

# An overview of the NFAIS 2016 Annual Conference: *Data sparks discovery of tomorrow's global knowledge*

Bonnie Lawlor

Guest Editor, NFAIS Honorary Fellow, 276 Upper Gulph Road, Radnor, PA 19087, USA  
E-mail: [chescot@aol.com](mailto:chescot@aol.com)

**Abstract.** This paper provides an overview of the highlights of the 2016 NFAIS Annual Conference, *Data Sparks Discovery of Tomorrow's Global Knowledge*, that was held in Philadelphia, PA from February 21–23, 2016. The goal of the conference was to examine how data has risen in importance and is transforming all aspects of research – from funding policies through to reporting, publication, and archiving policies. Data literacy is an essential skill in today's digital world and even new career paths have emerged – data scientist, data engineer, data librarian, etc. The conference raised both practical and philosophical issues regarding data management, use, and reuse, and provided a glimpse of what information services should look like in the future.

Keywords: NFAIS 2016 Annual Conference, Big Data, data management, artificial intelligence, content globalization

## 1. Introduction

Over the past decade data has emerged as the driver in the creation of new knowledge. In their 2013 book entitled *Big Data: A Revolution that will Transform How We Live, Work and Think*, Viktor Mayer-Schönberger and Kenneth Cukier state that our fixation with data is a continuation of humankind's quest to measure, record, and analyze the world. And today's digital technology makes this possible in ways that we have not experienced to date. Data can now be captured and preserved in a variety of formats. Anything that can be measured *is* being measured and sensors are driving an exciting new area of innovation called the Internet of Things [22]. According to a recent McKinsey Global Institute Report, "if policy makers and businesses get it right, linking the physical and digital worlds could generate up to \$11.1 trillion a year in economic value by 2025" [16]. Mining data could be as economically rewarding as mining gold!

It is no wonder then that research funders around the globe now require that data be managed and curated as a key aspect of the research process. Future researchers need to build on current science, therefore it is critical that data be preserved and made accessible in order to provide a continuum in developing new knowledge. As a result, academic institutions, societies, and corporations both small and large are struggling, not only with the daily concerns regarding data management, reproducibility, and re-use, but also with the philosophical ideologies that underlie the policies that will be used to control those day-to-day activities.

To learn how others are dealing with issues related to data gathering, analysis, and management 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 *Data Sparks Discovery of Tomorrow's Global Knowledge*. Attendees considered and discussed the following: the expanding globalization of data supply and demand; the potential impact of artificial intelligence on information discovery and the user experience; the enhancement and use of content through new innovative tools and platforms; emerging business start-ups and business models; and new policies and practices surrounding data and the implications of such policies for strategic development.

## 2. Setting the stage

The conference began with an opening keynote given by **Steven Miller**, from IBM's Global Leader Academic Programs. He noted that after decades of data scarcity, we are finally arriving in an era of data abundance. He said that data is driving industrial transformation and noted the top trends that Gartner believes are shaping the future: 1) Information of Everything; 2) Internet of Things; 3) Advanced Machine Learning; 4) Autonomous Agents & Things; 5) The Device Mesh; 6) Ambient User Experience; 7) Advanced System Architecture; 8) Mesh App & Service Architecture; 9) Adaptive Security Architecture; and 10) 3D Printing Materials [7]. Data is driving these trends and as a result new analytics and new professions such as Data Scientist, Chief Data Officer, Data Engineer, and Data Policy Officer are emerging.

With the Internet of Things (sensors and actuators connected by networks to computing systems) we can sense and measure just about everything. For example, we can determine if we turned off the oven or locked the front door; we can find lost keys; check on an elderly relative; monitor pollution levels; identify equipment about to fail, etc., and we are only getting started! Today, "smart cities" gather data from smart devices and sensors embedded in its roadways, power grids, buildings, and waterways – anything that can be sensed. Their goal is to deliver improved health, more efficient infrastructure systems, cleaner environments, and safer and more inclusive societies.

He believes that for better decision-making the only way we can make full use of the huge amount of data now available is with the help of cognitive systems, such as IBM's Watson cloud-based cognition as a service platform. That system is currently being used by doctors and it is used to gather data, create a set of possible diagnoses, and identify the diagnosis and treatment for a specific patient. IBM has established a new business unit, Watson Health, and has developed "IBM Watson for Oncology." It provides evidence-based suggestions to support oncologists' decisions by 1) combining patient data with massive volumes of medical literature, including journal articles, physicians' notes, and the National Comprehensive Cancer Network (NCCN) guidelines and best practices and 2) providing ongoing learning from new oncology techniques, treatments, and evidence (see: <http://www.ibm.com/smarterplanet/us/en/ibmwatson/watson-oncology.html>).

Miller said that today data is a core business asset that must be curated and protected and that there is a growing demand for data policy skills. He quoted from a recent *Harvard Business Review* article [4] that stated that the sexiest job of the 21st century is "Data Scientist" and that the demand for data science skills is on fire. Data Scientists represent a diverse field. There are human data scientists whose primary role is advisory (they make sense of any dataset(s); apply any form of analytics from descriptive to cognitive; they are visualization experts as well as Data Storytellers); and there are machine data scientists whose primary focus is writing advanced algorithms for such things as advanced robotics, self-driving cars, recommendation engines, virtual assistants, and systems such as IBM Watson. He stressed

the need for data literacy to be incorporated at every level of education in order to prepare students for their careers and cited a recent IBM report from a workshop that they held in conjunction with the EDC's Oceans of Data Institute (ODI) on this issue [1]. Miller's slides are available on the NFAIS website and a very brief article based upon his presentation appears elsewhere in this issue of *Information Services and Use*.

### 3. Data usage practices

In the next session, three speakers presented emerging data usage practices from their individual perspectives. The first was **Sayed Choudury**, Associate Dean for Research Management at Johns Hopkins University, who spoke on research data curation. He said that there are several definitions of Big Data, and that they all are based upon its "V's" – volume, velocity, and variety. His definition is more about the "M's" – methods or lack thereof – than "V's." For him Big Data is defined as data having a scale and complexity such that it overwhelms their respective communities' ability to deal with it using previous research methods.

Choudury's fundamental premise is based upon the assertion that research libraries should consider the potential consequences, need for interpretation, and degree of control to better allocate and optimize scarce resources for data management. He believes that by doing so, it may be possible to support broader goals of data management at scale, identify network effects through linked data, and highlight possibilities for partnerships, including partnerships with the corporate sector. He used the example of the Zika virus and how the public and private sectors could work together to track the potential migration of the virus before, during, and after the Olympics; e.g. data based upon airline ticket purchases, Google searches, etc. A real-life example that he gave was the United Nations' use of call detail records from mobile phones (corporate data) to depict the migration of individuals following the 2010 earthquake in Haiti in order to determine where recovery resources were most needed. He suggests that co-operative sharing of data across sectors will facilitate its broader utility for the common good. A paper based upon his presentation appears elsewhere in this issue of *Information Services and Use* and his slides are available on the NFAIS website.

The second speaker in this session was **Courtney Soderberg**, Statistical and Methodological Consultant at the Center for Open Science. Her presentation was entitled *The Open Science Framework: Increasing Scientific Workflow Transparency to Facilitate Reproducibility*. She noted that there is a science reproducibility crisis today and that we are finding that a lot of published research cannot be duplicated. Her premise was that researchers need to share their entire process – not just the results. There is a need to document everything, noting how their work may have changed over time. Many journals now have data sharing policies and that is a good thing. But the focus is on the end product and what gets published. The scientific process can be a long one, and Soderberg believes that the entire process needs to be documented *as it takes place*, not after the fact. Research hypotheses can change during the process as can analytic techniques. Researchers worry about the data at the point of publication. They backtrack to "find" the data to get it into a repository. What if they don't find it all or forget the nuances of changes? It can make a difference in reproducibility. A solution to this problem is the Open Science Framework that provides a scholarly commons to connect the entire research process. It provides free, open source, end-to-end support of a research project. This makes it easier for the scientist and actually provides incentives for them to do the documentation. A researcher has his/her own private cloud-based space in which they can store documents, write manuscripts, etc. and when they are finished with a

project they can choose to share all or part of their work. During the process they can provide various levels of access to their work by collaborators so they do not need to worry about having their research “scooped.” Version control is also available so changes are noted and dated. Perhaps most importantly, if a researcher is already using a system such as GitHub, Mendeley, or figshare they can continue to do so – they do not have to learn anything new. The Open Science Framework connects to all of these silos and will be adding more. Persistent identifiers are added so that the researcher can be cited and get credit for his/her work, and they are provided with information on how many times their work is accessed and/or downloaded. They even are given “badges” to put on their site to signal their sharing behaviors. To ensure that the posted work is discoverable, the Center for Open Science is working with the Association of Research Libraries in the SHARE Project – a higher education initiative whose mission is to maximize research impact by making research widely accessible, discoverable, and reusable. To fulfill this mission SHARE is building a free, open, data set about research and scholarly activities across their life cycle (see: <http://www.share-research.org/>).

The Open Science Framework is free and definitely worth a look (see: <http://OSF.io>). Soderberg’s slides are available on the NFAIS website.

The final speaker in this session was **Lisa Federer**, a Research Data Informationist at the National Institutes of Health Library who spoke on *Research Data Management: Roles and Opportunities for Librarians*. Like Sayeed Choudury, she began her talk discussing the “V’s” of Big Data – Variety, Volume, Velocity, and Veracity and she discussed the increase in born-digital data, from digital health records to electronic lab notebooks.” She noted that a 2012 IDC report said that from the beginning of 2005 to the end of 2020, the digital universe will grow from 130 exabytes to 40,000 exabytes, or 40 trillion gigabytes! [6] She asked if we as information professionals and researchers are ready for this exponential growth and her answer is that we are not, providing the following quote: “The exponential growth in the amount of biological data means that revolutionary measures are needed for data management, analysis and accessibility. . . . But curation increasingly lags behind data generation in funding, development and recognition” [11]. Federer said that now that government funding requires that data be shared and that data management plans be an integral part of the research process, we can only expect more data to find its way into the public domain. We are not able to curate this amount of data nor are the social behaviors for data sharing an integral part of workflow processes. A study done at NIH asked researchers to rank data-related tasks in two ways: 1) based upon relevance to their work and 2) based upon their personal expertise. In most cases the tasks were rated high in relevance, but there was not a correspondingly high expertise [5]. She said that there is a great opportunity for librarians and other information professionals to play a training role in helping researchers comply with the data management plan mandate at every stage of a project – from data discovery to sharing and reuse, to visualization, etc. In closing she added the role of “Data Librarian” to the list of new data-related professions mentioned during the opening keynote and provided a link to the NIH Library Data Services resources: <http://nihlibrary.campusguides.com/dataservices/>. A paper based upon her presentation appears elsewhere in this issue of *Information Services and Use* and her slides are available on the NFAIS website.

#### 4. Managing data and establishing appropriate policies

The final session of the day focused on how data usage activities and practices could potentially re-frame legislative and institutional policies that protect significant investment by funding bodies and ensure that data is appropriately identified, tagged, housed and preserved.

The first of three speakers was **Heather Joseph**, Executive Director of the Scholarly Publishing and Academic Resources Coalition (SPARC), who spoke on the evolving U.S. policy environment for open research data. She noted that the development of these policies, at least in the USA, are very much a community-based effort and that there are opportunities for everyone to participate. What is driving policy development? Certainly funding plays a role. She noted that funders invest tens of billions of dollars annually on basic and applied scientific research and that about \$60 billion per year is spent in the U.S. alone on publicly-funded research. The goals of the funding are to stimulate new ideas, accelerate scientific discovery, improve educational outcomes, fuel innovation, grow the economy and create jobs, and to improve the welfare of the public. Joseph contends that these goals can only be met if the outputs of the research are freely-accessible and re-usable. The premise is that policies that encourage open access to the results of this research, including data, will accelerate and significantly improve the research outcomes. The community at large increasingly sees additional benefits to opening up research data. These benefits include improved reproducibility, the additional productive use of data, the prevention of or improved response to crises (Zika, earthquakes, etc.), the enablement of large-scale collaborations, and improved research transparency/accountability. She reviewed some of the key U.S. information policies, beginning with the Freedom of Information Act in 1966 through to the 2013 Public Access to Federally-Funded Research Outputs Directive and noted that in response to the latter, draft or final policy plans for access to research data (and articles) have been released by fourteen of the nineteen U.S. science agencies as of February 2016. She noted that while the White House directive provided some guidance, all plans differ somewhat in their interpretation of the guidelines as well as in the implementation processes. She noted however, that there are commonalities across the plans. Most require data management plans; provide direction for approved locations for data deposit/storage; acknowledge the need for routine attribution for data; require data inventories and indices; support public/private collaboration; and recognize the need for long-term preservation. However, she pointed out that there is not yet a common set of standards for any of these policy components and more granular guidelines may be needed. She closed by providing suggestions to keep things moving in the right direction. She said that community collaboration is essential as well as interagency collaboration and that we must all recognize that policy development is an evolutionary process. We cannot build an effective, sustainable infrastructure to support vital national/public interests without the additional investment needed to support access to and use of research data. Joseph's slides are available on the NFAIS website and her paper appears elsewhere in this issue.

The second speaker was **Anita De Waard**, Vice President of Research Data Collaboration at Elsevier who spoke on encouraging infrastructures to promote data integration and reuse. She began her talk by reviewing the life cycle of research data using a chart taken from a recent JISC report [12]. The components are: data creation and deposit; managing active data; data repositories and archives; and data catalogs and registries. She provided examples of each stage with projects in which Elsevier is involved. For example, in data creation and deposit they work with a group called hivebench ([www.hivebench.com](http://www.hivebench.com)) to provide structure for data input – it is similar to the Open Science Framework that was discussed earlier. Elsevier is involved in “data rescue;” e.g. finding data that is not available because it is in someone's drawer or in an obscure electronic format. Elsevier has created an award for data rescue in the geosciences (see: <https://rd-alliance.org/international-data-rescue-award-geosciences.html-0>). They have also created a new journal, *SoftwareX*, that actually stores software. One of the journal's objectives is to support publication of research software in such a way that the software is given a stamp of scientific relevance, and is provided with a peer-reviewed recognition of scientific impact (see: <http://www.journals.elsevier.com/softwarex/>). The journal won an award for innovation in journal

publishing. They also have a journal, *Information Systems*, that includes not only software, but a virtual machine and the computational environment in which cloud-based experiments can be run. Basically the journal is supporting the publishing of reproducible formats.

Elsevier is also working on a number of forward-thinking projects such as one with the Research Data Alliance (RDA) to create a hub that will link data DOI's with the DOI's of relevant papers. The effort is a collaboration with CrossRef, DataCite, ORCID, OpenAIRE, the International Association of STM Publishers, the National Data Service, and RMap.

She noted that the current process of data sharing and publishing is as follows: 1. the researcher creates datasets; 2. the researcher then writes a paper and publishes it in a journal; 3. (sometimes), a dataset is posted to repository; and 4. the researcher reports (post-hoc) to the Institution and the Funder. The process takes a lot of work, data posting in a repository is not always mandated, and there is no link between the data and the published article. A proposed process is as follows: 1. the researcher creates datasets and posts to a repository (under embargo); 2. the Funder is automatically notified of the dataset posting; 3. The researcher writes paper and publishes it in a journal; the embargo is lifted and the data linked (note: this also allows release of non-used data for negative result and reproducibility); and 4. The Funder and Institution receive a report on publication and the embargo is lifted. In closing she noted the many organizations with which Elsevier is working to improve all aspects of the data life cycle, including Force11, the National Data Service, and the Research data Alliance. A paper based upon her presentation appears elsewhere in this issue of *Information Services and Use* and her slides are available on the NFAIS website.

The final speaker in this session was **Larry Alexander**, the Executive Director of the Center for Visual and Decision Informatics (CVDI) at Drexel University. CVDI was established by the University of Louisiana at Lafayette and Drexel University. It is a national research center funded by the industry/University Cooperative Research Center program of the National Science Foundation (NSF), by members from industry and government (the CVDI members), and by university matching funds. The goal is to get technology out of the universities and into industry. Industry provides financial support, chooses the projects, and provides guidance. The universities provide research expertise, infrastructure and students to get the work done. The CVDI is the only NSF Center with a "Visualization & Big Data Analytics" focus. It is jointly managed by the University of Louisiana Lafayette and Drexel University and is one of the six NSF centers in the country that has an international site (Tampere University of Technology in Finland). Its research areas are: 1) advanced analytics to mine information from multi-dimensional, multiple-data sources; 2) to render novel visual interfaces and visualization techniques to rendering complex data for intuitive interaction; and 3) high performance data management strategies to streamline data processing, storage and analytics.

The center has been in existence for three years and has supported fifty-four students by its research activities. It has generated thirty-six potentially-patentable discoveries and twenty-nine potentially-copyrightable discoveries. Sixteen projects have been completed resulting in sixty-two research articles. And CVDI has received more than \$1.1 million in additional NSF funding, including research support for undergraduates, teachers, and veterans.

Alexander walked through a number of CVDI's projects such as predicting the spatio-temporal evolution of Chicago crime hotspots; the development of an influenza forecasting model that uses environmental conditions (temperature, sun exposure) and influenza history; and the development of social media analytic tools to predict emerging events and detect drug safety signals. He said that data is reinventing how science is performed and it is becoming more and more powerful. His thoughts regarding policy are as follows. He believes that there are lots of reasons for data to be open as noted earlier today, but

there are also reasons for it to be closed such as for the protection of corporate trade secrets. He said that there is a need for metadata standards along with the software to automate the creation of metadata. He believes that we do not give as much attention to developing strategies for data preservation as is needed. He said that data is at risk- we need better technologies to prevent the hacking and misuse of data and we need to be conscious of data pedigree – where did it come from? has anyone had the opportunity to alter it? We need to be aware of data quality and use filters to identify possible errors. And finally he talked about the “Data Self” – data about each of us [10]. He included not only the usual – fingerprints, face recognition, and genetics, but also data generated from body sensors, the internet traffic we each create, etc., and said that we need to protect that data as well. For more information on the CVDI go to <http://www.nsfcvdi.org/>. Alexander’s slides are available on the NFAIS website.

## 5. Data is the new black

The second day of the conference was opened by **Ann Michael** who gave an excellent presentation on the business applications of data. She noted that data is such a hot topic in today’s world because of the scale and availability of data/digital information; because we have networks that are robust enough to disseminate that information; and because the computational power to crunch those data is now ubiquitous. She cited a recent *Scholarly Kitchen* posting on text and data mining [2] and why publishers need to support those activities – definitely worth a read! Michael noted that publishers need to care about data in order to: serve their customers better by improving current products; exceed their customers’ expectations with new products; provide better customer service; build their audience/customer base; and ultimately become more efficient. She went on to provide some examples of how publishers and media companies are using data. Article usage data generated by Altmetrics and Plum Analytics provides information on the importance of articles while Springer’s Bookmetrix provides usage data on chapters of the books that they publish (citations, downloads, mentions, etc.) and they can compare the relative success of their portfolio; HighWire Press started Impact Vizor to determine where articles that they reject end up – do they get published elsewhere and if so, where? It allows them to make better rejection decisions as some articles that they rejected did get published in other journals and were highly-cited. *The New York Times* is using predictive analytics and data to determine why people subscribe, what topics are of interest and need to be promoted via social media, etc. They use it to make business (but not editorial) decisions. And she mentioned a new tool, Tamr, that can pull together and catalog about 90% of a company’s data from disparate silos in a completely automated fashion. It removes duplicates, highlights errors and inconsistencies, etc. She noted that Wiley is using it for customer and author disambiguation and has found that it has increased their productivity (see: <http://www.tamr.com/>).

She encouraged everyone in the audience to make the most of the data that they have – capture it, unify it, analyze it, act on it – and ultimately repeat the process to create value not only for their customers, but also for themselves. Michael’s slides are available on the NFAIS web site.

## 6. Building value

The second session of the morning focused on building value through a portfolio of software and systems. The first speaker was **John Hammersley**, the Co-founder and CEO of Overleaf, a collaborative writing and publishing tool. He noted that the internet is fostering collaborative research. About 36% of the world’s papers are produced with more than one international author and the number of citations

per article increase as the number of countries that are collaborating on a paper increase [13]. But collaboration in writing papers can be frustrating. There are multiple versions of the same document, long email chains, formatting and typesetting issues, the maintenance of references, and long revision cycles. The ability to do cloud-based activities was a game changer and enabled the creation of Overleaf. Its original focus was on authors, but as usage grew there was an added focus on institutional installations with Stanford University as an early-adopter. Before the Stanford trial began there were three hundred and seventy-five confirmed Overleaf users at the University. Within two years the number of users increased by 450% to two thousand and sixty-five and the number of collaborative projects grew 620% from eighteen hundred and ninety-six to thirteen thousand six hundred and fifty-five! A side-benefit is that it has allowed university libraries to get a better awareness of the diverse collaborative projects in which their students and faculty are involved.

Overleaf has had a good growth trajectory to date. After three years there are three hundred and fifty thousand users in ten thousand institutions worldwide and four million documents have been created. They have established direct partnerships with academic publishers, institutions and other tools (Mendeley, zotero, etc.) in the research workflow to streamline writing, collaboration, submission and review. Overleaf now has custom submission links that provide automated transfer of files and metadata, automated pre-submission checks, and direct submission into systems such as Editorial Manager, Scholar One and eJournalPress. It is free to authors and provides added value to editors, reviewers and publishers. For more information go to <https://www.overleaf.com/>. Hammersley's slides are available on the NFAIS website.

The second speaker in this session was **Thomas Grandell**, President of Etsimo, a cloud-based visual content discovery platform that focuses on exploratory search. The company is a spinout from the University of Helsinki, Finland. Its technology is the result of three years of research that generated two patents. It was founded in 2015 and has six employees, three of which are internationally-recognized scientists. The company is backed by University of Helsinki Funds and TEKES, the Finnish Funding Agency for Innovation.

Grandell talked about the difference between traditional searching and exploratory searching. In traditional searching the search intent is captured in the initial query only. Such a search usually results in a long list of hits spanning several pages with the "important" (high-ranked) hits at the top and users seldom go beyond the first few pages. It works if you know what you are looking for, where to look for it, and how to formulate a query. It does not work if you want to learn/acquire new knowledge and/or search for unusual findings. He noted that according to Microsoft up to half of all searcher are exploratory. He went on to provide examples of the two types of searching and made a compelling story about how exploratory searching does not compete with traditional (lookup searching) because the use case is different. He noted that using appropriate tools for exploratory searching gives superior results and happier returning customers. You can do a live exploratory search on Wikipedia at <http://wikipedia.etsimo.com/>. A paper based upon his presentation is available elsewhere in this issue of *Information Services and Use* and his slides are available on the NFAIS website.

The final speaker in this session was **Elizabeth Caley**, Chief Operating Officer at Meta, a big data company for science and intellectual property that uses machine intelligence to structure the content within scientific papers so that it is more discoverable. Caley noted that since 1900 approximately 25,775,932 biomedical papers have been published and about four thousand new are added daily. In addition to using content freely-available on the web, Meta partners with twenty-nine major STM Publishers and has created a database with thirty-eight thousand titles (Books and Journals) comprising nineteen million closed-access, full-text articles. They have created the world's largest knowledge graph

– identifying concepts, authors, drugs, procedures, diseases, citations, etc. and have linked (and mapped) all of these entities via more than three-and-a-half billion connections. This has resulted in one of the largest commercial STM text-mining collections in the world.

Meta offers both free and fee-based services. Meta Science is a free, real-time literature discovery service that lets researchers stream and discover papers through the people and things they follow in their world of research. They have mapped papers over time based on citations, so researchers new to a subject can find the most important papers in that field. This service was actually discussed by the CEO, Sam Molyneux, at the 2015 NFAIS Annual Conference and its name at the time was Sciencescape. Use of that name in a Google search will take you to the Meta website.

On the fee-based side they offer a service entitled Bibliometrics Intelligence. This service plugs the intelligence of the Meta database into author workflow systems to transform manuscript triage, cascading, and journal planning. One of the benefits for publishers is that it helps editors manage the growing volume of manuscript submissions and helps to avoid rejecting potentially high-impact papers because it can predict potential future citation counts. It basically tells a publisher/editor what manuscripts should be given a priority review.

She asked a question: What if you could discover patterns of emergence and connections between technical and scientific concepts at a speed, scale, and comprehensiveness that exceeds human capacity? And then went on to demonstrate how Meta is doing this today. Four years ago, IARPA (the Intelligence Advanced Research Activity from the office of the Director of U.S. Intelligence) started a project to enable the early reliable detection and monitoring of technical emergence through machine intelligence. They invested more than sixty billion dollars to develop a stand-alone service for analysts and worked with SRI International, BAE Systems, and Raytheon for five years to be able to predict entities (researchers, drugs, genes, etc.) that demonstrate early signs of being able to have a major impact in three to five years. Meta is now the sole commercial partner for that effort with a service entitled FUSE Horizon Scanning. It will provide continuous monitoring and early awareness of emergent concepts, technologies, researchers and institutes via real-time analysis of the scientific and patent literature. Caley noted that while Meta began in the life sciences they plan to move into other disciplines as well – chemistry, physics, etc. For more information go to <http://meta.com/>. Caley's slides are available on the NFAIS website.

## 7. Creating value for external institutions and systems

The final session of the morning continued the theme of creating value. The first speaker was **Carl Grant**, Associate Dean, Knowledge Services, and Chief Technology Officer at the University of Oklahoma. He spoke on how libraries and information providers – indeed the entire information community – need to work together to eliminate information silos and facilitate the creation of new ideas. He offered five ways in which information providers can make this happen and the full text of his presentation is available elsewhere in this issue of *Information Services and Use*. It is worth a read to learn more about how a librarian looks at the current state of affairs – especially a librarian who also walked in the shoes of a commercial service provider and knows both sides of the equation. His slides are available on the NFAIS website.

The second speaker was **James King**, Branch Chief and Information Architect, from the Information Resources and Services Branch of the National Institutes of Health Library, who talked about how his group creates value for the Institutes and its researchers. The group of NIH Informationists was started

in 2001 with the goal of integrating information solutions into the workflow of the NIH and the Department of Health and Human Services (DHHS). The group focuses on topics such as clinical information, bioinformatics, data management, public health, and public policy. They deliver knowledge-based solutions, synthesize search results, provide support for bibliometric and portfolio analysis as well as support for systematic reviews, data management, and bioinformatics. Three areas of assessment in which they are involved are: bibliometric Analysis (publication metrics), collection assessment (subscriptions vs. publications), and custom information solutions (web-based services). King went on to describe each of these assessments in detail and how they try to learn more about their users engagement with content so that they can provide more personal and customized services, including digitization (the NIH Library has been able to digitize over two thousand publications from its own print collection and has partnered with other Centers to digitize thousands of additional documents). He also noted that the Library is attempting to build community within NIH, with one example being that since 2013, the Library has served as host and a co-planner for the annual Drupal GovCon (<http://drupalgovcon.org/>), a free event that brings together individuals from the public and private sectors that use, develop, design, and support Drupal. Last year more than eight hundred attendees enjoyed keynote talks, breakout sessions, training classes, and coding sprints.

In closing he noted that the technology trends of the past twenty years can be summed up as a shift away from collections and towards innovative, collaborative services. To learn more about what NIH is doing refer to King's paper that is available elsewhere in this issue of *Information Services and Use*. Also, his slides are available on the NFAIS website.

The final speaker of the morning was **Miranda Hunt**, User Experience Researcher at EBSCO Information Services, who gave a brief presentation on how they go about doing market research to make informed decisions about their products and services. She noted that they take a comprehensive, three-dimensional approach using 1) usage data, 2) secondary research and, 3) primary research, and then provided examples of what can be used in each category [19]. She referred to some research that they did to understand the digital lives of students and I refer you to a detailed paper that was published in *Information Services and Use* last year [15] by one of her co-workers. Hunt said that one of the most important lessons learned from doing market research is to understand the difference between attitudinal studies (what people *say* they do) and behavioral studies (what people *actually* do). The former requires focus groups and design workshops while the latter requires usability studies and prototyping. She also said that it is very important to ensure that you have access to the people you really need to include in any studies and that they have the time to allot to the study. You also need to manage their expectations. And she stressed the need for several practice runs especially for usability testing – things can go wrong... and do! Hunt's slides are available on the NFAIS website.

## 8. Members-only session: How readers discover content in scholarly publishing

Between the morning and afternoon sessions there was an NFAIS Members-only luncheon with an outstanding presentation by **Simon Inger** from Simon Inger Consulting Services. His presentation was based upon his company's 2015 survey on "How Readers Discover Content in Scholarly Publishing" which is a continuation of work carried out in 2005, 2008 and 2012. It provides in-depth reporting on the discovery behaviors of people working and studying across all disciplines (including the humanities, social sciences and STM). The survey revealed important trends in discovery, for example: a reduced reliance on professional search; some increases in the importance of library-intermediated discovery;

the continued emergence of professional and social networking sites; and a shift in importance from Google to Google Scholar. In addition, the survey has revealed some important perceptions about the abundance of free versus paid downloads that may influence library budgets in the future. A detailed summary article is available elsewhere in this issue of *Information Services and Use* and his slides are available on the NFAIS website. This is definitely worth a read. You can also download the full report at [http://www.simoningerconsulting.com/how\\_readers\\_discover.html](http://www.simoningerconsulting.com/how_readers_discover.html).

## 9. Globalization of content

The afternoon opened with a look at the globalization of content. The first speaker was **Stacy Olkowski**, Senior product Manager, Thomson Innovation and Thomson Data Analyzer, from Thomson Reuters, who discussed patent data. She said that patent searching and data are the beginning of a story that can answer questions such as: What is my competition doing? What technologies or innovations are out there? Where are people investing? Where are the pockets of innovation? She noted that from 1970 to 2010 the number of records in Derwent's World Patent Index increased from less than two hundred thousand to almost one-and-a-half million. And she said that it is important to remember that 70% of the text in patents cannot be found elsewhere.

She discussed the growth of patents in Asia, especially China. The Chinese patent office opened in 1984 and they are now number one in the world for the number of patent applications submitted (only 40% of which are granted, perhaps suggesting that quality is still lacking). They continue to experience a 12.5% annual growth in patent applications and 85% of the Chinese patent applications are from Chinese nationals. She noted that China is encouraging innovation (and patent subsidies help!). She closed by saying that she believes that the growth in Chinese patents will continue and she expects that the quality will improve. Her slides are available on the NFAIS website.

## 10. Globalization of content: Regional journals

The session continued with a presentation from **James Testa**, also from Thomson Reuters (Vice President, Emeritus, Editorial Development and Publisher Relations). He continued the theme of the globalization of content, noting that while in the past articles based upon scientific research were primarily from the United States and Europe, today countries such as China, South Korea, and Brazil – along with others that were minor contributors a decade or so ago – are becoming major players. He shared some data related to the coverage in Web of Science of journals and articles from a selected set of the following ten countries (in alphabetical order): Australia, Brazil, China, India, Italy, Poland, S. Africa, S. Korea, Spain, and Turkey). He noted that as of 2015, China led this group of countries in the number of papers published, but it ranks seventh in journal impact for that same group (Australia is number one in impact for the group). Testa also talked about questionable editorial practices that are used to raise the impact of journals and articles. The first and most common practice is the attempt to artificially raise Citation Impact scores through intentionally-excessive self-citations or by groups of journals working together in deliberate and artificial citation stacking. His presentation included some interesting numbers. To learn more refer to his paper that is available elsewhere in this issue of *Information Services and Use*. The paper also includes copies of his slides and they are available on the NFAIS website as well.

## 11. Globalization of content: The perils

The final speaker in this session was **Donald Samulack**, President, U.S. Operations, Editage, who talked about globalization and the erosion of trust in the literature and what, if anything can be done about it. He opened by saying that he would be less “politically correct” than Jim Testa because he himself, quite frankly, is scared about what is going on globally with regard to irresponsible publishing practices. He said that the world is so flat that it is beginning to curl at the edges and we have a tsunami of papers coming out of Asia, specifically China. He said that in the Western World, publishing is built upon a foundation of Trust; in Asia, that trust is being eroded. (*Note: the final keynote speaker talked about the importance of “Trust” in information systems – does that not also apply to journals and books?*). He said that in Asia there is an element of commerce in every facet of society and there is always someone in academia looking to provide services. He said that Asian authors have an angel on one shoulder encouraging them to do the work and follow good publishing practices and a devil on the other shoulder telling them that they can get the paper done for them. And with the cash incentives to publish that are promoted in China the decision is not without pressure. To get a PhD in China you must have at least two publications a year in a Western journal. For a clinician to practice in a Chinese hospital they must publish two to three articles a year in a Western journal. He said that it is extremely difficult to meet this requirement since a typical hospital, even in large cities such as Shanghai, has thousands of patients a day with a relatively small staff of doctors. The doctors simply do not have the time to write papers. He mentioned a highly-cited article (one that may not have been peer reviewed, however) in which a coin flip test was done. There were fifteen countries involved and fifteen hundred and thirty-nine participants. They were told that they would receive a \$5 U.S. dollar reward if they had a heads up and a \$3 U.S. dollar reward if it was tails. Based upon statistical probability, Great Britain was the most honest country and four Asian countries were the most dishonest (China was number one!) [9].

Samulack expressed concerned about the trajectory of Chinese publications (as noted by Jim Testa) as compared with their actual impact on research. STM numbers indicate that by the year 2020 the number of papers coming out of China alone will surpass the number of papers from the USA. He talked about “predatory practices” in publishing that are not unique to Asian countries. These include: Editorial Board solicitation; Peer Reviewer solicitation; Manuscript solicitation; Peer Review practices; Predatory Author Services; Authorship for sale; Plagiarism; Writing fraud; and Data fraud., He predicted that in two-years’ time an author will have a very hard time identifying what the ethical practices in publishing actually are.

He noted that China has taken action on these predatory practices. While circulated within China as early as September 18, 2015 by the China Association for Science and Technology (CAST), the official directive is dated November 23, 2015 and was publically released December 2nd. It forbids Chinese scientists from: using a third party to write journal articles, using a third party to submit articles, hiring a third party to substantially revise articles, providing fake peer review information, or giving authorship to scientists who have not substantially contributed to the research. There also now appears to be a secret anti-fraud unit sanctioned by the China government and operating in association with CAST that is working to identify fraudulent publication activities in China, with severe penalties. He went on to talk about predatory open access journals and the countries with the most predatory authors.

He closed with an overview of the Coalition for Responsible Publishing Resources and I refer you to their website at <http://www.rprcoalition.org/> for more information. This was an interesting talk whose message was echoed by others who also noted that China is not the only culprit in the use of questionable publishing activities. Refer to Samulack’s slides on the NFAIS website. They have a lot of data.

## 12. Miles Conrad Lecture

The final session of the day 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 **Deanna Marcum**, Managing Director, Ithaka S + R, and the complete transcript of her talk is published in full elsewhere in this issue. It gives an interesting perspective on the kind of leadership that she believes is required today now that libraries have completely entered into the digital era (special print collections notwithstanding) based upon her fifty years of librarianship – including eight years as the Associate Librarian for Library services at the Library of Congress. In her paper she applies ten descriptors of successful digital organizations to academic libraries and it is well worth a read. There were no slides for her presentation.

## 13. Shark tank shoot-out

The final day of the conference began with a “Shark Tank Shoot Out,” in which four start-ups (ranging between garage level and Round B funding stage) each had ten minutes to convince a panel of judges that their idea was worthy of potential funding (the “award” was actually a time slot on a future NFAIS Webinar). The judges were Kent R. Anderson, Founder, Caldera Publishing Solutions; Christopher Wink, Co-founder and Editorial Director, Technical.ly; and James Phimister, President, PHI Perspectives.

The first presenter was **Pascal Magnier**, CEO and Co-founder of Expernova, a company that provides global access to business expertise. The company was founded in 2008 and was launched officially in 2010 after two years of research and development. Expernova is the first cloud-based platform dedicated to Knowledge Networks Intelligence. Magnier said that the success of such networks depends upon the size of the network and the quality of the expertise that is provided. Expernova has developed a database of ten million experts in fifty-two countries. They have also detected fifty-five million collaborations among companies, academic laboratories, and individuals. Their search engine is customizable and they offer both competitive and talent intelligence. The company is located in the South of France and plans to open a U.S. subsidiary before the end of 2016. They have twelve employees. They currently have one hundred companies as customers (80% of their subscribers are corporate companies in twelve countries) and their business model is based upon annual subscriptions. They experienced 56% growth in 2015 with \$775K in revenue (they were profitable) and \$1.5 million was invested in product development. Expernova has won several prizes, including the Global Innovation Challenge of the Presidency of the French Republic and the Global Entrepreneurship Competition (Berkeley/INSEAD). An article on his presentation is available elsewhere in this issue of *Information Services and Use* and his slides are on the NFAIS website. For more information visit the Expernova website at <http://en.expernova.com/>.

The second presenter was **James Harwood**, Founder, Penelope Research. They create a manuscript publication tool for authors and editors with the stated goal of improving the quality of science that is published while making the process easier and faster. Harwood noted that the review process is flawed. In support of that statement he mentioned a study done by the *British Medical Journal* in which they planted nine major flaws in a manuscript and sent it to seven hundred reviewers. On average, reviewers found only three of the flaws [20]! In addition to manuscript flaws, the publishing process takes too long and frustrates everyone involved – authors, editors, and publishers. With Penelope, an author uploads his/her manuscript in docx format. It is screened for errors via machine reading and questionable areas are highlighted and the author is given comments and suggestions for improvement. The manuscript

is tested for completeness – citations and references, tables and figures, statistics, funding information, grant codes, etc. The company is a year old with two employees. Their preferred business model is that the service is free to authors and the publisher pays. It is currently supported by grants from the UK government and Digital Science. For more information go to <http://www.peneloperesearch.com/>.

The third pitch was made by **Alberto Pepe**, Co-founder, Authorea, a collaborative writing platform to write, share, and discuss research – all in real-time. It was created in 2013 by Pepe, a Harvard astrophysicist, and Dr. Nathan Jenkins, a UC Berkeley physicist, who met while working at CERN and were disappointed by the slow, inefficient, and obsolete ways by which research papers are written and disseminated. Pepe actually spoke on this topic at the NFAIS annual conference in 2013. His premise remained the same at the 2016 conference: scientists produce 21st century research that is written up with 20th century tools and packaged in a 17th century format (print) – a better tool, such as Authorea – is needed. Since he spoke back in 2013 the platform has grown in popularity and is currently used by about seventy-six thousand scholars across seventy countries. It offers users a collaborative online editing platform tailored for academic and technical writing – a word processor, which makes adding citations and equations and formatting references incredibly simple. It is built on a Github-style model and every document created is a Git repository. This allows users to track changes in documents in a very granular way and to easily integrate data into documents. Manuscripts can be formatted for specific journals (e.g. *Nature*, *Science*, etc.) with a click of a button. The company received \$1.5 million from Lux Capital earlier this year. The product is free for anyone creating public content and about sixty dollars a year if the content is kept private. At this point in time Authorea is pre-revenue. For more information go to <https://www.authorea.com/>.

The final presenter was **Ariel Katz**, CEO of Research Connection, an organization that connects students with potential mentors. It is a platform with which researchers can show their projects to students who want to do research in those areas. The idea for the company started when two undergrads were looking for research assistant positions at their school. They found that information about labs was scattered and that researchers had no forums that could be used for recruitment. They found that their problem was a common one and Research Connection was born in 2014. As of January 2016 they had forty-one university partners and one hundred and seventeen thousand users. Seven Angel investors have provided funding. They refused to discuss their business model. For more information go to <https://researchconnection.com/>.

After the break the judges announced the winner. They said it was very close and that the top two contenders were Expernova and Authorea. The latter was selected because at this stage of the game Expernova is a viable, profitable business. Investing in Authorea, which is pre-revenue, would allow the judges to argue for a larger equity position and they liked the freemium business model. The winner will receive a plaque and the opportunity to present their business in a future NFAIS webinar. The slides of all of the presenters are on the NFAIS website.

## 14. Leveraging your content

The final session of the morning focused on how businesses can leverage their own content. The first speaker was **Marjorie Hlava**, President, Access Innovations, who addressed the topic within the context of the application of Artificial Intelligence (AI). She noted that AI has long been the Holy Grail in information and computer science and went on to say where she believes we are in the quest for that Grail. She said that from her perspective there are three levels of AI. First, Artificial Narrow AI, which

is also known as weak AI. It often covers a single area or domain. It can, for example, play chess – and does so very well. Second, there is Artificial General Intelligence, which is also known as strong AI or Human Level AI. This involves more reasoning, problem solving, and the creation of complex thoughts based on experience. And finally, there is Artificial Super Intelligence where the computer is smarter than a human, or at least it thinks it is! She then went on to look at the current AI systems and what content providers should be doing to enhance the information that they own in order to increase customer satisfaction. Where does she think we are in the quest for the AI Holy Grail? To find out read her excellent paper that appears elsewhere in this issue – you won't be disappointed. Here slides are on the NFAIS website.

The second speaker in the session was **Daniel Mayer**, CEO for North America, Expert System Enterprise. In September 2015, that organization acquired Temis, an information service provider that developed platforms for the semantic enrichment of content with domain-specific metadata. Mayer assured the audience the Temis is still at it, albeit under a different name and in his presentation he discussed the building of a semantic information application to help people find content and make informed decisions – search, discovery, and intelligence. He said that the key is enriching content with a taxonomy to support faceted search [23] capabilities and provided several examples of how this can be done. For more information go to <http://www.expertsystem.com/>. Mayer's slides are on the NFAIS website.

The final speaker of the morning was **C. Lee Giles**, David Reese Professor and Interim Associate Dean of Research at Pennsylvania State University's College of Information Sciences and Technology. Giles participated remotely and spoke on Scholarly Big Data. He defined that data as *all* academic and research documents – journal and conference papers, books, theses, etc., along with related data such as academic/researcher/group/lab web homepages, funding agency and organization grants, records, reports, research laboratories reports, patents, presentations, experimental data, course materials, and social networks. He gave a few examples such as Google Scholar, Microsoft Academic Search, Publishers/repositories, CiteSeer, etc. Giles said that Scholarly Big Data is not well-organized, but rather it is a networked, heterogeneous map [8] and is of interest not only to scholars and scientists, but also to economists, policy makers, funding agencies, educators, social scientists, businesses, and governments. Just one of the many applications of this data is to identify new discoveries, directions and trends in research and he gave DARPA's "Big Mechanism" Program as an example [3]. He also mentioned the IARPA Fuse Program that was mentioned earlier in the conference by Elizabeth Caley. The Fuse Program analyzed forty to fifty million documents to see if it would be possible to enable the early reliable detection of emerging scientific and technical capabilities across disciplines and languages as found within the full-text content of the scientific, technical and patent through machine intelligence [18]. The program was successful and now has commercial applications run by Meta (see announcement at <http://finance.yahoo.com/news/meta-sri-international-announce-agreement-132450971.html>). Giles noted that the field of scholarly big data is taking off and that there are numerous conferences being held on the topic around the globe. He mentioned in some depth the work that he and his group are doing with CiteSeerX and automated metadata extraction as well as the identification and extraction of entities from text. His work is absolutely fascinating and more information can be found at <http://csxstatic.ist.psu.edu/about> and his slides are on the NFAIS website.

## 15. Closing keynote: AI and the future of trust

The conference closed with final keynote by **Stephane Bura**, Chief Product Officer and Co-Founder of Weave.ai, a London-based start-up startup that is building an alternative to Google Now on Tap that

can mine tweets for context and bring up relevant data in other apps on your phone [17]. The topic of his talk was artificial intelligence and the future of trust, and he used videogames throughout as examples of how trust is designed and reinforced. He said that trust is a guiding principle in videogame design. It is trivial today for most games to focus on being better than the player. The games are no longer just about winning. They are about providing experiences and generating emotions, and the whole design process is based on creating those two factors. Games are about losing with style and they are designed by observing the players motivations. Bura said that there are two types of motivations – extrinsic and intrinsic. The former are motivations outside of ourselves that are pushed on to us (e.g. the reason for choosing a product or service – maybe it is the best or only tool for a given job). These motivations do not generate better product loyalty because they come from the outside, and if something better comes along, the user will move along with it. Intrinsic motivations are the ones that matter – they are self-determined; e.g. the desire to be good at something (mastering) or to be the agent in one’s own life (autonomy), or finding one’s place in a community (relatedness). Today’s products are not very good with generating intrinsic motivations. Bura commented that while CityMapper makes his life easier, it does not make him a better person. He also used computerized personal assistants such as Siri as an example – they can predict and tell us what to do, but they are fallible. Even if they tried to create a more “personal” relationship, we really wouldn’t buy it! Bura then went through each of the three motivations – mastering, autonomy, and relatedness – to show how they are related to trust and how they can be used to improve information services.

## 16. Mastery

Mastery is the desire to be good at something. To build trust a service must be accessible – if the user needs it, it must be within reach. It must consistently and persistently reward the user by being reliable and by providing usable knowledge – the user does not want to relearn a system every time he/she goes to use it (that’s why customers get annoyed when changes to a system’s usability are made). And the user wants to be able to experiment with the system in a safe place (with games you can experiment and are in a safe place – you yourself will *not* get blown up if your “experiment” fails. This is what games are all about) [14]. Experimenting with the system allows the user to better understand the system. He said that information service providers must open the black boxes of their new systems and listen to user feedback, both good and bad. Our systems “punish” users for experimenting. They are complex enough to guide the user to solutions, but too complex to explain themselves to the user.

## 17. Autonomy

Autonomy is the desire to be the agent of your own life and be able to set your own goals. To build trust a system must be fair. It must be designed to help the user and to provide solutions for them and the user should not be punished if he/she does not understand how the system works. The user trusts that it will get a solution and that the solution will be true (e.g. accurate). Systems must not only be useful, but actually be meaningful to the user’s life. And finally, the user must be able to set his/her own goals when using the system and the system must ensure that those goals are met. In games autonomy is a given – playing the game is all about the user’s experiments and choices.

## 18. Relatedness

The final intrinsic motivation is relatedness – the desire to be a part of a community and to connect with others. With regard to systems this motivation breaks into two categories: 1) I know the system and what I think of it; The system knows me and what I think of it and; 2) I know others (through the system) and; Others know me (through the system).

Bura said that systems must be interactive and that the AI stack for user interaction comprises the following: 1. Understanding what the user wants and why; 2. Understanding who the user is; 3. Predicting and planning tasks; 4. Understanding what the user is asking and; 5. Predicting/correcting user commands and inputs. He asked the question: Wouldn't it be great to not only ask a computerized personal assistant a question and receive an answer, but also to be told *why* you are being given that specific answer? Our systems need to be able to let the system and the user reach answers together.

He said that his most precious possession is time and he wants to use systems that maximize his time – such as those that provide recommendations for things that interest him so that he does not go looking for them. And when they are right – it builds trust! Bura believes that building trust will be the next big thing in computer-generated recommendations. Now they are based upon aggregated information. But we as individuals move through life within our own unique “bubble of data,” and it is these bubbles that need to be understood and responded to with unique information – not information based upon the aggregation of the things that others like. Bura wants to be matched with people who share his tastes and he does not know of any organization that is working on the development of such a platform.

He discussed Minecraft – the most popular video game today. More than one billion people play it and it takes ten-year olds about two weeks to master it. He noted that videogames account for 96.6% of all of the top twenty YouTube views and Minecraft accounts for 41% of those. It is the most searched term after “music.” Microsoft just bought it for \$2.5 billion and will be releasing Minecraft education tools. The game lies at the very center of the three intrinsic motivations and information providers can learn a lot from it. Bura closed with a shot of one of Steven Miller's slides (the conference opening keynote speaker) where Miller discussed the competencies for data literacy (refer to Miller's paper elsewhere in this issue). Bura said that he agrees with Miller completely and noted that game players are already data literate and that they apply that literacy to every aspect of their lives. They are the next generation of information service users and publishers and information service providers owe it to them to develop systems that the next generation can trust.

Bura's slides are on the NFAIS web site. He has also written a brief paper on artificial intelligence and the next generation of operating systems that appears elsewhere in this issue.

## 19. Conclusion

As in the past, without intentionally doing so, the speakers at the conference reinforced one another in the identification of a number of industry trends and issues, for example: the importance of applying artificial intelligence to content; the increased use of software and other means to raise journal/article impact factors or even to write a totally false paper (see Marjorie Hlava's paper in this issue where she refers to software written by MIT students to generate manuscripts [21]); the changes in libraries; the major impact of Big Scholarly Data initiatives such as the Fuse program; the proliferation of collaborative writing tools; the need for data literacy, and so on. All of the papers were interesting and educational, and they were book-ended between two very strong messages that librarians and information system

providers need to heed – Data Literacy is essential and on the rise and the systems of the future *must* offer the intrinsic motivations of mastery, autonomy, and relatedness and ultimately build the trust of users. Steven Miller’s plea was strong and Stephane Bura’s comments were eloquent and to the point!

Plan on attending the 2017 NFAIS Annual Conference that will take place in Alexandria, VA, USA from February 26–28, 2017. Watch for details on the NFAIS website at: <http://www.nfaais.org/>.

Note: If permission was given to post them, the speaker slides are embedded within the conference program at: <http://www.nfaais.org/2016-conference-program>. The term “slides” is highlighted in red).

### 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. She is currently an NFAIS Honorary Fellow. Prior to NFAIS, 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 part of 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 a Fellow and active member of the American Chemical Society and a member of the Bureau of the International Union of Pure and Applied Chemistry for which she chairs their Publications and Cheminformatics Data Standards Committee. She is also on the Board of the Philosopher’s Information Center, the producer of the *Philosopher’s Index*, and she serves as a member of the Editorial Advisory Board for *Information Services and Use*. She has served as a Board and Executive Committee Member of the former Information Industry Association (IIA), as a Board Member of the American Society for Information Science & Technology (ASIS&T), and as a Board member of LYRASIS, one of the major library consortia in the United States.

Ms. Lawlor earned a B.S. in Chemistry from Chestnut Hill College (Philadelphia), an M.S. in chemistry from St. Joseph’s University (Philadelphia), and an MBA from the Wharton School (University of Pennsylvania). Contact: [chescot@aol.com](mailto:chescot@aol.com).

### 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 world’s knowledgebase through enabling research and managing the flow of scholarly communication.

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

## References

- [1] Building global interest in data literacy: A dialogue, <http://oceansofdata.org/our-work/building-global-interest-data-literacy-dialogue-workshop-report> (last checked June 22, 2016).
- [2] T. Carpenter, "Text and data mining are growing and publishers need to support their use – An AAP-PSP panel report," *Scholarly Kitchen*, February 11, 2016, see: <https://scholarlykitchen.sspnet.org/2016/02/11/text-and-data-mining-are-growing-and-publishers-need-to-support-their-use-an-aap-psp-panel-report/> (last checked June 24, 2016).
- [3] P.R. Cohen, DARPA's big mechanism program, *Physical Biology* **12**(3) (2015), <http://iopscience.iop.org/article/10.1088/1478-3975/12/4/045008> (last checked June 25, 2016). doi:10.1088/1478-3975/12/4/045008.
- [4] T.H. Davenport and D.J. Patil, Data scientist: The sexiest job of the 21st century, *Harvard Business Review* (2012), <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/> (last checked June 22, 2016).
- [5] L.M. Federer, Y.-L. Lu and D.J. Joubert, Data literacy training needs of biomedical researchers, *Journal of the Medical Library Association* **104**(1) (2016), 52–57, 2016. doi:10.3163/1536-5050.104.1.008.
- [6] J. Gantz and D. Reinsel, The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the Far East, December 2012, Sponsored by the EMC Corporation, see <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf> (last checked June 23, 2016).
- [7] "Gartner Identifies the Top 10 Strategic Technology Trends for 2016," *Business Wire*, October 6, 2015, <http://www.businesswire.com/news/home/20151006006831/en/Gartner-Identifies-Top-10-Strategic-Technology-Trends> (last checked June 22, 2016).
- [8] C.L. Giles, Scholarly Big Data presentation, see slide #4, [http://research.microsoft.com/en-us/um/redmond/events/fs2015/speaker-slides/july9/Giles\\_Lee%20\\_scholarly-big-citeseerx.pdf](http://research.microsoft.com/en-us/um/redmond/events/fs2015/speaker-slides/july9/Giles_Lee%20_scholarly-big-citeseerx.pdf) (last checked June 25, 2016).
- [9] S. Griffiths, "How honest is YOUR country? People in the UK are most likely to tell the truth while the Chinese are the most deceitful," *Daily Mail*, November 16, 2015, <http://www.dailymail.co.uk/sciencetech/article-3320606/How-honest-country-People-UK-likely-tell-truth-Chinese-deceitful.html> (last checked June 25, 2016).
- [10] R. Horning, "Notes on the 'Data Self'," *The New Inquiry*, February 2, 2012.
- [11] D. Howe, M. Costanzo, L. Hannick, W. Hide, D.P. Hill, R. Kania, M. Schaeffer, S.St. Pierre, S. Twigger, O. White and S. Yon Rhee, Big data: The future of biocuration, *Nature* **455**(7209) (2008), 47–50. doi:10.1038/455047a.
- [12] C. Ingram, How and why you should manage your research data: a guide for researchers. An introduction to engaging with research data management processes, JISC, January 7, 2016, see [https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data?mkt\\_tok=3RkMMJWWfF9wsRonuqjMZXonjHpfsX56+4pW6S+IMI/0ER3fOvrPUfGjI4ATMRml+SLDwEYGJlv6SgFTrLHMalizLgNUhA=](https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data?mkt_tok=3RkMMJWWfF9wsRonuqjMZXonjHpfsX56+4pW6S+IMI/0ER3fOvrPUfGjI4ATMRml+SLDwEYGJlv6SgFTrLHMalizLgNUhA=) (last checked June 23, 2016).
- [13] Knowledge, networks and nations. Global scientific collaboration in the 21st century, the Royal Society, March 2011, <https://www.snowballmetrics.com/wp-content/uploads/4294976134.pdf> (last checked June 23, 2016).
- [14] R. Koster, *A Theory of Fun for Game Design*, Paraglyph Press, November 6, 2004, ISBN 1-932111-97-2.
- [15] K. Lawrence, Today's college students: Skimmers, scanners and efficiency-seekers, *Information Services and Use* **35**(1–2) (2015), 89–93, see <http://content.iospress.com/journals/information-services-and-use/35/1-2> (last checked June 24, 2016). doi:10.3233/ISU-150765.
- [16] J. Manyika, M. Chui, P. Bisson, J. Woetzel, R. Dobbs, J. Bughin and D. Aharon, *The Internet of Things: Mapping the Value Beyond the Hype*, McKinsey Global Institute, 2015, <http://www.mckinsey.com/business-functions/business-technology/our-insights/the-internet-of-things-the-value-of-digitizing-the-physical-world> (last checked June 22, 2016).
- [17] R. Metz, Artificial intelligence that makes your smartphone smarter, *MIT Technology Review*, July 6, 2015, <https://www.technologyreview.com/s/539056/artificial-intelligence-that-makes-your-smartphone-smarter/> (last checked June 25, 2016).
- [18] S. Reardon, Text mining offers clues to success, *Nature* **509** (2014), 410. doi:10.1038/509410a.
- [19] C. Rohrer When to use which user-experience research methods, Nielsen Norman Group, October 12, 2014, <https://www.nngroup.com/articles/which-ux-research-methods/> (last checked June 24, 2016).
- [20] S. Schroter, N. Black, S. Evans, F. Godlee, L. Osorio and R. Smith, What errors do peer reviewers detect, and does training improve their ability to detect them?, *Journal of the Royal Society of Medicine* **101**(10) (2008), 507–514. doi:10.1258/jrsm.2008.080062.
- [21] R. Van Norden, Publishers withdraw more than 120 gibberish papers, *Nature*, February 24, 2014, <http://www.nature.com/news/publishers-withdraw-more-than-120-gibberish-papers-1.14763> (last checked June 25, 2016).
- [22] Wikipedia: Internet of Things, [https://en.wikipedia.org/wiki/Internet\\_of\\_things](https://en.wikipedia.org/wiki/Internet_of_things). (last checked June 20, 2016).
- [23] Wikipedia: Faceted Search, [https://en.wikipedia.org/wiki/Faceted\\_search](https://en.wikipedia.org/wiki/Faceted_search) (last checked June 25, 2016).