

# Summary report APE 2022

## The Future of the Permanent Record

*By Maaïke Duine*

### **DAY ONE: Tuesday, 11 January 2022**

The **17<sup>th</sup> Academic Publishing in Europe Conference: The Future of the Permanent Record** was opened by **Eric Merkel-Sobotta** (Managing Director, Berlin Institute for Scholarly Publishing (BISP), Berlin). He stressed that the APE conferences have always been and will remain to be an open, neutral, and collegial space, where all stakeholders, publishers, funders, suppliers, researchers, and students can engage with current topics of scholarly publishing.

**Dr. Caroline Sutton** (Incoming (February 2022) CEO, The International Association of Scientific, Technical and Medical Publishers (STM) and Director of Open Research, Taylor & Francis Group, Oxford) started off with a personal story that demonstrated the importance of trustworthy publications. She said that publishers can provide trust with a robust Version of Record (VoR). The VoR is a manifestation of the publisher's accountability for this robustness, showing the importance of this topic.

Dr. Sutton added that the changing nature of research, with new methods, technologies, and OS practices, has consequences for the VoR. The linear publishing workflow is being broken up and undergoing a structural shift. According to Dr. Sutton, this raises several questions: Is the VoR a drag on the system? What does versioning mean for the permanent record? Do these early sharing outputs require the same level of standards? Can we bring all outputs together? Is that the new VoR? Who will be held accountable?

Dr. Sutton said it is important to recognize which actors are responsible for validating the information. Trust can be eroded and there is a growing number of stakeholders who question publishers' motives. When moving away from the printed version to multiple digital versions of the VoR, a universal way to stamp items, to place trust and assign accountability and responsibility, should be discovered. Dr. Sutton concluded that there is no simple way to do this and input from different stakeholders and from multiple perspectives is needed.

The two keynotes were introduced by **Liz Ferguson** (Senior Vice President, Research Publishing, Wiley).

In the first keynote: **Quality and Equity in Academic Publishing**, **Prof. Dr. Maria Leptin** (President, European Research Council (ERC), Brussels) introduced [Review Commons](#): a project

to enable journal-independent Peer Review (PR) for preprints in the life - and biomedical sciences. The reviewed manuscript can be submitted to 17 participating journals. This enables a new form of the VoR, whereby all versions are recorded and a refereed reprint carries the entire file of the PR process. The traditional 'version of record' is being replaced by a 'record of versions'. Prof. Leptin stressed that high-quality publishing does not come for free, selectivity is expensive; it takes time to reject papers. Prof. Leptin said that journals add value at many levels, e.g., selection for quality, transparent PR, plagiarism and image manipulation checks, copy-editing, verification of data deposition according to community standards, scientific consultations, structured presentation on platform, OS platform development etcetera.

She continued by saying that partly as a consequence of funder mandates, publishing is gradually shifting to an Open Access (OA) system, mostly financed by Article Processing Charges. There are other approaches, e.g. transformative agreements. Alternative approaches include funder supported journals and publishing platforms (e.g. eLife), models supported by consortia of libraries and/or funders (e.g. PLOS Community Action Publishing) and Subscribe to Open initiatives.

Prof. Leptin concluded her speech with some thoughts on what (else) funders could do to prevent inequalities: ensure that research grants come with necessary OA funding for high-quality journals and books, develop mechanisms to support OA after the end of the grant, provide adequate training, focus assessment criteria on the intrinsic quality of the proposed project, and consider supporting societies/institutions that publish their own journals/books.

In the second keynote: **Building a Framework for the Future of the Record of Science**, **Todd Carpenter** (Executive Director, NISO (National Information Standards Organisation), Washington, DC) looked at how the shift to electronic publishing is changing the article of the future. He said that the future scholarly record needs to be much more than a PDF. It needs to be interoperable, machine-readable, and available in multiple formats, with an adaptive design, accessible for everyone, transformable, preservable, linkable, trackable, and with high-quality metadata. Carpenter added that these components rely on standards, standards like ISSN, ISBN DOI, MARC records etc. With the current transition to large scale data-driven science, there is a bigger focus on these standards.

He added that metadata quality needs to be improved: "It's the product's front door". High quality metadata is key to discovery and online interoperability. There are changes in the ecosystem that support this transition, but funding for these infrastructures is a problem, funding for maintenance is an even bigger problem.

Carpenter showed how formats have changed over the past 30 years, from print-focused in 1992 to primarily, almost exclusively, digital distribution in 2022. From just two identifiers, ISBN and ISSN, to multiple identifiers and a focus on data management. Carpenter said that the "article of the future" is so much more than a single object, it is a network of inter-related content forms. In order to support the new "Record of Science", the ecosystem needs to evolve; who is responsible for the scholarly record of the future and necessary metadata creation?

Researchers? Librarians? Publishers? How do we build a notification system to connect the disparate pieces?

Carpenter said that technology changes far faster than the cultural changes that are required to drive widespread adoption of a new ecosystem. People will only change the systems they use if they are motivated to do so, either by better benefit to them, or because they are forced to. When people get meaningful recognition for data sharing, they will. He also stressed that although technology drives change, it does not determine the direction or eventual destination.

In his talk: **Metadata Quality in Time of Diverse Research Outputs**, **Dr. Martyn Rittman** (Product Manager, Crossref) showed that although a proliferation of research outputs (preprints, datasets, online commentaries, peer reviews etcetera) clearly has benefits, it can also be problematic. What happens to a research output when a related output is corrected or retracted? E.g., when a dataset is corrected, should the article be adapted as well? And who is responsible? And what about research integrity and trustworthiness? Dr. Rittman said there is no coherent answer but like Carpenter, he emphasized the importance of metadata.

He introduced several Crossref services related to research integrity. One of them is [Crossmark](#): it shows the current status of content; if it has been updated, corrected, or retracted. Another service is [Event Data](#): this functionality creates links between related outputs, to provide context and commentary to research, e.g., citations on Wikipedia, mentions on social media, blogposts etcetera. Rittman added that since its implementation, the number of events has grown but there is work to do, e.g., in updating the infrastructure.

Currently, Crossref is prototyping a relationships API: you can put in a query for an item (e.g., a DOI) and get back the relationships to other items. Through the relationships API, it is possible to show the research nexus: all the outputs that come within the scholarly sphere and the relationships that exist between them. Rittman concluded by saying that Crossref can't solve all problems related to research integrity. Metadata plays a keyrole and [complete metadata from members](#) is needed, as well as feedback on [roadmap items](#).

The session: **The Version of Record under Attack! The Dark Side of the Scholarly Publishing Universe** was introduced and moderated by **Dr. Liz Marchant** (Global Portfolio Director (Life, Earth & Environmental Sciences), Taylor & Francis Group, Oxford).

**Prof. Dr. Christian Behl** (Editor-in-Chief, Journal of Cellular Biochemistry and Institute for Pathobiochemistry, University Medical Center of Johannes-Gutenberg-University, Mainz) started his presentation **The Need for Trust in Science and how to (try to) preserve it** by saying that integrity is the non-negotiable prerequisite for the general acceptance of scientific facts; scientific data must be trustworthy. He added that even though scientific mistakes, misconduct, data manipulations, and fraud are not new, more recently, numerous journals have been the target of multiple attempts of fraud, by so called "paper mill offices".

Prof. Behl said that the paper mill papers contain no real data but fake and fabricated data, and need to be recognized upon submission or retracted if they have already been published. The unusual increase in submissions due to paper mill attacks has created awareness in the community, and initiatives like [retractionwatch.com](https://retractionwatch.com) have been established.

Prof. Behl added that although costly, measures of quality control need to be consequently executed. In November 2021, the Journal of Cellular Biochemistry published a supplemental issue with retractions only and has implemented several additional measures, such as a new editorial board, new author guidelines, and new workflows, to maintain scientific quality.

Prof. Behl showed some examples on how the new workflow enables automatic checks on plagiarism and image integrity. He ended his presentation by saying that scientists have to exemplify scientific integrity and teach working conscientiously with data. He added that cooperation and knowledge exchange on novel movements of scientific fraud are key to bring back, protect and preserve trust in science.

In his speech: **Paper Mills: Global Knowledge Contamination by industrial-style Fake Science Publishing, Prof. Dr. Ph.D. Bernhard Sabel** (Editor-in-Chief, Restorative Neurology and Neuroscience and Medical Faculty, Otto-von-Guericke University of Magdeburg) continued on the threat of paper mills. He said they are professional, illegal organizations that produce and sell fabricated or manipulated manuscripts. Their services are used by students and scientists for different reasons, such as an unrealistic publication pressure and/or for promotion/graduation reasons. There are different types of paper mills that offer different kind of "services" (ghostwriting, "experiments", data falsification, sell complete articles) for different prices.

Prof. Sabel said Restorative Neurology and Neuroscience (RNN) carried out an analysis to find out if RNN was a victim of paper mills. They identified 23 (15.3%) "probable" Fake Papers (FPs). Next, they screened a random mix of 3,500 publications from 35 basic and clinical neuroscience journals and identified 376 "suspects" (10.7%). When extrapolating this number to PubMed publications, they estimated 450.000 medical FPs and 1.4 million FPs for all 14 million science/technology publications.

Prof. Sabel said that all stakeholders should be concerned when looking at this scope. Everyone must act, even if these actions will lead to (temporary) revenue loss. Employers should not put pressure on scientists. Publishers should support the journal editors to filter out FPs, provide reviewers with incentives, help develop tools to eradicate FPs from the permanent scientific record. Prof. Sabel stressed that publishers are responsible to help clean up the mess. The FPs impact science and society, create risks, and damage trust of the public in science. He concluded: It's time to address the problem.

**Dr. Alicia Wise** (Executive Director, CLOCKSS, Stanford, CA) started her presentation: **Time is a Thief of Memory** with an introduction on [CLOCKSS](https://clockss.org): a collaboration of academic publishers and research libraries, established to ensure the long-term survival of digital scholarly content. Dr. Wise stressed that long term digital preservation is an ongoing challenge and requires

active management; diversity in organization, governance and geography; security, forward migration and legacy. Dr. Wise stated that formal publications are better looked after than e.g. preprints but added there are also cases of journals that disappeared completely from the web. [Jasper Journals Preserved](#) is a project to preserve Open Access journals. It was launched in 2020 and spans all disciplines and all parts of the world. Dr. Wise added that disappearing journals are not exclusively an OA issue.

She said that with book preservation, the challenge is even greater. There are no standards for tracking: e.g. How many ISBNs are there? How many of those for eBooks? How many of those are preserved? These are questions that need to be answered. Dr. Wise announced that Springer Nature expanded the partnership with CLOCKSS, all SN titles published since 1815 have been added to the archive.

Dr. Wise concluded her presentation with the challenges around the core infrastructure that need to be solved: discovery tools, unique identifiers and knowledge graphs are needed to ensure the long-term availability of the record of scholarship. She added that publishers have a crucial role to play.

In his presentation **The research integrity arms race and the role of specialists**, **Tim Kersjes** (Research Integrity Manager, Springer Nature, Dordrecht), spoke about Research Integrity issues from a publisher's perspective. He said that due to the increasing number of Paper Mills submissions, Springer Nature has invested in human expertise, tools and AI to prevent fake publications and is proactively investigating patterns of publication misconduct. This is done through Advisory and Assurance (training resources for internal staff, external editors, authors and reviewers) and an issue handling team (to provide advice on handling research integrity and publication ethics issues).

Kersjes said that SN has had numerous paper mills cases over the years, across various fields. He shared some examples of manipulation from different fields: image -, authorship -, and Peer Review manipulation. In the mathematical field, this has led to nearly 40 retractions. More recently in the geosciences, 44 articles filled with gibberish have been retracted.

Kersjes said that paper mills are getting more sophisticated, the goal is to spot trends and patterns. New types of paper mills get flagged, automated additional checks in Quality Control have been added, e.g. plagiarism checks. Projects around image integrity have been initiated. Kersjes added that it is difficult to predict where it will end. He concluded that SN is fully committed to protect the literature but added that more automation may not always be the answer; investing in specialists and streamlining workflows are important too. This is labor-intensive but will lead to more effective improvements to correct the literature on a large scale.

**Dr. Hylke Koers** (Chief Information Officer, STM Solutions, Amsterdam) said in his presentation **Tackling misconduct through technology and collaboration: the research integrity collaboration hub** that the STM logo text "Advancing trusted research", represents the key values of the industry: quality, ethics and research integrity, and that technology is taking center stage within these key values. STM has spearheaded work on safeguarding research

integrity e.g., on image alteration and duplication detection. Dr. Koers said that collaboration on these issues is needed, but difficult because of poor interoperability between publishers, anti-trust regulations, and other reasons.

STM wants to take collaboration to the next level with the [Research Integrity Collaboration Hub](#). Trusted algorithms can access content from participating publishers to scan for research integrity issues during the editorial workflow. Two things are critical: 1: Publishers maintain control and 2: The infrastructure supports a broad variety of use cases, e.g., image alteration, paper mills, plagiarism, simultaneous submissions, etc. Dr. Koers added that developing this system collectively has many benefits: more agile, lower barriers for smaller and mid-sized publishers, benefits from network effects, and cost reduction.

According to Dr. Koers, STM solutions is ideally positioned to develop this system because it is governed and trusted by members, dovetails work that is already being done and can help foster collaboration with other stakeholders in the ecosystem. In developing the system, they follow an agile, use-case driven approach. Seven publishers are currently participating in the first use case: "simultaneous submissions". Dr. Koers invited other publishers to join.

**Dr. David Crotty** (Chef Cook of 'The Scholarly Kitchen' and Senior Consultant, Clarke & Esposito) presented **The APE Lecture: Life in a Liminal Space: the journey shapes the destination**. He said that a liminal space is the time between the 'what was' and the 'next'. Scholarly communication is leaving print behind, and the end goal, Open Science, is known. However, the path towards the end goal is still unknown and hard to predict.

Dr. Crotty said the journey to date has been marked by waves of consolidation and during this journey it has become evident that bigger publishers are better able to adapt to changes in the ecosystem. The growth of big publishers is driven by several factors: uncertainty e.g., due to the pandemic, transformative agreements, technology/reporting burden (complex landscape, expensive without inhouse expertise e.g., on metadata), but scale is the most essential component.

He added there is a current shift of publishers becoming workflow providers: the big publishers are building big portfolios of services encompassing all aspects of the research workflow, whereby it seems that most grants are for new developments. Dr. Crotty quoted Kurt Vonnegut: Another flaw in the human character is that everybody wants to build, and nobody wants to do maintenance. According to Dr. Crotty, that is a big problem in scholarly communication too. He said that the landscape is still dominated by subscription packages and even though change is happening, it is happening at a different pace, geographically and across disciplines. With the structure of the APC Gold OA model and Transformative Agreements, costs of publication are no longer spread among a large number of readers. Dr. Crotty said that institutes with low research output can expect to see a cost-saving, but that research-intensive organizations will see an increase.

Dr. Crotty saw an increased emphasis on publishing in quantity. For the OA long term, this would mean that low-volume flagship journals cost too much. There is a changing attitude from publishers toward these flagship titles, with an emphasis on lowering acceptance

standards and publishing more articles over time. According to Dr. Crotty, this means for the long-term future, that there is a drive for low-cost, high-volume bulk publishing. But there is still a demand from authors for high prestigious journals, and Dr. Crotty expected a few will stay in-house but other less prestigious journals will be run by societies.

Dr. Crotty said that this is the path we are currently on: Business organisms have adapted accordingly to OA. But he asked: is this acceptable for long term? We should not assume this is the path we should stay on, or we get stuck in a system where quantity is key to success. Are there other options? Dr. Crotty concluded: At the end of the day, we know where we want to go, what's the best way to get there?

## **DAY TWO: Wednesday, 12 January 2022**

**The Panel Discussion: The Version of Record - an Anchor for Innovation Upstream and Downstream** was moderated by **Anne Kitson** (Managing Director and Senior Vice President, Cell Press and The Lancet, Elsevier, Oxford) and introduced by **Prof. Lisa Janicke Hinchliffe**, (University Library, University of Illinois, Urbana): **Why is the VOR worth it for the Research Community?**

Prof. Hinchliffe said that all core functions of scholarly publishing (registration, certification, dissemination, preservation) are historically tied to the VoR. She continued with the formal definition of the VoR: A fixed version of a journal article that has been made available by any organization that acts as a publisher by formally and exclusively declaring the article 'published'. Hinchliffe stressed that it is key that this declaration is made by the publisher, not the author, funder or institution.

Prof. Hinchliffe explained some potential confusions around the VoR; it is not necessarily the printed version, or most recent version, or the gold OA article. Publishers are posting Preprint PDFs, the VoR is not always peer reviewed. Different article versions exist, this leads to a 'record of versions', and all these versions are eligible for a DOI.

Hinchliffe said that, for the majority of researchers, the article VoR is considered the most authoritative and credible source. They prefer not to use preprints for teaching purposes for example. The VoR is also a funder and rewards structure priority. The VoR is currently often the locus of compliance, e.g., with regard to OA requirements. Prof. Hinchliffe said because the VoR is central to OA publishing models, central to upstream and downstream innovation, it is so important to discuss why the VoR is worth it for the research community.

**Dr. Bernd Pulverer** (Chief Editor, EMBO Reports, Head of Scientific Publications at EMBO Press, Heidelberg) referred to Carpenter's talk on the article of the future: it is a network of different objects. These objects are not interchangeable, there is a need to establish an optimal linear process to avoid redundancy. Dr. Pulverer said that it is for example not possible to apply PR to all objects, but more quality control is important. Currently, only the VoR matters for research

assessment, therefore the VoR is crucially important but should be synchronized with other objects, and not be seen as the only valuable object.

**Dr. Niamh O'Connor** (Chief Publishing Officer, PLOS) believed the VoR is not the right way and not valuable. She said it is all about trust and signals of trust. A record of versions is better: the ability to link different parts of a piece of research within an open scholarly ecosystem. To force a historical format into the current system does not help address the issues around reproducibility for example. She concluded: Publishing is an integral part of the research process and publishers' systems should evolve and create systems of trust that work in an OS ecosystem.

**Prof. Dr. Ulrich Dirnagl** (Director, Quest Center for Transforming Biomedical Research and Dept. of Experimental Neurology, Berlin Institute of Health, Charité and MDC, Berlin) declared the VoR dead, because: 1. There are different timepoints in research: there is no final version. 2. The historic publisher centric view only made sense decades ago. 3. The VoR is supposed to guarantee trust, but this is a misconception and dangerous. Trust depends on how researchers design research. 4. The growing problem to cope with and evaluate all information that is available. 5. There is a pressing problem to safeguard quality in different stages. What can we devise to establish trust?

**Prof. Dr. Christian Behl** (Editor-in-Chief, Journal of Cellular Biochemistry and Institute for Pathobiochemistry, University Medical Center of Johannes-Gutenberg-University, Mainz) continued on the issue of trust. He said it is important to define quality for all different outputs, both on the publishing and editorial site. When looking at the retraction numbers, it is crucial to ensure that publications are correct. Awareness around paper mills should be increased and retraction policies should be strict. Quality takes time. According to Prof. Behl, the core problem is the publish or perish culture.

The subsequent panel discussion started with the question if moving away from the VoR onto various versions would create a sustainable business model for publishers. Prof. Hinchliffe saw the VoR as the primary one version that generated revenue. Dr. Pulverer stated that the VoR could still be the financial modal point and publishers could start charging for services, e.g., PR and OS activities. Dr. O'Connor felt that OA payments are not for the product or VoR: they are for the publishing services. She added that PLOS has a different approach with the Community publishing model and that it is important that publishers think about business models differently.

The discussion continued around the role the VoR still has for career assessments. Both Dr. Pulverer and Prof. Dirnagl agreed that the academic community is conservative and stuck in an ecosystem where too much value is still placed on the VoR. Ways to valorize other outputs should be found and funders and institutes should play a more proactive role in this.



Finally, more issues around trust were discussed. Prof. Behl said that Peer Review is key for quality control and trust. Prof. Ulrich disagreed: he has given up to trust articles just because they have been peer reviewed. He said that research is too complex to have just two reviewers. There is no label for trust; scientists have to make their own judgements. Prof. Sabel said that he does see improvement of the PR process and it should be more embedded into the academic system, e.g., by giving reviewers more credit. Prof. Hinchliffe was of the opinion that quality control should take place across the whole ecosystem. The industry should think realistically about resources, not just about money but also about researchers' time. Dr. O'Connor agreed and said that trust has nothing to do with the format. It is about transparency and being open about the processes. The output of the PR process does not give an absolute truth.

The session **Towards an IDEAL World: How to foster Inclusion, Diversity and Equity in Scholarly Communication** was moderated by **Dr. Christene Smith** (Chair of the Editorial IDEAL Committee at De Gruyter, Berlin).

**Dr. Nicola Nugent** (Publishing Manager, Quality and Ethics, Royal Society of Chemistry, London) showed in her presentation **Collective Action to make Things better, together** how the Royal Society of Chemistry has brought together 48 publishing organizations to set a new standard to ensure a more inclusive and diverse culture within scholarly publishing. The [Joint commitment for action on inclusion and diversity in publishing](#) was launched in June 2020. It was developed following a workshop in which RCE shared their Framework for action: a practical guide to reducing bias in publishing activities, e.g., by increasing transparency, reflecting the research community (e.g., more women in editorial boards), and improving training and encouraging intervention.

Dr. Nugent said that the joint commitment discusses the areas in which publishers can collaborate to move forward to understand and reflect the diversity of the research community. In order to ensure action is taken, working groups have been formed, and workshops have been organized, e.g., on author name changes and tackling harmful (historical) content. These workshops resulted into minimum standards, among them: work to define and communicate specific responsibilities and to publicly report on progress on inclusion and diversity in scholarly publications at least once a year. Dr. Nugent said that so far, best practices have been shared, hundreds of researchers have been helped to make name changes, and many publishers have paved the way for diversity data collection.

**Nancy Roberts** (Founder and CEO of Umbrella Analytics, Bury St Edmonds, Suffolk) explained **Why Data is the only Way to really improve Diversity**. She said that diversity is increasingly important in all industries. Companies invest a lot of time and energy on well-intentioned interventions. But what happens next? This is what Roberts called the diversity disillusion curve: at first, diversity becomes a board level priority, after that there is a bit of uptake, but then there are less commitments which result in a plateau of passivity and the actual business

diversity is not changing over time. Roberts said that causes for this curve are lack of planning, lack of clarity over goals, targets and outcomes, reliance on good intentions of committed people, insufficient resources, no budget, no time. And most significantly: no data!

Roberts said that companies should look at how progress is measured in other areas and the same rigor should be applied to improving diversity. What gets measured gets done, it is the same as any other business problem. Because of a disconnect between what employees want and what actions are taken, companies should carefully consider their strategy, e.g., who should collect the data? What kind of data should be collected? There should be a mix of quantitative and qualitative data, transparency is important. Roberts concluded: With data you can analyze, make predictions and set targets, and benchmark across the industry. Only this way, measurable improvements regarding diversity can be made in publishing.

**Ylann Schemm** (Director, Elsevier Foundation and Corporate Responsibility, Elsevier, Amsterdam and Chair of the Research4Life Executive Council) explained in her speech **Inclusive Research and the Global South** how Elsevier started 6 years ago with a gender equity taskforce. They saw the need to move beyond gender and to focus on multiple aspects of diversity by using the SDGs as a framework. Gender equality and contributions of researchers from the Global South (GS) matter to achieving all SDG goals. The Elsevier foundation focuses on inclusive research and developing partnerships to align strategies. Schemm said that this can be done on different scales. An example is celebrating women in science, another example is the green and sustainable chemistry challenge whereby solutions to tackle SDG challenges in the GS are developed.

Schemm said that partnerships such as the Research 4 Life initiative have been fantastic for supporting inclusion. But with the growth of OA, the scholarly landscape is changing and grown more complex. The focus on equitable participation in the research ecosystem should be increased and a more inclusive research ecosystem with the GS should be built. Publishers should acknowledge that researchers in developing countries move from being consumers of knowledge, to producers of research. Schemm added that to foster this development, publishers must examine their own role and take measures to develop equitable OA policies, create editorial boards with geographic diversity, recruit reviewers from low- and middle-income countries and publish robust research from the South as well as the North. Schemm said that to do this, publishers need to work together and create best practices, such as the APC waiver guidelines. Schemm concluded: Nobody can do it alone.

The panel discussion: **Ensuring Public Access to Research Data** was introduced and moderated by **Nick Campbell** (Vice President Academic Affairs, Springer Nature). This session focused on the progress on Data Management that is being made from the perspectives of three academic research librarians: **Michael Levine-Clark** (Dean of Libraries, University of Denver), **Judith C. Russell** (Dean of University Libraries, University of Florida) and **Rick Anderson** (University Librarian, Brigham Young University).

Levine-Clark said the current ecosystem is about rewarding publishing papers; the underlying data should be rewarded in the same way. Researchers' willingness to share data is influenced by the college department, but also by publishers, funders and other factors such as discipline, services, tools and repositories being used. Levine Clark added that it is also important to look at the differences and similarities between universities when looking at data management. To effectively manage data, differences in research activity, full time faculty, number of librarians and professional staff, research expenditures, number of articles published/year and disciplines should be taken into account.

Anderson explained what can be done to accelerate data sharing. Challenges that need to be addressed are about leadership, recognition, policy, data types, preservation, infrastructure, disciplines, sponsors, concerns about data security and scientific openness. He said that, when looking at its historic role, the library would seem like the obvious place for data management. However, historically, content has been published in limited formats. Data may seem similar, but functions like compliance monitor, content analysis, and misuse monitoring, require a redirection of resources within the library.

Russell added that Research Data Management is supported by an increasing number of tools (locally developed and external) and initiatives, e.g., DataCite, Crossref, CHORUS. She explained how [CHORUS](#) brings together funders, publishers and institutions to share knowledge, develop solutions, advance innovation and support collective efforts. CHORUS delivers dashboards, reports and APIs, monitors for OA on publisher Sites, helps stakeholders to improve metadata, participates in OS pilots, connects datasets to content, and is interoperable with other solutions such as publishing platforms.

The three panelists showed how institutions can make use of CHORUS to manage data more effectively and showed the dashboards for their universities. Levine-Clarke showed how the University of Denver could do better in identifying datasets. Russell said that for the University of Florida dashboard, although the scale is different, the pattern is similar.

The subsequent panel discussion started with collaboration and partnerships on data initiatives. Russell said that at the University of Florida, colleges and libraries have specific roles, e.g., on maintaining information on funders, implementing DM plans, curation assistance to researchers, enhancing metadata and linking datasets. She added that, because of the scale, automated solutions are needed, e.g., to generate notifications when a deadline to make a data deposit is approaching.

Levine-Clarke said they use the CHORUS dashboard to understand author behavior, what can be done to automate the processes as it is very labor intensive. Russell added that one of the challenges is that authors are not using their university email addresses and are not matched with their departments. The use of ORCID iDs should enable automated identification.

A Menti poll showed that four publishers in the audience are tracking authors' compliance with OA requirements, and five responded negatively. Anderson said to look at the publishers' incentives to track author's compliance. Levine-Clarke added that there is a need to build an infrastructure around connecting funders, article, data; if there is an incentive for publishers to

track compliance, this might increase. Campbell added that in SN's submission system, it is possible to enter the FunderID but most authors don't use this option.

Campbell posed the question how the different libraries are addressing the challenges of understanding multiple funders' requirements? Russell said that at the University of Florida, they document requirements of various funders, assist authors with compliance, use tools to identify publications, encourage use of ORCID iDs. Anderson said that at Brigham Young University, the library has not been involved in the relationship between funding and research. Levine-Clarke stated that at the University of Denver, more needs to be done, in general to help understand publications and data linking. Standards and infrastructure to connect different data types should be implemented to make processes easier. Campbell concluded: for Data Management to be part of the culture and standard, we all need to work together.

**Dr. Niamh O'Connor** (Chief Publishing Officer, PLOS) concluded the second day of APE 2022 with her presentation: **Publishing as a Process: Open by Design**. She started by saying that the move from physical to digital has been a huge change within the publishing industry but did not lead to the level of disruption as was anticipated. Dr. O'Connor stated that publishing should no longer be seen as a separate event but as a process. That way, publishing can become a truly integral part of the research process. She continued on the definition of Open Science and showed some examples of how the transition to OS is taking place. She added that OS is more than publishing an Open Access article. She showed how within PLOS Biology, publishing is seen as a process. This process can include protocols, preprints, peer review reports, data, methods, code etc.

Dr. O'Connor said that instead of talking about an article, we should talk about a *linked digital knowledge stack*, and to enable this multiplicity of outputs, metadata quality should be improved. The decision on what is valuable shifts more from the creator towards the user. Selectivity needs to be redefined, and the focus should be on the research question.

What does that mean for the VoR? Dr. O'Connor thought that articles will have value in the future, but they should not be outvalued over other outputs. She pointed out that many current publishing systems and business models are built on the VoR. This hampers research advancement. In collaboration with libraries, consortia and funders, publishers should move beyond the article and APCs and develop new business models. Selectivity does not have to mean that costs are enormous. Dr. O'Connor said that in order to move to a system that is truly open, the change has to be radical.

### **DAY THREE: Thursday, 13 January 2022**

The panel discussion **How can we foster Entrepreneurship and Innovation in Scholarly Communications?** was introduced and moderated by **David Worlock** (Chief Research Fellow, Outsell Inc., London). Panelists were **Sami Benchekroun** (Co-founder and CEO of Morressier) and **Ijad Madisch** (Co-founder and CEO of ResearchGate).

Worlock kicked off the session by asking the panelists about their drive to innovate and invent. Madisch said that he and ResearchGate's cofounding team shared the same drive to reinvent – not disrupt – the status quo, to invent ways in addition to old ways to move science forward by using technology. Benchekroun added that they saw the problem of online conferences and with Morressier they had the drive to wanting to change something and improve online conferences.

When asked about how to avoid being sidetracked from achieving the main goal, Madisch's advice was to do what you want to do, and not to think too much of what you need to do. He added that even though at ResearchGate the scientist is in the center, they are also thinking about the future, e.g., how can RG help scientists to become more productive? Madisch said that new challenges and problems do come up while working out the main problem. Benchekroun added that it is almost inevitable to solve different problems, when tackling the main problem in this area. To bring offline conferences online, for example, so many steps are involved; you need a holistic approach.

The discussion continued on the risk of starting a start-up. Madisch said that he does not think in risks but in opportunities; that's how you become an entrepreneur. He added that he underestimated how you monetize, from an entrepreneur led organization to a more professional organized enterprise. You must think how you want to monetize without sacrificing your product. Change culture, change leadership, and ask yourself if you are still the right CEO. Benchekroun added that a lot of startups forget to innovate themselves, even though the market develops further. An innovation mindset is important, as well as keeping the conversation going with scientists, society, and publishers.

Worlock asked about the route for growth. Is it generic, or through collaboration, or acquisition? Benchekroun replied that within the scholarly communication community, there are a lot of opportunities to grow. There is a clear need in scholarly communication to collaborate but also to grow from within. He said that not enough is being done. When there would be more start-ups within the sector, the speed of innovation could be brought up to the speed of change. Worlock said there might not be more start-ups because growth rates are not as fast as other markets. Madisch said that to be ready for growth senior leaders and non-scientists should be brought in for fresh ideas. He said the problem with scholarly communication is twofold. First, the fact that growth is not big and interesting enough for investors. Secondly: Pseudo innovation. Transforming offline to online is not real innovation. New ways of working need to be created and it is crucial to get closer to the creation of research.

The session **New Dotcoms to Watch** was presented by **Drs. Eefke Smit** (STM Director of Standards and Technology, Amsterdam). The following start-ups in scholarly communication were presented:

**Andrew Preston** introduced **Cassyni.com**. He started by saying that academic seminars have existed for a long time; offline and in person and they have always been a good way to

collaborate. There usually was no record. With the pandemic, things have changed, seminars went online and onto Zoom. Problems around these online seminars started to emerge: finding, managing and scheduling seminars, use of online promotional material, keeping the webinar page up-to-date, making recordings available etc. Preston added that online seminars are here to stay because of the opportunity to connect with so many peers from around the world.

With Cassyni, it is possible to engage directly and globally with the community and increase awareness and impact of a journal's research. For existing seminars, it is possible to save time, reduce complexity and improve experience. Preston added there are lots of additional features, such as DOIs for the seminars, and different use cases: journal seminars, university department and research group seminars. Cassyni's vision is to build a rich diverse and inclusive ecosystem. Since the public launch in September 2021, Cassyni has been picked up by the community showing that it is a product that fits well into the research ecosystem.

**Keith Kallmes** introduced **Nested-Knowledge.com**. He stated that reviews drive clinical practice guidelines and biopharmaceutical therapy development. Therefore, enhancing systematic literature review is critical. He stated that currently, less than 15% of the literature is covered by up-to-date clinical reviews. Within the current paradigm, the major problem is that you end up with a static PDF. With Nested Knowledge, at the end of conducting a review, there is a living document including data access, dynamic visualizations etc.

Clients use the Nested Knowledge platform to organize, visualize and analyze research while compressing timelines and improving outputs. Pharmaceutical and device companies utilize the NK reviews for FDA interactions, CER literature reviews, and other purposes.

Kallmes concluded his talk by stating there is a need for a mindset change, towards adopting the living knowledge paradigm. With the NK software, living reviews are possible, people will have access to the most up-to-date knowledge, and it will enable publishers to move forward to a more dynamic and systematic paradigm.

**Andrew Chalmers** introduced **CiteAB.com**. He stated that life scientists experience problems identifying the right antibody for their experiments. This leads to waste of funding and time. To address this problem, CiteAb focuses on identifying product citations and the antibodies that have been used. They do this with text mining technology and human reviewing to generate high quality citation data. Researchers can use the search engine, find antibodies of interest, and choose the one they want to take back and validate.

CiteAb is expanding on data collection, with a focus on quality. They moved from being a search engine to a data provider. Data around the products has value for reagent suppliers, it informs their strategy and help drive sales.

Chalmers said they are just scratching the surface of what can be achieved. There is more data to collect, and researchers and suppliers that can be helped. CiteAb is also exploring partnerships with publishers. Potential benefits are to access novel content and revenue, collect data from publications and then return with added value.

**Francesco Mazzotta** explained more about **Pilloxa.com**: a platform to digitally engage with patients, learn from them and empower them by identifying what they need. Pilloxa was founded to make people healthier through technology.

Pilloxa offers different solutions: one of them is the patient-facing mobile app program. Clients, like university hospitals, can tailor-make the app to support their patients during their treatment. This can for example include sending information, e.g., journal articles at the right time. But also, to provide patients with information on medication and motivation tools. This can be done through add-on smart hardware. A university hospital can for example learn how to better equip patients to motivate them to stick to the treatment plan.

Mazzotta said that Pilloxa aims to make complex issues simple and fast, whereby the patient's needs are put first. By bringing information together, this will accelerate learning and knowledge creation.

**Liat Belinson** introduced **aipatents.com**: a unique search platform for systematically analyzing long text documents. Currently, the problem with analyzing lies in the fact that the same idea can be described in multiple ways, terminology changes over time, and patent documents contain hundreds of pages so using keywords is not sufficient. Current solutions are inefficient and ineffective, search queries use long complex syntax, results are not sorted by relevancy and there are just too many results.

Aipatents enables a content-based search: matching is done on all content, not on keywords. It uses a unique Natural Language Processing approach for analyzing long text documents. It gives comprehensive ranked results, sorted by relevance, and connects different types of documents: patents, standards, literature, product information, etc.

AiP has worked with leading integrators in the patent space and enables diverse applications. A PubMed search query showed that AiP is able to find highly relevant articles. It is not only able to identify the articles but can also find articles that contradict, with other claims. Belinson concluded: Big data is here to stay and being able to connect different types of data is the future.

All panelists replied positively to the question if they would be willing to collaborate with publishers. Belinson said it is about adding more value and insights for researchers and finding the right home for the technology. Mazzotta agreed: by working with publishers, we can help spread and build knowledge, improve health literacy, and make sure information is accessible at the right time. Preston said that Cassyni is already working with publishers in organizing seminars; it is an opportunity to build relationships with the community. Kallmes said that journals could adopt the living review paradigm and move forward with enabling living documents. Challmers said that collaboration would be in both the publisher's and scientist's interest, data can be used to help scientists read their papers.

At the end, the audience was asked to vote for the different dotcoms by answering questions: which of these dotcoms would you like to collaborate with, in which dotcom would you invest your money, in 3 years who will be the most successful dotcom? At the end Smit declared Nested-Knowledge.com as the winner.

**Auf Wiedersehen! Goodbye!**

Eric Merkel-Sobotta closed APE 2022. He thanked APE's founder Arnoud de Kemp, the program committee, all speakers and sponsors, and extended a special thanks to Marta Dossi, without her the APE Conference would not have taken place.

*Please note: APE 2023 will take place 10-11 January, 2023.*