NFAIS Newsletter 27(2):44-52 April 1985

MAPPING THE MULTI-DISCIPLINARY WORLD OF NFAIS!

Eugene Garfield, Ph.D.
President
Institute for Scientific Information

Lynne Neufeld informs me that the Miles Conrad Lectures were published by Information Resources Press in a volume entitled Abstracting and Indexing in Perspective.[1] This confirms Garfield's fourth law -- or Murphy's Nth Law. This law states that you will be invited to give the 17th Annual NFAIS lecture immediately after the first 15 have just been published.

The Miles Conrad Memorial Lecture is supposed to be on a topic "suitable to the field of abstracting and indexing information, but above the level of any individual service." What that means is, "Don't talk about yourself or the institution you represent." Since my fifth law states that this is impossible, consider these opening comments off the record. If, 16 years from now in 2001, Saul Herner doesn't want to include my remarks in the second volume, I won't be offended. But there is simply no way for me to begin a Miles Conrad Lecture without getting personal.

Since I have a poor memory for certain kinds of events, I was not exactly sure when I first met Miles Conrad. So I checked my files on Miles -- that has a nice ring! But that didn't help. So we called Hazel Philson. She confirmed that I met Miles when he was still at the Library of Congress, before he came to BA in 1953.

About this event I do remember saying to myself, like Sancho Panza in Man of La Mancha, "I like him." He had a soft, deliberate way of speaking, and his voice had a special, deep resonance. Phyllis Parkins may recall, however, that I knew about Biological Abstracts (BA) first through John Flynne. That was during the time I was still at the Welch Indexing Project, because I did not start my career in Philadelphia until September, 1954, when Ted Herdegen of Smith-Kline asked me to take a temporary assignment there.

Like Ted, Miles had known me from the John Hopkins project. He had heard that I was working as a Documentation Consultant. At that time, I was considered a "machine maniac." The word "computer" was barely part of our vocabulary. Miles asked me to do a study of the systems and procedures at BA. This was one of my early free-lance assignments. Out of this work came an unpublished report called "Biological Abstracts in an Era of Automation."[2] The report described the use of perforated paper tapes -- so symbolic, in those days, of automated procedures. It is difficult for this generation to realize what a painful transition we made from the punched card or Hollerith card to the Flexowriter paper tape, then to the hybrid card containing both, and eventually to magnetic tape.

Miles Conrad Lecture, National Federation of Abstracting and Information Services 27th Annual Conference, Arlington, VA, March 3-6, 1985.

But this report also tells us a great deal about the agonies of moving from one technology to another. It deals with the eternal human fear of the unknown, the concerns for job protection, the fear of change and how unnecessary that fear usually turns out to have been. Automation has created more jobs than it abolished. Nevertheless, if you were one of those who was temporarily displaced, it was painful.

The BA report also helps us focus on the changing perception of intellect. There was a time when human editing included functions that today are routinely performed by a word-processor. Although the future potential of artificial intelligence is often grossly overstated -- as was the case for mechanical translation in the 1950's -- it is clear that our perception of human intelligence has changed considerably in the past 30 years. I can recall how proud I felt to publish a paper on "The preparation of subject heading lists by punched-card machines."[3] The younger generation may find it hard to believe that this was a task considered totally beyond the capability of a machine.

Miles Conrad and many of his contemporaries who could endorse what I am saying are no longer with us. But there are enough of us young upstarts around who can still remember those halcyon days and the agonies and the ecstacies we shared. This is not merely nostalgic self-indulgence. It is important to realize that there were fundamental socio-economic factors that led to the separate foundation of the National Federation of Science Abstracting and Indexing Services (NFSAIS) and the Information Industry Association (IIA).

Two decades later, the non-profit/for-profit dichotomy may seem less important than the database producer/online vendor dichotomy. As with nation-states, today's adversaries may become tomorrow's allies. Historically, it may seem to have been pointless in getting too worked up about these periodic expressions of territoriality. But like the person who lost a factory job to a robot, it makes no difference to the database publisher who goes broke whether it was a non-profit foundation or AT&T that administered the ultimate coup de grace. So the microhistory of information science, like any of the areas of human endeavor, can only be understood in terms of the politics then prevailing.

With this long introduction, what can I contribute that is neither personal or self-serving for ISI[R]? You have all either read or heard about the sooth-saying prognostications of John Nesbitt, Alvin Toffler, and Peter Drucker (as recently as the January 9 issue of the Wall Street Journal[4]) about the "informatization" of society. Whereas 32 years ago, Saul Herner could ask the question in Science Monthly, "Technical Information -- Too Much or Too Little?"[5] his kid brother Garfield was moved just a few weeks ago to talk in Current Contents (CC[R]) about information overload.[6]

We are all swimming in a sea of information. This is not a new metaphor. When we organized the first Symposium on Machine Techniques for Scientific Documentation back in 1953 in Baltimore, Lowell J. Reed, vice-president of Johns Hopkins University, said, "We are drowning in a flood of information."[7] That classic remark was picked up by the wire services and led to one of the most significant cross-disciplinary links in the history of information technology: the cross-breeding of legal information methods (Shepard's) with scientific information methods and the ultimate creation of the first large-scale multidisciplinary unified index to science. I say "large-scale" multidisciplinary index because there were many earlier, but smaller,

predecessors, when science itself was still small. And certainly Index Medicus, CA, and BA were, and remain, multidisciplinary, but each is positioned as discipline-oriented.

Every new generation seems to be drowning or swimming in a progressively larger sea of information. It is now so large that we have created a whole new specialty within information science, variously called scientography, the geography of knowledge, or the mapping of science -- geo-epistemology, if you wish. You can't follow a ballgame without a scorecard. And you may get lost in the sea of information if you don't have a map of science. This is what we used to call classification, but in its latest incarnation, it has taken on some new qualities.

I will not break the rules of the game by describing to you one of the future information systems of ISI. But I will show you a few pictures to entertain you and at the same time stress my main point. In fact, this theme was one I proposed for an ISI ad campaign, but it was never executed. So let me try to put you on the map -- by which I mean, of course, let me try to put your discipline on the map of scholarship.

Maps of the literature tell us many interesting things about the ebb and flow of information specialties. Each of you in this room can benefit from this because your discipline orientation demands that you know something about the history -- past, present, and future -- of your discipline. ISI does not have an exclusive on the many ways that one can map knowledge. For instance, recently in Physics Today, the American Institute of Physics published its revised classification system. It is the result of an enormous human effort that boils down the deliberations of hundreds of experts.[8] Each and every classification system we can name -- from MeSH to BA's Cross Index -- is the result of enormous systematic intellectual effort. Each one has its own particular shortcomings, but we eventually conclude that we must be doing something right. So the advantages of conventional classification systems will be combined with such techniques as co-citation clustering. Indeed, our work has stimulated other efforts such as, for example, the use of PASCAL indexing terms to create cluster maps. I have no doubt that eventually a hybrid system will emerge. The Science Citation Index® (SCI®) started as a pure citation index, but then added permuted title word indexing.

NFAIS members produce or process over two million abstracts per year. You can't say there is a shortage of abstracts. We also index in a variety of ways those millions of articles and books we abstract. In addition to this, there are at least 50 to 100 thousand review articles published each year. And here I mean review articles of at least ten pages with an average of 100 or more references, such as you would find in Annual Reviews. So there is no paucity of reviews. Added to this, we hear that in the near future, thousands of primary journals containing at least 500,000 articles will be accessible by full-text online, or on laser plates, compact discs, or some other medium.

So with all these information riches, why are we suffering? Why so much feeling that we cannot cope? That is why I mentioned earlier the phenomenon of information overload. While we have traditionally relied on the truism that all previous generations since the invention of printing have complained about the flood of information, we always have to consider the possibility that someday there may come an "information winter." The nuclear holocaust is not the only

cataclysm that can destroy mankind. However, I can also foresee the implementation of self-correcting mechanisms to prevent the information holocaust -- unless politicians intervene to guarantee that we continue on course to the Information Winter. The move towards standardization has increased, by virtue of the online era. It will be accelerated as full-texts go online. But that in itself won't prevent problems of information overload -- it will only increase them.

I don't read much science fiction, but it is easy to imagine all the world's scientific information stored in a small vault of electronic information, not unlike the old vision of Memex, or the World Brain of H. G. Wells. Then, suddenly, the power is turned off.

My personal micro-system went down recently and I felt a sense of impending apocalypse. I could not find a phone number; I could not locate a needed reference; I was trapped while waiting for the disc-fix. The information holocaust had occurred, and I yearned for my old-fashioned 3x5 cards and file folders. But it was too late: I couldn't make the call in time and I couldn't cite the reference I wanted in time for it to be useful.

Recently, I've often discussed the transition from the age of bibliographism to the age of encyclopedism. I think we NFAIS members have done a pretty good job of gaining "bibliographic control," as we used to say. The NFAIS statistics speak for themselves. But in the next phase of the information revolution, to simply produce the electronic version of what we have already done is not sufficient. There needs to be a new vision of our role as information condensers.

While we have not exhausted the possibilities of our existing intellectual procedures, the new technologies present opportunities that were impractical before. Information technology is merging with instrument technology. Hardware and software cannot be separated.

Whereas research directors could, in the past, passively regard the use of information as optional, the computerization of information has raised our industry to the level of scientific instrumentation. So you must aggressively market your wares in the face of competition from hardware, materials, and so on. This means that you will increasingly become commercial, and the non-profit label will have less meaning. It is to be expected, therefore, that there will be increased cooperation, or what the private sector calls "joint ventures." And this is why I believe we need to pull together to avoid an information winter -- and use NFAIS and IIA to do so. And if society is not so foolish as to indulge in nuclear holocaust, we can certainly avoid information disaster.

Passing from the philosophical to the more mundane, I would now like to take you on a National Geographic-type expedition to explore the wonders of the world of science. I guess you thought I'd never get to my stated theme -- mapping the world of NFAIS.

I'm well aware that these snapshots of past and present science are imperfect. The SCI and its companion, Social Sciences Citation Index^{R} (SSCI^{R}), together cover only about 700,000 published articles per year. We conclude from the NFAIS statistics -- and in spite of considerable overlap -- that we could easily include another 300,000 articles per year. This might change some of the

details or increase the number of smaller topics we now omit. That's like omitting a subject heading from you thesaurus for some minor topic.

Nevertheless, I think that most of you would agree that it is valid to speak about the significant literature of science. And I believe that the significant journals from most of the NFAIS disciplines are represented in the ISI databases. So the picture we obtain of worldwide scholarship is reasonably accurate, in spite of its minor omissions. Of course, our Soviet, Third World, or even our French colleagues might assert that we don't include enough of their literature. But when I lectured in the USSR recently, audiences confirmed that we had portrayed Soviet research accurately. Indeed, later, you will see some of the most active research fronts we identified. And I've had similar experiences in India and elsewhere, because so much of Third World research is reported in the international journals of science.[9]

How do we make a map of science? There are many different methods one can use, not the least of which are the old-fashioned ones. If you look in older sources, you will occasionally see maps of the literature. But I don't have time to give you an historical account of scientific mapping.

One of the most modern methods of mapping the quantitative relations between objects -- whether it is in statistics, geography, or biology -- is called multidimensional scaling, or MDS. In our use of MDS, we want to portray the world of science in terms of the quantitative outputs and connections between articles and books. You can do this in many ways -- by studying author linkages, word frequencies, journal networks, etc. We have found that the most helpful method is, first, to use ranked citation frequencies and then co-citation clustering to identify where the action is. Since this is essentially a language-free method, we can do this on a multi-disciplinary basis with minimum difficulty. The details of these procedures have been widely reported. So let me just show you the ultimate result of these methods.

Imagine you are looking through a telescope and you are going to gradually zoom in on a very narrow point. We will start by looking down from outer space, at the "world of science." In a quantitatively based system, it should not be surprising that the central, most active area of this world is a continent called "Biomedical and Physical Science." Let's call it Natural Science -- accounting for easily 75 percent of the scholarly literature. There is also another smaller continent called "Social, Behavioral, and Biological Sciences." These two continents are connected by a thin, short strip of "land" -- much shorter, for example, than the one connecting agriculture to either of them. These strips are not plainly visible from far out in space, but when we zoom down, we will be able to observe what they are.

Zooming in, if we ask the computer to show us the map for "Social, Behavioral, and Biological Sciences," we see it includes two sub-continents -- Social Sciences and Systematic Biology -- connected by an isthmus. Before I tell you the name of the isthmus, try to guess what it will be called! What do social scientists and systematic biologists share most in common today? These disciplines are linked mainly by the mathematical and statistical methodologies they share, thus the isthmus is called "Multivariate Analysis." On hindsight, this may not be surprising, but the purpose of the exercise is to exercise a little foresight.

Look at it another way; a Japanese woman scientist, Junko Matsubana, noted the wide spacial separation between Demography (Human Population Studies), on the far left of the map, and Animal Population Studies, at the top right. She asserted that these scholars seemed rarely to talk to one another. The map would seem to confirm her subjective impression. Were we to examine the content of a good abstracting service that reports on all types of population studies, one might get a very different picture.

Let us now zoom in further on the sub-continent of the Social Sciences. Here you can see some of the areas that are covered by such services as Psychological Abstracts, Sociological Abstracts, and so on, e.g., social psychology, women's movement, behavioral disorders, political systems, etc.

On the other hand, if we zoom in on the large continent of the Natural Sciences, we can observe the Life Sciences on the left, then chemistry, physics, computer science, over on the right.

Now let's zoom in on the sub-continent of physics. This, of course, presents a somewhat different picture of physics than we get from the AIP qualitative classification system.

We can continue our zooming exercise by selecting from the physics subcontinent an area we'd like to examine in more detail. So we select the point and press the button for the area called General Physics. As it turns out, within this map we'll find the area that will prove to be one of those in which our Soviet colleagues are very active.

To illustrate that, let's zoom in on point #50, the subcontinent on the left of the physics map called Solitons. From here we can now select one of several different research fronts on solitons. The research front for one of the soliton topics includes several core papers by Soviet authors. However, the total number of core papers in this research front is 27. This demonstrates vividly how much use is made of Western literature by Russian scientists, and vice versa.

We can also look at the ranked number of papers published by different countries on this subject, and here we see the high degree of commitment to this field. Of 245 core papers for another sub-topic in soliton research, "Methods and applications of soliton solutions for nonlinear equations," 59 are by Soviet authors, 52 by USA, 31 France, 31 Federal Republic of Germany, and 21 Italy.

We have to stop the guided tour here. However, I should point out that the map is a dynamically changing one. It is analogous to a global weather map that is constantly changing.

The map of physics and chemistry that was obtained back in 1974 has vastly changed by 1983. By now, molecular orbital science has become a huge, invisible discipline in its own right, even though we rarely speak about molecular orbitalists. Even though there are such persons, it is remarkable that there is no journal devoted solely to molecular orbital science. But that is one of the peculiarities of specialty development. In 1974, we recorded about 50 core papers, whereas in 1983, it would be thousands (a core paper being one that is cited above a given threshold).

Apart from the fascinating pictures we can make by these methods, how do they relate to NFAIS? Very simple -- if NFAIS were one centralized service like PASCAL or VINITI, we could use these methods to help rationalize and divide the multidisciplinary abstracting chore. NFAIS doesn't work like a centralized service such as VINITI. Nevertheless, these maps permit NFAIS to identify precisely those areas relevant to each of its members' chosen areas -- and, in particular, areas of overlap. Through the classification system implied here, you might be able to eliminate the costly chore of screening ever larger lists of nonrelevant papers on your borders. This was the essence of a suggestion I made here in Washington at the International Conference of Scientific Information in 1958. I proposed a Unified Index to Science in which all abstracts would be recorded so that we could uncover the duplicates and/or the gaps.[10] This could also be incorporated into the SCI, so that a library could tell which NFAIS service had abstracted a particular document.

In recent years, we have conducted several experiments based on using the central, unified SCI database to produce such discipline-oriented files in computer science, mathematics, earth sciences, polymer science, biochemistry and pure chemistry -- not to mention the biomedical file we still have mounted on DIMDI. Each one of these experiments has taught us something about the process of automatic classification.

Of course, some of you could correctly argue that it is by no means completely automatic. Human editors must use our computer-generated data to assign names to the 10,000 research fronts we identify each year. But we have made much progress in all this, and I suggest that it can be used to make any existing database even more useful. We are interested in cooperation with any database producer who wishes to apply these techniques to complement existing methods of classification.

We performed this mapping exercise for our recently developed Chemistry Citation Index. This is a special file of articles in pure chemistry. We identified thousands of active research fronts and clustered them at various hierarchical levels. Zooming in hierarchically, first you see the overall scheme for chemistry. Then at each succeeding level, you zoom down to the next level of detail. Eventually, you come to a group of core researchers and their papers.

Time does not permit me to show you the data for each of the NFAIS disciplines, but it can be done -- and in the process, we can all learn more about the process of specialty identification. This intellectual task used to be the exclusive domain of editors. But science is changing so rapidly these days that it becomes more and more difficult for abstracting editors to follow the rapidly changing boundaries of science, even if they are omniscient or polymathic. So we need artificially intelligent systems to do so.

Derek Price used to refer to these maps as "command and control" maps. By appropriate statistical manipulations, you can follow the growth and decline of fields within countries, institutions, or even professional groups. That is why he and others visualized it as a tool for science policy analysis. Out of this has grown the field of scientometrics, much as econometrics grew out of economics.

It will be fun to come back in ten years to see how these maps have changed, and to observe to what extent the disciplines of science have become intertwined in ways that we can only speculate about right now. In 1958, a minor field called molecular biology began to emerge; in less than ten years, it had become a major discipline, and today it has ramifications all over the map of science. This can be seen on the earlier map of natural science, where molecular genetics is seen as central to biomedicine and is linked to chemistry through the discipline of protein science. Genetic engineering is a subset of molecular genetics today, but by the end of the decade, it may emerge at the disciplinary level.

I have used the occasion of the Miles Conrad Memorial Lecture to cover a wide range of issues, from a brief review of the old days and the unwarranted fear of machine methods to the unnecessary preoccupation with political issues, the age-old problem of information overload, and, finally, the eternal problem of changing classification systems. The irony is that we need more categorization at the same time that we need it less, because the boundaries are constantly changing. Further, if we aren't careful, we could find ourselves dealing with an information holocaust; whether we like it or not, education and marketing information services are no longer the exclusive provinces of academic or private commercial organizations. I have tried to focus a bit on the original concept of the unified index of science, the World Brain, as visualized by such early pioneers as H. G. Wells, Paul Otlet, and many others. In the past two decades since the death of Miles Conrad, we have witnessed many changes that have been exciting, indeed. But the future lies ahead; that is where the challenge always will be.

REFERENCES

1. Neufeld, M. L., Cornog, M., & Sperr, I. L. Abstracting and indexing services in perspective: Miles Conrad Memorial Lectures, 1969-1983. Arlington, VA: Information Resources Press, 1983.

2. Garfield, E. Biological Abstracts in an era of automation. Unpublished report, 15 July 1958.

3. _____. The preparation of subject-heading lists by automatic punched-card techniques. Journal of Documentation 10:1-10, 1954.

4. Drucker, P. F. Out of the depression cycle. Wall Street Journal, 9 January 1985, p. 26

5. Herner, S. Technical information--too much or too little? Science Monthly 83(2):82-6, 1956.

6. Garfield, E. Scanners of the world unite: you have nothing to lose but information overload. Current Contents (44):3-9, 29 October 1984.

7. Reed, L. J. Symposium on Machine Techniques for Scientific Documentation. Speech presented at Johns Hopkins University, Baltimore, MD, 3 March 1953.

8. Physics and astronomy classification scheme -- 1985. Physics Today 37(12):PACS 1-PACS 52, 1984.

9. Garfield, E. Third World research. Parts 1 & 2. In Essays of an information scientist (Vol. 6, pp. 253-275). Philadelphia: ISI Press, 1984. Reprinted from Current Contents 33:5-15, 15 August, 1983 and 34:5-16, 22 August 1983.

10. _____. A Unified Index to Science. Proceedings of the International Conference on Scientific Information, 16-21 November 1958, Washington, DC (Vol. 1, pp. 461-474). Washington, DC: NAS-National Research Council, 1959.

...Contined from page 43

VU-TEXT INFORMATION SERVICES, INC.

Houston Post on VU/TEXT

The Houston Post is now available on VU/TEXT, Knight-Ridder's information retrieval service. The Post joins other top regional newspapers such as the Boston Globe, Chicago Tribune and Philadelphia Inquirer in making VU/TEXT the electronic library for its reporters, editors, and librarians.

Among specialized areas of the database's in-depth coverage are:

-Energy
-Trade
-Aerospace
-Petrochemicals
-Medicine

The Houston Post database contains editions from January 1, 1985 forward, with the full text of articles available 48 hours after publication.

For information contact: Donna Willmann (215) 665-3306 or Kathy Foley (713) 840-5606.

WRITING CONSULTANTS

Synonym Finder

Writing Consultants announces a new thesaurus program called Synonym Finder. It may be integrated with existing word processor programs, and provides an online dictionary useful for confirming both word meanings and spelling. Over 9,000 keywords may be entered to call up an average of over 10 synonyms each. The synonyms in the database were compiled by a team of lexicographers and reviewed by Purdue University English Professor Roger Schlobin, Ph.D.

The program is available for the MS-DOS, PC-DOS, and CP/M-80 operating systems for WordStar Versions 3.0 through 3.3, and for MS-DOS and PC-DOS versions of Multimate. Versions compatible with other systems are planned. For further information, contact Writing Consultants, 11 Creek Bend Drive, Fairport, NY 14450.

MAPPING THE MULTI-DISCIPLINARY WORLD OF NFAIS!

Eugene Garfield, Ph.D.
President
Institute for Scientific Information

Lynne Neufeld informs me that the Miles Conrad Lectures were published by Information Resources Press in a volume entitled Abstracting and Indexing in Perspective.[1] This confirms Garfield's fourth law -- or Murphy's Nth Law. This law states that you will be invited to give the 17th Annual NFAIS lecture immediately after the first 15 have just been published.

The Miles Conrad Memorial Lecture is supposed to be on a topic "suitable to the field of abstracting and indexing information, but above the level of any individual service." What that means is, "Don't talk about yourself or the institution you represent." Since my fifth law states that this is impossible, consider these opening comments off the record. If, 16 years from now in 2001, Saul Herner doesn't want to include my remarks in the second volume, I won't be offended. But there is simply no way for me to begin a Miles Conrad Lecture without getting personal.

Since I have a poor memory for certain kinds of events, I was not exactly sure when I first met Miles Conrad. So I checked my files on Miles -- that has a nice ring! But that didn't help. So we called Hazel Philson. She confirmed that I met Miles when he was still at the Library of Congress, before he came to BA in 1953.

About this event I do remember saying to myself, like Sancho Panza in Man of La Mancha, "I like him." He had a soft, deliberate way of speaking, and his voice had a special, deep resonance. Phyllis Parkins may recall, however, that I knew about Biological Abstracts (BA) first through John Flynne. That was during the time I was still at the Welch Indexing Project, because I did not start my career in Philadelphia until September, 1954, when Ted Herdegen of Smith-Kline asked me to take a temporary assignment there.

Like Ted, Miles had known me from the John Hopkins project. He had heard that I was working as a Documentation Consultant. At that time, I was considered a "machine maniac." The word "computer" was barely part of our vocabulary. Miles asked me to do a study of the systems and procedures at BA. This was one of my early free-lance assignments. Out of this work came an unpublished report called "Biological Abstracts in an Era of Automation."[2] The report described the use of perforated paper tapes -- so symbolic, in those days, of automated procedures. It is difficult for this generation to realize what a painful transition we made from the punched card or Hollerith card to the Flexowriter paper tape, then to the hybrid card containing both, and eventually to magnetic tape.

Miles Conrad Lecture, National Federation of Abstracting and Information Services 27th Annual Conference, Arlington, VA, March 3-6, 1985.