Information Services & Use 34 (2014) 49–64
DOI 10.3233/ISU-140726
IOS Press

Making information a business: The voice behind the curtain

NFAIS Miles Conrad Memorial Lecture, February 23, 2014

Marjorie M.K. Hlava
President, Access Innovations, Inc., PO Box 8640, Albuquerque, NM 87108, USA

Abstract. Gathering from her perspective on over 2000 information projects and products completed over the last forty years, Marjorie Hlava offers a fascinating view of database creation and search using developing technologies, and information technology standards. As an early implementer of most of the technologies we use and products we all sell today, and having worked behind the scenes at one time or another for most of the organizations in NFAIS, she has seen a wide variety of data and applications. She has thrilling tales from her adventures in the field, along with a unique perspective to share. This development in the trenches under a continuing onslaught of new options has honed her knowledge on what has worked well, and what will work in the future. In this talk she will draw from her varied background, bring in statistics from our broad marketplace, and share what she sees on our horizons.

1. Introduction

I want to share with you some tales from the field; some adventures that I've had along the way – which has mostly been a tremendous amount of fun. A few adventures I will review include the case of the missing abstracts, working in Russia, the U.S. Patent Office, some Getty adventures and the Vatican Bibles. It's amazing what you can do and learn in the field of information.

Then I will share with you my thoughts on some of the driving forces we face today as an industry and finally, thoughts about where we might be going from here on out.

2. Tales from the field

You all know that when you are on the cutting edge or the leading edge of a project, you must first figure out what the client needs. Then you figure out the specifications to accomplish these needs, and you get client approval of the specifications (which is sometimes a back-and-forth process). Next, you figure out how to deliver the data according to those specifications. Then you do the work, massage the data as appropriate, and deliver it . . . except when life happens.

One tale is "The Case of the Missing Abstracts". Chemical Abstracts Service (CAS), an organization that is a huge force in our industry, had done numerous tests in 1985 and figured out that just searching the indexing terms that were related to their abstracts did not provide the full answers that the users wanted. CAS wanted to be able to search the titles and the abstracts as well. They did quite a bit of negotiation and found out that Dialog Information Systems and Orbit would be willing to put the data up on their servers if CAS provided it. At the same time, they decided that they would implement a

new platform to continue to make the abstracts available in digital form. The system they developed was called Messenger. (It's the same one that they supplied to the U.S. Patent Office at the time.) The specifications were written and the test file was approved. We were all ready to go. All we had to do was convert their photocomposition tapes. They went to the vault to package up the 9-track tapes and send them to us, but found that the tapes had been destroyed, along with the 792,000 abstracts stored on them. Someone had wanted to make room in the archive vault for other things and CAS didn't have the data to send to us. Heartbreaking!

Instead, we decided to take the data from 1970–1982 as it was in print. It was chemical information, and those chemical names are not always spelled intuitively. They wanted 99.998% accuracy, because the users had made it pretty clear that they would accept; nothing less. It needed to be in a left-tagged ASCII format. We decided that we would triple-key it and programmatically check the keying streams. Wherever there was not a match, we would make corrections and then double-proof it to meet the required level of accuracy. We needed to find two full sets of the volumes for all thirteen years: one on which to use a guillotine cutter to remove the spines of the books for parceling out the pages to individual keyers, and one to use for the proofreading set. There were eighteen volumes printed every six months, each volume being 8.5×11 inches and about two inches thick. If any of you have ever searched those in the physical form, you know that they were huge! CAS came up with the two required sets, but it wasn't an easy task for them.

3. The CAS adventure

We had at that time established an office in Mexico City called Access de Mexico, and that is where we were going to have the keying done. I signed the contract and called the office. Omar Alvarado Diaz (the manager of Access de Mexico) and Jay Ven Eman (the CEO of Access Innovations) were there. They boxed up all of the printed volumes, and Omar took them as luggage on Aero Mexico (they did not charge for luggage back then) with him to Mexico City. It was late when he arrived, so he loaded the suitcases into his car, drove home, parked his car in the garage and went to sleep.

It was a lovely place. But at 7:17 AM on September 19, 1985, there was an earthquake in Mexico City that measured 8.7 on the Richter scale. The eight-story building where our office was located collapsed into one-and-a-half stories of rubble (see Fig. 1).

Fortunately for us, it happened during a shift change and only one person from our team was killed. The documents were still in Omar's car, under the caved-in garage at his house. The news was full of stories about the disaster. We couldn't get in contact with anyone. We didn't know what was happening at all. The runways at the airport were so badly buckled that no planes could get out, and telecommunications were disrupted. Jay and I traveled down there two weeks later. Here is a photo of our building. At the office, we found one $8''$ floppy disk that was undamaged. Fortunately, we had only sent one set of volumes to Mexico City. The second set of volumes was in our office. The volumes were eventually dug out of the garage – undamaged – and we brought them back with us, although we managed to key one-half year's worth of data there.

We did have a deadline for the project, so in early October we moved the data to another keying facility, this time in Jamaica, also a lovely place! I don't know how well you remember that year, but there was a major hurricane that struck Jamaica. The amount of damage from Hurricane Kate in November 1985 was considerable; there were FOUR FEET of water in the computer room and no power on the island. That wasn't so good for keying or for computers, so Jay went to Jamaica and got the books and took them to the Philippines. (Laughter from audience.)

Fig. 1. Impact of earthquake on Access de Mexico Office Building. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/ISU-140726.)

Now, this wasn't funny at the time. (More laughter.) You might remember that in 1983, the Philippines' President, Benigno Aquino, was assassinated, and by 1985 his wife Corazon was running for President. The books again had been transported as luggage because we had brought them back to Albuquerque from Mexico City after the earthquake, and the duplicate set had been shipped to Jamaica. Those now needed to be retrieved. We loaded them into the back of a limousine to take them from the airport to the keying facility in the Philippines. On the freeway, the limo got caught up in a "peace march". Hundreds of thousands of people with yellow armbands marching in Manila in support of Corazon Aquino stopped traffic and flooded the freeways. There was political unrest, but it did not adversely affect the project getting started. But, if you remember, there was another storm – Hurricane Dot. Power was out on all of the Philippine islands for weeks. The storms were so bad that they had to evacuate Clark Air Force Base. And we still had that deadline.

After that, we moved the data to China in November of 1985. This was a lot of moving around for the CAS data. You might remember what happened in November of 1985 in China. . . . NOTHING! Nothing happened. We finished the project on time, at the promised accuracy level, and slightly under budget. The project manager said "You know, when I read your contract, I thought you had some unusual "Act of God" clauses, but I didn't expect you to use every single one of them!". That was one adventure.

4. To Russia with love

Not too long after that, we became involved in Russian information, thanks to book publisher Maxwell Macmillan and Ron Dunn. VINITI, The All-Union Institute for Scientific and Technical Information, teamed up with Maxwell to make Russian scientific information available throughout the world. We

went over to Russia. The project was to be basically a trade of technology for publications, but it was not going well. I was sent over to find out why not. It turned out that Maxwell was sending microfilm machines, but no batteries for those machines, and they were sending microfilming machines, but no cameras. The Russians were providing digital copies of their information, but Maxwell wasn't providing the keys to transmit or decode that information.

Once the Maxwell deal came to an end, largely because of too much mistrust on both sides, we ended up opening an office in Moscow. We had to take the payroll in cash in our shoes because otherwise, it would "disappear". We learned a great deal. I learned about automation, automatic translation, artificial intelligence, and many other things. The Russians were light years ahead of us in programming and hundreds of miles behind us in any kind of computer operation. We established Access Russia in 1988 and sold it in 1999.

One of those adventures, as we moved forward and stopped dealing with VINITI itself, was to go to other offices in Russia. One of them sure didn't look like an office to us when we got out of the car. The lot was littered with little dug-up areas. We wondered if maybe they were burying something there – we hoped it wasn't bodies! We took a deep breath to steady ourselves and went down three flights of a metal staircase into the old KGB subway out of the Kremlin. To our surprise, we found a very well-equipped keying office. They keyed abstracts for us and did a wonderful job for many years.

5. Patent archives

Another rabbit hole that we went down was Iron Mountain. One mile back into those caves in Boyers, Pennsylvania is the U.S. Patent and Trademark Office (USPTO) archive, where original U.S. patents are stored. At the time, there were 5.5 million patents (see Fig. 2).

We needed to create 9-track tapes of the data. The USPTO wanted to make their operation more efficient and computerize the operation. We needed machines for scanning so that we could digitize the patents and make them available electronically for everyone who wanted to use them. We ended up building the scanning machines from scratch, because the machines we needed just did not exist. The data was fragile. We saw Thomas Jefferson-signed patents for Ben Franklin. We had to air feed the patents and scan both sides in one pass through the machine because of the fragility of some of

Fig. 2. USPTO archive at Iron Mountain, Boyers, PA. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/ISU-140726.)

the papers. We went to the then-new, advanced, 6250 bpi tapes on this project (up from the standard 1600 bpi tapes). The scans were pretty high density because the data was to be displayed at 300 bpi (bytes per inch) density, but we only scanned and delivered the OCRed text at 97%. Then we had to test them and do some quality control algorithms, and figure out how to display the images. After much debate, I finally agreed that we could use the dirty OCR because statistically your term was likely to be spelled right at least once in the average thirty pages of a patent and, therefore, findable.

6. The Bible adventure

Another adventure was the conversion of Bibles from the Vatican. They have two large collections: one is in the Vatican Library in Rome, and the other is in Perugia, in northern Italy. We worked with the collection in Perugia with Digital Equipment Corporation (DEC). The plan was to capture and convert the data for distribution on 12-inch video disks. I still have one, but nothing to play it on. The library of antiquity was a very fancy looking place. The Bibles had a lot of marginalia, notes written in by the popes themselves in their copies of the Bibles (see Fig. 3).

They were horrible to work with! They were written on parchment and leather-hide bound, made entirely by hand and some were centuries old. We built a tent scanner because some would not open all the way. There was an amazing source of pure research data in those volumes. We supplied 12-inch video disks that at the time were very advanced technology, but were later superseded by CD-ROMs (see Fig. 4).

Fig. 3. Page from a scanned Vatican bible. (Colors are visible in the online version of the article; http://dx.doi.org/10.3233/ISU-140726.)

Fig. 4. Equipment used to scan the original Vatican bibles.

7. Other adventures

Another adventurous endeavor that we completed was digitizing the British Library map catalog collection. We took the old maps – those published prior to 1850 – from the printed catalog and converted and added them to the digital catalog. That was really fun to do. It was a great big folio-type book when we started the work to digitize it.

Another project we did involved the Getty Art & Architecture Thesaurus, which had been done by the Getty Research Institute, and we used that as a basis to create a thesaurus for the Art and Archaeology Technical Abstracts (AATA). I can tell you that art history is *not* art conservation! Going from a faceted vocabulary to one that meets the NISO standards is not a trivial activity, but we got it done and got the entire collection indexed and digitized. Our work was complicated by the fact that the project involved participants in three countries, with many differing opinions on how to accomplish the work.

We have had a lot of exciting adventures. We acquired Tamil print and screen drivers above an Indian market in a bazaar in Singapore. It was scary climbing the stairs to a dark marketplace with all eyes silent and following us. We got some photocomposition keys – the nuggets that I needed to unscramble the data – in a bar under a bridge in Chicago, and we designed the Chicago Research and Trading (CRT) desk that you now see in photos, with all the terminals above and below rather in the shape of a boat, to easily show all those data feeds. CRT was the first one to implement those. It's been fun.

8. Knowledge theory

All of these projects have used classification in some way or another. We have to organize the job. We have to organize the information. Once things are digital, they are even harder to find than they were in print, so we have to tag them. I am a big advocate of thesaurus-based, or controlled vocabulary-based, tagging. Applying some kind of notation, either textual or alphanumeric, is not necessary. However, the tagging does need to be content-aware, something that really reflects the content. We need to keep in mind that the classification is never done as long as you have an ever-changing database. Of course, we need standards. We need standards to make databases portable, flexible, and interoperable. That is why I still chair the NFAIS Standards Committee and why I still sit on the NISO Content and Collection Management Topic Committee.

Underlying all of this, we have a very, very long theoretical basis. If you look at outlines of knowledge from Thomas Aquinas, John Knox, Francis Bacon, Mortimer Taub, the Cutter system, the COSATI (Conference of Scientific and Technical Information) with Alvin Weinberg (our response to Sputnik) and the Cranfield Institute, particularly papers by Cyril Cleverdon, Jean Aitchison and Brian Vickery, I think you will find that we have an uninterrupted line, over thousands of years, that keeps us on track for classifications. We need to organize knowledge. As mankind grows and expands, its knowledge base *has never* diminished.

Organization of knowledge started back in the earliest days for which we have any kind of written knowledge. The early philosophers wrote about some kind of tagging for keeping track of different kinds of knowledge. They talked about characteristics in common; how it was in the particulars – that most specific *thing*, the *thing* that is the same object – and it wasn't the same object if they didn't share the particulars. John Locke, in his Classification of Kinds of Knowledge – also outlined the organization of knowledge that we follow today.

In the 20th century, we had many different approaches and systems that began to divide the world of information. Our world had enough people living in it to group them into individual specialties. Instead of just being "a philosophy of knowledge", information science was divided into many additional fields. If you want to keep track of the broad view of what we're doing, you have to read much more broadly than just the information science literature.

Modern classification itself, which I really think started with Charles Ammi Cutter and his classification system, is the modern system that we follow today. We took a bit of a detour with Melvil Dewey and his Dewey Classification System. Parallel to this kind of thinking, I have found, to my surprise, that other nations also classify stuff. Imagine that!! We Americans were not the only ones. Part of the reason, I suspect, that we Americans don't understand, for example, the Russians, particularly well is because they are taught from birth to think differently than we think. Lenin wrote The Outlines of Knowledge, which he called the *Rubicon* and the *Rubricator*, and the associated classification system that is used for organizing not just Russian information, but also the education system, governmental departments, and so forth. The *Rubicon* as an outline of information heavily influences the way Russians (still) think about life – and it is far differently than the way that Aquinas and Locke and some of the other thinkers did. And that means that Americans and Russians do not think the same way. Nor do we think the same way as those in India. American traditional thinking followed largely in the British tradition, but in India, Ranganathan developed his classification system. Ranganathan's system is an interesting combination of two ways of thinking – an Eastern way of thinking and a Western way of thinking.

Back to Cutter. Cutter and Dewey were friends and colleagues. The Cutter System was ousted from the library in a rather bloody verbal battle and replaced with the Dewey Decimal System in 1911. The two men never spoke to each other again. The expansive classification system that Cutter used has seven levels of classification, each with increasing-specificity, as he believed that things should be ever more specific. I personally like reading Cutter and I recommend him to you. He has a very musical style of writing and the way that the NFAIS membership of abstracting and indexing services index our own materials and collections is based on Cutter's work.

Some outlines of knowledge follow a single point of knowledge: the biblical tradition of Eve and the Apple, for example. The concept that everything derives from the first organism, taxonomic systems like the Linnaean system, Lenin's *Rubricator* system, and those by John Knox himself, as well as those by Dewey, are all based on systems that expand from a single point of knowledge. However, in our lives today, we have multiple points of knowledge. We live in a heavily interdisciplinary world, in which multiple fields come together. There are differing views on whether or not they should be treated

separately or together. You find that the interdisciplinary fields of engineering or physical biochemistry or education are all represented by having multiple broader terms or multiple points of origin. Cutter supported this. Alvin Weinberg's COSATI in 1964 supported the interdisciplinary approach, and the resulting Thesaurus of Engineering and Scientific Terms (TEST) is the basis for many thesauri we use today.

We have different kinds of approaches to information science. Unless you are dealing with an organization that has a single patent and everything derives from that patent and nothing else, these days it's pretty hard to find a single point of knowledge information corpus. But information is changing. We had the teletype and then we had the fax machine. I remember when we got our first fax machine. My Dad used to be able to read the ticker tapes on teletypes with his eyes closed. My husband still carries around Hollerith punch cards. He says they make the best note cards. We have online, and we have downloading, and we have the Internet. I remember the times when if things were going slowly at a meeting, all you had to do was say 'download' and suddenly Art Elias of BIOSIS would verbally erupt in the back of the room and you would have a big lively lecture about the perils of downloading.

9. Driving forces affecting us

The state of the art is ever-changing. It's a part of our fun. The players are changing as well. We used to have stand-alone publishers, aggregators, serials, book vendors and hosting services. Now we have the Cloud and disaggregation and now everyone can be an author and anyone can host a database. Whether we have appropriate levels of quality, accuracy, and peer review are matters for another talk. We definitely have a great deal of change in our lives every day.

The information delivery vectors that we use have become very, very different. We are using webinars and blogs, tweet-ups and meet-ups, databases, articles, or just standing around the water cooler talking about things for a while. Everybody has a preferred way to get their information delivered. New formats just keep emerging. We started with some digitization of data and photocomposition markups. We got SGML because no one wanted to be handcuffed by photocomposition vendors anymore. Then we got XML because SGML was just too hard, and now XML has been digested down to little tiny URL code.

9.1. Technology

We had great big iron (the mainframes) and then we had server clusters and now we have Amazon Cloud and other services that make data extraordinarily accessible and simple to start up. The cost of entry has gone way down. The iPhone itself has 240,000 times the memory power of Voyager, which is transmitting some beautiful pictures back to earth, but in just a little bit of time computing power has increased logarithmically. The technology that we all deal with every day has exploded. We can't get away with carrying just one device any more. Everybody I know has multiple devices – desktops, laptops, tablets, iPhones and other phones. Some people still have their flip phones and they think of themselves as Luddites.

We all know that we need to continue supporting the technology. It was a difficult time for us when we worried about the move from print indexes to online and what that was going to do to our revenue streams and how it would erode them, etc. Then we had those pirates that were downloading our online stuff, and that was pretty scary. Then CD-ROM technology emerged, and the worry was how to deal with that. What is going to happen with our online revenue? What is going to happen to our print revenue? Then, it wasn't long before additional technologies were added. It seems to me that we haven't gotten rid

of any of those old technologies; instead we just keep adding more. Distribution points and distributing methods are extremely diverse.

9.2. Search

Search is also, finally, changing. It used to be that we had Boolean search with Dialog, SDC Orbit, BRS (now Ovid) and Medline. Then we were overtaken by relevance ranking, Bayesian, vector-based, neural nets, etc., and statistical approaches led by Autonomy, Google and some others. Those didn't work very well by themselves. Now we have a general hybrid approach of the two, presentation layers on top of Solr Lucene. We have the rules-based approach that gives us the best of both worlds. Some of them are free; some of them cost money; some are using five or six different technologies under the hood, and I think that is good for us.

In my mind, there are currently two major kinds of searches. The first is entity search, which looks for people, places, and things. This type of search, linguistically speaking, behaves differently than the second type of search – conceptual search. Most of our journal articles and books and proceedings are represented well by conceptual approaches, whereas a search for 'restaurants near 16th and Walnut' is an entity search. I think these two approaches need different kinds of search engines and they need different kinds of data linked to them behind the scenes. Entities need a lot of synonyms. So, if I want to know if 14th Street in Philadelphia is also known as Broad Street, for example, I need my system to know that synonym. Search results should be the same no matter which one I feed the system. The genome system is another example. We just downloaded the genome system and made a huge automatic indexing project out of it and it only took about three days to do it. The reason that it was so easy is because genomes are really entities. The National Library of Medicine (NLM) Genome Trial has about 38 synonyms for every term in that corpus. It's a huge synonymy, which makes it really easy to build a system. Geography searches are done in the same way. You want to be able to know what the region is, what the state is, what the neighborhood is, what the streets are and, if you layer those terms hierarchically, it's pretty easy to build a system to support them. It is harder to build a conceptual engine, because with the conceptual engine, you have to know the whole concept that surrounds the "thing" for which you are searching. Then you want to have links from your search "thing" out to other "things" in other areas that would impact it, depending on your point of view. You render them differently in an ontology, but it works in a very similar way.

Graphic user interfaces (GUIs) for search are becoming more elaborate, and more helpful to searchers. The web service layer and the search engine underneath have both seen improvements in order to serve up a good answer to the user. People who build both kinds of search into a single system need to organize the information infrastructure so that it will support both. Lumping them together does not give good results.

9.3. Tagging and classification

The tagging – how we should engineer that? – is still debated. Permuted indexes like those seen in the printed *Biological Abstracts* and *Chemical Abstracts* passed to online, but now we have drop-down, permuted, type-ahead listings that work off of a permuted list of index terms.

John Blossom talked earlier about the Signal Economy (see John's paper that appears on pp. 17–25 in this issue). But you need to be able to process the signal. Some kind of tag on that data is needed (whether it's a tag that tells you how to respond to a gesture or something that enables you to connect it

to another piece of data), or the signal economy is not going to work very well. We need to get that data tagged.

Classification systems are coming around again. We haven't thrown out the Library of Congress classification system, although it is being realigned. We are broadly embracing thesauri, taxonomies, ontologies, and other controlled vocabularies. The differences between those are fairly subtle.

The inverted files of search are still with us and increasingly robust. Now we hear a lot about triples and triple stores as well. The horizons are just more complicated. We have field – formatted data; we have relational databases; we have SQL databases; and we have object-oriented systems. Java and XML really lend themselves well to object-oriented systems.

We had the semantic web approach, which didn't really work very well because, like SGML, it was very complicated and perhaps over-engineered. Now we have linked data, which is being widely embraced. Triple store systems that we heard about earlier today are the new kid on the block and are just beginning to be implemented.

Telecommunications is struggling to keep up. I remember party lines. In my youth in our rural community our ring was three longs and a short on a 13-member party line (a phone line used by multiple households). Then direct-connect lines came along, with more privacy and expanded use options. Trunk lines were in use early for direct connections. This meant that when you started your "private" conversation, half the time it was actually private, and half the time you got in on someone else's conversation as well. The switching got better over time. Then we moved to long line and fiber optics. Now we have cell towers disguised as fir trees, at least in my area, and wireless is ubiquitous.

9.4. And new things continue to emerge

We have all kinds of apps. They are becoming socialized, interactive and stand-alone. We have more legal disputes and more patent wars. These are not gentleman's disputes, either. They are vicious, take-no-prisoners wars. We also buy and merge companies and people at an amazing pace. If you really want to know what Google is doing and where they are headed, look at the patents held by the people they just hired or the companies they just bought. They are not listed as Google patents, but rather as patents for the individuals hired. By hiring those people, Google is getting the expertise that led to the patent. Steve Arnold of ArnoldIT is especially good at tracking Google initiatives via patent acquisition.

There is a lot more interest in open data and linked data. But open applications and open data generate a lot of worries and concerns, particularly for corporations, because they are not sure what kind of vector allows the use of that data. What potential security problems are hidden in that open source code? Are these really Trojan horses or can we trust them?

Search analytics is a brave new frontier, but in my mind it is not yet very well developed. Streaming content is taking extraordinary amounts of bandwidth and leading to net neutrality (which is anything but neutral), but still helping with those ever-interesting disputes. The names of the participants can be very misleading.

We have a lot more options. The content budget just can't keep track of all the options. Because we have so many more options and so many more devices, in the aggregate, it's a lot more definitive than it used to be. We begin to classify them: This is an Android app, or a iPhone app, or it is content that will run under any app. The beauty of Java as a development platform is that it is operating system neutral. You can compile it under practically any operating system.

We have a great blending of our offline and our online worlds. Our friends keep track of us by Facebook, LinkedIn, texting and email, but seldom with a letter. The handwritten letter is an historic option. They don't even teach cursive writing in schools anymore!

We can now geo-tag people at places giving geo-contextual data. Are you going to allow me to track where you are when I interface with you on this particular app? And there are voice interfaces – Siri and other tools that give you all kinds of options, back and forth. It's really fun! But sometimes we reach the level of technology exhaustion and some people opt out entirely.

The interactive learning potential, combined with the new Common Core standards for education that the U.S. Government has promulgated on behalf of the States, has lent itself to an extraordinary avalanche of new kinds of learning options that we are only now beginning to see.

Then there are massively open online courses (MOOCs), which certainly affect the universities' revenue stream and provide wide lifelong learning options for the masses.

We have more analytics. We have more big data. We have an Internet of things we can port anywhere. We have many predictive pursuits and personalizations, whether we know it or not. Media has changed. I think that many of us have lived through the entire migration. I just threw out another batch of 3.5″ disks; I don't think I have any machines to read them on anymore. Social media is a great social equalizer. Seventy-two percent of Americans used social networking in 2013, up from 8% in 2005. The users of social media do not adhere to social boundaries – 67% are high school dropouts and 72% are college graduates. This media is a great social equalizer. You don't have a "have and have-not" situation in today's information availability. Historically, only those people with some money could really access this technology, and now everyone seems to have it. The third world has leapfrogged the rest of us with the implementation of cell phone towers, instead of laying long physical lines for telecommunications.

The kinds of indexes that we pursue have multiplied. New types of indexes have not necessarily replaced old ones. We still have back-of-the-book indexing, we still have subject headings, but now we have added machine learning and neural nets and rule systems tao the mix.

Our cozy little world has changed. The landscape is shifting in very profound ways and with it the funding models are changing. Who is going to pay for information is changing. The systems and the way that we manipulate knowledge are changing. The user wants and needs are changing, because users are becoming aware of all kinds of new technologies that they previously didn't know they could have. It is similar to impulse buying. In a store, you see people shopping and buying things they didn't know they needed until they saw them. This means that trends can change overnight in our mobile interconnected world. In the old days, we had industries and we worked within those industries. Products were sold to specific markets. You knew what your business model was. You could pursue it – like a 5-year business plan. You could make assumptions that were pretty solid. Professional associations represented those target markets and they knew exactly what they did and where they fit, and the world moved at a reliable, predictable pace.

10. Future directions

Now we have to challenge all of those assumptions. The future changes and future trends are very different, and innovation is absolutely constant. To find those growth opportunities is ever more challenging. We communicate in 140 characters in Twitter and convene a flash mob. We have small screens, and perhaps our thoughts are smaller, too, but we are always connected by social networking.

An Australian research study done recently says that 65% of preschoolers will work in jobs that have not been invented yet. That's sobering. During college, my daughters majored in subjects that I didn't necessarily know existed, and this situation is going to be more extreme by the time that today's preschoolers graduate from college.

We have new fields, new knowledge architecture, knowledge managers, and we have knowledge organization systems. We have new associations and a new company every time we turn around. The systems manipulating that knowledge are very interesting. We have ways to trick those systems and to trick the people that see them and predict the results. You can manipulate results so that a squirrel dying in front of your house might be considered more relevant than thousands of people dying in Africa.

10.1. Predictive systems and profiling

Building predictive systems that can lead people in one way or another for thought is actually still a straight-forward process. Profiling, although it has gotten a bad name lately, is still something that all of us do every day. We look at the way in which someone dresses, we look at the kind of car that they drive, we note the way that they talk, and we form an impression. We have a mental image of that person. It's just automatic human nature to profile. When you drive into a neighborhood, when you notice that a lawn is nice or that a car is up on blocks, you immediately form an impression. You also have an idea of the kinds of things that the residents might want to buy. Door-to-door salesmen always did very predictive analytics as to whether someone was likely to be a customer for them or not. I think that continues to be the case.

In building systems and delivering information to the right group, the hard part is not in building the systems; the hard part, to me, is finding the right group – finding the persons who will be your customers. I've talked with many people in the information industry that have said, "Where do I go for my customers now? Where do I meet them?". Special Libraries Association meetings are not the meetings they used to be. The American Library Association has a narrow profile of customers. International Online has disappeared. It's an interesting thing that as people search for new meetings to attend, they also search for new profiles on both the customer side and the vendor side. People are trying to figure out where to go to find people like them, people that meet their profile, so that they know where to congregate and learn.

Here is an example of providing different information to different people based on an automatically-generated profile. You can just look at this pictorially. Say, for example, that two individuals, Scott and Dan, each do their own search. Because they each have a personal profile digitally-generated from their previous searches and views in Google, the search results that they are served radically differ. If you search on two computers in your house, try to search for different kinds of things on them; otherwise, you'll find that the results are skewed in strange ways. The system is trying to figure out what *you* want – what kinds of things would be best served to *you*, what is most relevant. For example, let's say Scott and Dan both enter a search query about prices in Egypt. Scott's results include information about the protest in 2011 and other social and economic matters. Dan, using the same exact search string, received information about travel, vacations, and the CIA facts book. This shows radically different results from exactly the same search query based on a profile. We tend to prefer to read what we want and believe, and skip the items that present a view different from our own. We can read the *Wall Street Journal* or the *The New York Times* and the same news is presented from very different perspectives.

If you are of one persuasion and you read the *WSJ*, you love it. But if you are of the other persuasion and you read it, your reaction is YUCK. For balance you might need to read *The New York Times* or *The Washington Post* or a publication of that ilk, to try to get a balanced view of what is happening in the world. I have a fear that, just like in George Orwell's *Animal Farm*, we are going to be led into group think without being aware of it at all. It's the publishing community, more than the media community, I think, where we need to be careful. We can ruin a reputation by putting metadata behind a record which

will sink the ratings or move them to a bucket they don't want to be in. Put in a photo of an individual and a nice caption, but behind that add nasty affiliations like rapist and terrorist or something. The search engine will put that right at the bottom of the returned results.

10.2. Politics

We have our own political landscape to deal with. There is a "Y2K" type event coming up for health care in the United States, with the mandatory adoption of the ICD-10 coding system in the near future (ICD-10 is the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD), a medical classification list by the World Health Organization (WHO)). It codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases. Currently, the healthcare industry uses ICD-9, which has about 15,000 codes for medical diagnoses combined with the American Medical Association's CPT codes for procedures and HCPCS codes for supplies. These codes are all used for insurance billing purposes. The move to ICD-10 will offer 178,000 codes to choose from. When Europe implemented ICD-10 some years ago, productivity fell 50%, and they are still only at 70% of the patient encounters they accomplished prior to ICD-10. Coupled with increased regulation and Obamacare, a medical Y2K is coming to us all.

There are net neutrality battles going on, privacy and security concerns. It seems to be true that regulations cannot keep up with the pace of technology. I don't think that will change any time soon.

The world population is exploding, as we know, and the number of people that we have to deal with is amazing. The amount of money that is spent is not evenly distributed among the populations of the world. The United States accounts for 25% of the global economy and 20% of global manufacturing – that is a pretty good sized portion of the world.

10.3. Finances

On the other hand, as a country we have a huge debt, huge expenses, and not enough revenue to cover daily expenses. We grow that debt at $2.44 billion dollars each and every day. The Federal income for the United States is usually thought of as taxes, and this is one way that you can look at the government income. Actually, there is a significant amount of additional fees and other streams of income in the hidden taxes that go into the Federal budget. We need to keep an eye on all of those because the 17% that remains is where a lot of our customers' revenue comes from. If your customers are academics or not-for-profits, that's where your money is coming from. I believe that the funding models will change because the Federal budget is changing. If you look at the funds that the University of California spends, the amount that comes from the government is 18%, plus 11%, and plus their medical centers. Not exactly a revenue stream you can count on.

Last year in the State of Illinois, 18.9% of the University revenue came from the State, but that decreased 9.4% from four years earlier. In the State of New Mexico, in 2011–2012, we received 25% for the universities from the State and in 2013–2014, that number is down to 12%. We need to look for other ways to raise money or cut expenses.

The demographics are increasingly challenging, and using eminent domain and growth as drivers isn't going to continue to work for us.

10.4. Professional associations

For association memberships, when an employer pays for membership, no problem. Of course the employee would love to join. When that same employee has to pay personally for membership, they

often don't think it is worth the expense. Organizations are cutting back on membership sponsorships and travel expenses. The government has not paid many memberships and is cutting travel for next year. So the distribution of knowledge that comes from memberships and associations is in jeopardy. Knowledge distribution and membership models will have to change.

I don't think that associations are decreasing in value, but I think the way they perform will have to change. Many associations make a significant portion of their revenue from the annual meetings, and I think the annual meeting is no longer going to be quite the revenue-generating event for associations that it used to be. I think they will have to find other ways to reach out to members and other ways to supplement their budgets. Member recruitment and member retention will become more important. Member recruitment is connected with a well-known set of activities. As for member retention, once you have members, you've got to give those members reasons to remain and to keep coming back. Things like databases of information on the field, webinars, certifications and continuing education options are going to help hold people to an organization.

10.5. Librarians and Publishers

For libraries, there is a double whammy in that there are increasing numbers of people depending on Google and not going to the library. In addition, there is pressure on library funding, making it very difficult for them to know what the best model is to offer their patrons. "Ask the Experts" sites are very popular. In 2005, 15% of people used them. In 2013, 83% of people used them. That's where the library need is being satisfied.

We find that publishers are consolidating like crazy. Ebooks are surging. There are a lot of open access initiatives coming to the fore, plus we have less peer review and more self-publishing. The peer reviewers themselves are getting rather scarce, especially where the peer review is open and published. The diminishing use of secret ballots and blind feedback peer review is diminishing the number of people willing to contribute. These are massive changes for publishers, both in what they are doing and the ways they drive readers to content. Publishers have to be increasingly ingenious in order to make readers willing to move forward.

10.6. Technology

In all, we have had some pretty amazing battles. We have hardware and software, search and mobile devices. Behind these technologies are very big players with very big budgets battling for market shares. The growth in open source innovation and problem solving has pushed these capabilities into the hands of people all over the world. There are many more installations of Solr than there were just a few years ago. I was recently down in Brazil for a digital library conference. I found that conversations about technical innovation and access among Brazil, Uruguay, and Norway are extremely strong. We think of the United States as having most of the innovators, but actually, in the whole Java/Linux community, which is driving much of our business here, it is actually outside the United States where the basic innovation and application work is done.

Then we have weather. Now, when we have power outages, they are more of a problem than data breaches. In a power outage, we can't get in contact with each other; the phone is dead and the computer is down and we can't get onto the Internet. In a new map of the Internet by Martin Vargic, down in one quadrant is the old technology, showing it as a disintegrating continent, while other quadrants show growth areas. It is an extraordinarily detailed map and has practically every technology you can think of from the Internet.

10.7. Learning

The reality is that our future for meetings is set in learning. Learning is what people want to do most in the 21st century. We are going to have to find ways to support that, because learning is the way that we will be able to keep up. We have to replenish that knowledge. We have a universe of options. So what should we do?

We are changing the ways that we learn. We are changing the ways that we find things. We are finding it easier to manipulate what we know. We have comprehensive information, and it is increasingly invasive. And, remember that elusive end-user we all used to pine for? It's my mother now.

We need trusted sources, and we need to be able to trust the semantic enrichment on those sources. We need to be able to make data findable. We need to be sure that the data is trustworthy and replicable. We want discovery. We want precision and recall again, which was lost to relevance for many years. I am finding that people really want that more than they want auto-clustering. Auto-clustering frustrates some. They want additive search results. When they do the search they want what they got before plus the new stuff. They don't want a totally new search set returned. They want to be able to aid the human brain and automate everything they can for efficient processing.

10.8. The new path

For me and my team, our new path is taxonomies. I am immersed in taxonomies and metadata. They have finally come to the fore. Publishers are increasingly using sophisticated metadata and putting a semantic fingerprint on everything. Production has shifted heavily to the Web, so we don't have to depend so heavily on "big iron" in-house. We've put power into the hands of the users so the authors can submit and do their own tagging. We can channel the data to users directly by predictive analytics. We do much more semantic tagging to leverage and re-combine and re-process the data that we already have. Many content creators and providers are selling directly to the consumer. I think that publishers are going to embrace the Web for distribution. I think that sales to the government will decrease because, I think, the government won't have as much money. I do think that enterprise sales will increase and I think that medical coding is going to be a huge issue for us in the coming years. For reaching customers, I think that layout ads aren't going to do much for organizations because it is hard to reach the customer reliably. I think fewer people will attend conferences, so we will have to find another way to do a personal approach.

My path has been with Access Innovations. We do services to create content-aware data. We created a tool set for content enrichment that includes quite a lot of automation. My latest enterprise is Access Integrity, an initiative to automate medical coding. I was going to tell you about Mount Pinatubo, but that's a story for another time.

The future: information – any place – anytime. Everyone can create content, and it will be a great big mess unless we can control it. We need to tag it, clean it, weed it, and curate it, and we know how to do that. We are in a perfect position. We have those skills – and that's good because the information explosion has just begun! I am looking forward to it![1]

[1]This paper is based upon the Miles Conrad Award Lecture that was given at the 2014 NFAIS Annual Conference in Philadelphia, PA, USA on February 24, 2014. The Miles Conrad Award was established shortly after the death of NFAIS' first President and one of the original founders, G. Miles Conrad, in order to provide a fitting memorial to his accomplishments. The award has been presented every year since 1968 to an information industry leader who has made significant contributions to the Information Community and who has been a supporter of NFAIS. A complete list of awardees can be access on the NFAIS website

About the author

Educated as a botanist and trained by NASA as an information engineer, Marjorie (Margie) Hlava has worked behind the scenes for most of the major information organizations. In the early days she worked at NASA, logging up to 20 hours per week as an online searcher using such systems as the NASA Recon, Dialog, BRS and SDC. She was also the Information Director for the DOE National Energy Information Center and its affiliate NEICA. She rose to the position of Information Director before taking her team private as Access Innovations, Inc., in 1978.

Her abiding research interests center on speeding the human processes in knowledge management through productivity enhancements. She developed the Data Harmony software suite specifically to increase accuracy and consistency while streamlining the clerical aspects in editorial and indexing tasks. Her other developments include the XML Internet System (XIS) for the capture and creation of metadata, and both Thesaurus Master and Ontology Master for the management of thesauri, taxonomies, and ontologies. Through over 2000 engagements at Access Innovations, she has been at the cutting edge of technical innovations and their implementations for her entire career.

Margie has served as an officer and/or Board member of SLA, ASIS&T, NISO, ASIDIC, NFAIS and other organizations. She has worked on the development of standards, authored two books and 200+ articles, and holds two U.S. patents encompassing 21 patent claims. Her work has been acknowledged through numerous awards during her career, including ASIS&T's Watson Davis award, and recognition both as an SLA Fellow and as a Woman of Influence for Technology. Margie has said that she has no intention of resting on her laurels and plans to continue her adventures in Information Science and explore the boundaries of new technology and methodologies.

at: http://nfais.org/complete-miles-conrad-lecture. The slides related to the presentation can be accessed on the NFAIS Web site at: http://info.nfais.org/info/ANCO14_MCL_Hlava.pptx.