An overview of the NFAIS 2014 Annual Conference – *Giving Voice to Content: Re-Envisioning the Business of Information*

Bonnie Lawlor

Guest Editor, NFAIS Executive Director 2002–2013, 276 Upper Gulph Road, Radnor, PA 19087, USA
E-mail: chescot@aol.com

Abstract. This paper provides an overview of the highlights of the 2014 NFAIS Annual Conference, *Giving Voice to Content: Re-Envisioning the Business of Information*, that was held in Philadelphia, PA, on February 23–25, 2014. The goal of the conference was to take a look at the opportunities offered by current Big Data technologies (metrics, analytics, data mining, visualization, linking, etc.) and challenge content providers and librarians to attempt to free themselves from the constraints and biases of the past and re-envision what their products, services, and business models would look like if these technologies were fully utilized. Audience response devices were used throughout the meeting to gauge the opinions and perceptions of the attendees and these are also discussed where relevant.

1. Introduction

In 2013, a thought-provoking book entitled *Big Data: A Revolution that will Transform How We Live, Work and Think* by Viktor Mayer-Schönberger and Kenneth Cukier was released. In it they state that the move to big data is a continuation of humankind's quest to measure, record, and analyze the world. What is new is that current technology makes this possible in ways that we have not experienced to date. But to make the most of this technology requires a new mindset – one that allows us to look at the world not just as a source of information, but as oceans of data than can be explored at ever greater breadth and depth. The ultimate reward will be the "... extraction of new insights or the creation of new forms of value in ways that change markets, organizations, the relationships between citizens and governments and more".[1]

The authors make it abundantly clear that the Big Data revolution is not in the machines that calculate the data, but in the data itself and how we use it – how we identify and understand the relations within and among the pieces of information that we struggle to fully grasp and, if we are ultimately successful, will allow us to give voice to content!

So what is this new information mindset and what are the key behavioral and technological trends that are driving it? What approaches are being used to collect, manage and integrate diverse data sets?

---

[1]V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution that will Transform How We Live, Work and Think*, An Eamon Dolan Book, Houghton Mifflin Harcourt, 2013, p. 6.

How are big data techniques being applied to those datasets in order to create new knowledge? How will the application of big data techniques change the creation, delivery and use of information and what does it mean for publishers, database producers, aggregators, distributers, librarians and users? To answer these questions a group of researchers, publishers, librarians and technologists met earlier this year in Philadelphia, PA, when the National Federation of Advanced Information Services (NFAIS™) held a two-and-a-half day conference entitled *Giving Voice to Content: Re-Envisioning the Business of Information*.

The purpose was to highlight examples of how the application of Big Data technologies such as data mining, linking, analytics and metrics can expose, enhance and create new information and insights from data already at hand.

2. Setting the stage

The opening keynote was given by **Hilary Mason**, a Data Scientist in Residence at Accel and Co-Founder of HackNY, an organization that seeks to complement traditional computer science education with student hackathons and connect students with start-ups in New York. She noted a quote made by Richard Hamming back in the early 1960s, "The purpose of computing is insights, not numbers", and said that gaining insight is really what Big Data is all about. Computing should be perceived as a tool for enhancing cognitive theory (a perspective to be re-enforced throughout the conference) and now that the cost of technology is far less than that experienced during Hamming's era more than fifty years ago, those insights are within our reach. She talked briefly about the emergence of "data scientists", a new job title coined about four years ago, who are responsible for obtaining the data, cleansing it, exploring it using techniques such as visualization and clustering, and then modeling and interpreting the data. She asked members of the audience if the cultures of their organizations were data-driven (conference attendees had been provided with devices to answer such question throughout the meeting). Less than half (44%) said yes, 37% said that they were just beginning and 19% responded no. When asked if they had a plan to use data for extracting more value from existing assets, the responses were 33% yes, 41% said that they were just beginning and 26% said no. She then encouraged all of the content creators in the audience to be aggressive about applying the "Big Data" mindset to their content in order to identify and seize new business opportunities.

3. Mining digital traces

Hilary was followed by a session entitled "Enhancing policy decision-making with large-scale digital traces" that provided a practical example of how insights can be extracted from data. **Vanessa Frias-Martinez**, Assistant Professor from the College of Information Studies at the University of Maryland, discussed how she has used cell phone data to assist in the gathering of socioeconomic data. She chose to use cell phone data because of the amount of data that cell phones generate. There are approximately 5.9 billion SIM cards (smart cards that store data for GSM cellular telephone subscribers) in use, representing 2.2 billion unique users (45% of the world's population). The cards gather data such as user identity, location and phone number, network authorization data, personal security keys, contact lists and stored text messages. The core goal of her ongoing research is to see if human behavioral information can be automatically extracted from mobile digital traces in order to augment or replace the current

manual efforts (surveys, interviews, etc.) that are being used. She provided two case studies. The first attempted to develop socioeconomic maps that are used by National Statistical Institutes to determine where services such as schools, hospitals, etc. are needed. These maps are expensive and labor-intensive to create and as result are not updated as often as needed. The second study attempted to determine the success of a government health policy.

The first study utilized six months of cell phone data from five hundred thousand users in a Latin American city that had 920 call areas. This data was combined with information on socioeconomic levels that had been manually gathered for one specific region within that city. The goal of the study was to see if information gathered by mining the two data sets could be used to develop a model that would allow the socioeconomic levels (SEL) of other regions in that city to be accurately predicted. The results were 86% accurate when three SELs were involved, but less accurate (68%) when six SELs were used. More work will be done to refine the models, but there is hope that this methodology will allow for the automatic generation of such maps, thus reducing the manual work currently required.

The second study was done to see if a health policy instituted by the Mexican government actually worked. When the government faced the breakout of an epidemic several years ago, they first issued a general alert to the public, followed later by the closing of schools, and then by the issuance of a non-travel policy to minimize the spread of disease. The study took cell phone data that had been gathered before, during, and after the three stages of the emergency to see if people actually did stay home or travelled less. They found that there was a 10–30% reduction in travel and that the peak of the epidemic was actually delayed by forty days, allowing the medical community to strengthen their response. Dr. Frias-Martinez left the audience with the following take-away from her research – the fact that geo-located data can be used to model human behavior, measure behavioral changes, and predict and classify external sources of information.

4. Leveraging a historical archive

The next session was a fascinating look at how *The New York Times* (*NYT*) is attempting to leverage the history and heritage of its vast archive in order to create new value. **Evan Sandhaus**, Director for Search, Archives and Semantics at the *Times*, noted that the newspaper has been continuously published since September 18, 1851. As of February 22, 2014 the paper had published 58,669 issues consisting of 3.5 million unique pages and 15,171,586 articles. Over the years there have been four approaches to making the data accessible and usable. The first was the morgue (the physical archive). This is a warren of file cabinets located three levels below the streets of Manhattan and consists of newspapers, clippings filed by subject, and millions of analog photos that have never been digitized. Its search interface is a librarian and to this day the files are used (Bryan Cranston who won the 2014 Tony for leading actor in a play used the files to prepare for his role as Lyndon B. Johnson in *All the Way*). But access is controlled and limited. The second approach was the creation of *The New York Times Index*. It actually is as old as the newspaper, with the first edition written by hand. It was made commercially available in 1913 and is an index to each individual year of the paper. To search it comprehensively one has to gather the information from each year – there is no cumulative index across years. The third approach was *The New York Times Information Bank* that was launched in 1972. It was launched as a digital index accessible to the *Times* newsroom and to libraries via a subscription fee. There was no text searching, rather specific codes needed to be used. In the end it was not profitable and was sold after ten years to the

Mead Corporation who in turn sold it to Elsevier ten years later. The content still appears in LexisNexis. The fourth and most recent approach is called the *Times Machine*. It leverages the content that has been in digital format since 1981 as well as the scanned material from 1851 through 1980. The goal is to take the existing archives – all 163 years of the paper – and to give users an experience that is as close as possible to actually reading a newspaper. This new service is in the early stages, but a free sample can be accessed at: http://timesmachine.nytimes.com/timesmachine/1904/06/16/issue.html.

5. A new data mining business model

Kalev Leetaru, a Yahoo! Fellow in Residence at Georgetown University, continued the theme of creating value through the application of Big Data techniques in his presentation "The virtual reading room: Enabling secure data mining of vast content archives". He opened his presentation by asking the audience if they have ever provided exports of their data to academic research projects for data mining. The response was 60% yes and 40% no. He noted that while the text of books is being digitized, the related images have been relatively ignored. In one digital library there are three hundred billion words, but no one knew how many images (photos, maps, illustrations, etc.) were associated with the text. The images had been extracted and identified, but only to separate them from the text. A project was developed to identify the images, extract them and the associated text, and create an index to five hundred years of imagery. The net result was a searchable database of sixty million images from one million books – no new technology was used, it was just a new way of looking at content that already existed. The database facilitates the retrieval of imagery related to keywords such as "telephone". One of the search results on that topic was a photo from 1926 of a young girl talking on the phone. The caption held a prediction that in the future, teenagers would be the highest users of telephones! Another project was the data mining and mapping of the Internet Archive's television news collection. This is the largest television news archive on earth, with 609,992 shows recorded, 488,118 shows with closed captioning streams, including 1,417 programs from 63 channels and 2.7 billion words of closed captioning ($\sim$110 words per minute). The end result was an interactive map of the geography of TV news, with the countries getting the most news being highlighted. He then asked the audience if they are exploring ways to make their content more available for data mining. The responses were 65% yes and 35% no.

He noted that there are currently three approaches to data mining and all have problems: (1) Special extracts of content shipped to researchers on USB drives (great for the researchers, but either not sustainable or not permitted); (2) Page/Article-level APIs (most data mining tools need access to the whole corpus for normalization. The *NYTimes* has shrunk 50% over sixty years and researchers need to know this to trace a keyword over time); and (3) Custom extracts like "n grams" (most data mining tools need access to the text to see the words in context and in order, not just how many times a word was used by year). He then discussed a fourth data mining model for publishers that could minimize, if not eliminate, the fear of intellectual property being "stolen", and that is the use of hosted cloud computing where researchers run their codes directly on the content holder's servers. He told the audience to think of digital libraries like archives and not libraries. With libraries the material is checked out and read at home – the content travels to analysis. With archives the user physically travels to the reading room, analyzes the material there, and can only take their notes when they leave – analysis travels to content. This is the model that was used on the Internet Archive Reading Room project. He geocoded the 2.7 billion words of closed captioning content. The TV material did not leave Internet Archive's physical premises.

Instead, geocoding algorithms were submitted to the Internet Archive and they ran on a server physically located on IA's premises. Only the final list of locations (such as "Kiev, Ukraine") were returned for analyses – not the full corpus. The Virtual Reading Room is a three-part model: technical, methodological, and legal. Publishers could use this same model to enable data mining of their collections. Essentially a researcher would get a login to a secure computer running in the publisher's data center. The computer has access to all of the material to which the researcher's institution subscribes and data mining tools can run on all of the content, but none of the content can leave, only the codified research results. Data mining access could be made an option on the subscription package. For an incremental additional fee, subscribing institutions could receive a set of logins for their researchers. As a result researchers would have a legal and authorized platform for all of their data mining needs, the data is even in data mining-friendly format ready to go, and the publisher has a new revenue stream.

Before closing, he asked the audience if they were exploring making more sophisticated kinds of search available, like geographic, topical or entity search. The response was 65% yes and 35% no. His closing message was that there has been a huge shift in how content is used because current researchers think in terms of information, not its containers. They want to jump right to the piece of information they need. And successful content providers will not be those who just provide the most content in one place, but those who provide the best access to that content.

6. The British Library and digital scholarship

The final speaker of the conference's opening day was **Mahendra Mahey**, Project Manager of the British Library lab. The lab is funded by the Mellon Foundation and has the goal of getting the Library's collections more widely used, especially by researchers in the arts and humanities. The Library holds more that 150 million items (14 million books, 60 million patents, 8 million stamps, 4 million maps, 3 million sound recordings, 1.6 million musical scores, 0.3 million manuscripts and 0.8 million serial titles). They want to digitize as much as possible to improve accessibility. In partnership with Microsoft they have already digitized 68,000 volumes from the seventeenth, eighteenth and nineteenth centuries and a quarter of a million books have been digitized in partnership with Google.

The lab ran its first annual competition designed to attract scholars and software developers over the age of eighteen who see the potential for new and innovative research and development opportunities in the Library's digital collections. Contestants were invited to submit a proposal and if their idea was selected the Lab would work with them over a four to six month period to bring their idea to fruition. Prizes of up to £3,000 were awarded. Twenty proposals were submitted for the 2013 competition and there were two winners. The first was Pieter Francois, a Post-Doctoral Researcher at the University of Oxford, whose proposal was for the development of Sample Generator for Digitized Texts. He uses statistical methods to support text analysis and the tool that was developed produces representative samples of texts based on search criteria (see prototype at: http://goo.gl/YFnZmu). The second was Dan Norton, a PhD researcher at the University of Dundee and Artist in Residence at the Hangar Centre for Art and Research in Barcelona, Spain. His proposal, "Mixing the library: The disc jockey and the digital collection", applied the model of information interaction to develop an innovative prototype for working with multiple content sources and data-representations from British Library digital collections. Funding has been approved to run the competition again in 2014.

You really should take a look at the very fascinating results of another competition run by the Library that resulted in a mash up of Library maps and music by second year students at De Montfort University, Leicester, UK (http://youtu.be/SPY-hr-8-MO) – definitely impressive!

7. The signal economy

The second day of the conference was opened up by **John Blossom**, President of Shore Communications, with his plenary session on the Signal Economy. John defined the term signal as "clear status and actionable indicators derived from complex inputs that provides highly-actionable information at the right time and place". He stated that signal is the most abundant knowledge resource today because everything can generate signals due to electronics (this statement was repeated by the closing keynote speaker). Consider that the basic smartphone incorporates a number of sensors that monitor voice, location, identity, etc. (cell phones have already been discussed as an important source of data), and that thirty-two billion new personal sensors will be added in 2014. He noted that the Cisco Corporation has put an economic value of $14.4 trillion on the signal economy and they believe that the Internet of Everything has the potential of growing global corporate profits by 21% in aggregate by 2022.[2] It is the generation, collecting, organizing and analysis of signals that will drive economic activity predictively at unprecedented scale. As a result of this new economy, publishers will need to move from information to predictive services; from auto-categorization to auto-contextualization; from building data sets to signifying signal sets; from extracting entities to mapping realities; from computing to thinking machines; and from analysis to tailored actions. John asked two questions of the audience to see how they are reacting to the changes in the information economy. He first asked if they were integrating big data analytics into their publications or client platforms. The response was 43% yes, 45% no and 13% uncertain. He then asked the audience if they believed that the provision of predictive services would add value to their current or prospective clients. Here the response was 72% yes, 17% no and 12% uncertain. For more information on the signal economy refer to John's paper on the subject that appears elsewhere in this issue.

8. Product and processing changes

The second part of the morning took a look at some of the changes that publishers are making as they re-envision their content. The first speaker was **Mike Teets**, Vice President of Innovation at OCLC, who discussed how the information industry has changed over the past quarter century. He noted that in 1986 people were concerned with counting the number of records and databases. By 2005, this stopped and publishers began processing data in a new way due to the increased scale of content being produced. This growth in content was due in part to the advent of open access publishing and to the emerging globalization of publishing. He noted that in 2005 there were 1,643 open access journals, with this number growing to 9,759 in 2013. In the year 2000 there were approximately 30,000 scientific papers published by Chinese authors, with this number more than doubling by 2006! In just the special sub-discipline of "strength training" that he found in PubMed, the number of articles published annually had grown from about 400 in 1999 to just under 1,800 in 2011. In the latter case he noted that the number of new articles in that field equated to one being published every hour during a normal forty-hour work week, making it difficult for any one reader to keep current. And making data curation and processing even more difficult. In order to make this growing body of information more findable and accessible OCLC has moved to an "entity" view of their content. They decompose information to its elemental forms and derive relationships and links to make it more accessible. They first define the elements,

---

[2]The Internet of Everything (IoE): The Value at Stake in the IoE Economy, CiscoIBSG 2013, http://www.slideshare.net/CiscoIBSG/internet-of-everything-ioe-economy.

note and link the relationships between those elements, and store the entities as "Triples" (the term for the most granular relationship possible between discrete pieces of information). They expect to have 31 billion Triples stored by the end of 2014. As a result of this new form of processing, OCLC has had to change 80–90% of their processing infrastructure over the past two years. They have moved from an Oracle database of record (because of its inability to scale in a financially-viable way), to the use of Hadoop which provides greater flexibility. In closing, Mike asked the audience if their organizations supported entity views. The response was that 26% currently publish entity views, 36% are actively considering it, 18% said that they are not doing it, but expect others to extract and related entities from their databases and 20% said that they do not want their data presented as entities.

9. Interactive 3D publishing

Another major change was noted by **IJsbrand Jan Aalbersberg**, Vice President of Content Innovation at Elsevier, who discussed Elsevier's move to include interactive 3D visualization into the online journal articles that are hosted on the Science Direct platform. This was done in partnership with Kitware, a visualization technology company, in response to customer demand, and was announced in October 2013. The first release was in the field of neuroscience, and the 3D enhancement has since been expanded to other disciplines. The details of this project are discussed in an article that appears elsewhere in this issue, but it is interesting to note the following two responses from the audience when Dr. Aalbersberg asked some questions. He first wanted to know how the audience believed people would choose to read their scientific articles three years from now. The response was that 60% believed people would read the articles on a tablet, 17% said that readers would use a smartphone, 14% said that readers would use a computer-browser (HTML (5)), 7% said that they would use a computer and read a pdf file and 2% said that readers would still use print. He also asked the audience how important 3D content and 3D visualization was to their readers. Only 17% believe that it is not important; a total of 83% gave 3D content and visualization some level of importance (16% very important, 19% important and 49% somewhat important).

10. Assigning trust to content

After the morning sessions ended, NFAIS members attended a members-only lunch during which Dr. **Carol Tenopir**, Chancellor's Professor and Director of the Center for Information and Communication Studies at the University of Tennessee, discussed the results of her recently-completed study on how researchers assign trust to scholarly content. Dr. Tenopir has summarized some of the key results in an article that appears elsewhere in this issue, but I want to highlight the responses to the questions that she posed to the audience. She asked if they believe that the criteria for "trustworthiness" are the same when applied to how researchers choose what to read, what to cite, and where to publish. The responses were 41% yes, 49% no and 10% had no opinion. She then asked if they believe that the criteria for "trustworthiness" of information varies by subject discipline. The responses were 80% yes, 16% no and 4% no opinion. The audience was then asked what they believe is most important in their publications to help readers trust. Brand name was as the number one response at 50%, articles by well-respected authors received the second rating at 28%, metrics such as journal impact factor came in third with 13%, abstracts followed at 6% and the lowest ranked was article level metrics (altmetrics) at 3%. Read her article to see how the audience responses matched the survey results.

11. Crowd computing – A new workforce model

The first session of the afternoon took a look at emerging technologies. **Adam Devine**, Vice President of Product Marketing and Partnerships at Crowd Computing Systems, Inc., discussed crowd computing as a means of augmenting an existing workforce. He pointed out that users want more content, but that existing systems and staff are often maxed out from a processing perspective, especially when production is cyclical – with peaks and valleys. He said that the major problems that publishers face are rising costs and shrinking margins and that the roadblocks are twofold: a financially unscalable and inflexible workforce (one that cannot expand and contract easily) and technology in that the latter requires time and money to build the new systems and infrastructures that are needed. He asked if those in the audience have crowd sourced their content/data production. The response was that 58% have not, 20% have experimented with it, another 20% have crowd sourced some of their work and 3% said that they have maximized their use of the crowd. He also asked if it was true that their automation tools, distributed workers and subject matter experts were not integrated into the same workflows. The response was that this lack of integration is a reality for 60% of the audience. With regard to this response he said that this is the most common problem – that wall exist between automation, crowed-sourced labor and the regular staff.

What his organization does is unite these three silos in the cloud, creating a hybrid workforce solution that utilizes WorkFusion, their proprietary, intelligent work-management platform. This approach maximizes the effectiveness and output of automation and the crowd, leaving the regular staff to do only the highly-skilled tasks. He used Thomson Reuters as an example. One of their divisions is responsible for creating 40,000 tear sheets a year. These are one-page summaries of public companies that provide an overview of business segments, recent operating results and key fundamental analysis metrics. It took a year for a staff of fifteen to create these sheets and the staff was maxed out. Using the services of CrowdComputing, the staff was reduced to four, the use of automation went from 0% to 94%, and the processing time went from twelve months to one month. Adam noted that his company is the largest employer on the face of the earth as they have access to sixteen million freelance employees. They must be doing something right, because about two months after the conference, on April 29, 2014, they issued a press release that they had just secured $15 million in Series B Funding[3] and renamed their company WorkFusion because of the brand recognition that their platform had gained.

12. Big Data technologies for publishing

Bradley Allen, Chief Architect at Elsevier, talked about the role of Big Data in Scientific publishing. He noted that scientific publishing is the act of compressing a universe's worth of data into small pieces of content that people can consume and, in essence, this is the ultimate big data problem. But it is one in which until recently publishers have played a very simple role – a fact that is in the process of changing. Publishers can now create more useful content by enhancing it with data extracted from content. They can make the researcher's life better by exploiting data about how content is used to improve his/her experience of using their online applications. And they can enable research itself by supporting the care and feeding of experimental data at scale (refer to the 3D publishing discussed earlier). He asked the audience what amount of data they routinely manage to create their online services.

---

[3]PR Newswire Press release, WorkFusion Secures $15M Series B Funding, April 29, 2014, http://www.prnewswire.com/news-releases/workfusion-secures-15m-series-b-funding-257149621.html.

The responses were Terabytes (54%), Gigabytes (21%), Petabytes (19%) and Exabytes (7%). Brad said that in today's world most publishers are dealing with Terabytes. He then asked what use of Big Data is most important for their organization and he listed three: (1) extracting data from content, (2) improving user experience through usage analytics and (3) managing experimental data. The highest-rated response at 36% was all of the above. The second highest was number 2 at 29%. The third highest was number 1 at 24%. Number 3 came in fourth at 7%. And 3% of the audience said that none of them were most important for their organization. He then discussed the key tools currently used to manage Big Data – Hadoop, LexisNexis HPCC, Twitter Storm, etc., and noted that all of them make it easy to perform routine Big Data Processing – they break the data into many chunks, each chunk of which can fit into memory on a given machine, then send each chunk to a machine where it is processed into chunks containing intermediate results, ultimately combining the intermediate results into a single aggregate data set which is then cleansed and the whole process repeated. He listed the major Big Data Technology issues as being talent acquisition; best practices and design patterns for big data; the proliferation of Big Data platforms (Cloudera, MapR, HPCC, etc.); data formats and practices for scaling; and numerical computing frameworks for optimization. He closed his presentation with a thought-provoking question: can Big Data be used to enable new business models and optimize the value exchange between author, publisher and reader? He believes it can, but that to do so will require all of us to walk away from legacy preconceptions and ask ourselves "do we do it this way because it is the most effective way OR because it was done this way on paper?".

13. Miles Conrad Lecture

The final session of the afternoon was the Miles Conrad Lecture. This presentation is given by the person selected by the NFAIS Board of Directors to receive the Miles Conrad Award – the organization's highest honor. This year's awardee was **Marjorie M.K. Hlava**, President, Access Innovations, Inc., and the complete transcript of her talk is published in full elsewhere in this issue. It gives an interesting perspective on the past, present and future direction of the information industry based upon Margie's wealth of industry knowledge and experience, and is well worth a read.

14. New business models

The final day of the conference opened with a session on business models. The first speaker was **Mitchell Davis**, Founder and Chief Business officer of Bibliolabs. He opened his presentation by asking the audience how wide they believe the gap is between a consumer mobile media user experience and those delivered by libraries. The number one answer (58%) was that the gap is large and that it is a huge risk to the future of libraries. The second highest response (28%) was that there is a gap, but people will put up with it. Only 4% did not believe there is a gap; and 10% had no opinion. He then showed data that said that while e-books comprise 60% of total book sales, 94% of library patrons have never checked out an e-book, and he talked about the competition that is given to libraries by organizations such as Amazon and Netflix, both of which provide elegant user experiences. His organization has created BiblioBoard to help libraries provide a better user experience and eliminate the complexity associated with navigating the information silos within a library. In fact, The British Library worked with them to make hundreds of historic works from the Library's popular 19th Century Historical Collection available in a series of curated Anthologies on BiblioBoard, a free iPad® app that offers a platform to explore places, events,

people and themes from across the centuries. More details on this successful platform can be found in an article by Mitchell Davis that is published elsewhere in this issue.

An equally interesting presentation was given by **Nader Qaimari**, Senior Vice President of Marketing at GALE (a part of Cengage Learning) who talked about their newly-released purchase-model, Usage Driven Acquisition (UDA). Using this model libraries can access a collection of up to two thousand reference titles for a full six months, and then ultimately own into perpetuity the tiles that were most used. More details on this new business model can be found in Nader's article that is also published elsewhere in this issue.

The final speaker in this session was **Judy Luther**, President of Informed Strategies, speaking on behalf of one of her clients, Knowledge Unlatched, an organization that is helping stakeholders to work together for a sustainable open access future for scholarly books. Judy noted that because of the market adoption of an open access models for journals, scholarly books are being driven to offer a similar model for their readers. These forces have become even stronger since the Association of Scientific, Technical and Medical Publishers (STM) has stated that they recognize that open access is no longer an "if", but rather a "when". But in those disciplines where books are not well funded or for which the demand is limited, publishers simply cannot afford to take risks. Print runs are declining and the costs are being spread over a smaller pool. Also, the "discoverability" of books is difficult. Unlike journals that have abstracts, DOIs, links, and are covered by abstracting and indexing services, the finding of books is driven mainly by MARC records.

Knowledge Unlatched (KU) looked at the first-copy cost and asked what if a publisher knew up front that there was a large enough customer base at the point of publication. This core market could support the costs. Thus a pilot project was initiated to see if the idea was feasible. They sought the cooperation of 200 libraries and more than 250 signed-up by the closing date of February 28, 2014. Thirteen publishers, mostly academic presses, joined the project and submitted one hundred titles of which twenty-eight were selected because of where they were in their publication cycle. The cost per library was projected at $1,680 ($60/journal) for the projected 200 libraries. The actual cost dropped to $50/title since more libraries participated than originally projected. Publishers are paid a pre-agreed fee in return for making the books available on a Creative Commons license via OAPEN and HathiTrust as a fully downloadable PDF immediately upon publication. It became a global initiative and Knowledge Unlatched worked with LYRASIS, JISC Collections and the Max Planck Society to facilitate library sign up. Those who did sign up for the pilot are charter members and will have a say in the future management of the initiative. It should be noted that as of the date of this presentation, nineteen countries had responded and there was more participation from libraries outside of North America. There are three steps to the project: the pilot which ran from October 2013 to February 2014; a second phase to obtain additional funding to increase the scale of the project; and finally, the development of a sustainability model. KU receives a membership fee and a percent of the title cost. It is very much a community-based project that benefits readers, authors, libraries and publishers. For more information visit the KU site at: htpp://knowledgeunlatched.org.

15. Policies, practices and procedures

The final morning session was an interactive panel discussion on the policies, practices, and procedures that are required to facilitate, rather than inhibit, the use of content. The panelists included the moderator, **Chris Kenneally**, Director of Business Development at the Copyright Clearance Center, **Britt Mueller**, Senior Director, Qualcomm Library and Information Services, **Maryann Martone**,

Executive Director Force 11 and Professor-in-Residence, Department of Neuroscience, University of California, San Diego, **William Trippe**, Director of Technology, MIT Press, and **October Ivins**, Principal at Ivins e-Content Solutions. Chris Kenneally's transcript of that panel discussion appears elsewhere in this issue, but I include here the answers of the questions that he posed to the audience. First, he asked about the future of open access publishing (OA) over the next five years, giving four options from which to select: (1) OA is inevitable for all scholarly publishing (19% said yes); (2) OA is unsustainable as a business model (25% said yes); (3) OA will be required by government and funder mandates (25% said yes); and (4) All of the above (the winner at 32%). He then asked what members of the audience considered the most important requirement of content within the context of their work environment. 21% said that content must be actionable, 15% said that they must be able to share it, 6% said that it must be secure, 2% required mobility and 56% said that all of those factors must be met. He also asked what the audience considered to be the greatest threat facing the information industry today. Number one (35%) was listed as disappearing business models; number 2 (25%) was information overload; number 3 (19%) was a resistance to technological change; number 4 (8%) was piracy/copyright infringement; and 13% said that none of those listed were the greatest threat (the audience response devices did not allow for write-in alternatives).

16. A new era of computing

The closing keynote of the conference was given by **Kerrie Holley**, an IBM Fellow and Chief Technology Officer, IBM SOA Center of Excellence, and his theme was the next era of computing. He stated that there are four major trends that are driving this new era: the globalization of the workforce and businesses; demographic shifts (a decline in the labor population and birth rates); mobility and social networking; and the explosion of data (Big Data). He re-enforced John Blossom's focus on the importance of sensor data and agreed that it will be the major driver of the future. He also stated that while many of us may view social networking as the hedonistic and narcissistic behavior of the young, it is much more than that and will definitely shape the future of information. Indeed it is the social and sensor data that is growing the fastest. Some say that we are hindered by too much data. Holley says just the opposite – we are hindered by too little data or by simply not knowing enough about the data that we have. He compared a piece of information to a piece of a puzzle. What it represents is unknown – there is no picture on hand. Some pieces are duplicates, missing, incomplete, of low quality, or have been misinterpreted. Some pieces may even be professionally fabricated lies. Until you attempt assembly, you do not know what you are dealing with. More data finds data – hence the true value of analyzing Big Data. He then talked about the three eras of computing. The first was the tabulating era with mechanical devices doing simple tasks; the second was the programmable era with a focus on the Von Neumann architecture (a stored program concept), and the new era as being one of cognitive computing where the computer takes on a more human-like characteristics (e.g. artificial intelligence, natural language processing, machine learning, etc.). He said that it is possible to build computer intelligence into a traffic light that could detect the potential for an accident to happen and warn the vehicles involved or alert emergency vehicles to arrive. He also briefly discussed IBM's Watson (remember the computer on the TV program Jeopardy?). It is now being utilized in the healthcare industry to augment the physician's performance. He also said that while 96.7% of Google's revenue is from advertising it would be irrelevant to ask if they were in the advertising business. Their business knows no boundaries – they are aggressive in being essential and ubiquitous in every industry. Amazon is no different – they are into books, newspapers, grocery delivery, etc.

He discussed at length the importance of API's as digital assets to make products and services more accessible and usable. In fact, he stated that not having API's today is like not having a website in the 1990s.

He said that a major trend is the emergence of new, born on the Internet enterprise software companies that provide new models for collaboration and life cycle management for applications and code; that use the Cloud for Internal IT and product delivery; that deliver customer experience and business experiences through software as a service; and that use a self-service sales model with scalable subscription pricing models. The API-centric, as-a-Service delivery is disrupting the consumption of business services just as the Cloud disrupted the IT consumption model. He closed saying that with the trends and applications that are fueling the creation of the cognitive computing era, businesses face significant pressure to leverage technology and differentiate like never before.

17. Conclusion

The presentations in this conference only re-enforced my belief that the information landscape has moved into a totally new era. Technology has freed content from its silos and given it voice, allowing previously hidden meanings and relationships to be revealed. The technology is available and affordable (to most) and content providers must be aggressive in leveraging that technology to the benefit of their users and for their own long-term viability. They must provide answers, not reams of information, and they must provide tools that allow users to analyze those answers. Digital journals must allow researchers to replicate results from the original experimental data and facilitate their interaction with that data. As put forth by the authors of the book I mentioned earlier, *Big Data: A Revolution that will Transform How We Live, Work and Think*, a new information mindset has indeed emerged. The ability to unearth or create new layers of knowledge using big data techniques such as data mining, linking, analytics and metrics challenges publishers and librarians to ensure their relevance in a highly-competitive digital information world. How? By re-envisioning both their content and its use to give researchers what they need – information products that provide the answers that they seek along with the policies, practices, and business models that will facilitate – not inhibit – the use of that information. But as Bradly Allen noted in the conclusion of his presentation, to leverage this opportunity requires a willingness to embrace the new information mindset *and* move beyond the comfort zone of current practices and processes.

Note: If permission was given to post them, the speaker slides are embedded within the conference program at: http://nfais.org/event?eventID=537.

About the author

Bonnie Lawlor served from 2002–2013 as the Executive Director of the National Federation of Advanced Information Services (NFAIS), an international membership organization comprised of the world's leading content and information technology providers. Prior to this, Bonnie was Senior Vice President and General Manager of ProQuest's Library Division where she was responsible for the development and worldwide sales and marketing of their products to academic, public and government libraries. Before ProQuest, Bonnie was Executive Vice President, Database Publishing at the Institute

for Scientific Information (ISI – now Thomson Reuters) where she was responsible for product development, production, publisher relations, editorial content, and worldwide sales and marketing of all of ISI's products and services. She is active in the American Chemical Society, and a member of the Bureau of the International Union of Pure and Applied Chemistry. She is currently on the Boards of LYRASIS, one of the major library consortia in the Unites States, and the Philosopher's Information Center, the producer of the *Philosopher's Index*. She has served as a Board and Executive Committee Member of the Information Industry Association (IIA), and as a Board Member of the American Society for Information Science & Technology (ASIS&T).

Ms. Lawlor earned a BS in Chemistry from Chestnut Hill College (Philadelphia), an MS in chemistry from St. Joseph's University (Philadelphia) and an MBA from the Wharton School (University of Pennsylvania).

About NFAIS

The National Federation of Advanced Information Services (NFAIS™) is a global, non-profit, volunteer-powered membership organization that serves the information community – that is, all those who create, aggregate, organize, and otherwise provide ease of access to and effective navigation and use of authoritative, credible information.

Member organizations represent a cross-section of content and technology providers, including database creators, publishers, libraries, host systems, information technology developers, content management providers, and other related groups. They embody a true partnership of commercial, nonprofit and government organizations that embraces a common mission – to build the worlds knowledgebase through enabling research and managing the flow of scholarly communication.

NFAIS exists to promote the success of its members and for more than 50 years has provided a forum in which to address common interests through education and advocacy.