An overview of the 2020 NISO plus inaugural annual conference: *A grand experiment*

Bonnie Lawlor*

Guest Editor, NFAIS Honorary Fellow, 276 Upper Gulph Road, Radnor, PA, USA

Abstract. This paper offers an overview of the highlights of the 2020 NISO Plus inaugural conference that was held in Baltimore, MD from February 23–February 25, 2020. This conference replaced what would have been the 62nd Annual NFAIS conference. However, NISO and NFAIS merged in June 2019, resulting in the conference being renamed NISO Plus and taking on a new format. The goal was to continue some of the best traditions of past NFAIS conferences while building in time for discussions. With two and a half days of networking and education on tap, attendees had the opportunity to learn about emerging and exciting areas of change and development such as Artificial Intelligence and Machine Learning and new content types such as Augmented Reality/Virtual Reality. There were very practical sessions that focused on the status of standards, current issues and problems, with the goal of working towards innovative solutions and developing plans for moving forward.

Keywords: Metadata, information discovery, open access, open source, artificial intelligence, scholarly communication, data integrity, electronic resource management systems, standards, augmented reality, virtual reality, digital preservation, seamless access, JATS, machine learning, repositories, KBART

1. Introduction

At the 61st Annual Conference held by the National Federation of Information Services (NFAIS) in February 2019 it was announced [1] that NFAIS would possibly be merged into NISO, pending the membership approval of each organization. This approval was attained and the merger became official on June 30, 2019 [2]. Since NFAIS had a long history of annual conferences, it was agreed that the tradition would be continued, but that the both the name and the format would change. The conference would combine the thought-leadership of NFAIS with the hands-on practicality of NISO - hence the new conference name, NISO Plus, and in his opening remarks, Todd Carpenter, NISO Executive Director, labeled the 2020 inaugural conference as a "Grand Experiment". He said that the goal was not to have a "sage on the stage" talking to the audience, but rather to have a dialogue among experts - both those on the stage and in the audience. Since the estimated two hundred and fifty meeting attendees represented all of the major stakeholders in the global information community - publishers, librarians, service providers, technologists, etc. from across all market segments, Todd encouraged everyone to contribute to the in-depth discussions (and they did!).

---

*E-mail: chescot@aol.com.

The conference was multi-track, with four parallel sessions on diverse topics running at the same time. Attendees were encouraged to participate in those of most interest to them and they could move among sessions at will.

Todd noted that throughout its history NISO standards have made an impact, e.g., the machine-readable cataloging standard (MARC), the international standard serial number (ISSN), etc., and that he hoped that the net outcome of the conference would be yet another major initiative. The measure of success for the meeting was neither to be the number of sponsors nor the fact that it was sold out. The measure of success would be the generation of creative ideas that would eventually be practical solutions to the shared challenges faced by all members of the information community.

I can attest that at least for the sessions that I attended, the discussions were lively and did generate ideas. Be forewarned - although I am pretty good at multi-tasking, even I cannot attend two meetings simultaneously. This overview does not cover all of the sessions, but will provide a glimpse of what transpired and perhaps motivate you to attend next year's meeting. However, a summary of the ideas that were generated by *all* of the sessions was published in July and is available on the 2020 NISO Plus website [3]. It is also reproduced with permission in this issue of *Information Services and Use.*

2. Opening keynote

The opening keynote presentation was given by Dr. Amy Brand, Director of the MIT Press and co-founder of the MIT Knowledge Futures Group. She spoke about the importance of diversified distributed information infrastructures and how seemingly divergent stakeholders and models ideally interoperate in what she termed our print-optional, post-truth world. She asked the following rhetorical questions: (1) how do we avoid creating monocultures within our research and publishing ecosystems? And (2) how can we better anticipate the future consequences of the business models and policies that we adopt? She noted that we have made a great deal of progress on open access and open data, but added that "open" for its own sake is insufficient for a thriving knowledge ecosystem. We have yet to think through the longer-term consequences of an all-open world and the struggle for control over information and knowledge looms large everywhere that we turn.

Brand said that when she talks about information infrastructure she is referring to the underlying systems, standards, and practices; i.e., licenses, business models, the internet, even the cloud. The infrastructure is what connects research labs, data, computers, and people with the ultimate goal of creating and sharing knowledge. It is the technologies that are driving the transformation of knowledge and the tools in our research communication ecosystem. Our infrastructures and how we use them to build knowledge will further shape our futures. She added that systems and networks vary according to how centralized or decentralized they are, both in terms of architecture and in terms of ownership and control. She stressed that owning technology is a form of control and said that she is not alone in believing that institutional leaders today need to explore the implications of commercial control of research data, analytics, and infrastructure, along with the potential for community-owned alternatives. She recommended reading a recent SPARC white paper that makes related points [4], and added that while commercial investment and ultimate ownership is one way that innovative start-up solutions (e.g. Authorea, Figshare, Altmetric, etc.) in publishing become sustainable, and perhaps even profitable, there is a lot going on in the nonprofit and open source tool space. She noted that MIT Press issued a Mellon-funded report on the estimated eighty open source publishing tools in existence at that time, and said that there are even more available today [5].

Brand went on to provide some examples of these tools; e.g., the Center for Open Science (see: https://www.cos.io/) that has developed a variety of software tools, workflows, and data storage solutions called the Open Science Framework. They also promote the use of open science badges, such as for open data materials and protocols. Another important infrastructure initiative is COAR, the Confederation of Open Access Repositories, which aims to build a sustainable, inclusive, and trusted global network based on OA repositories (see: https://www.coar-repositories.org/). There is even the Research on Research Institute (RoRI) that was established in September 2019 by the Wellcome Trust, Digital Science, and the Universities of Sheffield and Leiden to organize and support more meta-science research across stakeholder groups (see: http://researchonresearch.org/).

She then went on to discuss what the MIT Press is doing to contribute momentum towards open source advances. She said that while she runs the publishing company that outputs about three hundred books and forty journals a year, she also partners with the academic community to create the knowledge tools and services of the future. She noted that academic book publishing was once a lot more simple. MIT monographs used to sell thousands of copies because libraries used to buy print books. Until recently, they were able to afford their own warehouse and distribution services in partnership with two other universities in the Northeast. But beginning July 1, 2020 the MIT Press will be distributed by Penguin Random House because an efficient supply chain in the Amazon era means scale and interoperability at a whole new level. She added that business models are formative structure, too, and MIT Press is actively experimenting there around open access. They just received a grant from the Arcadia Fund to develop a durable financial framework for OA monographs.

Brand said that their work on new models extends into technology as well. They have a Knowledge Futures Group (KFG) that has a twofold goal: (1) to incubate homegrown solutions; and (2) to spark a movement towards greater institutional investment and knowledge infrastructure. One core service in the KFG today is the PubPub platform (see: https://www.pubpub.org/), a turnkey open publishing solution with collaborative editing, rich media, annotation, and versioning. It is intended to support publishing as a community-driven activity. They are also establishing the MIT Open Publishing Services (MITOPS), a new operating unit that supports community publishing at MIT and beyond.

Again she reinforced that the concept of "Open" is not enough and that the research community must be alert to potential unintended consequences. They need to invest in alternative solutions and efforts to avoid future monopolies over research content, infrastructure, and analytics. Brand said that she recommended reading a forthcoming article in the Knowledge Futures publication, "The Common Place". It is by Sarah Kember, a professor of new technologies of communication at Goldsmiths College and director of Goldsmiths Press, who says that "Commercial platforms represent the next phase in the capitalization of knowledge, and tend towards replacing old monopolies for new - the giants in the commercial journal publishing world with the tech giants such as Amazon and Google [6]."

Brand closed by saying that she is confident that the best solutions for a sustainable, secure knowledge future will arrive through multi-stakeholder coordination, shared infrastructure, and open standards.

A video of Dr. Brand's presentation, along with her slides and a transcript are available on the 2020 NISO Plus website. An article related to her presentation appears elsewhere in this issue of *Information Services and Use*.

3. Electronic resource management systems

The first session that I attended after the opening keynote focused on emerging trends and opportunities for vendor/publisher partnerships with regard to Electronic Resource Management (ERM) systems. The first speaker was Lola Estelle, Digital Library Specialist, SPIE (see: https://spie.org/). Her key message was that libraries invest a great deal of money and time into their ERM systems, and that the end user experience is very much dependent upon publisher data being correctly represented within them. Estelle laid out the initial problem and showed the ERM life cycle from a librarian's perspective - acquire, provide access, administer, provide support, and evaluate/monitor the system. She highlighted the requirements of a library system that is to be truly user-centric [7]:

- Systems, like the libraries to which they provide services, must be completely re-architected to *center on the user*.
- Systems must be completely re-architected to enable *the facilitated collection*.
- Library systems must be completely re-architected to *integrate effectively* on a service and data layer with other systems that enable research, teaching, and learning.
- Library systems must be completely re-architected to provide modern *business intelligence* capabilities for individual libraries as well as their consortia.

She stressed the importance of KBART ("Knowledge Bases and Related Tools"), a NISO Recommended Practice that facilitates the transfer of holdings metadata from content providers to knowledge base suppliers and libraries. Knowledge bases are widely-used to support library link resolvers and electronic resource management systems (see: https://www.niso.org/standards-committees/kbart). A key problem is that librarians often do not know which package they need to link to their system and therefore often do not always "turn on" all of the files for which they have paid. She noted that packages appear differently across discovery systems and that platform migration can cause problems. Selecting and enabling the correct package within an ERM, discovery layer, or link resolver is difficult because there are multiple similarly-named packages. Librarians seek guidance from vendors, but vendors themselves often do not know which package to enable and may not have a way to see the selection screens in ERMs and related systems. She said that content providers who wish to provide guidance around knowledge base content selection must ensure that their KBART files are properly named and that they provide sufficient documentation and training for librarians. Estelle provided several examples of good and bad naming practices and briefly discussed the NISO Open Discovery Initiative (ODI), a technical recommendation for the exchange of data, including data formats, method of delivery, usage reporting, frequency of updates and rights of use. It offers a way for libraries to assess content providers' participation in discovery services and provides a model by which content providers can work with discovery service vendors via fair and unbiased indexing and linking. The recommendation was officially published in June 2020 and is available on the NISO website [8].

The second speaker in this session was Peter McCracken, Electronic Resources Librarian at Cornell University. For those who think the name is familiar, McCracken was one of the co-founders of Serials Solutions (1999) and has since launched ShipIndex.org, an online research database focused on maritime and vessel history which he discussed at the 2019 NFAIS Annual Conference [9]. McCracken provided an overview of an Electronic Resource Management System being built under the FOLIO umbrella. FOLIO is an open source project that "aims to reimagine library software through a unique collaboration of libraries, developers and vendors. It moves beyond the traditional library management system to a new paradigm,

where apps are built on an open platform, providing libraries more choice and delivering new services to users," (see: https://www.folio.org/).

He said that the system is truly being designed by librarians and that there are special interest groups (SIGs) that provide constant feedback to the product owners and developers. He noted that anyone can participate in the development and/or view recordings of the SIG meetings. Cornell University is an active participant and they went live with the system in October 2019. At the time of the conference they were populating the system and expect full implementation in July 2021. He noted that the initiative is extremely important to him because it is a true paradigm shift in library software development. There is continual interaction across libraries, vendors, developers, product managers, etc. There is a true community committed to the success of the project. The community is now large enough to survive if some choose to withdraw their support or participation and it is continually growing - anyone can implement the system. And, most importantly, it is a financially responsible and viable ERM solution.

After the two presentations there was a discussion around the KBART Recommendations and any gaps that exist that exacerbate problems. Some gaps mentioned were lack of uniformity of names, not all content types are covered, and KBART does not handle e-books very well. It was noted that KBART is a perfect example of where vendors, publishers, and librarians need to work more closely together and that a conversation is needed because not all subject disciplines are involved.

The slides from this session along with notes are available in the repository on the 2020 NISO Plus website.

4. Artificial intelligence and machine learning

With more and more information providers exploring the implications of Artificial Intelligence (AI) and Machine Learning (ML) for their businesses and services, this session offered an overview of the issues that need to be considered in order to ensure that those new to these fields make the right decisions for their future.

4.1. Artificial intelligence 101

The first speaker was Jason Chabak, Vice President for Business Development at Yewno (see: https://www.yewno.com/) who provided a basic introduction to AI and ML. He said that more data has been created in the past two years than in the entire prior history of the human race. He added that by 2020, about 1.7 megabytes of new information will be created every second for every human being on the planet and that by 2025, our accumulated digital universe of data will be about 160 zettabytes, or 160 trillion gigabytes [10]. Chabak added that information does not equate to knowledge and that it would take more than three trillion years for a team of ten thousand analysts to read all of the unstructured information that is currently available. He went on to say that Artificial Intelligence stands in a long line of human innovation in the creation, recording, storing, codifying, and dissemination of cultural knowledge - from the first written language in 3500 BC, to Egyptian papyrus rolls in 500BC, to the Library of Alexandria in the 300 BCs, all the way through to Natural Language Processing in the 1950s, the Web in 1989, etc.

In doing some research for another article I found a concise one-sentence definition of AI which is as follows: "Artificial Intelligence is the science and engineering of making computers behave in ways that, until recently, we thought required human intelligence [11]". The article from which that definition was extracted went on to define "Machine Learning" as "the study of computer algorithms that allow computer

programs to automatically improve through experience [12]". ML is one of the tools with which AI can be achieved and the article goes on to explain why the two terms are so often incorrectly interchanged (ML and AI are *not* the same. ML, as well as Deep Learning (DL), are subsets under the overarching concept of AI.

Chabak went on to describe where we are in the pursuit of AI. He said that we have successfully replicated three hundred and two neurons, but since the human brain contains one hundred billion neurons, to have true AI today's computing capacity will need to be exponentially increased - so we still have a long way to go. He chose to continue his presentation discussing Machine Learning and "narrow" AI that focusses on a single task such as a self-driving car.

He said that as of today AI, ML, and DL all have shortcomings; e.g., they cannot offer 100% accuracy; they are not 100% unbiased because humans write the algorithms; they require human intervention; they can have adverse effects; they cannot scale due to constraints of the current technological infrastructure; and they cannot take all of our jobs nor take over the world…yet. However, narrow AI and ML do have a number of pluses; e.g., the have the ability to perform repetitive, mundane tasks - and improve over time; they can provide intelligent augmentation of current workflows; they can reduce (not eliminate) errors; and they can ingest and analyze orders of magnitude more data than any human - or group of humans - can do. He stressed that the quality of the data that emerges from AI-based studies is only as good as the quality of the data that is used in those studies and that the human element is of the utmost importance.

4.2. AI/ML from the publisher perspective

The second speaker was Brian Cody, CEO and Co-founder of Scholastica (see: https:// scholasticahq.com/). He said that the appeal of AI and ML from a scholarly publishing perspective is that they: (1) provide efficiency at scale; (2) can lower production costs; (3) can create higher-quality output; and, (4) can provide a competitive advantage. Some examples of the use of ML in publishing are (1) natural language processing and (2) use in the peer review process to identify likely reviewers and to identify manuscripts that are the most-likely to be published so that they can be expedited through the process. He said that smaller publishers have concerns about AI and ML. They question how much they can trust their use; they fear a loss of control over decision-making and a loss of personal relationships; they are unsure of potential consequences such as their use becoming an unintended administrative burden; and they are concerned about the time costs.

Cody said that he has his own concerns about the use of AI and ML. Predictive analytics are only as good as the data that is used. Since history is used to predict the future, the predictions are only as good as the data set that is used and there is room here for human bias (Chabak made the same point and it was stressed again by the conference's closing keynote speaker, danah boyd). Also, expertise is required to interpret the results and decision-making by humans needs to be preserved - not handed-off to AI and ML for convenience. There are ethical considerations that need to be balanced with efficiency and convenience. In closing, Cody suggested that attendees take a look at Amazon Rekognition, (it uses ML to automate video and image analysis) just to see what ML can do (see: https://aws.amazon.com/rekognition/?blog-cards.sort-by=item.additionalFields.createdDate&blog-cards.sort-order=desc). If you are interested in the use of AI and Ml in scholarly communication, I recommend that you take a look at some of the articles from the NFAIS conference on just that topic that were published in *Information Services and Use* [13].

4.3. Quality datasets are key to quality AI

The final speaker in this session was Huajin Wang, a Librarian at Carnegie Mellon University. As already mentioned by the previous two speakers, Wang stressed that the essential ingredient for successful AI and ML is accessible, high-quality labeled data. She noted that high quality data is hard to find, but that there are technical solutions that can help: (1) the use of search engines to find existing data; (2) automation in evaluating data quality and integrating datasets; (3) automation in data curation; and (4) model transfer and data augmentation. Wang said that the reason that datasets are hard to find is primarily because an overall discovery layer is missing and the datasets are distributed. Repositories contain primarily structured data, while the Web is ninety-nine percent unstructured data. Other examples of unstructured data include publications and sets of images. She added that structured data is important as it makes it easier for search engines to find it [14]. This fact was reinforced by a later speaker, Carly Robinson from the Department of Energy, who noted that DOE just recently implemented Google's structured data guidelines in anticipation of Google's new service, Google Dataset Search, which was launched in January 2020 [15].

Wang said that if you are looking for structured datasets you can: (1) use a search engine powered by AI; (2) perform a simple keyword search for datasets across the web; (3) do searches over embedded metadata; and (4) perform searches over metadata from data providers. Wang referred attendees to Schema.org, a collaborative, community with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond (see: https://schema.org/).

Locating unstructured datasets first requires that the datasets have already undergone metadata tagging and data linking. She said that the human factor - sound data stewardship - is key for data discovery and reuse - it all starts at the beginning. There must be responsible data collection and documentation; Best Practices for data management must be put in place and followed consistently; and high-quality metadata must be created and data standards followed. Wang added that ideally we should follow open science practices such as sharing data, code, and workflows; following the FAIR principles (data should be findable, accessible, interoperable, and reusable); building easy-to-use and robust tools for data sharing and reuse; and participating in interdisciplinary collaborations. Researchers, tool developers, data curators, repositories, and other institutions all need to work together to build a healthy data ecosystem.

In closing, Wang referred attendees to the Artificial Intelligence Data Discovery and Reuse Symposium (AIDR) that was to be held in May. It will now be a virtual meeting scheduled to take place October 20, 2020 (see: https://events.library.cmu.edu/aidr2020/).

After Wang's presentation there was a general discussion on AI and ML and an issue emerged of which I was unaware. As a chemist I am well aware of the practice of Green Chemistry and the ideals of sustaining our planet, but I really never heard of Green Computing - with the goals of reducing the use of hazardous materials, maximizing energy efficiency during the product's lifetime, etc. [16] What followed was an interesting discussion of the energy costs related to AI. After the conference NISO, posted a blog that reported on the carbon footprint of AI processing that said that "the process can emit more than 626,000 pounds of carbon dioxide equivalent - nearly five times the lifetime emissions of the average American car (and that includes the manufacturing of the car itself) [17]". NISO also posted a link to a chapter on Green Computing Algorithms that is technical, but worth a read if you are interested [18].

Note that the slides used by Chabak and Wang are available on the 2020 NISO Plus website.

5. Seamless access

Tim Loyd, CEO of LibLynx, an organization that provides an API-based solution for managing user authentication and authorization (see: https://www.liblynx.com/), gave a presentation on the "seamless access" initiative - what it is and its current status. He said that library use of IP [19] recognition was developed when off-site access to electronic resources was in its infancy and little has changed in the intervening years. Today, after twenty years of IP authentication, there are now better alternatives.

Lloyd said that remote access to content needs to be improved. With IP authentication researchers are forced to start from, or at some point circle back through, the library's web site to find a proxy-prefixed URL, and this is extra work that simply deters users. He said that is not how researchers work and we need to aim for delivery at the point of discovery. Also, the usability of access workflows needs to be improved. Current issues include the numerous clicks to reach content behind an authentication barrier, and the numerous user credentials that are scattered over a multitude of platforms. Users face multiple, inconsistent access experiences and can quickly feel confused and overwhelmed. In fact, access is currently so complicated that even fully-entitled end-users are turning to questionable alternative resources, such as SciHub [20], ResearchGate [21], etc.

Lloyd briefly recapped the work that has been done to date on this problem, specifically the Resource Access in the 21st Century project (or RA21) that was initiated in 2016, initially to explore the challenge of remote access. It involved stakeholders from the publishing, library, software, and identity communities, and took input from sixty organizations over a three-year period. RA21's conclusions were published as a draft NISO Recommended Practice in April 2019. The draft received more than two hundred comments that helped identify further areas for investigation and confirmed the value of testing a beta service. A final NISO Recommended Practice was published in June 2019 [22].

SeamlessAccess (SA) was created in July 2019 as a community-driven effort to enable seamless access to information resources, scholarly collaboration tools, and shared research infrastructure. Founding members include NISO, GEANT, Internet2, ORCID, and The International Association of STM Publishers. To summarize, the RA21 project developed and piloted ideas from the end of 2016 until last June, when it wrapped up. SA is now in the process of testing these ideas in the light of community feedback and developing best practices around the use of federated authentication. It is still in the beta stage and they are currently working on issues such as the terms & conditions for service providers; the user consent workflow; access to Identity Provider choices prior to authentication; personal data; and feature requests such as customization.

Loyd's slides are available on the 2020 NISO Plus website and a detailed paper based upon his presentation is included elsewhere in this issue of *Information Services and Use*.

6. Linked data in the library and publishing ecosystem

The two speakers in this joint "conversation" session were John Chapman, Senior Product Manager, Metadata Strategy and Operations, at OCLC, and Philip Schreur, Associate University Librarian for Technical and Access Services at Stanford University. Chapman helps direct the linked data strategy for OCLC and is the manager of a two-year data infrastructure grant from the Andrew W. Mellon Foundation [23]. Schreur is most interested in the transition of traditional Technical Services workflows from MARC-based to linked data-based counterparts. He believes that we will be living in a hybrid environment (MARC/linked data) for quite some time and that we will need to carefully assess which

functions are best retained in MARC and which are best approached as linked data. This session was truly a conversation between the attendees and the panelists. There was no defined structure for the session - it was just introductory comments followed by Q&A.

So what is Linked Data? According to Wikipedia, Linked Data "is structured data which is interlinked with other data so it becomes more useful through semantic queries. It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages only for human readers, it extends them to share information in a way that can be read automatically by computers. Part of the vision of linked data is for the Internet to become a global database [24]". On their website, OCLC has a brief, visual introduction to what linked data actually is. The introduction is labeled "Getting Started with Linked Data [25]".

The Library of Congress (LC) has been working in this area with its Bibliographic Framework Initiative (BIBFRAME, see: https://www.loc.gov/bibframe/). A comment was made by an attendee that a common language is needed to link things and that BIBFRAME is key. They added that the more people use the same ontology and the same format, the better the links and that while BIBFRAME remains LC's baby, it should become an open standard released to the world.

Also briefly mentioned in the conversation was the Linked Data for Libraries (LD4L) project with which Schreur is very much involved. This is a collaboration of Cornell University Library, Harvard Library Innovation Lab, Stanford University Libraries, and library researchers at the University of Iowa, originally funded by a nearly $1.5 million dollar two-year grant from the Mellon Foundation (see: https://ld4l.org/).

In closing, Chapman and Schreur said that a lot of the value of participating in Linked Data initiatives comes from the discipline of structuring data in ways that it can be easily processed by machines (echoing Wang's earlier comments). Thinking this way leads to a new perspective on the critical importance of consistency, collaboration with partners, and participation on the web.

The speakers did not use slides, However, John Chapman kindly submitted an article based upon the session and it appears elsewhere in this issue of *Information Services and Use*.

7. Metadata

This, too, was a conversation session and the speakers were Patricia Feeney, Head of Metadata at Crossref, Jonathan Ponder, Metadata Librarian Manager at Ithaka (JSTOR), and Laura Paglione, Project Upholder, Metadata 2020. They all agreed that good metadata is essential to the information infrastructure - from modern languages and medicine to music and movies! Metadata is data about data; i.e., it is data that is used to describe another item's content. It can enhance the scope, speed and clarity with which scholarly communities can curate their outputs for optimal discovery and reuse. All agreed that there needs to be metadata standards and JATS was the standard that was most frequently mentioned (see more on this in a later section in this article). Also briefly discussed was the Metadata 2020 initiative. It is a collaboration of publishers, librarians, researchers, funders, and technologists that advocates richer, connected, and reusable, open metadata for all research outputs, which they believe will advance scholarly pursuits for the benefit of society (see: http://www.metadata2020.org/).

Laura Paglione, a member of the core team leading Metadata 2020, kindly wrote a paper that takes this session and springboards some of the points that were made to highlight the importance of metadata within the context of global pandemics such as Covid-19. The article appears elsewhere in this issue of *Information Services and Use.*

The speakers in this session did not use slides.

8. Preservation of digital data

This session had two speakers, Leslie Johnston, Director of Digital Preservation, National Archives and Records Administration (NARA), and Cliff Anderson, Associate University Librarian for Research and Digital Strategy, Vanderbilt University. Johnston opened the session and noted that NARA is an entity unto itself and that she would not focus on her institution's specific issues, but rather would focus on the more general issues in digital preservation with regards to both born-digital and digitized items. She noted that both types are just files and both require the same amount of description and care.

8.1. Key challenges

Johnston said that, in general, the key challenges for the archiving and preservation of born-digital research and scholarship, and digital collections are as follows: heterogeneity, technology, complexity, scale, and serving multiple communities and purposes, including ourselves.

Heterogeneity: the issues here are that: (1) research within and across disciplines utilize very different methodologies, equipment, software, and hardware. Outputs range from publications to websites, A/V, textual and numeric datasets, and software needed to process the results. This also applies to electronic records and general digital collections; (2) there are literally thousands of variant versions of file formats over time, and they just keep changing. And we cannot identify every legacy format with certainty; and (3) there are dozens of carrier formats - floppy disks, hard drives, CDs, DVDs, thumb drives, tapes, etc. - and we need to be able to read the files off them in order to preserve them.

Technology: The issues here are that: (1) with heterogeneity comes a wide variety of ever-changing tools and workflows needed to process, describe, preserve, and provide access to born-digital scholarly research; (2) storage can become a concern when you consider scale and the need for preservation replication; (3) with scale also comes stress on local networks and the limiters of moving files using web protocols; and (4) machines used to process born-digital materials will require increasingly more storage and memory and higher bandwidth network connections.

Complexity: The issues here are: (1) digital materials do not exist without a context and a provenance which must be recorded and maintained; and (2) scholarly output and electronic records are increasingly complex, comprised of multiple or multi-part or containerized files that require all their components, have relationships to other files, or are bundled with software that is necessary for research to be reusable and replicable.

Scale: The issues here are: (1) there are thousands of researchers, students, and prominent individuals associated with any university and its community whose files will be collected by universities or other cultural heritage institutions over time; (2) there is a massive amount of observational data and research datasets that are created in scientific research. Often research data preservation policies require that the organizations with which the researchers are affiliated must potentially retain and preserve such data; and (3) some types of collections - audio, video, film, email - produce both huge files and huge numbers of files to preserve. And finally,

Serving multiple communities and purposes, including ourselves: the issues here are: (1) if it's not accessible, we have not preserved it; (2) it's not just about the files and the technology, it's about people. There is no single community of creators, nor of users, and new communities will emerge; (3) as with all our collections, we will never know all the uses that our digital files will serve for research or the

public; and (4) we will need to change our own organizations to meet the needs of our collections and our communities.

Johnston then went on to discuss some of the successful strategies that should be part of every digital preservation program. These are guidance for content creators; ongoing risk assessment; the prioritization of basic levels of control; a scalable and flexible infrastructure; and collaborations and partnerships. She then expanded on each strategy.

Guidance for content creators: Remember that: (1) the digital preservation life cycle starts with the people creating the files, not when the files come over the transom to libraries and archives.; (2) there is no such thing as the ability to completely enforce what is created or what is collected, because the work requires whatever the appropriate tools or formats are. But guidance on data management strategies, appropriate storage criteria, preferred and acceptable formats, and minimum metadata make long-term preservation more likely; and (3) examples include Research Data Plans, Format Statements, the Federal Agencies Digital Guidelines Initiative (FADGI - see: http://www.digitizationguidelines.gov/), the National Digital Stewardship Alliance (NDSA) Preservation Storage Criteria and Levels of Digital Preservation (see: https://ndsa.org/publications/levels-of-digital-preservation/), etc.

Ongoing risk assessment: You need to: (1) identify and document the format risks and risk triggers associated with the digital materials, and make feasible plans for taking preservation actions, such as storage and format migration; (2) identify "essential characteristics" AKA "significant properties" for different types of files that provide testable success metrics for content fidelity in format migrations; and (3) remember that the goal is always to preserve the content of the files. Persevering the full look and feel and user interactions is just not always possible, and that's OK.

Prioritization of basic levels of control: Note that: (1) it is deceptively simple to say that an organization has to know what it has, where it is, and who it belongs to when it comes to the preservation of research output, but that's the place to start; (2) the priority should be getting files from wherever they are into a single managed environment if possible - hopefully a single preservation repository; and (3) if that is not possible, document the location, level of risk, and who has the responsibility for management and preservation.

Scalable and flexible infrastructure: Remember that: (1) the Cloud can provide geographical distribution and replication, and is generally easier to scale for processing and storage than on premise data centers; (2) Machine Learning applications can assist with processing and description. But be aware that training ML systems is a non-trivial effort; and (3) Back-ups are not archives nor are they preservation. Have a disaster preparedness plan for your infrastructure and systems of record and a preservation repository and test those systems for recovery on a regular basis. And finally,

Collaboration and partnerships: Be aware that: 1) there is a growing community that can provide resources for planning and executing digital preservation programs, share best practices, share access to equipment, and collaborate on shared collection development and preservation projects. Community examples include the NDSA, the Digital Preservation Coalition (DPC - see: https://www.dpconline.org/about), the Digital Curation Centre (DCC - see: https://www.dcc.ac.uk/), etc.; (2) there are services including the Hathi Trust, APTrust, Portico, Ithaka, etc.; and (3) there are dozens of mature, open tools for all aspects of preservation workflows, from BagIt [26] for transfers to BitCurator for processing (see: https://bitcurator.net/), to the DuraSpace systems (see: https://duraspace.org/), for processing and preservation.

In closing, Johnston said to remember that it is not only about technology, but also about people - the communities that drive what we do, both the creators and the users. It is they who create the digital scholarship that we should preserve, who guide us in identifying other digital content to collect, who tell us how and where they discover our collections, and who tell us how they make use of what we collect and preserve. We do not fail if we cannot preserve it all. She cautioned attendees to not try to do it all because no single institution can. She advised to do what you can and that there is no one right way. You need to do what makes sense for *your* organization.

8.2. *TV news preservation*

Cliff Anderson then spoke on the very specific issue of preserving TV news. He gave a brief history of the Vanderbilt Television News Archive that was established in 1968 with the goal of recording and preserving national news programming on the three major networks that existed at the time (ABC, NBC, and CBS). The archive has faced several challenges as it evolved (it now covers representative news from the Fox and CNN cable networks) - most notably financial and legal issues such as, who really "owns" the news? Even today archiving digital news remains financially and legally challenged as the number of news networks increase and privacy laws emerge. Anderson touched on the many ongoing preservation issues that need to be addressed; e.g. the number of copies to be created, the importance of metadata (see earlier discussion about metadata), the technical requirements, what qualifies for preservation, and the ever-present issue of sustainability. But what resonated with me personally was that this preservation initiative was undertaken because it was believed that the perspective of New York-based news directors biased the national discourse about the contentious cultural and political events of the sixties - more than fifty years ago and an issue that is even more obvious today!!! The story goes that Paul Simpson, a Nashville insurance executive, instigated the creation of the Television News Archive after seeing an interview on the evening news in 1967 with Timothy Leary. Members of Congress shared the concern that news directors might be "rigging" television news programs for political reasons. Simpson testified at congressional hearings in 1972 on the possibility that news networks were "staging" news events [27]. But moving from suspicion to evidence required an archive. An absolutely enlightening and fascinating history!

The slides from both speakers are available on the 2020 NISO Plus website and both speakers have written articles based upon their presentations that appear elsewhere in this issue of *Information Services and Use.*

9. More on preservation - ask the experts

This half hour session featured two preservation experts, Stephanie Orphan, Director of Publisher Relations at Portico, and Craig Van Dyck, Executive Director, CLOCKSS Archive. There were no formal presentations or slides. The experts first fielded questions from the moderator, Wendy Queen, Director of Project Muse, who asked about preservation challenges, the role that scholars should play in having their works preserved, the preservation of outputs from thought-leadership conferences such as this one, standards across publishing that create a burden for preservation, etc. In the remaining time they answered questions from the audience.

A brief article based upon this Q&A session appears elsewhere in this issue of *Information Services and Use.*

10. Digital humanities and standards

The speakers in this session were Michelle Urberg, a Metadata Librarian at Proquest (see: https://about.proquest.com/), and Daniel Fisher, Project Director at the National Humanities Alliance (see: https://www.nhalliance.org/). They opened the session with a definition of "Digital Humanities". Urberg defined it as follows:

"A broad-shouldered field of study, defined more by methodologies and approaches to content than by the scope of the content studied. Scholarly outputs are intentionally designed to be "born digital".

While Fisher offered the following definition:

"The humanities encompass disciplines and methods for investigating human decision-making in all cultures and times. This takes place in educational and research institutions and in public through research, interpretation, discussion, curation, etc. Computational tools and methods can open new horizons for this work. These tools and methods also have the capacity to bring the institutional and public spheres together, empowering mutually-beneficial collaboration and communication. *That is the Digital Humanities*".

They agreed that it is both a discipline and a practice and said that the objectives of the session were to: (1) raise awareness about the complexities of metadata creation for digital humanities projects; (2) begin a conversation for supporting the digital project ecosystem - from idea to discoverable and usable product; and (3) reveal potential challenges faced by researchers interacting with ecosystem stakeholders. The stakeholders were defined as researchers, funders, librarians, publishers, and members of the public who engage with the various projects and digital objects.

The importance of metadata was emphasized (yet again!) and they put forth three questions for people to think about when creating metatdata for a project: What types of metadata does your project require? What constitutes "good" metadata for your project? and, What tools are at your disposal for creating metadata? They noted four types of metadata [28]:

- *Descriptive:* "Aboutness" information or description for discovery and identification (e.g. title, abstract, author, keywords).
- *Administrative:* Contextual information about a resource (e.g. rights or acquisition information).
- *Preservation:* Information designed to preserve an object.
- *Structural:* Information linking together compound objects.

They also discussed the essential characteristics of good metadata. It needs to be:

- *Compatible*: Interoperable and machine readable.
- *Complete*: Designed for long-term preservation.
- *Curated*: Information appropriate to the intended users.
- *Credible*: Controlled vocabularies, unique and persistent identifiers.

The above are principles emerging from the Metadata 2020 Initiative and you can find more on the subject in an article by Laura Paglione *et al*. that appears elsewhere in this issue of *Information Services and Use*.

In closing, it was noted that the metadata is part of a larger issue with choosing the right software tools in making a project a reality and that there is information at the New York University web page that highlights the wide variety of tools that offer solutions for building out these projects see: https://guides.nyu.edu/dighum/tools).

If you are interested in Digital Humanities I highly recommend that you take a look at the slides and notes from this session that appear on the 2020 NISO Plus website. Also, a paper by Michelle Urberg based upon this session appears elsewhere in this issue of *Information Services and Use*.

11. Data publishing

The next session that I attended was on data publishing with a focus on the use of Permanent Identifiers (PIDs). The first speaker was Carly Robinson, Assistant Director for Information Products and Services, Office of Scientific and Technical Information (OSTI), U.S. Department of Energy. She began with some background information on OSTI. Its Mission is to advance science and sustain technological creativity by making R&D findings available and useful to DOE researchers and the public. Its core functions are collection, preservation, and dissemination, and it provides persistent identifier services to help with the discovery of research results, the tracking of research impacts, and the linking of research objects. The DOE invests twelve billion dollars in R&D each year resulting in more than fifty thousand scientific and conference papers, theses/dissertations, scientific and technical software, patents, workshop reports, videos, and data sets.

Robinson defined data publishing as "the act of releasing research data in published form for (re)use by others. It is a practice consisting in preparing certain data or data set(s) for public use thus to make them available to everyone to use as they wish [29]". The DOE requires that DOE National Laboratory contractors (researchers) "announce" publicly-available scientific research datasets to OSTI. "Announcement" means providing metadata describing datasets to OSTI to facilitate data discovery, and when datasets are announced, a DOI is assigned to each set. She added that the DOE does not host the data. However, they do require that there is a Data Management Plan (DMP) for making all of the research data open, machine-readable, and digitally- accessible to the public at the time of publication and that the published article should indicate how these data can be accessed. She stressed that the DOE encourages the use of persistent identifiers such as Digital Object Identifiers (DOIs) that facilitate the accurate linking between publications and data products. She added, that in most cases the DOE can provide DOIs free-of-charge for datasets resulting from DOE-funded research through its Office of Scientific and Technical Information (OSTI) DOE Data ID Service.

Robinson said that OSTI.GOV (see: https://www.osti.gov/) is the primary search tool for DOE-funded R&D results. The service includes journal articles, data, software, technical reports, and more. The use of Google Dataset Search can help researchers find what they need. The service was officially out of beta testing as of January 2020, and it indexes repositories that use the schema.org structured data guidelines. To facilitate this, OSTI implemented Google's structured data guidelines (largely schema.org) for datasets, using both JSON-LD and microdata representations (remember what Wang said about data sets earlier!). The structured data was tested and validated using Google structured data tools. All changes were implemented well in advance of the Google Dataset Search release, ensuring that DOE-funded datasets would be discoverable immediately.

Robinson noted that OSTI is a member of Crossref, ORCID, and DataCite, and is exploring a relationship with the Research Organization Registry (ROR). She said that their goal is to create connections throughout the research lifecycle, from grants to researchers to research outputs.

In closing, Robinson said that there is a culture shift in the assignment of persistent identifiers to data and software, and then citing these objects. She noted that there is a need for more persistent identifier connections such as output DOIs, ROR, and grant/award DOIs. They need to make sure that relationships (related identifiers) are captured in the metadata, not just siloed in a separate infrastructure, and that there needs to be a common understanding between organizational roles and relationships across DataCite, Crossref, and Scholix. (Note: Just as an FYI I recently found out that the Research Data Alliance (RDA) has a Working Group looking at the development of a persistent identifier that will let researchers know exactly

what piece of equipment was used in an experiment - see: https://www.rd-alliance.org/groups/persistent-identification-instruments-wg).

12. Dataset citing - the DOI blackhole!

The second speaker in this session was Shelley Stall, Senior Director, Data Leadership, at the American Geophysical Union (AGU). She, too, gave some background on her organization. She noted that the AGU is the largest earth and space science society with sixty thousand members representing one hundred and thirty seven countries. It also covers much more that "geophysics". AGU is the largest Society Publisher in the earth and space sciences with twenty-two peer-reviewed journals that published more than sixty-seven hundred papers in 2019. It does outreach to government leaders and the public and has partnerships with the European Geosciences Union, the Japan Geoscience Union, and others. Stall said that the AGU's position on data is as follows:

> "*Earth and space science data are a world heritage, and an essential part of the science ecosystem. All players in the science ecosystem - researchers, repositories, publishers, funders, institutions, etc. - should work to ensure that relevant scientific evidence is processed, shared, and used ethically, and is available, preserved, documented, and fairly credited*".

She built on Carly Robinson's presentation regarding the creation of datasets and noted that most datasets are not cited in journals. The DOIs of datasets do not make it to Crossref - somehow they get "lost" between when a manuscript is submitted to a journal and when the journal information is pushed to Crossref. She noted that Crossref and DataCite started to collaborate on exchanging links between Crossref DOIs and DataCite DOIs - many of which are links between articles and data. In a recent study it was shown that as of March 2018, there were more than 870,000 links between Crossref DOIs and DataCite DOIs. The majority of these links - more than 850,000 - originated from DataCite DOIs, compared to about 22,000 links originating from Crossref DOIs. The number of data citations that can be found via links between Crossref DOIs and DataCite DOIs was very low - only 3,657 [30]. Another study looked at the data repositories where the cited data are hosted separately for each publisher. The Dryad repository received 4,538 data packages in 2017. Because Dryad *only* hosts datasets that are associated with published articles, this should have led to 4,538 data citations being passed to Crossref in 2017, yet the total of data citations noted in Crossref in March 2018 was only 3,657! [31]

Stall showed examples of AGU articles that included the dataset DOI's, but where the Crossref record for that article did not. She said that the workflow from start to finish is not transparent.

The current flow for an AGU paper is as follows:

Step 1. A paper is submitted to AGU; if accepted, that data citation is validated for accuracy
Step 2. The paper goes to Wiley for preparation
Step 3. The paper goes to SPi Global for format conversion to Crossref XML Schema
Step 4. The paper goes to ATYPON for publishing and the record is pushed to Crossref
Step 5. Crossref receives the record and then pushes the data citation to DataCite
Step 6. DataCite receives that data citation and pushes it to Scholix.

In closing, Stall asked - Where is the Blackhole for dataset DOIs?? And then she let out a call for HELP!

There was, without a doubt, a long and lively discussion at the end of this session (take a look at the notes at http://bit.ly/NISOPlus_Preconference). It was noted that we do not have a culture of citing data nor is there a culture of data re-use in some disciplines. Indeed what is the incentive to re-use data when

"novel" research is what gets rewarded? Also, there is an issue of Trust - is the data valid (would Blockchain technology help here?)? This is definitely an area that requires a conversation among the stakeholders and perhaps NISO can make that happen.

Both Stall's and Robinson's slides are available on the 2020 NISO Plus website.

13. Augmented reality, virtual reality & 3D

The next session that I attended was a joint presentation by Carl Grant, Interim Dean of Libraries, Oklahoma State University, and Chad Mairn, Librarian, Innovation lab, St. Petersburg College. The two speakers provided an overview of new forms of information that have emerged in libraries over the past five years; e.g., vast repositories of 3D objects have emerged that are increasingly coupled with metadata, facilitating their access and subsequent use in traditional forms, including new and sophisticated applications known as Extended Reality (XR) tools [32]. The X is a variable that is used to describe the assortment of immersive technologies that are available today, which includes 360-degree imagery, Virtual Reality (VR), Augmented Reality (AR), Mixed Reality (MR), and 3D content. The speakers noted that with XR, there has been a shift from creation, transmission, and consumption of information to the creation, transmission, and consumption of experiences. These experiences can help enhance learning and foster more engagement throughout the entire research and learning processes. XR has the potential to bring people, places, and shared experiences closer together than ever before. The speakers brought devices with them so that attendees could put them on and actually experience what the speakers were describing.

Carl Grant has spoken at past NFAIS conferences and is a very experienced, innovative librarian. Prior to his current position at OSU he was their Chief Technology Officer and he spoke in 2018 about the changes he made at OSU, including the use of the technologies he and Marin discussed in this session [33]. He noted that as of today, all thirteen colleges within the OSU system are using Virtual Reality (VR). These new forms of "information" are used to provide access to things that are hard to reach or hard to see or that are fragile (fossils) or distant (Syrian ruins). OSU's law school uses VR to allow students to experience what is being said about a crime - could that actually happen as described? Students learn 3D math in 3D. Medical students can experience working with bones and muscles before handling real, live patients. If you want to take a look at Oklahoma State University's mixed reality lab and some of the projects, go to: http://trcf52.okstate.edu/x/index.html. Also, Grant and Mairn recommended that attendees read a recent report published by the Council on Libraries and Information Resources [34] on how these technologies are now being used in academic libraries.

In closing Grant and Mairn both stressed the fact that these new forms of information present both opportunities and challenges and they discuss these in a paper based upon their presentation that appears elsewhere in this issue of *Information Services and Use*.

Grant and Mairn's slides are not available on the 2020 NISO Plus website.

14. Search/retrieval/discovery of information: what does the future look like?

This, too, was a joint presentation given by Christine Stohn, Director of Product Management, Ex Libris (a ProQuest Company), and Alex Humphreys, Director, JSTOR Labs, Ithaka. Stohn said that the presentation would touch on three topics: the diverse resource types that are currently available; glanceability; and search that isn't search.

With regard to the first topic she noted that there has been a steep increase in the amount and diversity of material that is available - a veritable discovery nirvana of patents, journals, books, music scores, videos, etc. There are also more data sources and more types of data sources; for example, there are publishers, institutional repositories, aggregators, Open Access repositories (especially for data), libraries and library catalogs, digital repositories, websites, and physical objects from museums. Also, user expectations are constantly changing and evolving. They are used to receiving recommendations from Amazon and they are used to busy search interfaces. And user perceptions are increasingly influenced by social media and consumer platforms.

Faced with this diversity of information there are challenges to be considered: How do users search and find material beyond the article and the book? Are there different user stories for each resource type? What metrics and parameters beyond peer review and citation count are there? How should we index - what schema is appropriate for the content type (Dublin Core [35] versus the Metadata Object Description Schema (MODS, see: http://www.loc.gov/standards/mods/) versus the Encoded Archival Description (EADS, see: https://loc.gov/ead/), versus home-grown)?

Humphreys noted that a lot of valuable information gets lost. For example, historical documentaries often feature "witness interviews" with the people who had a front-row seat - or even contributed to - the history being documented. Unfortunately, most of these interviews end up on the cutting-room floor in order to create a streamlined and cohesive movie. As a result, a lot of valuable first-person perspective gets lost. He noted that JSTOR has created a prototype Interview Archive in order to preserve these interviews for education and research. The Interview Archive lets you search and browse the full-length interviews that contributed to the documentary, including hours of material not included in the movie itself (see: https://labs.jstor.org/interview/about).

Moving on to "glanceability," Stohn defined the term as the art of making relevance visible and she stressed the importance of tagging content in a consistent manner. She highlighted a service called SciRide Finder that creates context via data mining and can identify cited statements embedded in text [36]. To see a demo go to http://finder.sciride.org/ - it is quite interesting. Humphreys briefly talked about JSTOR's TopicGraph that allows you to search for terms within books and go directly to the pages that discuss the topics that you are researching (see: http://labs.jstor.org/topicgraph/) - also fascinating. He also mentioned that JSTOR and Portico are building a text and data mining platform aimed at teaching and enabling a generation of researchers to text mine. It is tentatively called the "Digital Scholar Workbench" – see: https://tdm-pilot.org/.

For the final topic - search that isn't search - their hypothesis is that in the ocean of data, information, and diverse material, search alone is no longer sufficient and that serendipity is as important as knowing precisely what you are looking for. They said that some ways to find new discovery paths are (1) to follow the citation trail; (2) let others inspire you (e.g. recommender services); and (3) stand in front of the book shelf and do a virtual browse to discover visual treasures in a collection. Humphreys added another path - use your own documents to find related material using JSTOR's Text Analyzer. I saw this demonstrated at an NFAIS meeting a few years ago and found it fascinating. Go to https://www.jstor.org/analyze/ and give it a try. He also mentioned a new service that is in beta testing called the JSTOR Understanding Series where you enter text or a passage from a book and it will retrieve articles that quote the text/passage (see: https://www.jstor.org/understand/).

Stohn and Humphreys closed the session with three questions for consideration: (1) How can we create more context? (2) How can we create different discovery paths and what is useful for which user story? (3) How can we help users add to the set of methods that they use when researching? Should we?

The slide deck used by Stohn and Humphreys is not with the other speaker slides on the 2020 NISO Plus website. The link to the slides is embedded in the notes from the session that can be accessed at: http://bit.ly/NISOPlus_SearchRetrieveDisc.

15. Miles Conrad lecture and panel discussion

A significant highlight of the former NFAIS Annual Conference program was the Miles Conrad Memorial Lecture, named in honor of one of the key individuals responsible for the founding of NFAIS, G. Miles Conrad (1911–1964). His leadership contributions to the information community were such that, following his death in 1964, the NFAIS Board of Directors determined that an annual lecture series named in his honor would be central to the annual conference program. It was NFAIS' highest award and the list of Awardees reads like the Who's Who of the Information community (see: https://www.niso.org/node/25942).

When NISO and NFAIS became a single organization in June 2019, it was agreed that the tradition of the Miles Conrad Award and Lecture would continue. The first Awardee under the NISO Plus umbrella was James Neal, the University Librarian Emeritus at Columbia University, who served as University Librarian between 2001 and 2014. He is a past president of the American Library Association (2017–2018). In 2019, he was appointed a Senior Policy Fellow at the American Library Association, with a focus on copyright and licensing. Neal also served as a member of the NISO Board of Directors and led NISO as Chair from 2006–2008. Neal could not attend the conference as he had a prior commitment in South America. His lecture was videotaped in advance as was the panel discussion that immediately followed his lecture.

In his presentation, Neal provided a brief look back at some of the key information industry challenges of the past four decades, but more importantly he highlighted the challenges facing the community today such as the democratization of creativity; the born-digital explosion; policy chaos, including privacy; market monopoly; global intellectual property and intellectual freedom; the challenges of diversity, equity, and inclusion; and human-machine symbiosis and blended reality. In his lecture he called out five commandments to which all stakeholders in the Information Community need to adhere in order to be successful moving forward together: (1) Thou shall preserve the cultural and scientific record; (2) Thou shall fight the information policy wars; (3) Thou shall be supportive of the needs of your users and your readers; (4) Thou shall cooperate in new and more rigorous ways; and (5) Thou shall work together to improve knowledge creation, evaluation, distribution, use, and preservation.

Neal's presentation was both thoughtful and provocative, and an article based upon it appears elsewhere in this issue of *Information Services and Use* as does an article based upon the panel discussion that followed the Award Lecture.

Neal's slides and videotapes of both the lecture and the panel discussion appear on the 2020 NISO Plus Website.

16. JATS, BITS, STS: Keeping Things in a "Family"

This session was a nuts and bolts presentation and discussion of the NISO Z39.96-2019 Journal Article Tag Suite (JATS) that defines a set of XML elements and attributes for tagging journal articles and describes three article models. The speakers were Jeff Beck, a Technical Information Specialist at the National Center for Biotechnology Information at the U.S. Library of Medicine (NLM), and

B. Tommie Usdin, President of Mulberry Technologies, Inc., a consultancy specializing in the design of XML vocabularies for prose documents. Usdin also co-chairs the JATS Standing Committee and is a member of the NISO BITS (Book Interchange Tag Suite) Committee and the NISO STS (Standards as Documents) Committee.

The speakers provided a history of the JATS standard. It is an extension of the work done on the NLM Document Type Definition (DTD) which was derived from the PubMed Central DTD. A DTD is a schema language for defining XML documents - basically, a set of rules for what can be in a document, what must be in a document, and the order of things if order is desired.

In 2006 the British Library and the Library of Congress announced that they would adopt the NLM DTD and it became a *de facto* standard, but they were concerned that it was not being maintained by a formal standards body. In 2008 it was agreed that NISO would take it over, but first NLM cleaned up some minor housekeeping issues that had been put off and created version 3.0 of the DTD in November of that year. The draft NISO Z39.96 JATS v 0.4 was released in March 2011 and the official standard was released in August 2012.

JATS is all about articles - documents that have very clear beginnings, main bodies, and an end. While chapters of books are similar to articles, the front matter is slightly different. Hence the Book Interchange Tag Suite (BITS) was created. It has book-specific metadata and is based upon the version of JATS that is used for archiving. BITS is more flexible than JATS because there is more variety in books than in articles. BITS was released in 2012.

Around that time the International Standards Organization (ISO) needed to reduce the cost and timeline required to produce and publish their standards. The process at that time was word- processor based, expensive, slow, and error-prone. In addition, the final publication of a standard could take months, if not years, after the standard was actually completed. Standards documents have some specific metadata requirements, but the narrative sections of the document are similar to articles and book chapters. ISO developed their own process, Standards as Documents (STS), for internal use and selected JATS as the foundation. They replaced journal metadata with standards-descriptions and local tracking and added standards-specific structures (e.g., Notes, Examples, TBX term and definition model). They did not remove anything from JATS, and while they made the process public in 2012, it was not released as a standard.

NISO developed NISO STS as a standard and it is based on the ISO STS. NISO added structures used by a variety of standards organizations as well as book-like structures (Table of Contents, Index) from BITS. NISO also made the metadata richer, more flexible, and optional. The standard was released in 2017.

It was noted that the JATS "family" is growing. It was designed to be customized and extended and people find it convenient to start with guidelines that work and that saves them time and money not having to start from scratch. JATS provides a common base and familiarity for creators/users of multiple document types. However, tag extensions can cause problems - some work, some do not, and it is difficult to predict the outcome in advance. Hence, JATS Compatibility Guidelines were developed. The Guidelines provide design principles in which compatible models as a whole must match JATS and also offers compatibility properties in which each modified model (element or attribute) must match JATS. The goal is to allow users to customize JATS to meet their needs; to use existing JATS tools and infrastructure; and to operate smoothly with other JATS documents.

The speakers stressed the fact that JATS was developed and maintained for quite a while before these principles were developed and that when documenting these guidelines, they became aware that JATS

does not completely conform to these rules. They hope future version of JATS may "clean up" what they now see as inconsistencies.

In closing the speakers said that if you find the guidelines helpful, use them and if you want smooth integration with existing JATS, consider them. But if your top priority is to create the best possible vocabulary for your specific use do that! On the other hand, if your documents will co-exist with JATS-based documents and your users have JATS-like habits, you cannot ignore the guidelines. If you do, you can expect slow-motion (possibly difficult to identify) chaos such as weird formatting, odd search results, and misleading error messages.

The slide deck used by the speakers is available on the 2020 NISO Plus website.

Note that on page 251 of this issue of *Information Services and Use* there is an article on JATS4R – the volunteer-run organization by Melissa Harrison titled *JATS4R - working together to apply the standard standardly* that produces recommendations for how people should use JATS. It discusses what JATS4R does, how the standard is maintained and updated, what the oversight group has achieved since it was established in 2013, and what the future may hold.

17. Privacy considerations for library and information professionals

The speakers in this session were Qiana Johnson, Collection Assessment Librarian, Northwestern University and Laura Paglione, Partner, Spherical Cow Group. The conversation was driven by a series of questions such as "Who is at risk as we move forward in collecting data in order to provide enhanced services?" and "What does it take to create an environment where privacy is a core business value?" There were no cut and dried answers, but everyone agreed that we are all better off when privacy is a core value of any business or institution. Indeed, respecting users' privacy is critical for all organizations in the information industry. The General Data Protection Regulation (GDPR - see: https://gdpr-info.eu/), the new California Consumer Privacy Act (CCPA - see: https://www.oag.ca.gov/privacy/ccpa), and other legislation mean that privacy protection is not only an ethical consideration, but also a legal requirement.

Three resources were recommended by the speakers:

1. NISO Privacy Principles: https://www.niso.org/publications/privacy-principles
2. ALA Privacy and Confidentiality Policy: http://www.ala.org/tools/challengesupport/privacypolicy
3. Foundational Principles of Privacy by Design: https://iapp.org/resources/article/privacy-by-design-the-7-foundational-principles/.

The slides used by the speakers only list the questions used to drive the discussion plus the three recommended resources. However, a brief paper based upon this session appears elsewhere in this issue of *Information Services and Use.*

18. Publishers and repositories: Opportunities for cooperation

An interesting session was led by speakers Angela Cochran, Managing Director and Publisher, American Society of Civil Engineers (ASCE) and Michele Mennielli, International Outreach Representative, LYRASIS. The purpose of the discussion was to see what the publishing and repository communities have in common and what opportunities exist for collaboration.

Cochran opened the session with a brief overview of ASCE. She said that on an annual basis they publish more than thirty thousand journals and between sixty to seventy books. She did a recent review of the publishing landscape and found that there are currently thirty-three thousand English language

journals and ten thousand non-English language journals in publication today. ASCE is in the midst of a huge legacy backfile project since they are digitizing their issues back to 1872. She noted that five of their journals serve very small communities, so even though they have been in existence for twenty-five years, they do not have an Impact Factor. Cochran said that publishing in journals became a currency in the academic world for tenure. In the beginning practitioners were publishing in ASCE journals, not researchers. Now the reverse is true and they receive complaints from practitioners because they view the journals as being too theoretical and technical. She noted that ASCE is a self-published society, one of a dying breed. They want to keep their content on their own platform and they want people to come to them to use their resources. Downloads are important; they need that count. But she admitted that both the "sharing" culture and government mandates for the deposit of funded research are changing what they do. She added that at first publishers supported the archiving of author-supplied manuscripts in repositories. Such deposits did not pose a threat as early on it was very hard to find those papers. However, today Google Scholar will point users to a free version and publishers are feeling threatened. Also, there are multiple versions of resources out there and she said that this is very concerning to those who are trying to protect a "pure" peer-reviewed version. She worries about what is truly the authoritative version of record for a particular body of research particularly when Google Scholar refers users to "free" versions and to preprint services. When asked how a researcher can a be sure that he/she has found the correct version of the output, Cochran said that we could use "badges" or require a label across the top that indicates the status of a work as not yet peer-reviewed. She said that she believes that we are living in a period of disruption.

Menielli said that he views it differently. He believes that we are living in a period of constant adaptation. He believes that scholarly output should be put in a repository and handled as an information resource. He noted that there are thousands of repositories out there. He admitted that his experience is with open source repositories, which have a very slow change process. The problem is that the repositories are silos - not necessarily interoperable - although that is changing. Menielli mentioned the Confederation of Open Access Repositories (COAR) - an international association that "brings together individual repositories and repository networks in order to build capacity, align policies and practices, and act as a global voice for the repository community" (see: https://www.coar-repositories.org/). COAR was also mentioned by Amy Brand in the opening keynote. Menielli also mentioned the European Open Science Cloud (EOSC) Partnership that aims to enable a trusted, virtual, federated environment in Europe to store, share and re-use research data across borders and scientific disciplines [37].

He admitted that there needs to be a process through which information in a repository can be updated or retracted. Once a researcher finishes with something they move on and unless repositories are incorporated into the research workflow updates will be haphazard. He also noted that more and more Societies are creating their own preprint servers to ensure that their material does get updated. This echoed a comment made at the 2019 NFAIS Annual Conference by John Inglis, Co-Founder of bioRxiv and medRxiv, and Executive Director of Cold Spring Harbor Laboratory Press, who said that the number of preprint servers is proliferating across all disciplines and sub-disciplines, each with its own technology and polices (for an excellent overview of preprint server growth see the summary of a presentation on preprint servers that was given by Shirley Decker-Lucke of Elsevier at the 2018 NFAIS Annual Conference) [38].

Menielli also noted that publishers and repositories say that they serve different purposes, but he doesn't believe that is true. Both groups are trying to disseminate and preserve the scholarly record and serve the needs of researchers across disciplines. Cochran said that her vision of a journal article is that users will be able to see everything related to it, but that everything does not have to be hosted on the publisher site. That network is being built, but is not yet obvious. She would love to see a grant application with a DOI on it so we can persistently link to them from the research results (a comment also made by Carly Robinson

in an earlier session). If there were more common persistent identifiers (PIDs) back and forth, it would be easier to correct these things.

Menielli agreed that PIDs for any content in a repository would be useful, as he believes that the history of the resource could then be tracked. He said that as long as repositories are not incorporated in the researchers' workflow they will not consistently update their content. He purported that the role of repositories should be changed to make them a tool for researchers, authors, and institutions, with which to create, preserve and share resources. He noted that publishers are now cooperating with repositories to manage data deposits and linking and that he would like to see the two communities cooperate in other aspects as well. He noted that both have the same shared purpose at a higher level.

Before closing the session, it was noted that in a recent analysis of data repositories, that only about two percent of researchers are reusing the data and that ninety percent of that two percent are the original authors who created the data and who are in the process reusing it for other purposes. Why? Probably because reproducibility is not valued in the same way as original research and because there is a matter of trust - what is the quality of the data? This reminded me of the comments made by Shelley Stall earlier in the conference with regards to the citing of datasets. She noted that we do not have a culture of data citation nor is there a culture of data re-use in some disciplines. She, too, questioned the incentive to re-use data when "novel" research is what gets rewarded? And she, too, raised the issue of Trust .This is definitely an area that requires a conversation among the stakeholders and perhaps NISO can make that happen.

No slides were used in this session.

19. Closing keynote: The legitimacy of data

The final speaker of the conference was danah boyd, a Partner Researcher at Microsoft Research Lab, the Founder and President of Data & Society, and a Visiting Professor at New York University. She gave an excellent presentation that focused on data, its uses, and how it can be manipulated to meet specific objectives - both good and bad. She noted that there is a problem with data - the moment it has significant power people will try to mess with it and this problem has existed forever. She provided real-life examples - many of which you will recognize - that reinforce the importance of ensuring that a dataset is unbiased and unflawed; e.g., does the data set represent the relevant subject matter or is data missing? She covered topics such as data quality, data voids [39], data infrastructures, alternative facts, and agnotology [40] (a new word for me!) She stressed that data become legitimate because we collectively believe that those data are sound, valid, and fit for use. However this not only means that there is power in collecting and disseminating the data, but also that there is power in interpreting and manipulating the data. She said that the struggle over data's legitimacy says more about our society - and our values - than it says about the data itself.

Boyd really gave me a lot to think about and I strongly recommend that you read the paper that is based upon her presentation - it appears elsewhere in this issue of *Information Services and Use*.

A video of Dr. Boyd's presentation, along with her slides and a transcript, are available on the 2020 NISO Plus website.

20. Conclusion

Overall, I found the conference to be quite interesting. I do I admit to being frustrated at not being able to attend all of the sessions (NFAIS did not do multiple track programming), but that is the case when I

attend most other conferences so I am not complaining. Indeed, my frustration is a compliment to NISO since when looking at the program before the meeting there seemed to be so many interesting sessions that I wanted to attend. As I had invited speakers in advance to submit articles for this issue, I chose to attend the sessions for which I had no firm commitments, knowing that the others would be covered.

As you know if you have read this article all the way through, the topics covered during the conference were quite diverse. But in retrospect, as I wrote this I found that there were several common threads that tied the speakers together. Similar points were made throughout the two and a half days and these include, among others, the fact that:

- Creating rich metadata is essential to facilitate information discovery and preservation.
- Using structured data is best to facilitate discovery via search engines.
- Preserving digital information is a complex effort and Best Practices need to be followed.
- Citing and reusing datasets requires a cultural and behavioral shift among researchers.
- Using Permanent Identifiers (PIDs) needs to be expanded to things such as grants and the content of repositories in order to track research from start to finish.
- Ensuring that datasets used for Machine Learning and Predictive Analytics are complete, unbiased, and relevant to the project at hand is absolutely essential to quality output.
- Developing standards to handle common problems is a worthwhile investment.

I mentioned earlier that Todd Carpenter, NISO Executive Director, called this conference a "Grand Experiment" in his opening remarks. From the perspective of an attendee, I believe that the experiment was successful. Most of the sessions that I attended were interesting, informative, timely, and several - especially the opening and closing keynotes and the Miles Conrad Lecture - were very thought-provoking. Discussions, for the most part, were lively and informative, although, not unlike other conferences, much depended on the level of effort the individual speakers put into preparing for their session and their use (or lack thereof) of slides.

Sara Rouhi, Director of Strategic Partnerships at the Public Library of Science (PloS), did several wrap-ups during the conference and she has written one that appears elsewhere in this issue. Take a look so that you can see what she took away from the meeting.

In closing, I must say that as a chemist I am quite familiar with experiments and I am also used to tweaking them to improve results. As successful as the meeting was, it, too, should be tweaked. It was called "NISO Plus" because of the merger of the practicality of NISO with the thought-leadership of NFAIS. From my perspective, there needs to be a little more of the thought-leadership added to the experimental conditions.

Having said that, what made the NFAIS conferences so interesting and valuable over the years was that NFAIS provided a *neutral* venue in which controversial issues could be discussed productively and with respect for differing opinions. This success factor was front and central to the NISO Plus conference and was in line with the fifth commandment put forth by James Neal in his Miles Conrad Lecture:

"Thou shall work together to improve knowledge creation, evaluation, distribution, use, and preservation."

My congratulations to Todd and his team for a job well done!! If you want to learn more about NISO - its history, mission, and activities - and how you can become involved even if you or your organization is not a member, read the first article in this issue of *Information Services and Use.*

Note: The only information available at this time on the 2021 NISO Plus conference is that it will be held February 21–23, 2021 in Baltimore, MD. Watch for details on the NISO Plus website at: https://niso.plus/. I hope to see you all there!

Additional information

If permission was given to post them, speaker slides used during the 2020 NISO Plus Conference are accessible in the repository on the NISO Plus website. Notes taken during the sessions are embedded within the conference program (available at: https://niso.plus/program-2020/) - just click on the session in which you are interested.

Also, as a reminder, NISO published a summary report that included ideas generated during the conference. That report is also on the 2020 NISO Plus website and has been reproduced with permission elsewhere in this issue of *Information Services and Use*.

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. She is a Fellow and active member of the American Chemical Society and an active member the International Union of Pure and Applied Chemistry for which she chairs the Subcommittee on Publications. 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*.

About NISO: NISO, the National Information Standards Organization, is a non-profit association accredited by the American National Standards Institute (ANSI). It identifies, develops, maintains, and publishes technical standards and recommended practices to manage information in today's continually changing digital environment. NISO standards apply to both traditional and new technologies and to information across its whole lifecycle, from creation through documentation, use, repurposing, storage, metadata, and preservation.

Founded in 1939, incorporated as a not-for-profit education association in 1983, and assuming its current name the following year, NISO draws its support from the communities that is serves. The leaders of about one hundred organizations in the fields of publishing, libraries, IT, and media serve as its Voting Members. More than five hundred experts and practitioners from across the information community serve on NISO working groups, committees, and as officers of the association.

Throughout the year NISO offers a cutting-edge educational program focused on current standards issues and workshops on emerging topics, which often lead to the formation of committees to develop new standards. NISO recognizes that standards must reflect global needs and that our community is increasingly interconnected and international. Designated by ANSI to represent U.S. interests as the Technical Advisory Group (TAG) to the International Organization for Standardization's (ISO) Technical Committee 46 on Information and Documentation. NISO also serves as the Secretariat for Subcommittee 9 on Identification and Description, with its Executive Director, Todd Carpenter, serving as the SC 9 Secretary.

References

[1] T. Carpenter, NISO and NFAIS Announce Plans to Merge, The Scholarly Kitchen, February 14, 2019, https://scholarlykitchen.sspnet.org/2019/02/14/niso-and-nfais-announce-plans-to-merger/, accessed August 6, 2020.

[2] Merger of Major Information Industry Associations Finalized, NISO Press Release, July 1, 2019, see: https://www.niso.org/press-releases/2019/07/merger-major-information-industry-associations-finalized, accessed August 6, 2020.

[3] J. Griffey, A. Meadows, N. Lagace and T. Carpenter, NISO Plus 2020: Outputs and Next Steps, July 7, 2020, see: https://nisoplus.figshare.com/articles/online_resource/NISO_Plus_2020_Outputs_and_Next_Steps/12622193, accessed July 15, 2020.

[4] C. Aspesi, N. S. Allen, R. Crow, S. Daugherty, H. Joseph, J.T. McArthur and N. Shockey, SPARC Landscape Analysis, March 16, 2019, see: https://infrastructure.sparcopen.org/landscape-analysis, accessed July 18, 2020.

[5] J.W. Maxwell, E. Hanson, L. Desai, C. Tiampo, K. O'Donnell, A. Ketheeswaran, M. Sun, E. Walter and E. Michelle, , Mind The Gap: A Landscape Analysis of Open Source Publishing Tools and Platforms, Simon Fraser University, MIT Press, September 2019, see: https://mindthegap.pubpub.org/, accessed July 18, 2020.

[6] S. Kember, Distributed Open Collaborative Scholarship, The Common Place, Knowledge Futures Group, March 15, 2020, see: https://commonplace.knowledgefutures.org/pub/rqaoavw9/release/1, accessed July 18, 2020.

[7] G. Evans and R.C. Schonfeld, It's Not What Libraries Hold; It's Who Libraries Serve: Seeking a User-Centered Future for Academic Libraries, An Issue brief from OhioLINK and Ithaka S+R, January 23, 2020, see: https://sr.ithaka.org/publications/its-not-what-libraries-hold-its-who-libraries-serve/, accessed July 20, 2020.

[8] NISO RP-19-2020, Open Discovery Initiative: Promoting Transparency in Discovery, June 22, 2020, see: https://www.niso.org/publications/rp-19-2020-odi, accessed July 20, 2020.

[9] P. McCracken, ShipIndex.org: A database with a narrow focus, but broad application, *Information Services and Use* **39**(3) (2019), 215–219, see: https://content.iospress.com/articles/information-services-and-use/isu190046, accessed July 20, 2020.

[10] D. Reinsel, J. Gantz and J. Rydning, Data Age 2020: The Digitization of the World from Edge to Core, November 2018, updated May 2020, IDC, see: https://www.seagate.com/files/www-content/our-story/trends/files/dataage-idc-report-07-2020.pdf, accessed August 4, 2020.

[11] R. Iriondo, Machine Learning vs. AI, Important Differences Between Them, Data Driven Investor, October 15, 2018, Available at: https://medium.com/datadriveninvestor/differences-between-ai-and-machine-learning-and-why-it-matters-1255b182fc6, accessed October 12, 2019.

[12] T. Mitchell, *Machine Learning* McGraw Hill, 1997, 414 pages, ISBN 0070428077, for more information see: http://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/mlbook.html, accessed October 12, 2019.

[13] Papers from the NFAIS Conference, Artificial Intelligence: Finding Its Place in Research, Discovery, and Scholarly Publishing, held May 2019 , Information Services and Use, 39, (4) 2019, see https://content.iospress.com/articles/information-services-and-use/isu190068, accessed July 21, 2020.

[14] J. Kennedy, A Monopoly on Structured Data, Boston Web Designers, see: https://www.bostonwebdesigners.net/news/structured-data-and-local-seo/, accessed July 21, 2020.

[15] Google Dataset Search, Wikipedia, see: https://en.wikipedia.org/wiki/Google_Dataset_Search, accessed August 4, 2020.

[16] Green Computing, Wikipedia, see: https://en.wikipedia.org/wiki/Green_computing, accessed July 21, 2020.

[17] N. Cohen, Researchers show glare of energy consumption in the name of deep learning, *Techxplore* (2019), see: https://techxplore.com/news/2019-06-glare-energy-consumption-deep.html, accessed July 21, 2020.

[18] K. Pruhs, Green computing algorithms. in: *Computing and Software Science* Springer/Naturepp. 161–183., first posted online October 2019, see: https://link.springer.com/chapter/10.1007/978-3-319-91908-9_10, accessed July 21, 2020.

[19] IP Address, Wikipedia, see: https://en.wikipedia.org/wiki/IP_address, accessed August 4, 2020.

[20] SciHub, Wikipedia, see: https://en.wikipedia.org/wiki/Sci-Hub, accessed July 25, 2020.

[21] ResearchGate, Wikipedia, see: https://en.wikipedia.org/wiki/ResearchGate, accessed July 25, 2020.

[22] Recommended Practices for Improved Access to Institutionally-Provided Information Resources: Results from the Resource Access in the 21st Century (RA21) Project", NISO Recommended Practice published June 21, 2019, available at: https://www.niso.org/publications/rp-27-2019-ra21, last accessed June 7, 2020.

[23] OCLC Awarded Mellon Foundation Grant to Develop Infrastructure to Support Linked Data Management Initiatives, Press Release, OCLC, January 9, 2020, see: https://www.oclc.org/en/news/releases/2020/20200109-oclc-awarded-mellon-grant-linked-data-management-infrastructure.html, accessed January 25, 2020.

[24] Linked Data, Wikipedia, see: https://en.wikipedia.org/wiki/Linked_data, accessed July 26, 2020.

[25] Getting Started with Linked Data, see: https://www.oclc.org/research/areas/data-science/linkeddata/linked-data-getting-started.html, accessed July 26, 2020.

[26] BagIt, Wikipedia, see: https://en.wikipedia.org/wiki/BagIt, accessed August 4, 2020.

[27] P. Simpson, Inquiry into Alleged Rigging of Television News Programs (Getzville, New York: U.S. Government Printing Office, May 1972). As Neil Postman and Steve Powers observe, this kind of "staging" persists in television news; see Neil Postman and Steve Powers, *How to Watch TV News*, Revised, Updated edition (New York: Penguin Books, 2008), Chapter Seven.

[28] J. Riley, Understanding Metadata: What is Metadata, and What is it for? A Primer, NISO, January 1, 2017, see: https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf, accerssed August 21, 2020.

[29] Data Publishing, Wikipedia, https://en.wikipedia.org/wiki/Category:Data_publishing, accessed July 21, 2020.

[30] K. Garza and M. Fenner, Glad You Asked: A Snapshot of the Current State of Data Citation, DataCite Blog, June 1, 2018, see: https://blog.datacite.org/citation-analysis-scholix-rda/, accessed July 21, 2020.

[31] T. Vines, What's up with Data Citations? The Scholarly Kitchen, May 28, 2018, see: https://scholarlykitchen.sspnet.org/2018/05/28/whats-up-with-data-citations/, accessed August 4, 2020.

[32] B. Marr, What is Extended Reality Technology? A Simple Explanation for Anyone, Forbes, August 12, 2019, see: https://www.forbes.com/sites/bernardmarr/2019/08/12/what-is-extended-reality-technology-a-simple-explanation-for-anyone/#3748ddb17249, accessed July 22, 2020.

[33] C. Grant, We are the Change we want to see, *Information Services and Use* **38**(1/2) (2018), 45–59, see: https://content.iospress.com/download/information-services-and-use/isu180011?id=information-services-and-use%2Fisu180011, accessed July 22, 2020.

[34] J. Grayburn, Z. Lischer-Katz, K. Golubiewski-Davis and V. Ikeshoji-Orlati, 3D/VR in the Academic Library: Emerging Practices and Trends, Council on Libraries and Information Resources, February 2019, see: https://www.clir.org/pubs/reports/pub176/, accessed July 22, 2020.

[35] Dublin Core, Wikipedia, see: https://en.wikipedia.org/wiki/Dublin_Core, accessed August 4, 2020.

[36] A. Volanakis and K. Krawczyk, SciRide Finder: A citation-based paradigm in biomedical literature search, *Nature, Scientific Reports* **8**(6193) (2018), see https://www.nature.com/articles/s41598-018-24571-0, accessed July 22, 2020.

[37] The European Open Science Cloud (EOSC) Partnership, May 2020, see: https://ec.europa.eu/info/sites/info/files/research_and_innovation/funding/documents/ec_rtd_he-partnership-open-science-cloud-eosc.pdf, accessed July 25, 2020.

[38] B. Lawlor, An overview of the NFAIS 2018 annual conference: Information transformation: open, global, collaborative, *Information Services and Use* **38**(1-2) (2018), 8, https://content.iospress.com/journals/information-services-and-use/38/1-2, accessed July 25, 2020.

[39] M. Golebiewski, boyd, d., Data Voids: Where Missing Data can be Easily Exploited, Data & Society, May 2018, see: https://datasociety.net/wp-content/uploads/2019/11/Data-Voids-2.0-Final.pdf, accessed July 25, 2020.

[40] Agnotology, Wikipedia, see: https://en.wikipedia.org/wiki/Agnotology, accessed July 25, 2020.