

An overview of the NFAIS 2017 Annual Conference: The Big Pivot: Re-Engineering Scholarly Communication

Bonnie Lawlor

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

Abstract. This paper provides an overview of the highlights of the 2017 NFAIS Annual Conference, *The Big Pivot: Re-Engineering Scholarly Communication*, that was held in Alexandria, VA from February 26–28, 2017. The goal of the conference was to examine the scholarly record and its current evolution in a digital world – both in how it functions and how it serves the information and scholarly research communities. The program stressed how in today’s environment, new and innovative advances in information technology are drafting a blueprint that will optimize the ways in which users create, access, and use data and information. New government mandates and policies continue to be implemented on a global basis to facilitate open access to research outputs while in parallel alternative methods for peer review and measuring impact are being utilized. Within the context of these changes, the conference attempted to look at where this blueprint may lead the information community over the next few years.

Keywords: Scholarly communication, open access, open data, data management, library cyberinfrastructure

1. Introduction

Over the past five years or so digital data has continued to emerge as *the* major force in shaping changes within the information community. It has created new needs and demands for the use and re-use of information, raised user expectations with regards to technology platforms, launched new legislation, spawned the development of archives and repositories, generated new forms of citations for information, and probably has raised more questions than have been adequately answered.

According to the news source, *insideBIGDATA*, while there are diverse sources that predict exponential growth in digital data by 2020 and beyond, they all tend to broadly agree that “. . . the size of the digital universe will double every two years (at least!) – a fifty-fold growth from 2010 to 2020. Human- and machine-generated data is experiencing an overall ten times faster growth rate than traditional business data, and machine data is increasing even more rapidly at fifty times the growth rate.”¹

This unbelievable growth in information requires that the information community examine its traditional methodologies for information creation, dissemination, preservation and use, along with the technologies and policies that serve as a foundation for those traditional methodologies. The NFAIS Annual Conference themes have been looking at the data deluge closely over the past five to six years;

¹“The Exponential Growth of Data,” *insideBIGDATA*, February 16, 2017 (see: <https://insidebigdata.com/2017/02/16/the-exponential-growth-of-data/>), cited July 3, 2017.

e.g. from *Taming the Information Tsunami* in 2011 to *Data Sparks Discovery of Tomorrow's Global Knowledge* in 2016, and in retrospect it is clear from an examination of those programs that the information community *is* changing. But it is equally clear that the process for change is slow compared to the growth of digital content. For example, the need for sustainable new business models remains. It was a hot topic for discussion from 2011 through 2015 when potential new models were regularly put forward and discussed, but there has been relative silence on this topic for the last two years other than to say that they are needed. Today business models have become yet another entry on a very long list of issues and questions that must be addressed in both the sciences and the humanities.

In an attempt to learn how others are attacking this list of issues, a group of researchers, publishers, librarians, and technologists met earlier this year in Alexandria, VA when the National Federation of Advanced Information Services (NFAIS™) held its annual two-and-a-half-day conference entitled *The Big Pivot: Re-Engineering Scholarly Communication*. The conference specifically looked at the following questions: How will open science and open data change the face of research in the near and long term? How is the digital era impacting and changing the scholarly record for the sciences and the humanities in terms of fostering data curation, reproducibility, historical preservation, the globalization of and access to content as well as the role of the research library/librarian? And how are information platforms and tools bringing content to life differently or otherwise influencing the user experience?

Were all the questions answered? No, if they had been the future would be clear. But it did provide much food for thought as to how publishers, librarians, researchers, developers – indeed all of the stakeholders within the information community – *can and need to collaborate* in order to build a new information infrastructure that will accommodate the evolving requirements of the new digital information order.

2. Setting the stage

The conference opened with a keynote presentation that was totally unexpected. In the past, this session usually provided an overview of issues related to the conference theme. This year, **Dr. David Fajgenbaum**, a physician-scientist at the University of Pennsylvania and co-founder/Executive Director of the Castleman Disease Collaborative Network (CDCN) gave a fascinating and thought-provoking presentation on his experience as both a patient and a researcher battling the rare and deadly Castleman disease² (iMCD) with which he was diagnosed while in his third year of medical school. He spent nearly five months hospitalized with multiple organ system failure, received his last rites (which he considers the start of “overtime”), and needed multi-agent chemotherapy to save his life. Dr. Fajgenbaum said that there are seven thousand rare diseases. They do not get a great deal of attention from pharmaceutical companies because of the relatively few people (as compared to cancer or heart disease, for example) who suffer from them. He noted that there are five thousand patients that have been diagnosed with Castleman disease of which there are three types.

What does this have to do with information? Everything – because it was the lack of information on Castleman disease as well as the lack of standard terminology used in the limited information that was available that inhibited the development of its proper treatment. The lack of a common vocabulary made it difficult, if not impossible, to make advances against the disease as researchers could not build on prior studies. Dr. Fajgenbaum said that he was treated by a doctor who was the world’s expert on this specific

²“About Castleman Disease,” the Castleman Disease Collaborative Network, see: <http://www.cdcn.org/about-castleman-disease>, cited July 3, 2017.

disease, but sagely noted that if information is not readily accessible or findable, then the world's best doctor will obviously not have it! He also noted that with regard to research funding, especially for rare diseases, you must hope that the right person applies at the right time – it is simply the luck of the draw!

Because of his first-hand experience, Dr. Fajgenbaum co-founded the Castleman Disease Collaborative Network (CDCN – see: <http://www.cdcn.org/about-us>) in 2012. It is an international networked community of four hundred plus physicians and researchers that takes a crowd-sourcing approach to prioritize research. They use fundraising to recruit experts to perform research studies and produce samples for further use. They translate results to existing drugs when relevant and measure effectiveness with a registry that is shared. Basically, the CDCN is changing the way in which medical research is done for rare diseases through their founding principles:

- Facilitate collaboration
- Promote innovation through crowd-sourcing
- Focus research investment into strategic, high-priority projects
- Facilitate open-source markets for the sharing of research samples and data
- Create systems to quantify effectiveness

The CDCN has developed a new model of pathogenesis through the synthesis of the published literature and has also enabled a unifying terminology system. However, he noted that both “Castleman disease” and “Castleman’s disease” are used in the medical literature. When someone searches PubMed for one of the spellings, then only papers with that spelling are included in the results – meaning approximately one-half of all papers are left out of the results for any search by a physician or researcher into Castleman disease. This can have major consequences as new diagnostic criteria or data on treatment options may not appear for a physician needing this information to treat the disease

In the five years since CDCN was founded they have made significant progress, including the establishment of a Scientific Advisory Board comprised of thirty-two experts from eight countries and the development of a collaborative partnership with Janssen Pharmaceuticals. Dr. Fajgenbaum is currently in his longest remission ever thanks to a precision treatment that he identified through his research at Penn that had never been used before for iMCD.

Dr. Fajgenbaum’s slides are not available on the NFAIS website. However, an article based upon his presentation appears elsewhere in this issue of *Information Services and Use*.

3. Evolving status of open access

The second, and final speaker on the first day of the conference, was **Deni Auclair**, Chief Financial Officer and Senior Analyst, at Delta Think, a consulting firm that helps publishers, membership organizations, and information providers anticipate, create and manage change (see: <https://deltathink.com/>). They have just launched a new product series entitled, *Delta Think Investigations*, in which they “deep dive” into special topics and Open Access is the first topic being addressed. She noted that there is lots of information on this topic, but that it is scattered, often nebulous and confusing, and not always unbiased. Delta Think’s goal is to serve as a neutral third party and to provide a centralized source that meets the needs to access information quickly so that funders and other stakeholders can make informed decisions.

Delta Think used both qualitative and quantitative data for their analysis. Qualitative data includes information gathered through more than thirty interviews with publishers, funders, archive managers, institutions, and thought leaders. This information gathering will be ongoing throughout the coming

year with plans to cover fifty to one hundred publishers, twenty funders, ten repositories, twenty institutions, and ten thought leaders. In addition, they will cover relevant conferences, webinars, and podcasts. Quantitative data includes data on twenty-five thousand journals, publisher data gathered via interviews and questionnaires, and public data sources, including websites, formal reports, and white papers. They look at the data for patterns and all data is curated. Confidential information remains confidential, but is used for benchmarking purposes. Public data from Scopus is used as a starting point, but any inconsistencies that are noted are adjusted with data from publishers. Other sources were used as well, including SCImago (see: <http://www.scimagojr.com/>) and the Web of Science.

They first looked at the growth of all articles (not just Open Access) and with their adjustments based on the lag time for an article to make it into a database, believe that the current rate of growth is 6%.

They looked at the Open Access market and believe that the total revenue in 2015 was \$374 million dollars, growing to \$419 million in 2016, and they project a 10%–15% growth in the next few years. Growth in this market they believe is driven by funding mandates which vary geographically around the globe. Europe is pushing towards Open Access (OA) with a mix of business models that support OA, suggesting that policy and politics are centralized. The UK is also centralized, but funders such as the Wellcome Trust take a more balanced approach towards business models. Auclair noted that the U.S. is the least centralized with the government more focused on *public* access rather than *Open* Access. The U.S. is not concerned about immediate access to research. However, private funders in the U.S., such as the Gates Foundation, are adamantly pro-Open Access. She noted that countries such as China and India are just out of the gate. Their focus is on rewarding authors who publish in high-impact journals (Impact Factor of five or higher), not on access to research information.

Auclair said that the Open Access market is highly-consolidated. The top two hundred and fifty publishers of Open Access journals account for 80% of the output, while there are about seventy-five hundred publishers in the field. Indeed, she said that the top fifteen publishers account for 50% of the market; the top five account for one-third, and the top fifty account for two-thirds. She noted that many factors, such as the length of an embargo, content type, the publisher's mission, competition, etc., impact the pricing of OA journals. She also noted that Open Access journals account for 16%–18% of total article output, but only 3% of total revenue. She said that while she has presented the audience with a lot of high-level data, Delta Think can drill down to a more granular level, including by country and subject matter.

In closing Auclair said that there is no single source of information on Open Access, but that Delta Think is building that resource. She noted that Open Access growth is slowing, but remains strong; that funder mandates and journal reputations continue to influence authors; that Open Access remains an incremental revenue model for established publishers; and that Open Access journals are becoming part of “Big Deal” subscription models.

Auclair's slides are available on the NFAIS website.

4. The physical record: Storage, curation and discovery

4.1. *Advances in manuscript processing*

The theme of the second day of the conference was the Evolving Scholarly Record. The first session, focused on the physical record, was opened by **Bob Kasenchak**, Director of Business Development at Access Innovations, Inc. (see: <http://www.accessinn.com/>), who gave an excellent presentation on

advances in manuscript submissions in scholarly publishing using text analytics. He reported that there are 28,100 peer-reviewed scholarly journals in the STM arena (he did not have figures on publishing in the humanities) that produce an estimated 2.5 million articles per year [8]. These figures are for published articles, not total articles submitted for publication, and does not include conference proceedings. The annual growth per year is estimated at three percent.

He said that the sheer volume of manuscripts that must be processed involves a great deal of work: matching manuscripts to editors and reviewers; fraud detection; screening for questionable research practices; and predicting which manuscripts have a high probability of being published so that they can be given a priority in processing. He noted that the tool provided by Access Innovations, Inc. uses article metadata/taxonomies, text mining techniques, and algorithms to address these work-related issues.

To match manuscripts with editors and peer reviewers, the system uses semantic indexing and subject taxonomies. Using the same taxonomies that the publisher uses to index, the system indexes incoming manuscripts at the point of submission and predicts the topics covered by the manuscripts. They are then matched against an index/database of reviewers and editors who are tagged with their areas of expertise. This facilitates the routing of papers, either automatically or manually, to those with the expertise relevant to a specific manuscript.

With regard to fraud detection, especially for papers that are generated by SCIgen, an online application that uses context-free grammar to generate “spoof” or nonsense papers based on a few user inputs, including references, examples, and an abstract,³ their system detects the papers at the point of submission as they have reverse-engineered the SCIgen algorithm.

Kasenchak used a case study related to problematic cell lines in order to demonstrate how their system detects dubious research. Several organizations publish lists of “known problematic cell lines” that have either been misidentified or known to be compromised or corrupted. He stated that seventy-five percent of researchers do *not* check their research against this list. Access Innovations is partnering with the Global Biological Standards Institute (GBSI), a non-profit organization promoting good research practices, and the International Cell Line Authentication Committee (ICLAC) to produce a tool for authors to check whether or not the cell lines used in their research are on the ICLAC list of known problematic cell lines. He said that the same process will be available to publishers to scan incoming manuscripts and previously-published papers.

His final point of discussion was the use of analytics to predict which papers should move to the top of the pile. Factors include country of origin, the number of authors, the topic, and the sample size.

For more information, see Bob’s slides on the NFAIS website and turn to a paper based upon his presentation that appears elsewhere in this issue of *Information Services and Use*.

4.2. Libraries and big data

The second speaker in this session was **Zhiwu Xie**, an associate professor who also directs the digital library development team at Virginia Tech Libraries. He began by saying that Big Data is no longer unique to Big Science. “Big” is a moving yardstick and Big Data sets have become more common. He noted that the 1000 Genomes project generated 200 TB of data in four years and that the Sloan Digital Sky Phase I and II generated 130 TB in eight years. Today, however, a small lab can produce Big Data sets in shorter periods of time and libraries need an infrastructure that will facilitate their use.

³“SCIgen – An Automatic CS Paper Generator,” Retrieved on April 30, 2017 from <https://pdos.csail.mit.edu/archive/scigen/>.

Examples of Libraries already supporting the use of Big Data are the Library of Congress with their Twitter Archive, the Digital Preservation Network (DPN), the HathiTrust Research Center (HTRC), the Digital Public Library of America (DPLA), and SHARE. All are well known, but he noted that even the average library can handle Big Data, and that this is especially important for academic and research libraries who have the responsibility to support their institution's intellectual property for use and preserving it *for future reuse* – including the application of analytics and data mining. He repeatedly stressed that preservation is not just for the sake of preservation; it is for the purpose of the use and reuse of data. He commented that a Big Data set often offers more value to a researcher than a “smart” algorithm and such sets need to be preserved.

Xie then discussed Virginia Tech's initiative to support the creation, use, and preservation of Big Data. They have a long-term strategic initiative – Beyond Boundaries: a 2047 Vision – and are looking to the future and what they will need in twenty to thirty years. The library is playing a major role to ensure that it remains relevant to the Vision.

One example is the Goodwin Hall Living Lab that was initially started by two faculty members. It was designed as a multi-purpose living laboratory that will provide opportunities for multi- and cross-disciplinary exploration and discovery. It is a one hundred and sixty thousand square feet new building that opened in 2016 and is wired with more than two hundred and forty different sensors. The sensor mounts were directly welded to the structural steel during the building construction and are strategically positioned and sufficiently sensitive to detect human movement. More than forty researchers and educators in various disciplines (music, math, etc.) and institutes expressed an interest in using the data that has been developed by VA Tech and the library has been given the task of building the digital libraries to manage the data and support these activities. The volume of data generated on an annual basis is more than thirty terabytes. Xie also provided two other examples of VA Tech's digital library initiatives.

Because of their long-term commitment to support their institution's Big Data initiatives, the library is developing a library cyberinfrastructure strategy for Big Data sharing and reuse. They have been given a two-year IMLS National Leadership for Libraries grant that started in June of last year. It is a collaboration between the VA Tech Libraries and the departments of Mechanical Engineering and Computer Science, with an emphasis on leveraging a shared cyberinfrastructure (e.g. such as amazon cloud) while also using VA Tech's small high-performance computing center. The questions that they are addressing are: What are the key technical challenges? What are the monetary and non-monetary costs (time, skill set, administrative, etc.)? Are there any cost patterns or correlations to the cyberinfrastructure (CI) options? What are the knowledge and skill requirements for librarians? What are the key service and performance characteristics? And how can the answers to the above questions be consolidated to form an easy-to-adapt and effective library CI strategy? They are addressing these questions within the context of VA Tech's three major Big Data Initiatives: The Event Digital Library and Archive, Share Notify, and the Goodwin Hall Living Lab.

To date the group has identified the network bandwidth as a key bottleneck in the bridge pattern. They have analyzed data loading, its acceleration techniques, and the tradeoffs in the network pattern. They have also participated in building VA Tech's mass storage facility and the tenG campus network

For more information, especially on some of the technical challenges see Zhiwu's slides on the NFAIS website and turn to a paper based upon his presentation that appears elsewhere in this issue of *Information Services and Use*.

4.3. Process for sharing supplemental information

The final speaker in this session was **Jeff Lang**, Assistant Director, Platform Development, American Chemical Society (ACS) who spoke on how and why ACS integrated the figshare DataStore and Viewer into its full text journal platform. He first provided a background on ACS Publishing initiatives.

The ACS Publications Web Platform is an integrated web publishing system that supports ACS journals, books, and its flagship member magazine, *Chemical and Engineering News*. ACS produces fifty journals (about one million plus original research articles) across chemistry and allied fields. It has an archive holding ACS articles from 1879–1995 (465,000 articles).

The ACS eBooks Collection includes both *Advances in Chemistry* which is an archive-only product (1949–1998) and the ACS Symposium Series that includes an archive (1974 to the prior year) and current year subscriptions. The book collection has fifteen hundred titles and thirty thousand chapters. The collection also included the *ACS Style Guide Online*.

The third area holds *Chemical & Engineering News*, ACS' flagship magazine, and includes both the archives (1923–2014) and the *C&EN* Global Edition with more than one hundred thousand original news stories.

ACS Publications delivers supplemental material when provided by authors and each file has its own DOI. The amount of supplemental material has been growing, from about twenty-two hundred and fifty files in January of 2009 to more than four thousand files in mid-2014. Usage of those files has also grown, from just under nine million accesses in 2009 to twenty-five million in 2014.

In early 2016 they decided to utilize figshare (see: <https://figshare.com/>) to provide access to the supplemental information for all but one of the ACS journals. The information is included on both the ACS platform and on the figshare site where it is publicly available. Why? Lang said that ACS had an intrepid journal editor who used Slideshare for videos that were provided as supplemental material and it was getting a lot of use. But when the ACS lawyers read the Slideshare terms of agreement, they nixed the initiative. ACS then decided to build their own version, ACS LiveSlides, using both open source and custom-built software. When presented to ACS editors, they found that all of the editors wanted to have such a feature for their journals, but ACS knew that the service had been costly to build and maintain and could not be cost-effectively scaled to multiple journals. Another piece of information being gathered in parallel was feedback from young researchers who said that they not only wanted access to high-quality information, but that they also believed that they have an obligation to share that information. They want it findable, accessible, searchable, reproducible, and capable of being downloaded and shared.

ACS' solution was to make supplemental material available on their own platform in a visual way and also on figshare (note that LiveSlides was given a new life as a video on the ACS platform and is now scalable). They are not yet sure how much of a positive impact that figshare has had, but they do know that there has not been a negative impact. [Figshare.com](https://figshare.com/) is not a significant source of new traffic. Each month, total usage on figshare.com that has not been referred by pubs.acs.org is less than one percent of the overall use of ACS supplemental information. Incomplete data shows that usage of the supplemental material has been maintained and may have possibly grown. This will be an area of further study in the coming year along with research on reader engagement (clicks, views that lead to downloads, etc.).

In closing, Lang said that the relationship with figshare has provided ACS Publications with a platform on which they can do more research regarding the level of user engagement with ACS material and help that engagement to grow. They are interested in knowing if such engagement will encourage authors to provide their own supplemental information in interactive formats rather than as PDF files, for at the

moment at least sixty percent of supplemental information files are submitted as PDF's. Bottom line: future research will determine if they are meeting the needs of young researchers.

For more information access Lang's slides on the NFAIS website.

5. Shark tank shoot-out

The final session of the morning was 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 session Moderator was Eric Swenson, Director, Product Management, Scopus, Elsevier and the Judges were Kent R. Anderson, Founder, Caldera Publishing Solutions; James Phimister, Principal, PHI Perspectives; and Jason Rollins, Senior Director of Innovation, Clarivate Analytics.

The first speaker was **Laurence Bianchini**, COO, MyScienceWorks (see: <https://www.mysciencework.com/>). Founded in 2010 by two young graduates with complementary profiles, Virginie Simon, a biotech engineer with a PhD in nanotechnology, and Tristan Davaille, a financial engineer with a degree in economics. They have fifteen employees and a presence in three countries – the USA, France, and Luxembourg. Their company offers three main services that are intertwined. *MyScienceWork.com* is a search engine platform providing access to sixty-six million publications and a sharing and profiling platform for research. It is their freemium service geared to individual researchers. *Polaris*, a software-as-a-service product, provides research institutions and publishers with turnkey repositories in which they can store their documents, videos, etc. and through which they can provide access to their open access material and showcase their research. And, finally, SRIUS, that specializes in high added-value semantic and data science services such as bibliometric studies and reports, science metrics, and research mapping. In 2016 they had five hundred thousand dollars in revenue, ninety percent of which was from Polaris licenses. Currently, their market is primarily in Europe and on the West Coast of the USA. They have twenty-five institutional customers, including universities (e.g. Stanford), consortia (e.g. U. Sorbonne Paris Cité), research centers (e.g. SETI Institute), funders (e.g. ARC Foundation), and laboratories (e.g. LBMCC). They have raised five million dollars in venture capital funds and predict revenue of two million dollars in 2017 (fifty percent Polaris licenses and fifty percent SRIUS services), with profitability in late 2017.

The second presenter was **Ruth Pickering**, Co-Founder of Yewno, a company that offers a knowledge discovery search platform, Yewno Discover, that uses machine-learning and computational linguistics to analyze and extract concepts, and discern patterns and relationships in order to make large volumes of information more effectively understood through visual display (see: <https://about.yewno.com/>). It is a complement to traditional library search tools. The founders believe that Yewno enables the extraction of new meaning and value from content and increases the ability to promote and expose collections in an entirely new way as it is based upon concepts, not key words. Pickering said that the 2016 revenues for the traditional search market was one trillion dollars and that the 2026 revenue for next-generation search technologies such as Yewno Discover is predicted to be ten trillion dollars.⁴ She said they expect that their technology will follow the classic adoption curve and that they expect close to one hundred million dollars in revenue by the year 2020.

⁴*The State of Artificial Intelligence*, World Economic Forum, Davos, January 2016, see: <https://www.youtube.com/watch?v=VBceREwF7SA>, cited July 3, 2017.

The third presenter was **Simon Adar**, Founder and CEO of Code Ocean (see: <https://codeocean.com/>). He noted that more and more research results include actionable data or code, but that the dissemination of that code relies on individuals to set up environments to reproduce the results. Officially launched in February 2017, Code Ocean is a cloud-based computational reproducibility platform that provides researchers and developers with an easy way to share, discover and run code that is published in academic journals and conference proceedings. It allows users to upload code, data, or algorithms and run them with a click of a button. The platform enables reproducibility, verification, preservation, and collaboration without any special hardware or setup. Code Ocean provides next generation tools to facilitate digital reproducibility, where users can access a working copy of a researcher's software and data, configure parameters and run it regardless of the users' operating systems, installation, programming languages, versions, dependencies, and hardware requirements. The Code Ocean widget is free to individuals to access and download. Publishers are charged one hundred dollars per year per journal title. The company employs more than ten people and has raised two-and-a half million dollars.

The final presenter was **Lenny Teytelman**, CEO, Protocols.io (see: <https://www.protocols.io/>). He noted that published biomedical research results are often not reproducible simply because they lack the detailed instructions necessary for repeating the experiments. Protocols.io attempts to solve that problem as it offers an open access platform for sharing research recipes. Unlike traditional publications, with the use of versioning the methods on protocols.io can be corrected and optimized long after publication. These protocols are also interactive and can be followed step-by-step during experimentation on the web and on native iOS and Android apps. All public content on protocols.io is both free to read and free to publish, with the business model based on subscriptions to private groups and data services. They have received funding of \$2.38 million dollars and another round of funding is underway. They hope to be sustainable in two years.

Teytelman spoke at the 2015 NFAIS Annual Conference on the same topic and published a paper in *Information Services and Use* [7].

Later in the afternoon the judges announced the winner of the Shoot Out, MyScienceWork. The reasons given were that they listened to their market, they had revenue, they took a strong technical approach, and focused on key customers. 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 in this session are available on the NFAIS website.

6. Members-only lunch session: Washington's impact on the scientific enterprise

Between the morning and afternoon sessions there was an NFAIS Members-only luncheon with a presentation by Benjamin W. Corb, Director of Public Affairs, the American Society for Biochemistry and Molecular Biology. Corb shared from his perspective how the Trump Administration policies together with Congressional priorities are shaping and influencing the country's scientific enterprise, and whether that will hurt or help the nation long-term.

He concurred with David Fajgenbaum, the opening keynote speaker, that Big Pharma will not invest in drug discovery for cures that impact few people. They cannot afford the investment as there will not be a return and most likely will result in a loss.

He admitted that he has no idea what the impact of the Trump administration will be on science. He said that it is equally possible that he could wake up tomorrow morning and read that the NIH budget has been cut in half and equally possible that he could read that it had been given a fifty percent increase.

He did a quick look at federally-funded R&D (including the defense budget) as follows: investment peaked in the mid-1960's; dropped precipitously from the 1970's to the mid 1980's; they rose slightly in the mid-1980's and then plateaued through 2007. There was a mild increase from 2007–2009, followed by a plateau until 2011 when the Budget Control Act of 2011 was implemented and there has been a significant decline ever since. In the mid 1960's federally-funded R&D research accounted for 11.5% of the entire budget (5.8% if defense funding is removed). Today it stands at 3.1% (1.6% if defense funding is removed).

Corb said that we are now in a very different place, but that he believes that Congress will stand behind the scientific community. He also said that scientists must demonstrate to the new administration that government funds *do* result in the creation of new businesses and new jobs. Federally-funded research has resulted in many scientific breakthroughs, including GPS systems, the Internet, vaccines, bar codes, the microchip, wind energy, Goodyear tires, infant formula, and others, and such potential breakthroughs must continue to be funded. Bottom line message: Scientists need to defend their research and stand up to detractors. Note that Corb did not use slides for his presentation.

7. Miles Conrad lecture

The first afternoon session 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 **Judith C. Russell**, Dean of University Libraries, University of Florida, and the complete transcript of her talk is published in full elsewhere in this issue. It gives an interesting perspective not only on the evolution from print to digital information that Russell has experienced first-hand, but also on the type of collaborative relationship that can be forged between commercial and non-profit organizations, in this case between the University of Florida and Elsevier. Russell pointed out that some members of the information community are motivated by profits while others have the opportunity to deliver “free” (no fee) or not-for profit services, but noted that this does not make them adversaries, but rather “We are colleagues who can and do benefit from collaboration and learn from one another.” It is a message with which I concur and I recommend that you read her paper.

Note that there were no slides for her presentation. However, to learn more you can read a paper based upon her presentation that appears elsewhere in this issue of *Information Services and Use*.

8. Data as the scholarly record

8.1. Open science and researcher behavior

The final session of the day was opened by **John Wilbanks**, Chief Commons Officer, Sage Bionetworks (see: <http://sagebase.org/>), who addressed researcher behavior and how it has changed in an open science environment. Wilbanks first described Sage Bionetworks, a non-profit biomedical research institution that works in three areas – team science, open science, and participant-centered science – and said that when they are really successful, they combine all three areas. His presentation focused on “decentralization” rather than “open access” because he believes that the latter tends to get philosophical. “Open” is a means to an end, and “decentralization” (as John said is defined by Wikipedia) is “the process of redistributing or dispersing functions, powers, people, or things away from a central location or authority.” He believes that science actually works best when decentralized.

He noted that there are four factors at work in biomedical research: 1) the pervasiveness of networked information (e.g. NIH, NLM, etc.); 2) the cloud infrastructure; 3) the democratization of the research process (e.g. citizen science); and 4) funders and public pushing for more sharing of information. It is this combination of forces that is dragging things toward decentralization.

One of the initiatives with which Wilbanks is involved is the Accelerated Medicines Partnership (AMP).⁵ Launched in 2014, this is a public-private partnership between the National Institutes of Health (NIH), the U.S. Food and Drug Administration (FDA), ten biopharmaceutical companies, and multiple non-profit organizations. Their goal is to transform the current model for developing new diagnostics and treatments by jointly identifying and validating promising biological targets for therapeutics. The ultimate goal is to increase the number of new diagnostics and therapies for patients and reduce the time and cost of developing them. They are looking at three diseases – type 2 diabetes, Alzheimer’s disease, and lupus. Wilbanks is involved with the Alzheimer’s project. That specific project has received \$129.5 million in funding (NIH with \$67.6M and industry with \$61.9M). There are six labs involved, each using their own methods, their own data sources, platforms, algorithms, etc. He noted that this would be a mess if it were not an open-standards based project. You would never get all of these labs to agree on the same methodologies, etc. They work privately, but share work on a quarterly basis and combine evidence across teams. He said that the work is decentralized in order to maximize research, but it is ultimately shared. The focus is on provenance, not publication. So the publication process of this type of research moves more slowly (they are not stopping to publish at each step on the way), but the knowledge-base grows more quickly.

In addition to “decentralization” another form of “open” is collaboration and Wilbanks used the TCGA-Pan-Cancer Consortium as an example (TCGA stands for The Cancer Genome Atlas). The Pan-Cancer initiative compares the first twelve tumor types profiled by TCGA. Analysis of the molecular aberrations and their functional roles across tumor types will teach researchers how to extend therapies effective in one cancer type to others with a similar genomic profile. This group had the “pain” of analyzing the data and when Wilbanks got involved they were using an FTP platform. They, too, were moved to an open standards data sharing platform, their work was made less cumbersome, and they had a number of papers published in *Nature*. Their successes motivated others to become involved in this type of open research.

Another form of “open” that Wilbanks discussed was citizen engagement. This was done in partnership with Apple when Apple launched their Research Kit that facilitated the use of mobile devices in clinical studies. The initiative was entitled “mPower” and it was used to study Parkinson disease symptoms – dexterity, gait, balance, and memory. The study allowed researchers to better understand how the symptoms were connected to the disease and let patients recognize their own signs and symptoms. Data was gathered on a frequent basis (e.g. before and after taking medication), and twenty-two thousand patients participated. During the first six months sixteen thousand-five hundred and eighty patients consented to participate; fourteen thousand-six hundred and eighty-four patients actually enrolled; nine thousand-five hundred and twenty patients agreed to have their data shared broadly; and one thousand and eighty seven patients self-reported their Parkinson Disease diagnosis. It has been about a year since the project was launched and the data was released before any primary publication was published. There are more than eighty independent ‘qualified researchers’ analyzing it. He suspects that one of the three research groups studying the data will release a paper before his organization does, and he believes that is a good thing. If the data had not been released, there would have been a longer wait for research

⁵See: <https://www.nih.gov/research-training/accelerating-medicines-partnership-amp>, cited July 3, 2017.

advances to be made! One of the outcomes of this effort is the creation of a Parkinson Disease research community.

In closing Wilbanks said a lesson learned is when two laboratories are working to solve the same problem they should share their results with one another early and privately. Ninety-five percent of the work required to share data will have been done, so when the time comes to share the data with the world, it can be done quickly.

For more information, access Wilbank's slides on the NFAIS website and/or go to <http://sagebase.org/>.

8.2. *Humanities commons: Networking scholarly communication*

The second speaker in this session was **Kathleen Fitzpatrick**, Associate Executive Director and Director of Scholarly Communication of the Modern Language Association (MLA). She discussed the development of the Humanities Commons, an interdisciplinary scholarly communication platform developed by the Modern Language Association, as an example that demonstrates the potential that online scholarly networks might present both for researchers and for the societies to which they belong.

She began her presentation with a brief discussion of Elsevier's May 2016 acquisition of the Social Science Research Network (SSRN) – a place where researchers posted their papers before the completion of the peer-review process and before they were “locked-up” behind a fee-based firewall. The acquisition raised fears that the data would no longer be made available and that Elsevier would mine the data for commercial purposes. Concerns increased when some of the posted papers were removed for reasons of copyright, and it was suggested that users abandon the network. She noted that a similar mindset occurred in 2006 when a movement was started for scholars to leave the Academia.edu network when that network suggested that they might charge scholars for recommendations of their papers by their website editors. This occurred again more recently when the network exposed reader information to paying customers. There is a feeling that the networks upon which researchers have come to rely are becoming commercial in nature. She stated that there can be a disconnect in core values between the provider of a network platform and those of the scholars that use the network. Scholars must be sure that there is an alignment of values between them and the providers of any service that they plan to use for the long haul so that they can be assured that the services will evolve appropriately along with them. Fitzpatrick said that while networks may be open for researchers to deposit and share information, there is no openness with regard to the platform provider's business model and goals, and used the following quote: “If you're not paying, you're the product being sold.”

She noted that membership societies can be a better option as they foster openness and communication among their members and between their members and the broader external world. Plus, they are governed by their members so the business goals, models, and values are quite transparent. The only barrier to true “openness” is that such societies are open only to members and not everyone may be eligible to join.

Fitzpatrick then went on to describe MLA's Humanities Commons. In 2013 MLA launched its “MLA Commons” platform that was built on open source software. They soon began to receive requests from members to be able to connect with peers in other disciplines within the Humanities and from other societies within the Humanities who were interested in what MLA was doing. Hence, the concept of a broader network took hold. With funding from the Mellon foundation, they began a planning process and a beta process in partnership with the Association for Jewish Studies; the Association for Slavic, East European, and Eurasian Studies; and the College Art Association. Each society has its own *Commons* hub. The Commons was launched in December 2016 and is a nonprofit network where humanities

scholars can create a professional profile, discuss common interests, develop new publications, and share their work. The *Humanities Commons* network is open to anyone (see: <https://hcommons.org/about/>).

She said that in 2017 they will be looking to expand the number of society partners and over the next five years will shift from grant-based support to society-based collective funding. She expects that fund raising will also play a role in the future sustainability of the network. They are now working on the development of a new governance model in which both individual and institutional members are given a voice. The goal is to ensure that the network remains non-profit, is maintained by scholars, and that the principle value of membership is the ability to participate in conversations and processes that evolve into collective action.

For more information refer to Fitzpatrick's slides on the NFAIS website and/or go to <https://hcommons.org/>.

8.3. *Data first manifesto: Shifting priorities in scholarly communications*

The final speaker of the day was **Clifford Anderson**, Associate University Librarian for Research and Learning at Vanderbilt University. Anderson discussed the Data First Manifesto (<http://datamanifesto.info/>), which he co-authored in 2016 with Todd Hughes, a colleague at Vanderbilt University. He said that the Manifesto emerged from a growing frustration that digital scholarship projects were getting hung up or stalled for technical reasons that had nothing intrinsically to do with their scholarly goals and that they borrowed the Manifesto's form from that of the Agile Manifesto [2], a call issued by leading software engineers in 2001 to shift priorities in project management.

He noted that the Manifesto calls for four key shifts in scholarly communications and that scholars need to emphasize:

- Data Serializations over Databases: Make data available in a format that can be preserved, read, and reused independently of the database in which those data are stored.
- Application Programming Interfaces over Graphic User Interfaces: Starting with an API places priority on *communicating data*, not on presenting a slick interface.
- Curating Data over Summarizing Information: Digital scholarship projects should, so far as possible, allow the data to speak for themselves. While providing a narrative (or interpretative visualizations) of data can provide a helpful snapshot of what the data contains, emphasis should be placed on making data self-describing and perspicuous.
- Citing Datasets over Referring to Results: This principle speaks to scholars who engage with others' projects. Assuming that projects make datasets readily accessible, the preference is that scholars test their claims directly against the data rather than rely on others' claims.

Anderson went into some detail on each point, including potential challenges. He closed by saying that the goal of the Manifesto is to help kick-start a conversation among digital scholars, especially graduate students and postdoctoral fellows, about how to shift the emphasis away from flashy websites towards data curation, data publication, and data citation.

Anderson's talk was very logical and thought provoking. For more information, see his slides on the NFAIS website and turn to a paper based upon his presentation that appears elsewhere in this issue of *Information Services and Use*.

9. Our napster moment: Academic publishing, access and what's next

The final day of the conference opened with a Plenary session in which **Kendall Bartsch**, CEO, Third Iron (see: <http://thirdiron.com/>), discussed how the environment in the Academic Publishing Community is ripe for end users to create their own “Napster Moment,” and what initiatives are underway to stop it from happening.

Bartsch began with a brief history of the development of the MP3 format⁶ for and how it ultimately transformed the recording industry, chiefly as a result of the 1998 lawsuit that the Recording Industry Association of America (RIAA) brought suit against Diamond Multimedia Systems because the latter's MP3 recording device facilitated the creation of second-generation, or “serial,” copies of original recordings that could then be easily passed on to others. Ultimately, the court ruled against the RIAA, stating that the noncommercial copying of recorded files from a computer hard drive to the MP3 device was fair use. Basically, the ruling said that the device simply made copies in order to “space shift” or make files that already reside on a user's hard drive more portable.”⁷

The decision led to a rise in music sharing despite it being cumbersome. Then around 1998 Shawn Fanning developed the software, Napster, to bypass the problems in sharing. He and his partner, Sean Parker, saw the opportunity and founded their own company in 1999. By 2001, they had millions of users [9]. That same year their company closed following a successful lawsuit filed by RIAA claiming Napster infringed copyright. But it was already too late! User behavior had already changed and expectations had been raised. The music industry now had to change in order to survive.

Bartsch says that Academic Publishing is facing a similar crisis. As with music, the transformation of information from print to digital has raised user expectations, and barriers to access and use are making users increasingly frustrated. Among the frustrated users are those who have authorized use, but cannot access content because they are off-site and not using their campus network. He said that he estimates that 35% of the 4.2 billion access requests that are denied annually are likely to have come from those who are legal users. And, as with music, illegal sources have emerged, to satisfy these frustrated users, with Sci-Hub being one of the more notorious.⁸

He then briefly described two initiatives that have been started to make legal access content easier: 1) RA21: Resource Access in the 21st Century established in 2016 as a joint initiative between the STM Association and the National Information Standards Organization. It aims to “optimize protocols across key stakeholder groups, with the goal of facilitating a seamless user experience for consumers of scientific communication,⁹ and CASA (Campus-Activated-Subscriber-Access), a joint project between Highwire Press and Google, that aims to achieve goals similar to those of RA21, specifically seamless access to licensed content found in Google Scholar.

In closing, he said that, like the music industry, we can expect that additional solutions will also emerge.

For more information, see Bartsch's slides on the NFAIS website and turn to a paper based upon his presentation that appears elsewhere in this issue of *Information Services and Use*.

⁶MP3, see Wikipedia: <https://en.wikipedia.org/wiki/MP3>, cited July 3, 2017.

⁷Recording Industry Association of America v. Diamond Multimedia Systems, Inc., see: <https://cyber.harvard.edu/property00/MP3/rio.html>, cited June 26, 2017.

⁸Sci-Hub, see: <https://en.wikipedia.org/wiki/Sci-Hub>, Cited July 3, 2017.

⁹See: <http://www.stm-assoc.org/standards-technology/ra21-resource-access-21st-century/>, cited July 3, 2017.

10. Mobile design, personal assistants, technological change and chatbots

The first speaker in the next session was **Rosalind Hill**, Digital Publishing Director, Future Science Group, a small, London-based publisher medical and biomedical journals (see: <https://www.future-science-group.com/>). She gave an excellent presentation on how publishers must adapt the ways in which they reach their audience with the content they need, when they need it, and stressed that this means embracing digital technologies and trends beyond traditional print journals and online journal platforms.

Hill began with a quick overview of the current information landscape. She said that in today's digital environment the customer is king (or queen). They want personalization and targeted content. They primarily use mobile devices to go online, with app usage making up sixty percent of mobile usage online in 2015. Seventy-nine percent of those from the ages of eighteen through forty-four years have smartphones with them twenty-two hours/day, eighty percent of whom check their phone as soon as they wake up. People interact with their phones on average thirteen times per hour, and seventy-nine percent of consumers use smartphones for research purposes both for work or personal reasons. Sixty percent of Google searches now come via mobile and up to forty percent of visitors will leave a site if it doesn't load after three seconds. Seventy five percent of smartphone traffic will be video by the year 2020!

The key points she wanted to stress to publishers are the following:

- Mobile is first – it is how most people now connect with content
- Everyone is an editor – user-generated content is a driving force in the information industry; e.g., Facebook is a major player, but does not own any of its content
- Social sharing
- Visual presentation of content/data beats text
- International reach by Publishers is essential and has also increased competition
- Tech giants (e.g. Facebook, Google and even Uber) set the standards for User Experience standards

Hill then went on to describe the key characteristics of the various generations that populate today's workforce and their preferences in technologies and devices. She then talked about some surveys that her organization conducted to obtain more information on people that used their journals. Speaking with researchers, doctors, nurses, and marketing professionals via conversations, surveys, and Twitter polls they found that the audience was split on device usage with sixty percent preferring mobile and forty percent preferring desktop PCs. The social media usage in order of reference was: Twitter, Facebook, and LinkedIn. And their favorite apps were: SmartNews, Amazon, City Mapper, Google Maps, Twitter, BBC, My FitnessPal, Instagram, Podcasts, and Splittable.

They discovered that what is most important to this group are the following: networking, finding researchers' contact details, finding out about unpublished data, getting at customized solutions; access to peers, bridging the gap with pharma, and curation of relevant information.

The pain points were: Open access/legal challenges of accessing research, integration of big, diverse data sets; solo/silo working; data protection concerns; poor communication; and the biggest pain point – time (or lack thereof).

Hill said that her organization is utilizing usage statistics and user feedback in order to develop new apps and to further personalize their offerings. She closed by saying that the challenge for publishers and marketers is to accurately predict how an individual user consumes content and navigates the digital space in order to ease their burden in an era of information overload.

If you work in the publishing industry, you *must* review her slides that are on the NFAIS website and read the article based upon her presentation that appears elsewhere in this edition of *Information Services and Use*.

10.1. Laboratory data management

The next speaker in this session was **Matt Dunie**, President, LabArchives, who spoke on the use of Electronic Laboratory Notebooks (ELNs) for data management.

Dunie said that there has been a growth in research-related data. To re-enforce his point he added that that much of this research is ultimately published and that according to a report from the International Association of Scientific, Technical and Medical Publishers (STM) there are more than twenty-eight thousand one hundred English language peer-reviewed journals in publication, with an annual output of an estimated 2.5 million articles [8].

All of this data must be managed properly in order to facilitate quality research. Data Management Plans have been a requirement of various funding agencies for most of the past decade and such plans are now being enforced. He also noted that factors such as data integrity, data lifecycle, data security, perpetual revision history, permanence and unchangeable time stamps are growing in importance with regards to the management of laboratory research data. Perhaps even more important to researchers and their institutions is the ability to provide proof of research and discovery. Ultimately, the use of a Digital Lab Notebook can help prove discoveries, protect intellectual property, and provide the tools necessary to defend or audit research activities.

Dunie provided an example of how a lab notebook helped to win a lawsuit. Albert Schatz worked with Selman Waksman at Rutgers University on research related to the discovery of Streptomycin. At Waksman's request, Schatz signed over his rights to royalties from the U.S. streptomycin patent to the Rutgers Research and Endowment Foundation, and later signed over his foreign rights. However, Schatz began to feel that Waksman was playing down Schatz's role in the discovery, and taking all the credit and recognition. In 1949 it came out that Waksman, contrary to his public pronouncements, had a private agreement with the foundation that gave him twenty percent of the royalties. Schatz sued Waksman and the foundation for a share of the royalties and recognition of his role in the discovery of streptomycin.¹⁰ According to Dunie it was because of the content of Schatz's lab notebooks that the suit was settled in his favor.

Dunie noted that his direct experience with academic research labs indicates that more researchers are using paper notebooks to document their finding than anything else. But that they are also using some digital substitutes for paper and a plethora of home grown solutions. This is what led to the founding of LabArchives. He went on to describe the capabilities of the ELN that they offer (see: <http://www.labarchives.com/>), and noted that in today's world of global collaborative research, such a notebook is essential, not only for efficiency, but also for proof of discovery.

In closing, Dunie said that ELNs can support the Scientific Method in ways traditional paper notebooks cannot and that they also support institutional research policies and objectives and provide a platform for institutional data management and research support.

Dunie's slides are not available on the NFAIS website. However, an article based upon his presentation appears elsewhere in this issue of *Information Services and Use*.

¹⁰ See: [https://en.wikipedia.org/wiki/Albert_Schatz_\(scientist\)](https://en.wikipedia.org/wiki/Albert_Schatz_(scientist)), cited July 3, 2017.

10.2. *The next big paradigm shift in IT: Conversational artificial intelligence (CAPs)*

The final speaker in this session was **Rajan Odayar**, Vice President, Head of Data Science, Digital Insights & Global Enterprise Management Solutions, ProQuest who spoke about chatbots – computer programs that conduct a conversation via auditory or textual methods. We all have had first-hand experience with them (often frustrating experiences) as they are used for customer service (or lack thereof) or information acquisition prior to connecting to a human over the phone.

Odayar said that the introduction of chatbots into society has brought us to the beginning of a new era in technology: the era of the conversational computer interface. It's an interface that soon won't require a screen or a mouse to use. There will be no need to click or swipe. This interface will be completely conversational, and those conversations will be indistinguishable from the conversations that we have with our friends and family. (*Note*: I recently saw that four of the top fifteen artificial intelligence platforms in 2017 are voice-related: see: <http://www.predictiveanalyticstoday.com/artificial-intelligence-platforms/>).

Odayar believes that Conversational Artificial Intelligence platforms (CAPs) will be the next big paradigm shift in information technology. He noted that they are already on the market today, but more are coming and that they are likely be the strongest instigator of investments that exploit Artificial Intelligence for a decade or more. He said that eighty-one percent of shoppers look online before making a big purchase and that in-store payments of purchases will reach seventy-five billion dollars this year. But he predicts that by 2026, eighty percent of the buying process will occur without any human-to-human interaction.

Odayar noted that IBM is making waves in this area with “Watson Conversation,” and that they offer the opportunity for potential users to try it for free,¹¹ commenting that in the future Watson may be the center of customer enjoyment.

In closing he said that research in this area encompasses more than chatbots, virtual assistants, and messaging-based applications, and that he believes that the emergence of CAP will stimulate significant growth in the exploitation of Artificial Intelligence in general.

Odayar's slides are not available on the NFAIS website.

11. Authentication, privacy issues, and opportunities

The first speaker in this session was **Subhjit Basu, PhD**, Associate Professor in Information Technology Law, School of Law, University of Leeds, who focused how to create a balance between innovation and regulation. He said that “Big data” is becoming a core asset in the economy. It fosters new industries, processes, and products, and it creates significant competitive advantages. Across every field data is the bible – the answer to everything. But there needs to be a balancing act and unfortunately the law lags behind technology. He noted that the best regulation needs to be “informed,” and that “knee-jerk” reactions must be avoided, and noted that data-driven innovation poses challenges related to governance and policy as well as challenges related to public understanding and public trust. It also raises questions about privacy, consent, data ownership, and transparency.

Basu then presented a case study: Google DeepMind and Healthcare Data. DeepMind¹² is an artificial intelligence company that was founded in London, England in 2010 and was acquired by Google in 2014. It has been given access to the healthcare data of up to 1.6 million patients from three hospitals run by a

¹¹ See: <https://www.ibm.com/blogs/watson/2016/07/building-better-bots-watson-conversation/>, Cited July 3, 2017.

¹² Deep Mind, See: <https://deepmind.com/>, cited July 3, 2017.

major London NHS trust. DeepMind will launch “Streams,” an app for identifying acute kidney injury and it will trigger mobile alerts when a patient’s vital signs or blood results are flagged as abnormal. The NHS has used a loophole around “implied consent,” and does not require patient “consent” for direct care. The UK’s data protection watchdog, the ICO, is investigating complaints about the “Streams” app. (*Note: I have found that despite the controversy, more deals continue to be signed*) [4].

Basu said that this has raised a number of questions: Did the patients sign up for this? Where is the transparency and fairness? Can Google/DeepMind be trusted? What’s in it for Google, for DeepMind? Will the data be repurposed? Will it be linked to other data? What are the most important ethical and legal challenges raised by AI in healthcare? Who does (and can) own data anyway, and on what basis? How can we ensure the rights of patients, indeed of all individuals, are safeguarded? Does the current legal framework on data protection take into account the reality characterized by the development of data-driven innovation in healthcare? What is the role that technology can play to ensure that data-driven innovation advances in an optimal way, particularly in the context of the privacy of the healthcare data?

He noted that in the UK patients do not own the data – they simply generate it. He said that the missing “balancing act” may require a new regulatory framework to protect privacy while at the same time advancing medical research and healthcare, and that perhaps new ethical standards are required for the private sector if they operate within the context of healthcare. He added that we need to balance public benefits from research/deals against commercial gains and questioned how the increasing amounts of data in society can be best used for public good and with public support.

Basu closed his very thoughtful presentation by mentioning a book that he has co-written with Christina Munns entitled, *Privacy and Healthcare Data: ‘Choice of Control’ to ‘Choice’ and ‘Control’*. The book was first published in 2015, and their goals in publishing it are to encourage and empower patients to make informed choices about sharing their health data and to provoke new visions for the sharing of healthcare data.

Basu’s slides are available on the NFAIS website.

11.1. *Proof-of-publication using the bitcoin blockchain*

The second speaker in this session was **Christopher E. Wilmer**, PhD, Assistant Professor and Managing Editor of *Ledger*, University of Pittsburgh who talked about cryptocurrency, defined as a digital currency in which encryption techniques are used to regulate the generation of units of currency and verify the transfer of funds, operating independently of a central bank. He noted that the journal he currently manages, *Ledger*, is a peer-reviewed journal for publishing original research on cryptocurrency-related subjects.

He said that people in the audience may have heard of Bitcoin. Bitcoin is a cryptocurrency, and in many ways it is like any other currency (e.g., USD, Yen, Euros), but it is one for which some functions that traditionally rely on trusted authorities are replaced by the use of clever cryptography; e.g., sending money to someone via check/wire/online or controlling the money supply. Bitcoin is decentralized (no single point of failure). There is no Bitcoin company or person who controls/manages it. . . it just exists autonomously, like a language. . . or a virus.

He noted, however, that despite their name, cryptocurrencies are not just about money. They are about recording ownership of arbitrary data and facilitating the transfer of ownership to someone else without a central institution; e.g., trading stocks without a stock exchange; transferring the title to a house without a notary/court; filing a patent without a patent office; issuing concert tickets without a ticket office and timestamping scientific results/publications.

Wilmer said that in 2008, someone using the pseudonym Satoshi Nakamoto, published a paper describing Bitcoin. After stimulating some discussion among cryptographers, Satoshi released a working prototype in 2009. The currency units, “bitcoins,” were worthless for the first two years. . . then token trades started (bitcoins for pizza). Now six years later, it is a \$16 billion economy, each of the estimated sixteen million bitcoins is worth more than \$1,000 US dollars.

Wilmer noted that since Bitcoin, thousands of similar cryptocurrencies have been created with tweaks to the underlying technology (“Blockchain technology” is equivalent to cryptocurrency for all practical purposes). Today Princeton, Stanford, MIT, Duke University, and several others teach Bitcoin/Blockchain courses. The Bank of England, The Depository Trust and Clearing Corporation (DTCC), and other large institutions have created blockchain “strategy” or “think-tank” groups. The National Science Foundation (NSF) has already funded over \$3 million for cryptocurrency research and nearly one thousand papers have been published on the subject.

He noted that while the growth in research papers on the topic has been significant, there has been very little peer-review. Since for many academics, their research on cryptocurrency doesn’t “count” in the eyes of their peers unless it is published in a traditional academic journal, the founders of *Ledger* wanted to make it easier for academics to get involved in cryptocurrency research. They also wanted to raise the standard of research being done by scholars (not necessarily academics) already in the Bitcoin/cryptocurrency community. So they set out to create a traditional academic journal devoted to this topic. The name “Ledger” is derived from yet another synonym for “blockchain technology” which is “distributed ledger technology.”

He said that some members of the community wanted to leverage cryptocurrencies to create a futuristic decentralized “journal” for cryptocurrency research, and, in principle, advanced cryptocurrencies like Ethereum could probably allow for an entirely decentralized journal publishing platform. He noted that citation tracking, in particular, is well-suited to be done in a decentralized way using a cryptocurrency, but the founders of *Ledger* chose to leave all of that to others.

Wilmer said that this is the first peer-reviewed journal for publishing original research on cryptocurrency-related subjects. It was founded in 2015 and has a broad scope – Mathematics, cryptography, engineering, computer science, law, economics, finance, and social sciences. It is Open Access; e.g. free to view content (no subscription cost), and free to publish (no author fees) and it is published by the University Library System at the University of Pittsburgh. The first call for papers was in September of 2015 for the inaugural issue that was published in December 2016.

He noted that the similarities to most journals are that the journal has three article types: original research, reviews, and perspectives; that the Editors handle submissions, find, and contact reviewers (typically three), and make final decisions; that there is a single-blind review process (i.e., reviewers know the identity of author, but not the other way around); and that there are multiple review rounds if necessary.

The differences from most journals include a transparent review process: Reviews, including author correspondence, are published alongside accepted articles; once accepted, articles are digitally signed by the authors (they provide a user-friendly tool for this). This cryptographically proves that an article has not been altered; the signed document is timestamped by the Bitcoin blockchain which cryptographically proves that the article existed before a certain time; under exceptional circumstances, authors are permitted to publish under a pseudonym (the demand for this is less than anticipated. One pseudonymous cryptographer, whose research was on how to make an even more anonymous version of Bitcoin, decided to publish in *Ledger* under his real name).

In closing, Wilmer noted that proof-of-publication is done using Blockchain Technology. Blockchains as a data-storage mechanism, are well-suited to be used in scholarly publishing because they are extremely resilient, tamper-proof, practically indestructible database; there is no single point of failure or cost of operation; and there is an incontrovertible proof-of-publication date, even across countries and institutions whose incentives are not aligned (which is sometimes a point of contention for scientists racing to discover cure/new theorem/etc.).

Wilmer's slides are available on the NFAIS website and the text of the above section is totally based upon those slides and his oral presentation.

11.2. RA21 initiative: Improving access to scholarly resources from anywhere on any device

The final speaker of the morning was **Ralph Youngen**, Director of Publishing Systems Integration, American Chemical Society, who spoke on the RA21 initiative for user authentication that was mentioned by Kendall Bartsch during his presentation on the "Napster Moment" that was discussed briefly earlier in this article.

As a reminder, RA21, the Resource Access in the 21st Century, was established in 2016 as a joint initiative between the STM Association and the National Information Standards Organization. It aims to "optimize protocols across key stakeholder groups, with the goal of facilitating a seamless user experience for consumers of scientific communication" (see reference [9]).

Youngen said that the problem statement is as follows: Access to STM content and resources is traditionally managed via IP address recognition and for the past 20 years, this has provided seamless access for users on campus. However, with modern expectations of the consumer web, this approach has become increasingly problematic. Users want seamless access from any device and from any location, and they are increasingly starting their searches on third party sites such as Google and PubMed rather than on publisher platforms or library portals. As a result, they run into access barriers. A patchwork of solutions exists to provide off-campus access: proxy servers, such as VPNs and Shibboleth; however, the user experience is inconsistent and confusing. The lack of user data also impedes the development of more user-focused, personalized services by resource providers. Publishers are facing an increasing volume of illegal downloads and piracy, and fraud is difficult to track and trace because of insufficient information about the end user.

In addition, the use of IP addresses also poses a significant risk to campus information security as a significant black market exists for the sale of compromised university credentials, that are typically used to access university VPN or proxy servers. When fraudulent activity is detected, a publisher may block the IP address, which then may impact an entire campus. Compromised credentials imply that a university's student/faculty data is at risk. Youngen said that these issues clearly indicate that it is time to move beyond IP-recognition as the main authentication system for scholarly content while making sure the alternative is as barrier-free as possible.

The RA21 Draft Principles are as follows:

1. The user experience for researchers will be as seamless as possible, intuitive and consistent across varied systems, and meet evolving expectations.
2. The solution will work effectively regardless of the researcher's starting point, physical location, and preferred device.
3. The solution will be consistent with emerging privacy regulations, will avoid requiring researchers to create yet another ID, and will achieve an optimal balance between security and usability.

4. The system will achieve end-to-end traceability, providing a robust, widely-adopted mechanism for detecting fraud that occurs at institutions, vendor systems, and publishing platforms.
5. The customer will not be burdened with administrative work or expenses related to implementation and maintenance.
6. The implementation plan should allow for gradual transition and account for different levels of technical and organizational maturity in participating.

He said that the task force will not build a specific technical solution or an industry-wide authentication platform. Rather they will adopt a diverse, inclusive approach and achieve consensus across stakeholder groups; recommend new solutions for access strategies beyond IP recognition practices; explain the standard measures that publishers, libraries, and end-users should undertake for better protocols and security; and test and improve solutions by organizing pilots in a variety of environments for the creation of best practice recommendations.

At the time of his presentation (February 28, 2017), the corporate pilot was underway and the academic pilot was just getting organized. The pilots will run through the third quarter of 2017. In October, the RA21 taskforce will facilitate the sharing of the results and learnings that emerge from the pilots, and conclusions will be used to develop best practices which will then be made publicly available in December.

In closing, Youngen suggested that interested parties go to <http://www.stm-assoc.org/standards-technology/ra21-resource-access-21st-century/> for more information.

Youngen's slides are available on the NFAIS web site.

12. Closing keynote: Open science: Towards reproducible research

The conference closed with final keynote by **Julien Jomier**, CEO, Kitware Europe, who discussed the necessity of integrating the three components of Open Science: Open Data, Open Source, and Open Access.

He began by saying that the ultimate goal of publishing is to disseminate knowledge, but that he has issues with the traditional publishing process. He believes that it is competitive rather than collaborative; that the content is limited (primarily text); that the process from manuscript submission to publication is slow; and that the ability to reproduce results is not being enforced. He then went on to discuss the three overlapping pieces of Open Science: Open Access, Open Data, and Open Source.

Open Access is Online research that is free of all restrictions on access; that is free of many (but not necessarily all) restrictions on use; and that requires a new publishing business model; e.g. some journals may be openly-accessed only after some period of embargo.

Open Data has been enabled by high-speed internet. He noted that the data is heterogeneous (environmental, genomics, 3D, etc.); that it includes access to all data in an experiment – input, intermediate, and final results; and the data sets can be massive. He provided a couple of large data set examples such as the *Visible Human Project* [1] initiated by the National Library of Medicine (NLM) in the USA and the *Give a Scan* [5] project initiated by the Lung Cancer Alliance, also in the USA.

Jomier then went on to talk about Open Source which he says goes back to 1985 and the establishment of the Free Software Foundation (FSF), a non-profit organization with the worldwide mission of promoting the freedom of computer users.¹³ The FSF was reinforced by the launch of the Open Source

¹³ See the Freedom Software Foundation at: <http://www.fsf.org/about/>, Cited July 3, 2017.

Initiative in 1998. The “open source” label was created at a strategy session held on February 3, 1998 in Palo Alto, California, shortly after the announcement of the release of the Netscape source code.¹⁴ This initiative offers a variety of usage licenses; e.g., BSD, GPL, LGPL, etc. and has a well-known infrastructure: iPython notebooks, Github, etc.

Jomier went on to discuss the Open Source values: security, affordability, transparency, perpetuity, interoperability, flexibility, and localization. He noted that a well-known open source project, The Insight Toolkit (ITK) [3,6] was initiated in 2000 by the National Library of Medicine (NLM) in order to “standardize” the implementation and use of image processing in the medical field. The project has been a success and is currently used by academia and industry around the globe. Other examples were also provided.

He noted that the Open Science movement presents advantages as well as limitations. He believes that it is helping scientists in several ways, one of which is that scientists can build upon previous experiments, datasets, and software without starting from scratch. He also noted that Open Science needs to be improved. For instance, the infrastructure required to share and deploy datasets and software is not free and usually is built and financially-supported by large organizations and governments.

Jomier also mentioned two actions in recent years that have pushed for even more openness. The first was a journal initiated in 2015 to encourage scientists to publish negative results and the second deals with actually publishing the replication of previously published work, but the lack of “novelty” with regard to latter is inhibiting its adoption by publishers.

In closing, he said that he hopes that in the near future publishers update the current infrastructure to support reproducibility and that scientists and publishers collaborate to improve the ways in which scientific findings are published.

Jomier’s slides are available on the NFAIS website and an article based upon his presentation appears elsewhere in this issue of *Information Services and Use*.

13. Conclusion

The speakers at the conference reinforced one another in the identification of a number of industry trends and issues, with collaboration being the most redundant topic. From the opening keynote that discussed crowd-sourcing and sharing of research information to treat rare diseases, to the Miles Conrad Lecture that spoke of collaboration across market sectors, to John Wilbanks’ presentation on open-standards research, through to the closing keynote on Open Science – collaboration, and its corollary of information sharing – was highlighted as the key factor in the advancement of knowledge both in the sciences and in the humanities. Related issues such as the reproducibility of research results, the preservation of data for future use and re-use (and the implications thereof with regard to data management and infrastructure), and the continued overwhelming growth in information made this an often thought-provoking meeting. I will think about Subhajt Basu’s presentation on the balance act that is often needed when using data. As I said earlier, he pointed out that fact that data-driven innovation poses challenges related to governance and policy as well as challenges related to public understanding and public trust. It also raises questions about privacy, consent, data ownership, and transparency. So I close with one of his questions. “What is the role that technology can play to ensure that data-driven innovation advances in an optimal way. . .?”

¹⁴ See the Open Source Initiative at: <https://opensource.org/history>, Cited July 3, 2017.

Plan on attending the 2018 NFAIS Annual Conference that will take place in Alexandria, VA, USA from February 28–March 2, 2018. Watch for details on the NFAIS website at: <http://www.nfaais.org/>.

Note: If permission was given to post them, speaker slides used during the NFAIS 2017 Conference are embedded within the conference program at: <http://www.nfaais.org/2017-conference-program>. The term “slides” is highlighted in blue.

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 Clarivate Analytics) 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).

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 almost sixty years has provided a forum in which to address common interests through education and advocacy.

References

- [1] R.A. Banvard, The visible human project © image data set from inception to completion and beyond, in: *Proceedings CODATA 2002: Frontiers of Scientific and Technical Data, Track i-D-2*, Medical and Health Data, Montréal, Canada, 2002, see also https://www.nlm.nih.gov/research/visible/visible_human.html, cited July 3, 2017.
- [2] K. Beck et al., Manifesto for Agile Software Development, 2001, <http://www.agilemanifesto.org/>, cited July 3, 2017.
- [3] H.J. Johnson, M. McCormick and L. Ibanez, *The ITK Software Guide: Design and Functionality*, 4th edn, Kitware, 2015.
- [4] N. Lomas, DeepMind Health inks another 5-year NHS app deal in face of ongoing controversy, *TechCrunch*, June 22, 2017, see <https://techcrunch.com/2017/06/22/deepmind-health-inks-another-5-year-nhs-app-deal-in-face-of-ongoing-controversy/>, cited July 3, 2017.
- [5] Lung Cancer Alliance, Give A Scan © – The first patient-powered open access database for lung cancer research, <http://www.giveascan.org/>, cited July 3, 2017.
- [6] M. McCormick, X. Liu, J. Jomier, C. Marion and L. Ibanez, ITK: Enabling reproducible research and open science, *Frontiers in Neuroinformatics* **8** (2014), 13.
- [7] L. Teytelman, Protocols.io: Reducing the knowledge that perishes because we do not publish it, *Information Services and Use* **35**(1–2) (2015), 109, see <http://content.iospress.com/articles/information-services-and-use/isu769>.
- [8] M. Ware and M. Mabe, *The STM Report*, 4th edn, 2015, p. 6, retrieved on April 30, 2017 from http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf.
- [9] A. Winter, The Short History of Napster 1.0, *Wired*, April, 4, 2013, cited June 27, 2017, available at: <https://www.wired.com/2013/04/napster/>, cited July 3, 2017.