Miles Conrad Lecture 2008

February 25, 2008

CAS in the New Information Order

Robert J. Massie,
President, CAS

Thank you, Kevin.

To be selected as the Miles Conrad lecturer was an honor and a bit of a shock. William F. Buckley came to mind. As he said in his book, *The Unmaking of a Mayor*, when asked at a press conference about the first thing he would do if elected Mayor of New York, Buckley famously replied: “have the votes recounted." (1) I had something of the same reaction. I surely don’t see myself as a long-time information industry expert on a par with past lecturers. And I am well aware of the limited interest that chemical information may hold for most audiences.

One possible rationale was supplied by my good friend Mike Tansey, former president of ISI. He reassured me last week that, after 16 years at CAS, I have become “the old man of the industry."

I think Mike felt that this was some kind of compliment.

I suppose we might title my lecture “an old hand in the information industry talks about chemistry databases." That would be a crowd pleaser.

I wonder if it is legal in Philadelphia to lock the doors to a lecture hall? No, I suppose that would be a fire hazard…

In any case, you can imagine the pressure I feel. This is NFAIS’ 50th anniversary, and the program for this annual conference is at a very high level.

I would like to begin by heartily congratulating NFAIS itself, and its Executive Director, Bonnie Lawler as well as Kevin Bouley, the current NFAIS president. I would also like to recognize past Miles Conrad lecturers, Karen Hunter, Carol Tenopir, and George Schultheiss, all of whom are in attendance. I have reviewed their speeches and drawn inspiration from them.

CAS celebrated its 100th anniversary last year, so there is some symmetry here. There is even more symmetry in that the CAS Director at the time, Mr. Dale Baker, was one of the five founders of NFAIS. He also was a Miles Conrad lecturer and a true pillar of the information industry through the post-war era.

Dale oversaw the early computerization of CAS databases (in fact, CAS offered one of the first computer-readable databases in science). And Dale pioneered CAS as a global information provider in the digital age. While our print services were distributed globally for all of our first 75 years, Dale helped found the international STN® online service partnership in 1983.

That STN system, still in operation today, adumbrated the Web by offering disparately located databases, all mediated by a single software platform and application suite. We think that in fundamental ways "STN was the Web before the Web was cool."

CAS has built its operations around two tools admirably designed for user needs in our core subject area: STN for information professionals interested in intellectual property and patent information, and SciFinder® for chemists interested in researching chemical and related sciences.

STN was and remains a global partnership, anchored by FIZ Karlsruhe in Germany and CAS in Columbus, Ohio.

STN is a classic system in many respects. But it is powerful and meets the needs of a critical market segment. The US Patent and Trademark Office, for example, signed a three-year agreement just two months ago, to extend its subscription to STN well into the future.

Of course, CAS is first and foremost about chemistry and chemical information. We are a division of the American Chemical Society, the world's largest scientific Society. The vision of the American Chemical Society is "Improving People's Lives Through the Transforming Power of Chemistry."

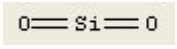
I will return to that theme at the end of my talk. But for now…

Let's do some chemistry…

The element silicon is

Si

and silica is just



But it is elemental silicon that is used to make computer chips; you get the silicon by melting silica. This is the element silicon as discussed recently on the

television program CSI: New York. It is perhaps the second most important chemical foundation for the “new information order,” as silicon chips lie at the heart of the computer revolution.

By the way, CAS was the first publisher to pioneer the transmission of structure information to handheld devices, three years ago.

I say that silicon is the second most important chemical substance because I think there is arguably a more important influence on the new age…

This is a structural representation of caffeine, the useful and ubiquitous molecule. It also can be seen at the cornerstone of the “new information order," at least in terms of powering the work of programmers, information consumers and, until recently perhaps, Starbucks stock valuations. What you are looking at is the CAS Registry^{SM} record for caffeine. The CAS Registry contains 33 million such references, all cross-linked to patents, journal articles, or other important information sources.

The Registry began its life in 1965 and has grown systematically since that time.

Caffeine, of course, is a xanthine alkaloid compound, one of the so-called xanthine family of stimulants. Realizing that it is late in the afternoon of a rather long day, I am hoping that if you focus closely on the molecule, you will experience a mild increase in heart rate, a warding off of drowsiness, keener mental focus, and even some increase in serotonin activity, to promote a favorable reaction to my speech.

On the other hand, nature originally created caffeine as a pesticide, its objective being to paralyze and kill insects before they could eat the germinating seeds of certain plants.

Maybe that is enough chemistry for the moment.

The topic of my speech is “CAS in the New Information Order.”

I want to avoid talking too much about us or our products, but listening to the program thus far I am struck that most of us face the same challenges, and it might be instructive to see how CAS deals with the new order.

It might also be interesting to reflect on how a 100 year-old organization like CAS has adapted, and continues to adapt, to changing conditions. My objective is to try to be broadly relevant, and not too parochial.

Of course, even 100 years of history won’t help us if there is no future for traditional publishers…

David Weinberger's keynote speech was pretty chilling in that regard; he states the elements of the new information order, Web 2.0 and beyond, are destroying authoritative sources.(2) But what if you think *you* are an authoritative source?

Is there room for traditional publishing in the new order?

To bring some structure to the speech, let me posit the key characteristics of the "New Information Order." I can take them directly from the conference materials:

The New Information Order is:

1. global, with new regional superpowers rising,
2. user-centric,
3. technology-driven, featuring virtual communities of techno-literate inhabitants,
4. collaborative, open, standards-driven… creating new content, and
5. giving rise to new business models

How does CAS match this checklist and indeed should we, or any organization, need to mirror or deal with all the elements? That is the prime question for the future!

When I was in consulting, we used to say that theory is great, but clients want to know "what am I supposed to do Monday morning?"

What, in a practical sense, should "traditional publishers" do in the face of these megatrends? I have already mentioned that, from day one, CAS acted on the global stage.

Here is a representation of the first abstract published by CAS in 1907; it summarizes an article in German about a laboratory apparatus.

You might be interested to see the range of subjects covered in that first issue, by the way.

Over its entire history, CAS has focused its content activities on global scientific output. During WW II, for example, CAS organized an informal network of scientists to ship journals out of Central Europe to the US through Switzerland!

Today our content derives substantially from global sources, and we have input centers around the world.

We keep things balanced, however, and have avoided chasing the low-cost labor market de jour. We have noticed, for example, that India has become decidedly less attractive in just the last six to twelve months, with currency appreciation and labor rate inflation as key drivers.

An axiom of CAS business life has been “pay close attention to global markets, trends, scientific output--anything that influences our work." It seems that this rule should apply to most information businesses. In fairness, science is especially global, while certain regulatory and legal markets are by definition somewhat more local.

Let’s return to our list of megatrends in the new information age.

To begin at the beginning--CAS was founded on two principles: First, as noted, we believed that the chemical sciences are global in nature and that the nascent American scientific community must participate actively in global information exchange. I have already pointed out the results of that global focus.

One of the ironies of our 100-year run is that CAS was founded, beginning in the late 19th century, partly out of a desire to bring US science up onto the world stage. Then, US scientists saw themselves as afterthoughts to the great European scientific powers.

As we turn into a new century, U.S. science, now essentially preeminent, is competing with both Europe and a resurgent Asian scientific-educational complex. And we are concerned about measuring up in the future against both this European and Asian competition.

Science and its competitive aspects were global 100 years ago, and they are global today in the “new information order."

The second great principle underlying CAS’ mission is that no individual scientist has the time to read and assess all the articles and patents in his or her field, and so summaries are needed to facilitate knowledge acquisition. Both of these notions, of course, are also thoroughly modern, or maybe timeless. And both are in line with the New Information Order list.

CAS began its existence totally and elegantly “user-focused” or “user-centric."

From the global to the local--what about that individual scientist? Well, of course, the notion of an “abstracting service” is at heart a time-saving device. Time pressures afflicted our forbearers as they do today.

In her 2006 Miles Conrad lecture, Dr. Carol Tenopir pointed out that the number of hours a researcher devotes to reading journal articles has essentially remained flat over the past twenty years.(3) That may be simply a reflection of natural human limits and time allocation opportunities. It is hard to know how much time late 19th century scientists could devote to reading, but we do know that already in that era there were complaints about the growth of information

and the need for summaries or other means to facilitate research against this wave of incoming new science.

How are scientists coping? Dr. Tenopir says:

“We first started noticing a strange phenomenon in the mid-1990s, where graph lines for articles read and time spent reading (that had pretty much followed each other proportionately) began to diverge. We witnessed more reading, but in not as much more time as you would expect. It seems that reading now is often likely to be skimming or scanning.”

So, how much has really changed?

My first conclusion, from the CAS perspective, is that in terms of “global reach” and “user centricity,” “the new information order” is different only in degree from the “old order,” if we can call 1907 one of the “old orders” from our perspective.

Science was global, and scientists were pressed for time in dealing with the output of their peers.

Being “user-centric,” however, also means getting the product right, and finding your core market or markets.

Nearly thirteen years ago, CAS pioneered online searching for the chemist community, which at the time may not have been a “virtual community” in today’s sense of the word, but surely was a constituency long sought after by NFAIS members. In fact, we all remember the debates of the 1980’s and early 1990’s as to whether the “end-user market” was illusory or real.

To demonstrate the dangers of predictions, let me quote from Ron Dunn’s Miles Conrad address from 1994. (4) He was speaking of what he called "our own frustrating pursuit of the elusive end user of scientific and technical databases.” Ron said that he had spoken already on this topic in 1985, and that "unless a miracle occurs” end users could not obtain what he called “excellent search results.” Then again in 1994, he said:

"Regrettably, I'm not aware that a miracle has occurred since 1985. As far as I know, no one in the NFAIS community has yet cracked the direct end-user searching market in a financially significant and profitable way. In fact, it remains to be demonstrated that commercially important end-user markets exist for our kinds of businesses at all.”

As it turns out, the “demonstration” was in fact about to happen.

Within months of this speech, CAS introduced the beta version of SciFinder, the end-user tool for chemists. It provided truly excellent search results, with

incomparable ease of use. And it has enjoyed double digit or near double digit growth every year since its launch.

We at CAS see SciFinder as an example of technology and content being combined to serve our mission and bring high quality service to a key user group. It is and remains every bit as "user-centric" as any Web service, because it was expressly designed to follow the thought and research patterns of chemists. And, SciFinder was built on the proposition that chemists then and today are "techno-literate" and perfectly able to conduct searches in their discipline. As we said at the time, we simply needed to create a tool, SciFinder, that eliminated the need for chemists to learn a new search system language.

What about enabling technologies and CAS?

Surely we live in an era more influenced than ever by technology advancements.

CAS was in the forefront of technology development for scientific information exchange throughout its history, using state-of-the-art technologies in every era and at times pioneering technologies where none existed. In the 1970's for example, CAS was one of only three or four centers for development of structure searching technology in the world. Others were at Hoechst, BASF, and Bayer. CAS cooperated with them and also the BASEL Information Center for Chemistry (BASIC), in Switzerland. A world renowned chemical nomenclature authority at CAS, Dr. Kurt Loening was a leader in building the foundation for structure searching, and worked closely with these organizations, IUPAC, and others.

CAS was way ahead of the curve for electronic services, and actually developed its own internal email system in the 1980's, under the leadership of Dr. Ronald Wigington. And, as already mentioned, CAS was a pioneer in computerizing databases.

Throughout the 1960's, CAS developed computer-assisted publishing for CA, which combined the value of the database with the efficiency of computers for editing, organizing, and preparing information for the press. It was a logical next step to perceive that this same information could be distributed in electronic form, long before online searching came about. The need to produce an index of chemicals led to the development of the CAS Registry System, so CA editors could see whether a substance encountered in newly published literature had already been indexed--if so, the Registry let them retrieve the structure, rather than redraw it.

CAS in 1961 produced *Chemical Titles*, considered the first periodical generated by a computer.

CT was also one of the first electronic products distributed on magnetic tape for batch mode searching, starting in 1962.

CAS' Chemical-Biological Activities (CBAC) was a pioneering, computer-based alerting service introduced by CAS in 1965. Even then, the leaders of CAS recognized the biological importance of what our database had to offer.

Reflecting on our own experience, my observation would be that compared to twenty or more years ago, the most advanced technologies for publishing are much more ubiquitous and available. Technology tends not to distinguish information providers today; rather, the combination of technology/content and its application or market specialization does.

What about items four and five on our list of trends or characteristics of the new information order? How are they affecting our business and yours?

This really is the heart of the matter….

But before I address it I do want to reflect again on history, because it so elegantly foreshadows today.

Across its first sixty years, CAS information collection was managed by editors, but the abstracts were written largely by individual contributors, the famous "volunteer abstractors" living and working around the world. Just as STN foreshadowed the Web, the CAS volunteer abstractors can be seen as an early version of Wikipedia contributors.

In those years, consigned as they were to the vagaries and slow speed of the postal services, the CAS abstractors nonetheless contributed to the edifice of science that we consult today, one document at a time, one letter at a time. Like rivulets and streams flowing to a great lake.

It was only when technology permitted and indeed encouraged the creation of a central processing hub that CAS moved aggressively to hire and train editors for the Columbus office in great numbers. Still, as mentioned, CAS retains global input centers and a global collaborative spirit and practice to build our databases.

And yet, new trends impact us today. These are very important changes…and we have to adapt.

Now, we come to the question of the hour…if all of this adds up to new business models, will traditional publishers be swept away into history?

In other words, even if CAS can both be a traditional publisher and also be "global, user centric, technology driven, serving techno-literate communities, and working collaboratively to build its resources," can it survive the challenge of "open, collaborative, standards-driven, new business models"?

In “Wikinomics,” Tapscott and Anthony argue that “Publishers of music, literature, movies, software and television are like proverbial canaries in a coalmine” and predict dire times for traditional businesses organized in traditional ways. (5)

For a statement that hits home even more explicitly, see Peter Murray-Rust’s recent article in *Nature* in which he states that CAS in the 21st Century is “incompatible with Web 2.0." (6) Murray-Rust is a distinguished scholar and thinker, so when he writes the following, we should sit up and take notice:

"Closed publications, binary software,and toll access databases are being swept away by the emerging technologies and software” and he states that “the chemistry establishment must adapt quickly or fracture." Much in his article is worth reading. The fact that he published in a so-called “closed subscription publication” may be an irony that escaped his notice.

But is his warning, to CAS and others like us, accurate and prescient? Or a bit overwrought? Or both?

Part of his argument is generational. He says that “many young scientists do not read or use closed systems, and are increasingly frustrated by out-of-date approaches.”

Here is where a broad generalization may do a disservice to an otherwise cogent argument…

Regarding the younger generation, I am not sure what Dr. Murray-Rust’s sample is, but last week the one thousand five hundredth university signed a contract for SciFinder Scholar™, our academic service. Graduate students in chemistry and allied fields continue to drive growth in usage in this version of SciFinder.

In fact, SciFinder is the number one object of attention also of the presumably young information pirates in China!

You might be amused to see this screen capture of a Chinese Web site providing instructions on how to break into universities and steal CAS information on our SciFinder platform. Apparently, young Chinese scientists are only frustrated that they cannot get *enough* of our closed system, not with the information or software itself.

I show this to demonstrate that the information scene can be a lot more nuanced than some observers would like it to be.

But seriously, is there really a binary and mutually antagonistic future for our business and Web 2.0?

So much in the new information order is and should be context-specific. All the broad trends and principles need to be seen in terms of your own business and markets.

That is why I would like to step back and remind us about the great trends in chemical information.

First, we continue to experience dramatic increases in information entering the CAS databases. Much of that represents an astonishing surge in publishing out of Asia, especially China.

This parallels economic growth trends…

For example, 52% of the economic growth in global GDP last year came from the developing world, not the U.S., Japan, and the Euro Zone. And nearly 19% of the growth came from China! (7)

Second, the future promises an even more astonishing era of discovery.

As you can see, one of the most distinguished chemists in America, Dr. Ronald Breslow, predicts that the number of molecules yet to be discovered may be more than 100 times what we have seen to date. (8) And with some 34 million molecules catalogued in the CAS Registry, and tens of millions more in our Markush and prophetic collections, the implications of this statement are profound.

And now this question: where will these discoveries be disclosed?

As you can see from this chart, in the last thirty years, there has been a marked movement towards embedding new molecular discoveries in patents. I call this the “monetization of chemical information” and it has important implications for information collection. While patents are freely available on the Web, processing is required to drive the chemical information out of the patents and make that information searchable and compatible with other collections. This is the essential work of CAS scientists and systems. And for the foreseeable future, we are not inclined to put down our tools in the hope that communities will self organize or standards and collective action will ensure that the information is collated and stored in a systematic way.

My own impression is that CAS and publishers like us can and should adapt to and adopt from Web 2.0. But we first need to know well what business we are in. And we first have to ensure that we meet our missions and provide the core value that our users demand.

Just as we at CAS adopted technologies and approaches that furthered our mission for our first hundred years, we hope to do the same in the coming

years. We see valuable hybridization of collaborative tools and practices and traditional approaches. Where we can add value to our core proposition--user-centric, quality controlled, secure and private research in comprehensive and timely databases-- we will do so.

Which brings us back to chemistry…

I began this talk by showing you a caffeine molecule, in the hopes that it would be a virtual pick me up. Now our thoughts will turn to the reception, and perhaps some wine.

So to ease your conscience, let me show you the molecular structure of the resveratrol molecule, the active ingredient in red wine that has salutary benefits for your hearts.

Here, by the way, is another example of improving people's lives through chemistry!

Thank you for your kind attention and for the honor of presenting this year's Miles Conrad lecture.

Notes

1. William F. Buckley, *The Unmaking of a Mayor*, Crown Publishing Group, 1977.

2. Dr. David Weinberger, "A Look At The New Information Order," NFAIS Annual Meeting, Feb 25, 2008.

3. Dr. Carol Tenopir, "Building Bridges to Information Products and Services," 2006 Miles Conrad Memorial Lecture, NFAIS Annual Conference, Feb 27, 2006.

4. Ronald Dunn, "Angst & Anticipation: How Will We Fit in the New Information Age?" 1994 Miles Conrad Memorial Lecture, NFAIS Annual Conference.

5. Don Tapscott and Anthony Williams, *Wikinomics: How Mass Collaboration Changes Everything*, Penquin: New York, 2006.

6. Dr. Peter Murray-Rust, "Chemistry for Everyone," *Nature*, Vol. 451 (7179), 7 Feb 2008.

7. David Hale, "Brave New Economy," *Wall St. Journal*, Feb 25, 2008.

8. Philip Ball, "Chemistry: What Chemists Want to Know," *Nature*, Vol. 442 (7102), 500-502.