma we face. Picture an information pentagon with five surfaces labelled: disciplines, mission accomplishment, problem-solving, commercial enterprise, and educational institutions. We can add additional surfaces and refine our list further; but for our purposes, let us stick with the five. In analyzing the EQ information problem, we can see that there is much difficulty in trying to create a truly dedicated EQ data system, unless we are prepared to duplicate massively the information holdings scattered throughout the pentagon. If we agree that we should take the logical course and seek maximum interchange, rather than create a monolith, the task of coordination is still immense whether we use teleprocessing or not. If all groups seek to carve out bigger slices of the action and quarrel about booty and boundaries, they may create an Augean Stable that will take more than a Hercules to clean out. It is my expectation that if anything leads us to national information systems development, it will come from our realization that well-coordinated information and data systems are needed to contribute to the solution of our complex societal problems, systems characterized with reasonable flexibility to cope with continuing change.

The 1970 Study of Critical Environmental Problems, cosponsored by MIT and other groups, identified eight critical global environmental problems, most of which are of interest to chemists. I will not have time to go into any detail or even read them, in full, but they involve long-lived toxic chlorinated hydrocarbons, toxic heavy metals in assimilable form, increasing carbon dioxide content, and particulate matter in the atmosphere, contamination of the stratosphere by transport aircraft, and the escape of petroleum oil and nutrients into the environment. You know them all; it is quite a list. I mention these problems because they cast light on the difficulty of filling information and data needs and the complexity of the multidiscipline, multimission interlock. We can ask—what are the proper roles of Government, disciplines, industry, and others in obtaining the requisite data needed if we are to improve the environment? How do chemists make their contribution to this team effort? It is too easy to state glibly, as some do, they can help by avoiding the use of chemical pollutants in the first place.

Yesterday, you heard Henry Kissman talk about his toxicological program at the National Library of Medicine. I do not know if he mentioned the excellent panel he led which looked into the EQ data programs of the Government. The SEQUIP project, an acronym which stands for Study of Environmental Quality Information Programs, deserves as much attention as the MIT report. It is a landmark effort which will interest all chemists, I know. OST is making a special effort to release this report, the findings of which have already appeared in the Congressional Record.

Now a word about another study group, which was chaired by Barbara Murray of the National Cancer Institute. Barbara was asked by COSATI to explore the feasibility of developing a cooperative program mutually beneficial to all Federal agencies that have information interests, to consider what steps are needed to bring agencies and the Chemical Information Service program closer together, and to elicit recommendations from Government officials on how to make progress leading to a stronger national and international chemical information system.

COSATI has just received the draft report from the *ad hoc* task group. It has not yet been analyzed and distributed to the COSATI members and observers for comments; hence, I am not at liberty to disclose its contents. It is a positive report, opting for action. It also contains a new Directory of Federal Chemical Information Activities that will be very useful. I would like to congratulate Barbara and her task group for the hard work they did to prepare this promising study.

Now, let me change gears and provide you with a few concluding observations. These are not official, I must point out, and if they coincide with the views of others, I commend them for their vision, wisdom, and perhaps nerve.

The so-called knowledge industries are going to grow in the next decade, whether we become a post-industrial society or not. The "for profit" information sector must be an integral, important part of the community. Its ability to market information and data that people want efficiently must be yoked to the programs of professional societies and Government agencies. The part of the action the "for profits" carve out for themselves will depend on their ingenuity and the economy and effectiveness of their services.

The proliferation of literature and the high costs of establishing and maintaining mechanized information systems are going to continue to be problems, even with an upturn in the economy. Government subsidies, such as page charges, should not be considered to be a desirable, *long-range* program, although they may be valuable during the period of transition we are going through.

I believe that one of the answers is in new management initiatives. For example, professional society abstracting and indexing groups might consider a new national service corporation operating under contract to achieve economies and excellent user service by means of high volume operations, expert management, and reduction of duplication and overlap, both of materials handled and overhead. We are moving in this direction; we could move faster. The new corporation would be built on present resources and use the best people now involved.

Information people must become much more interested in information utilization than they are or are permitted to be. The ex-

ploitation of knowledge must be considered a higher goal than mere information-handling, as important as efficient handling may be in the communication processes of society. Information analysis center development is one practical answer for our community. Re-structuring of our technical libraries into new viable service centers is another.

Well, ladies and gentlemen, for the last half hour I have been competing with your digestive and other life support systems for your attention. You have been most kind in listening to me. I will return the kindness by concluding my remarks with the hope that chemists will continue to be in the vanguard of information progress.

Factors Affecting Dissemination of Chemical Information*

DAVID E. GUSHEE
American Chemical Society
Washington, D.C. 20036

Received September 24, 1971

This paper demonstrates that the value of information and, hence, the design of information systems, whether for a company, a university, a learned society, or an individual, is in the early stages of a significant change in structure. We are struggling with a heritage of source orientation in an era of increasing orientation towards problem solving. The signs of response to changing circumstances are increasing, but are still limited in number and varied in nature.

Two major driving forces are changing the structure of chemical information systems. One of these is sheer volume; the other is economics. The impact of increases in rate of generation of scientific information has been well recognized for at least two decades. But the role played by economics is changing in that economics is becoming a more stringent constraint. Currently common assumptions about the value of information will thus be challenged and, to a significant measure, overthrown in the next ten years.

As things stand today, most institutions can be counted on to buy most books and journals published in their fields of endeavor. These institutions take pride in their library holdings and tend to measure their effectiveness, at least in part, by the ratio of information requests filled from stock, so to speak, to those handled by going outside their own resources through interlibrary loans, special purchases, and other relationships.

There has been little attention paid to operations research on such matters as frequency of use, cost per use, dead time, and other variables common in inventory and distribution and their control. One must, of course, distinguish between academia and industry in such statements, because the one has essentially all of knowledge as its field and scholarship as its output, while the other has clearly defined areas of knowledge to work in and new products and processes for growth and profit as its output. Industry thus has always had a greater basic ability to quantify the evaluation of its information resources compared with academia. Nonetheless, industry's growth and profitability have been high until recently and thus have not led to sufficient pressure to cause penetrating analyses and painful decisions to become general.

The literature in this field is scattered and data are hard to get. The nature of things to come is emerging gradually, and studies of cost effectiveness and the like are beginning to appear more frequently.[1,2] These studies show that the value of an information system can be quantified, but that there must be prior agreement among those concerned on such things as objectives in speed of response, comprehensiveness, and the like.

There has been some attention paid by librarians to the consequences of continuing current policies and practices.[3,4] Most of these studies have been stimulated by proposals to build new libraries, and most have actually led to decisions to proceed with such construction (although recently some of these have been delayed). From the early post-war period to the end of the 60's, average library acquisition costs increased at over 11% per year, total holdings increased at just under 5% per year, and unit costs at about 5% per year. The average university library was projected to have holdings of almost 3 million volumes in 1980 from just about half that in 1968.

Although university librarians and administrators have been concerned about these projections (acquisition budgets, for example, would reach an average of about $2 million per year by 1980 from their current $600,000 or so), none dared to propose as a matter of institutional policy that such increases in library costs were unacceptable. Rather, they were unavoidable, according to the then-current conventional wisdom, and, I infer, the administrators concluded that growing profits, Uncle Sam, or God would provide.

These projections were based primarily on extrapolation of data from a period of unprecedented prosperity and occurred prior to full recognition of the profound changes about to occur in the funding of the knowledge industry, of which science, and in turn chemistry, is a part.

University library acquisition budgets have not grown in the past two years at the rate of the previous period. Although I do not have enough data to quantify the situation, I estimate that acquisition budgets are at best holding their own against inflation, while publishers' prices are increasing at an average rate of 20% or so per year—the ACS is no exception, having had price increases over the past three years that average about that order of magnitude.

Again from estimates not quantitatively supportable, I feel that acquisitions budgets will continue to remain at about current levels for the next several years. The basis for this prognostication is a probable leveling off of Federal funding of science, except in medicine and biology, and an

* Presented in Symposium on Appraising Technology and Information Systems, Division of Chemical Literature, 162nd Meeting, ACS, Washington, D. C., September 1971.