



STM Tech Trends 2023

Recorded at STM US Conference, Washington, DC

For podcast release

Monday, April 15, 2019

KENNEALLY: A rock band or any musical group must stand or fall on the strength of collaboration. It's often said of the Beatles, for example, that the whole was greater than the sum of its parts. Their collaboration led them to innovations as well. John, Paul, George, and Ringo inspired each other to do their best, and they extended those collaborations to other musicians outside the group, notably Eric Clapton and Billy Preston. There are many kinds of collaboration in popular music. I think some of them are sort of the obvious mashups – imagine Queen and David Bowie on their performance of “Under Pressure.” And there were those that forced a kind of a double take. Who would ever put together Run-DMC and Aerosmith, Pearl Jam and Neil Young? And then, of course, there's Brian Eno. He'll collaborate with anyone – U2, Talking Heads, Jay-Z, and Grace Jones. Collaborations are essential for making music, because when a band works well together, their members inspire one another. They share ideas and they apply them. It's a great lesson for everyone, even if all you ever do is sing in the shower. If you come across a good idea, use it, whomever happened to think of it. And I hope you will do that today with any or all of these ideas you're about to hear from my group.

I want to introduce everyone. We'll start at the very far end. You've heard from IJsbrand Jan Aalbersberg before, but welcome IJsbrand Jan Aalbersberg. He is senior vice president of research integrity for Elsevier, responsible for new technology initiatives that safeguard the integrity of content. In this role, he is also responsible for user privacy.

And immediately to his right is Max Gabriel. Max, welcome. Max joined Taylor & Francis Group as chief digital officer in 2015. He is responsible for technology strategy, delivery, and operations for the group. Earlier, Max was CTO of Pearson India and Africa.

To his right is Gerry Grenier. Gerry, welcome. Gerry is senior director of content management for IEEE, the Institute of Electrical and Electronics Engineers. He leads the content engineering team responsible for content architecture and systems, and he serves on the board of the National Information Standards Organization.



To his right is Priya Arora. Priya, welcome. Priya heads the product and digital strategy for Wolters Kluwer Health's institutional platform. She is responsible for a portfolio of digital solutions, content, and services for clinicians, researchers, educators, and society members working in hospitals and medical schools across the US and around the world.

And then finally, immediately to my left is Sameer Sharif. Sameer, welcome. Sameer is founder and CEO of Impelsys. Impelsys is his second startup, and he has grown the company to a market leader in digital content and learning delivery solutions. Prior to Impelsys, Sameer founded – or sorry, co-founded – Medsite Inc., where as chief strategy officer and executive vice president of sales, he helped to become one of the largest Internet pharmaceutical marketing companies in the US. And in 2006, Medsite was acquired by WebMD.

I'd like to begin with you, actually, Sameer, talking about this new hit song of ours around open science, and I know you have some very important thoughts as a technologist about open systems and how they empower users. Tell us about that. Remember to grab the microphone there.

SHARIF: OK, so let me get started. No, I think – I mean, collaboration, open systems, I think where technology has evolved over the last 20, 30 years, and as it's evolved, it's become more and more complex. I think I talked about this on Society Day a couple of days ago. There's a lot more elements and sophistication in actually how you deliver solutions.

So as it becomes complex, the question for you as an organization, as a publisher, is can you actually manage all the different elements? Are you an expert in every area? And I think no, you can't. Technology is evolving, and it's evolving at a faster and faster rate. Take, for example, Apple. Apple, which has the iPhone – it's not completely done by Apple. If you think about it, Samsung is also involved in Apple. Apple's biggest competitor is Samsung. But in reality, they also collaborate on the phone. Why? Because it is complex. So I think in terms of building applications in the future, it's very important that you actually figure out who are your partners – best in class or strategic – to actually build out your solutions.

So now getting back to your question about open systems, I think also the evolution of how software is built – software traditionally was built as monolithic, individual systems. And if you look at how the whole infrastructure cloud all evolved, if you think about cloud infrastructure, servers and all were traditionally kept internally at your own – in your own office. It then moved into a central area, and then it went to the cloud. And it was something that people were against. But today, AWS or



Google Cloud or Microsoft Azure is the system where you can collaborate and you can leverage shared environments.

So on a software side, these monolithic systems that you build have really gone out. We are going to an open platform solution, where products are built as services, as microservices architecture, so that you can actually break these systems down into different elements. But also what it helps you do is also connect these solutions into other systems. And I think – I mean, Eefke, the band is a great idea. It's about collaboration. So the idea here is build solutions and software that is open, architected, so that you on one side can then connect into other platforms or services. On the other side, you can bring in solutions or services from best of breed and actually plug it into your ecosystem. So that way you can provide the most optimal solution for your end customer.

KENNEALLY: What's fascinating to me is the openness that we are just now able to sort of realize was there inherent in the system from the very beginning. I think some of the real thinkers around the Internet realized that even 20 years ago. David Weinberger, with *The Cluetrain Manifesto*, I think sort of hinted at that and continues to write about it. He has a new book called *Everyday Chaos* that is talking about how in our networked society, all organizations – any kind of business, any kind of research operation – needs to be not simply planning, but ready to respond. It's a responsiveness. This is the kind of openness that the technology makes possible. But it's also an approach, a cultural openness, would you agree?

SHARIF: Absolutely, yes. Cultural – I mean, at the end of the day, technology is an enabler. And I think it also has to – I've been in multiple conferences like this, and we are hearing more and more of the senior executive management at some of the bigger companies saying that it's an age where we have to collaborate. But the question is when the rubber hits the road, can these big commercial publishers and the society publishers – actually, the software can be open. But culturally, as an organization, can you actually openly collaborate from a business principals, business collaboration, share platforms, and not say that your platform is a primary platform? Can you actually share it together? Technology I think is the easy part. The hard part is corporate culture, corporate goals, profit motives. Those are the harder questions.

KENNEALLY: Gerry Grenier at IEEE, you were part of the Future Labs discussion in London, and I know you – the song you were singing that day anyway was around shared infrastructure. Tell us why that's critical in your vision of open science, and talk about what Sameer was just hinting at, which is there's a challenge around openness.



GRENIER: Right. So I think that the big challenge I see today is that as our revenues are under threat from open science, open access, Plan S, and as we move our focus from distribution of our content – of the traditional legacy content – to building tools to enable open science, there's so much money today in the scholarly publishing infrastructure wrapped up in the same things at this publisher, this publisher, this society, this society. And a first step that I think we need to take is while we're worried about our decreased revenues, we need to think about reducing our expenses on those legacy platforms. I think there's no reason why we can't declare that distribution platforms today, I would argue, don't offer a tremendous competitive advantage. We're spending the same money across the ecosystem on cybersecurity. We're spending the same money on storage, administration, the applications.

To get back to something that Sameer said, we do have a cultural problem. We think we all have the killer apps. We think we have the killer features on our platform. And I would argue that there are really two features that matter in the distribution platforms. That's uptime 24/7/365 and discovery. The other stuff I think is fluff, and the dollars are just floating away out of the scholarly publishing ecosystem.

Where we might realize some savings – I look at a continuum, right? The author begins his or her research here. It's distributed here at the publisher. The publisher is managing this distribution part. And the researcher is left to this big box of creating their own tools. I think there's a tremendous opportunity for us to get into that new box and pioneer new open science platforms and working together with the rest of the community – with the vendors, with other societies, and with other publishers. I hope I didn't veer too far from your question.

KENNEALLY: No, absolutely not. But I was going to bring it back and say so there are challenges. There are technology challenges. There are collaboration challenges. There's the challenge that Eefke Smit referred to in this very complex audience that has emerged recently. Talk about the ways that that approach to shared infrastructure would help address those kinds of complexities in the audience.

GRENIER: Yeah, well, I think that one of the things that diffuse platforms – one of the problems they cause is just the problem with human/computer interaction and the dashboard that we present – the multiple dashboards that we present to our end users. I think collaboration can bring together a single experience or maybe only two or three different experiences to our end users.

And then just going forward, getting off the platform rant for a minute, moving ahead to the open science platforms, I think that in reproducibility, for example,



there's so much work to be done in defining standards. What is reproducibility? What is replicability? And we cannot do that. When I say "we," publishers, societies alone cannot do that, and we need to work together. It's so important to work together towards new standards and new definitions. Not just technology infrastructure solutions, but standards behind reproducibility, replicability, open science sharing, citability, things like that. Policy. I think those are as critical, as Sameer alluded to, or more critical and more challenging than the technology.

KENNEALLY: Priya Arora, at Wolters Kluwer, you have, again, a very complex audience that you're trying to address. There are the researchers, but there are also the primary care physicians. And they have very different information needs. Expand on that a bit. Explain how you sort of meet their individual needs.

ARORA: Sure. So the end users for us, they're focused on the same content, which is the medical content, but the use cases can be very different from a medical student or to a teacher, healthcare professional. A healthcare professional will need content which is point of care, and it has to be on the bedside, and they want quick information, while a researcher or an author is actually doing deep research.

So the challenge for us is not dumping the content on our platform, but making it more personalized for the user to cater to their needs based on who the user is and what they're trying to find. So use machine learning to find out more about the user and basically deliver the content to them, as opposed to them having to find it, because there's just so much content out there that we need to help them figure out what their search can actually be. It's a needle in the haystack, and we want to reduce the number of layers between them and the needle so they can get to that information more quickly and more easily.

KENNEALLY: That's information they need if there's a bedside question. But there's also information that's needed in the classroom environment, and you've got some interesting approaches to that as well.

ARORA: Right. So in the classroom environment, what we are seeing is between the teacher and the students, it's not just simply about knowledge transfer. But we want to help them make their classes more interactive, and how can we do that? We basically offer expert solutions on our platform. So these are tools and services which helps them make the classroom interactions more – or encourage the students to become more creative and curious. So we give them tools like augmented reality in our anatomy product offering, where it becomes more – the student develops more curiosity, and they can pretty much dissect a body on their table in the classroom. So this is where –



KENNEALLY: Wait a sec. Now you have to tell us about that. (laughter) Dissecting bodies on tables sounds a little gruesome. But here this is in a way that not only moves them into that virtual experience, but gets them to explore, as you were just trying to say, much deeper than they might in a textbook or in a typical transfer of knowledge. They can go and they can find their own way. That's important.

ARORA: Right. And we found out through our surveys and scores that this way of teaching sticks with the students better than going through the textbooks or just reading chapters after chapters.

The other thing we have introduced is quizzing and self-assessment tools. So we use machine learning to see who the student is, and then based on what their status is and how they are performing, we give them – the next quiz is based on that. They can self-assess where they stand. And based on that, we recommend chapters and more reading to them. This information is available to the faculty as well, so they exactly know how to go about in their classroom teaching with the students.

KENNEALLY: It's a different sense of openness. It's opening up the information to them in their own individual way. It's a fascinating evolution of providing the teaching and the rest of it.

IJsbrand Jan Aalbersberg, integrity is important to your work and to open science. You have some thoughts, though, first on open – what the definition of that is. I think you said at one point that it is better to ask what does open include, not what is it – rather than trying to box it, sort of begin to give people an idea of what it can be. Tell us more.

AALBERSBERG: Yeah, openness is, of course, multiple ways. It's also open in APIs, for example, open that we present our content for researchers to build their own tools on top of it, APIs, like we do on Scopus, like on Mendeley, that our users can optimally build tools, build functionality on top of the services and the content that we provide. But openness is also to each other, that from the publishers, that we work together and that we share content together and standards together to make sure that the experience for our users is better than before.

A very classic example of that is of course what we did with Crossref, the linking of articles, where we really needed to be open to each other between the different publishers to work together. But also RA21 is another collaboration in open activity, and not only open to each other, but also to our users. RA21 also allows all sort of openness features to be built into our platforms.



Openness also is open with respect to making our content more transparent, and then we talk about openness of data, openness of peer review, openness of showing, for example, how we have checked the images for manipulations or openness of plagiarism checking, that we have done that. Openness in really looking behind the process of publication – like, yes, we have plagiarism checked here. And then we get into the openness of results – reproducibility batches and integrity batches. That those batches gives them openness, transparency, and in the different things that the publishers have done on the articles, like indeed, as you said, plagiarism check, image checking, reference checking, statistics checking. All those different things is part of making the process open, and that indeed relates to the integrity as well.

KENNEALLY: OK. Well, expand on that, because that openness, that transparency, is in a way a response to the concerns that openness can lead to a diminishing of quality, can open the door in fact to fake science over fact science. So this transparency it sounds to me can be a response to that.

AALBERSBERG: Yeah, I'm not completely sure whether openness immediately causes or leads to maybe more fake science. It's more the fact that open distribution – just wide distribution without any validity check – that indeed causes the potential for indeed fake science. So you need the validity check that is important that we need to introduce. But over the years we get more that our researchers, our users want to have more insight in that validity checks that we as publishers are doing, and that is where the openness comes in. They want to see why a peer review report was positive on a certain article. They want to see why a peer reviewer has said that a certain article could be checked. They want to see why the research was done in a certain way. And they want to see why the results were analyzed in a certain way. They want to see the data themselves. They want to redo the experiment partially.

So that is the way of openness, making it the whole research process – both what the researcher has been doing, and that's the open data, for example, but also what the publisher has been doing. That needs to be opened up.

KENNEALLY: Max Gabriel, this discussion around open is a fascinating one. I know you have some important thoughts to share on that. You see it again as a technologist, which I believe implies that open means to you interoperable, though of course it can mean other things. Tell us about that.

GABRIEL: No, you're right. I think this was sort of a key theme I heard over the past few days, and Olivier brought this out in terms of interoperability at the core of everything we want to do. Shared infrastructure, open infrastructure, doesn't mean getting everybody to a single platform. You want to make sure that each of the components of your platform can talk to each other.



A point on open that could be a bit controversial – I don't think your end customer cares whether it's closed or open. I think it's almost a reaction to our current state, rather than the customer sitting there saying I want an open system. Mainly because a customer wants something that's useful to them, contextual, and relevant. Our inability to offer that because of our closed systems that we have, systems that don't talk to each other, the debate has moved on to let's make the closed open. But if you have an open system that is still not useful, relevant, and contextual, we're not solving the real problem.

So I'm not against open or closed. If you look at Apple, yeah, they are a closed ecosystem, but they've figured out a way to make it really useful to the users across the globe and that seems to work. On the other side, you have Android, who have chosen an open model, and they've figured out a different way to make it useful. And I think we shouldn't lose sight of what, Priya, you were saying, about you got different audience and their needs are different. Their needs are contextual. So making things interoperable in terms of technology is super important, rather than this narrative between whether it's open or closed.

KENNEALLY: And interoperable, but also innovation can be interdependent, right? So you can have innovations in hard science and social sciences, and this requires STM publishers to think beyond their own immediate discipline or fields to see where there are other opportunities.

GABRIEL: Absolutely right. Every technology invention you've had over the past, I don't know, five, six decades, we fail to see the interdependence between them, right? So today we have a huge data problem. Well, that was because of the proliferation of mobile devices. Well, that was because of what Internet was built on. So the interdependence between these inventions are often overlooked. And actually, as Peter Senge said, the more interdependent we are, the less unaware we are about this, right?

So as much as we want to sit here and talk about how do we connect these pieces within publishing – Chris, to your point, I think there's a broader world out there, which is why I kind of disagree with Olivier's point yesterday saying let's move off of Google and figure out our own search and discovery. There's merits to that. But you know what? The rest of the world uses Google to find pretty much every other information out there. So we have to figure out how can we become more effective there, rather than trying to reinvent one more wheel for ourselves. I think we have to better appreciate the interdependence of our customers and the world that they live in, that they just don't live in a publishing world.



KENNEALLY: Well, Max, the next area that I think we ought to talk about is the user experience. For you, the user is – it’s obviously an identity. We all have our identities as users. This is where the real challenge is, and you have interesting thoughts on how we can address that, and there’s an example in China.

GABRIEL: Yeah, I was fascinated in terms of how user experience is ubiquitous in China, and that’s mainly because how they have looked at identity differently than everybody else. Of course, China came to digital identity much later than every other economy in the world, but that’s actually given them an advantage to create a single identity. And the identity is about the user, not about the product, not about the institution, and not about the publisher. So they linked it to the user’s identity, and because they’ve pinned it to the user’s identity, they pretty much told every other product company to say, work with this identity.

For example, WeChat, as you know, is the WhatsApp equivalent in China. There’s about 900 million registered users who have been verified by government records, and they’re digitalized into these interlinks. So the bank transactions, all their financial transactions, media transactions, everything is interlinked to that. So if you are a startup in China and you’re trying to solve a problem, you have to link it to the user’s identity. You don’t get to reinvent the identity. And I think we have that problem, mainly because we solved identity differently in the Western world, and trying to make cross-publisher identity working is still a step further removed, because the identity is about a researcher’s identity or a doctor’s identity, not Taylor & Francis or Wiley’s identity.

KENNEALLY: Sameer Sharif, the challenge there with this identity question is not only who the user is but where they are and which device they are on. This is a layering of the user experience becoming itself more and more complex. So we’ve gone from the desktop to the phone to the watch, and now to Alexa and other voice-based engagement. Talk about the implications for that – the move towards voice engagement.

SHARIF: I think user experience has evolved, and it’s getting better and better. I mean, 10 years ago all we did was engage with our desktops, and it was very impersonal. But as we got onto the mobile phone, it’s a better engagement. But if you think about what Priya said, your applications are at point of care. And if you think about how that nurse or the doctor actually engages at the point of care, they’re still picking up a phone, and they’re still tapping the screen for buttons. Is that the most efficient user experience? No. The most efficient user experience should be basically what we’re doing with Alexa. If the doctor or nurse wants to find a particular piece of information, you should ask that application – hey, I want this.



And that application should immediately go through AI and ML to that datapoint and actually either send it back through voice or send it back on the screen.

So user experience I think is breaking down. I think the voice interface experience is going to be tremendous especially in healthcare and professional settings. So that's really exciting. We are working with Wolters Kluwer on some of that in our innovation lab, and it's really exciting. Because you don't need to tap – there is a friction there between as you deliver your care and trying to get that information so that you can do your job.

The best user experience will be you're delivering care, you've got a question, quickly ask your – it's not a phone. This is a computer. This is not a phone. I mean, you're doing so many things. The phone is just an app in your computer. Literally, the phone is like 1/100 of what this is. This is the computer. So for the doctor, this is your assistant, right? So make your application and the software that you're building the best assistant possible. And from a user experience, voice interface, I'm super excited about that over the next few years, and we're going to see more and more of that kind of user experience.

KENNEALLY: Priya, then that's a nice tee up for you to perhaps share some thoughts on that – that integration of the individual and what we call a computer, but which is something so much more than that, to the network of information.

ARORA: Sure. So I think in medical, the stakes are a bit high, when we say that, OK, we are using machine learning to deliver the content in medical. What could happen is there is a greater benefit that we can save lives, but there is also a chance that somebody can get hurt. So the human to be in the loop, even if machine learning is involved, is really important where we have to – for example, a pharma company coming up with a new drug has to validate and test the efficacy of the drug so it doesn't hurt someone. So we basically have to have a human interaction or a human involvement even while using machine learning.

KENNEALLY: So we will actually get to a little bit more of the question around where the human being comes in or has to be out of the picture. Gerry Grenier, I wanted to ask you about this question that's posed on the slide here when it comes to getting used to multiple users. Can the user be a machine? Tell us about what that implies.

ARORA: Sure, so –

KENNEALLY: Oh, OK, I'm sorry. I was asking Gerry, but we'll get back to you. Oh, sorry.



GRENIER: Well, that's an interesting question. Can the user be a machine? I think to a certain point, it can. I think there's a certain amount of background sifting that can happen to our data that machines can do and lead a user to answers much more quickly. What comes to mind immediately is an opportunity before us that's so simple, yet it's eluded us for a number of years, and that's graphing all of the data in STM publishing, in the STM ecosystem. I think that that's an effort that some of us started with RMap. Because machine processing of information begins with clean data, begins with commonalities among datasets across the ecosystem.

So we're talking here today about machine learning and artificial intelligence, and the stark reality is that all of us still struggle with some very basic data problems, like affiliation IDs, author IDs, to get back to Max's issue. Simple institution IDs, our customers not having common IDs – there are a number of efforts across the ecosystem to come up with different affiliation ID schemes. Sometimes I feel schizophrenic in my job, where I'm approached by one unit that says, use this ID, and another unit says, use that ID, and so and so company is using that ID, and suddenly I'm collecting six different IDs depending upon the purpose. So I think that, yes, there is opportunity for that machine to become the user. But even the machines need good, clean input.

KENNEALLY: Priya, you did have a thought you wanted to share with us on that question about whether the machine can be a user – can be the customer.

ARORA: Sure. So when we talk about the machine being a customer, for us, it's basically having this tool or a service where we are giving it – it's running on its own, essentially. And we have this AI algorithm behind it where it's doing its job for the institute.

KENNEALLY: Well, there's another point on the – sure. Please.

SHARIF: The user being a machine – absolutely 100%. Part of your users are machines. I mean, we're end users. But if you think, what is Google? So if you are delivering a product, you want to get it discovered, who's using your product? Google is using your product to actually display your answer – the search results. Half of your users are machines, and as algorithms and all these applications are out there in the world, you've got to design your product to say that, yes, you've got individual users, like all of us, but then you've got applications and technologies and machines out there that are actually processing or looking at your product to actually create a solution or a thing. So absolutely. I mean, I've been in conferences where you have to design your product thinking that the use – there's the end user, but then there are the intermediate users, which are pretty much machines.



KENNEALLY: Sure, Gerry.

GRENIER: Could I embellish that a bit? Yeah, so I run into the problem constantly of dealing with third-party partners who are developing for us new discovery products. And I can remember 20 years ago when we invested in XML, we thought that XML was the answer. We're going to give everybody XML, and everything's going to be harmonious. I'm going to tell you, we give some of these discovery partners XML and they just look at it and go, what is – they're not unintelligent. But they're coming back to us and asking us for further massaging of that content and delivering things like RDF, delivering JSON to them, rather than the messy XML that drives our printing operation. And I think that – yeah, that's a great point Sameer, is that we are dealing with machines already. We just don't know it. It's one thing to send ambiguous data out to the end user human, but machines at some point require a little less ambiguity.

KENNEALLY: Sure, Max. Please.

GABRIEL: Sorry, I think you struck a very important chord here. Gerry, you're absolutely right. I think we have high tolerance for inaccurate data – humans. But I think when you have more machines starting to consume data, they want it in a certain – machines need to have it in a certain format, which will actually improve the overall data quality.

KENNEALLY: Yet as important as those machines are – increasingly important to this business, IJsbrand Jan – there is this point of accessibility here that we really don't want to miss discussing briefly, which is about, again, it goes back to this diversity of the audience. Expand on that for us.

AALBERSBERG: Yeah, I think accessibility is extremely important. I think that we often are not aware of how many people are using and how many variety of people are using our content. We have done at Elsevier some experiments. Yeah, we are really doing our utmost best to make our content available also for visually impaired, for blind people. It's surprising how many researchers do need all those accessibility hooks that we put in our content, and I think it's important that we make our content available to also those groups of researchers. Yeah, it takes an effort. It really takes an effort to do that. But it's a very important effort to do that.

Accessibility, I think – can I also make one comment? Back when we talked about identity, I think sharing an identity is fairly important, because if we know who the user is, we can offer the user a tremendous set of extra services of extra value. But it's always very important that we also have to allow the user to say I don't want



those services. I don't want to be known. I want to be private. I don't want my identity to be revealed.

I think that is equally important, and that's also one of the challenges for us as publishers – to really navigate in the one hand, creating the value, the personalized value, that we know we can create as much as possible if we know who the user is, but also making a system and making the environment in such a way that if the user doesn't want to reveal their identity, that we still allow for that as well, and also in that context, try to create as much value, but then of course less personalized. And I think in that context, yeah, we will have that RA21 project – we already mentioned that later today. That's really the balance that the RA21 access project is looking for. They want to allow for access in a way that people are not known to the publisher. They want to allow a service that is privacy-secure.

KENNEALLY: Well, I want to stick with you, IJsbrand Jan, because now we're on your favorite topic, which is integrity. It's your job. Yet as much as integrity is a human value, I'd like you to address the question of trust and the machine. It seems to me, to mix some metaphors here, the machine is the elephant in the room. It keeps coming up in all of these discussion around open science and collaboration and so forth. So the collaboration isn't just the members of the band, but perhaps it's the sort of ghost there doing the recording, if you will. So talk about that trust and the machine and what I guess can be sort of lumped together as bots and bias.

AALBERSBERG: Yeah, there is definitely – we can use the machine to increase trust, to increase the integrity by using the machine to allow for checking our content, to validating our content. So there is definitely a lot of help that we can seek in the machine to do so. But nowadays more and more we also get to machine learning, to deep learning, to artificial intelligence, that really takes content, that really takes our content in, to ingest our content to learn from, and then to indeed create new search results of new suggestions for research. And if that content that we put into those machines or that we allow others to put into their machines is not properly validated and is not properly curated, then, yeah, it will be a garbage in, garbage out system.

So the trust in the machine – you can put as much trust in the machine as you trust your content. Even with the machines coming in these days, content becomes or is actually getting more and more important than before. Because the machines digest the content, take it for truth, and it's our job as the publishers to make sure that the content can indeed be trusted. So the machine and the bias, it all starts with very reliable content.



KENNEALLY: Max Gabriel, on that point, there is some interesting side effects of technology that we don't anticipate and yet we see coming at us very quickly, and this is the question of trust and the machine. If I understand it right, you really feel that this can't be an afterthought. This has to be something that is considered from the moment of conception in the design process itself.

GABRIEL: Now, look, it's really encouraging that we're talking about ethics and AI and trust in machine early enough. We never spoke about Internet and ethics or social media and ethics early enough, and we are sort of at the tail end of realizing the side effects of our unintended consequences of those inventions. So it's super encouraging that we are having that dialogue right now, as Elsevier and a few others are setting up a trust and integrity committee around these things. Although I was a little disturbed – Google established their AI ethics committee, and they dissolved it very, very quickly.

KENNEALLY: Overnight, practically.

GABRIEL: Overnight, from what I heard. But the dialogue is absolutely valuable, although I do think we are more suspicious of the machine than the human experts, if you will – you know, day-to-day, there are so many human errors that are being tolerated or not measured. But I think there is this bias about how do I trust the machine and what happens inside? While it's suspicious about it, I think all the proactive steps in terms of what is going into the machine and what is coming out of the machine and being more conscious about it and having credibility around how those decisions are being made is absolutely valid. Otherwise, we'll be at the tail end of receiving the unintended consequences of these inventions.

KENNEALLY: Sameer, sure.

SHARIF: I think it's something that I've been thinking about quite a bit, this whole AI and ML. Can it go in the wrong direction? There's been debates between Mark Zuckerberg and Elon Musk. So I think about it a lot, because we're starting to do some of these things. I mean, we heard Springer just created a book, an AI-based – the computer generated the book. We're using it to create questions on the fly – let the computer actually generate questions, generate answers. Max, you're right. We need to early on create processes and systems to overlook what the machines are doing. I don't know, and we'll see how it transpires over the next decade.

But it is powerful. What the machines can do is quite powerful. Like you said, social media, we're seeing some of the ill effects of uncontrolled growth of social media. I think we as the owners of innovation and content and learning, we need to



make sure that there are systems and checkpoints in place so that it doesn't get uncontrolled.

KENNEALLY: Well, it's an interesting point around these machine-generated works, whether they are books or news articles or songs. A colleague of mine recently blogged about this and asked the question whether a machine can have copyright, which is an interesting question for us at Copyright Clearance Center, and he came to the conclusion that there was no intention. These machines don't have intention when they are creating whatever the work is. So without intention, there can't be copyright. That may be something that is a moving target, though, for the future. But it's a point that comes up here regarding education, and I know, Priya, you have some thoughts about that – that trust, integrity, do touch on issues like copyright and other questions of infringement or the legal right to use materials.

ARORA: So I believe that for our users, they follow basic instincts, so they are going to go to places where they can easily find the content. Max touched upon it a little bit, where we have to make it easier for them to find the content that they are looking for and give them the tools that they need to simplify it for them as well. So whether it be a machine behind it or not, the users don't quite care about it.

As far as the machine, I would like to use the example that if I'm playing chess with the machine and it beats me, then I don't trust it. (laughter) So yes, I want to put some process around fairness and bias, too, and some controls around measuring whether it's giving me – or whether it's producing the right results or not and make sure that we validate it.

KENNEALLY: All right. Well, let's move on to the last player in our band, the collaborators here, which is around infrastructure. Priya, talk about the importance of integration here. We were speaking about talking to machines, but the content also has to talk to its companions.

ARORA: So I think it's important for us to have collaboration in terms of linking for the users. So when they are using different access points to get to the content that they subscribe to, as providers, we have to come up with solutions for them where they can link between what they are looking for. So definitely A is linking and then discovery. Gerry talked a bit about discovery. But we are seeing discovery tools becoming more and more popular at an institutional level especially. Because there are so many resources a library could have from different platforms, and what a discovery tool does is bring it all together for them in one search result. So it's important for them to not lose the search results from one or the other resources that they have – or they should have access to.



KENNEALLY: Right. And, Gerry, there's great opportunity here. This is your part of the song. You really believe in this. There's great opportunity, great challenges, possible advantages to be had. Talk about that.

GRENIER: Advantages to be had in collaboration or –

KENNEALLY: In terms of the kinds of offerings that you could give to authors. Just seems to me –

GRENIER: Oh, right. So yeah, I think that there are two areas that we operate in, right? There's the – I'll call it the legacy area. That's the content distribution. And I think that absolutely we see that there are still opportunities there to collaborate. I think that whole ecosystem is balkanized. But I think coming down the trail pretty quickly are these open researcher tools that are – again, it's all new stuff to most of us. And we need to collaborate not only with ourselves, between ourselves as publishers – when I say ourselves, publishers – but also with the research community, with the library community, and try not to replicate the work that's going on in those different spheres, but rather bring it together and to reduce cost, to develop these new researcher platforms, whatever they may be.

KENNEALLY: Sameer Sharif, this is an opportunity for all players to be involved when there is this collaboration, when there is this shared infrastructure. It's not just the big players. It gives an opportunity to the smaller players as well.

SHARIF: Absolutely. And I think the great thing here is because there's – we live in a shared economy. I mean, if you think about it, Uber has made you have a personal driver. Only the rich people used to have personal drivers. Now we have personal drivers at a touch of your phone. But in terms of – for the publishers, it doesn't matter if you're a \$3 million society publisher or you're the Elseviers or the Wolters Kluwer or the Taylor & Francis or the IEEEs of the world. You can actually build sophisticated solutions cost-effectively because of shared technologies, collaboration, all of that.

And that's the marvels of today's technologies. If you look at where startups are, how come there are so many startups? Because the cost of actually getting an idea, building an application is so cheap. You don't have to get the servers. There's open-source software out there that you can actually build upon. So because of this shared collaborative world that we live in, you can actually build some great applications. Cost is not the – it's not a hindrance anymore if you do it smartly.

KENNEALLY: Right. IJbrand Jan Aalbersberg, the opportunity is there. You can get onto the platforms quickly. Sameer was just saying with startups there's the sort of



ease of starting. But it's the getting to where you want to be that's so difficult, because things are so complex.

AALBERSBERG: Yeah, I think that's – and by the way, I also think that comes back to collaboration and openness. I think if you talk about startups, I think it is very important that we as publishers look at open innovation, that we work with those startups, that we work with research groups even at universities, even before they become a startup, and that we work on open innovation – give assignments or give problems to universities and ask them in an open context to sort out those problems. Not only for a specific publisher, but really for the whole industry. The example of image manipulation detection – that's not something that we a single publisher needs to do, but it needs to be the full industry.

So working with startups, yes, that's extremely important. It is indeed also something that needs to be embedded in the way of working that we do in, for example, the standards and technology group. It needs to be together. It needs to be collaborative with the other publishers. It's not something that we should only focus on publisher by publisher.

KENNEALLY: Max Gabriel, I want to give you the last word on this. The point here is something that also relates to music, where it's about time, right? Getting together like this, it's about time.

GRENIER: Yes, absolutely. But I do wonder, though – I haven't gone to any conference where people disagreed with it. But in terms of – the progress in this area, just hasn't gone to a point where it needs to be. Maybe because of the point Sameer made, maybe it is a double-edged sword. Reinventing it yourself is not that expensive anymore, and people end up reinventing it. People end up building their own platform. And I was just trying to sort of – as a speculation, I was trying to wonder in terms of what is the root cause of this? I think publishing used to be very much a vertical – end-to-end vertically integrated model. It worked beautifully for them, so that you can optimize your cost base and profit margin, all of that.

And I think we haven't – compared to other industries, we haven't navigated, migrated, from the vertical infrastructure to a horizontal infrastructure, so that we don't have to own the end-to-end value chain. There are solutions out there with a combination of great startups offering pointed value-added services and borrowing technologies from other players. We really have to think about how do you make it horizontal and actually figure out an area you really want to compete. So you're not investing capabilities across the board. And we're still trying to resolve the multiple identity issue, because we're all trying to solve the same exact problem.



So I think to make this a reality, hopefully next year – same time next year, we’re still not talking about, yeah, shared infrastructure is a good thing. Each of the publishers and the other stakeholders have to think about what is the area they want to compete, what is the area they want to share, and be very specific about it.

KENNEALLY: And maybe to keep with the music idea, that maybe we tone down some of the competition and raise the volume on the concern for the customer experience.

GABRIEL: Very well said.

KENNEALLY: All right. Well, I want to thank my band here today – IJsbrand Jan Aalbersberg, Priya Arora, Max Gabriel, Gerry Grenier, Sameer Sharif. On behalf of STM and my own employer, Copyright Clearance Center, thank you very much, and I hope we passed the audition.

(applause)

END OF FILE