Re-envisioning the Business of Information: Policies, Practices, and Procedures

A panel discussion featuring

- October Ivins, Principal at Ivins eContent Solutions
- Maryann Martone, Executive Director, FORCE 11 and Professor-in-Residence, Department of Neuroscience, University of California, San Diego
- Britt Mueller, Senior Director, Qualcomm Library and Information Services
- William Trippe, Director of Technology, MIT Press

Moderator
- Christopher Kenneally, Copyright Clearance Center

KENNEALLY: The future, said the science fiction author William Gibson, is already here. It’s just not evenly distributed. And Gibson also said that the problem with the future is you can’t Google it. Now, we’ve all seen the Google Glass, but you can’t Google the future, although I did try to do that this morning. On my Google, the first result page was the bio for a rapper named Future. (laughter) It helped me to learn everything about him, from how many platinum albums he’s won, as well as what he pays in child support. But that’s as far as I think Google can take us when looking into the future.

To re-envision the business of information, which is the charge of this entire conference, it demands that we use our imaginations as well as draw upon our knowledge. Information looks forward. Knowledge is merely a record of the past – of discoveries and experiences gone by.

Now, according to the legend, the dog Nipper, who cocked his ear to the famous gramophone trumpet, as you have on the cover of your conference program – he was hearing his master’s voice. But the melancholy fact was that his master was dead. His new master – the brother of the first – was the painter who captured the endearing image. In 1899, when that scene was caught, technology was, as it is today, an instrument of the imagination. In 2014, if it ever seems to overtake us, then I submit it is only for lack of imagination.

So today our program, Re-envisioning the Business of Information, will focus on policies, practices and procedures. Again, at this conference, you’ve already heard about various new products and just recently looked at some of the new business models that are emerging. But re-envisioning the business of information does not stop with financial issues. One must develop policies, practices and procedures that facilitate – that do not inhibit – the use of content and maximize its value to the
users. Publishers and librarians are urged to move beyond the comfort zone of their current biases and mindsets.

So what are the issues that providers and librarians must consider as this mindset takes hold? What are the implications for content ownership, reuse and sharing of content and even for privacy? We’re going to ask those questions of our panel today.

I want to open the conversation by turning to Maryann Martone. And Maryann, you’re a scientist, you’re skeptical – I like that. I used to be a journalist and, you know, the first rule – the first day at journalism school, they give you an assignment. Your mother says she loves you. Check it out. (laughter)

So I want to check out this notion that the model of the scientific narrative is the core issue here, and to talk about the future, we have to go back to the past. I understand the perspective of Force 11 is that this all started going off the rails in the 1970s and ’80s, and a lot of things did at that time. Fashion sort of went right off the rails.

But scientific research, the way that it was taken on and the way that it was published – something went wrong. Tell us briefly what it was that went wrong, because that points to what the problem is and what the solution may be.

MARTONE: So I’m not sure this is necessarily the perspective of Force 11, because we’re very diverse, but I think a lot of our colleagues that joined Force 11 started to recognize that a lot of the problems that are plaguing science now – the number one being reproducibility, and that’s even starting to make it into the popular press – the fact that the way that we publish our scientific studies, the way that we fund our scientific studies, the way that we evaluate and reward the publishing of scientific studies is leading to the publication of a lot of things which then take many years to expose or refute when, early on, you knew you never had enough subjects to make this claim – you know, you didn’t do the analysis properly, you ignored the negative results, you did all of those things that, when we were in a phase, up until maybe the 1970s and ’80s, where careful observational work – very methodical work that maybe wasn’t completely mechanism-driven, it wasn’t splashy – was still valued and rewarded –

Whether that was the case that there was just as much lack of reproducibility back then and we didn’t know it, or that the types of science that really got a lot of press and were rewarded started to drive the publication of more of these types of studies, I’m not exactly sure.

But I think, if you go and you look at the way that science was conducted back then, you hear phrases like Carole Goble, who’s in computer science research, said,
well, it was more like a country club back then, where you’d get around. There was a lot of interaction. She says now it’s sort of like a sweatshop where it’s, you know, come on, come on, come on, come on – publish, publish, publish, publish. And when you put that sort of pressure on people, scientists, being humans, will respond to that pressure like everyone else does. They will do what gets rewarded. And if what gets rewarded are splashy papers that are mechanistic driven, hypothesis driven in Science and Nature, then that’s what they will publish.

But I think there’s now a reaction against that, saying why, if there’s no money for science and everyone’s complaining about scientific budgets, are we funding suboptimal ways of doing science? And it doesn’t make any sense. So I think that that at least was one of the motivations and one of the things that those of us who are advocating for more transparency, open access to data – it used to be a matter of altruism. It’s like, well, you should let people do this because they funded it and, you know, they should have access to it.

Now I think we have a pushback, which says, yeah, but look what you’re publishing here. OK? If 47 out of 50 major cancer studies cannot be replicated, if no spinal cord injury studies can be replicated, then we can’t defend our practices as optimally serving the domain. And I think that has to be one of the wedges that we use to open up this scientific communication to consider new forms of publications.

KENNEALLY: And one of the areas that you want to see opened up is data and –

MARTONE: Data.

KENNEALLY: – text mining. And here’s where the skepticism comes in. There are publishers and other owners of data who are concerned about opening it up entirely in the way that you would like to see happen. How are you going to get to your ideal state then? Is it just a matter of, you know, putting your foot down and saying this is what we need?

MARTONE: Well, I don’t think full stick ever works. There has to be a set of incentives. But I think – I’ve been sitting here, as many in this room might be, participating in some of the initiatives that are coming from NIH on data and data access and what makes sense for people to do. And so I’ve been writing a lot of this the last few days.

But we know that not all data is going to be equally accessible. But that accessibility should not be based on, well, I perceive this as a particularly valuable dataset, so I’m going to sell it to you. It should be on privacy concerns for public health. There’s certain animal data that is very sensitive and that cannot be widely
distributed. I mean there is data that we understand that’s not going to be open access.

I liked one of the speakers this morning, who said selling services on top of things rather than the content itself. I think that’s a model that just really needs to be explored, because people do put effort into these things, to making them more usable. It is extremely difficult to take things de novo off of instruments and sensors, put it out there and do anything with it. We know that effort needs to be put in there.

But the idea that you would never be able to get access to that underlying data, both for the purposes of verifiability – that you actually did what you said you did, for again, this idea of reproducibility – can anybody uncover mistakes and other sorts of things that are in the data – that’s one of the reluctances of people to expose it is they don’t want to expose their mistakes. But papering over them does not help. Right? Transparency helps find mistakes, correct mistakes. And that, I think, is absolutely part of science and needs to be done.

So I think, in this arena of what makes sense to make accessible, we have to investigate what it is that people are reluctant to do. Why is it? If it’s a purely economic decision, again, I don’t think that the models we establish can allow walling off of large areas of content. I think that that’s just sort of the antithesis of what we need. Whether there are, however, rights on top of that – value-added services and things – I think that’s a perfectly legitimate model to consider.

If it’s reluctance on the part of the researchers, again because they spent 30 years gathering this data and they want to mine it for the rest of their lives, that’s an academic reward issue and again must be addressed. They need to get credit for producing that data for people citing that data and using that data. Right now data producers in science are not valued commodities. They are viewed as sort of less than those who analyze this data and come up with a conclusion. I think that that’s also got to change.

So I think again it’s a multidimensional issue.

KENNEALLY: Right. Bill Trippe, you were talking about those biweekly meetings that you hold with your stakeholders in the – was it the CogNet project or the –

TRIPPE: It’s in the new one (overlapping conversation; inaudible).

KENNEALLY: I’m sorry – in the EducationXPress project. And you must be having discussions around some of these issues that seem very familiar to what Maryann was talking about. Is that true?
TRIPPE: Absolutely. You know, the spirit of the conversation among the researchers is they want the data to get out there. This isn’t sensitive data. They want it to be discoverable. They want to pair the data with useful tools. And they want the researchers to be able to collaborate around this.

KENNEALLY: That’s a challenge for you to respond to, though.

TRIPPE: It’s hard work, absolutely. And so that’s our – I mean great points, because our very question is OK, we really endorse the spirit of all of this, but we still need to make money. So what does our added value – what can our added value be?

KENNEALLY: Is there a sense, though – for this to move forward, all the various players are going to need to accept something less than an ideal situation. In the academic marketplace, that’ll be the case. Certainly, for publishers, you’re going to have to, you know, require that people pay for things and have certain behaviors that they might want to give up. Right? You have to really find compromise.

Is there a response from the academic community that you’re working in that they want to collaborate with publishers?

TRIPPE: Well, you know, what we hear from our management board – the professors – is they want us to experiment – to also help them understand what would work. And, you know, the MOOCs are actually a really fortunate –

KENNEALLY: And we should tell people it’s massive open online course.

TRIPPE: Oh, I’m sorry. Yes – massive open online courses – so edX, Coursera and so on.

KENNEALLY: So you have something like 60,000 students enrolled in a particular MIT class.

TRIPPE: Right.

KENNEALLY: But they’re all over the world.

TRIPPE: Yes. Thank you. Yeah, I’m sorry – being shorthand about that. So we very nervously post a free version of the textbook for 60,000 students to see. But guess what? They still buy the book. And isn’t that fortuitous? So we had the extra work of creating a free version of it. And we had no idea what would happen. But it looks like we don’t have to compromise in that particular circumstance.
KENNEALLY: Right. Well, Britt Mueller, you’re on the corporate side of things. And I wonder how collaboration as a policy and a best practice is an important part of your job.

MUELLER: It’s critical. We are in a corporate environment, but we have a massive corporate (sp?) R&D function, so there is a lot of science going on and a lot of research. On top of that, the business is aligned with the whole scientific and research effort. And that’s not unusual, and I don’t think that’s so different from a university. You have, in a university, a tech transfer process. That tech transfer process within a corporation is extraordinarily tight because, of course, we are producing value for our stockholders. But the real issue here is that scientific and technical content has value to the business side. And the business side certainly has value to the technical and research side. And again I just don’t think this is unusual to a corporation. This works in universities as well. I think that that cooperation is something where the content and the people who are producing that content or where we’re buying that content – and we are willing to buy content, we want to buy content that’s valuable, that’s open to us, that has less boundaries. And again I think that this is very equivalent. I don’t think these are so – such different environments, if you will.

And I think that, especially when you’re looking at analysis of that content, the value of finding out where new technologies are taking place, what those new technologies are, where they’re going, has implications for both sides of our organization. And the information is the same.

So we find that our engineers, who are researching and doing work, are very interested in market data. We know that our business folks are very, very interested in knowing where development of certain ideas are happening within the world, and where we can tap into that. And the technical literature supplies that. So it’s something that’s mined heavily.

KENNEALLY: Well, we saw in the poll of the audience that information overload was a critical concern for people. Talk about that at Qualcomm. How do you manage that fire hose that’s coming at people?

MUELLER: It is a fire hose. And there’s also a very limited amount of time that people have to consume this. We’re in a very tight cycle. So the information has to have value, it has to be quickly used, and we have to also be aware that there is this 80-20 – almost 60-40 rule at times.

When we’re looking in a business environment, we’re looking at saying what is the information that’s critical? What is the information that’s good to have just in
time? What’s the information that’s OK to have that we’re not quite sure about all that accuracy? And there are places for all of that.

So I would say, in terms of the fire hose, the library’s job in particular – and our content vendor partners’ job in particular – is to help us get that information in a targeted way at the point of need. Right? So if somebody needs market data, we’re not just going to throw them a bunch of reports in a database. It’s absolutely not going to happen. We’re actually doing analysis and we’re actually comparing market forecasts. Forecasting is a very important part of our organization, despite the fact that some people feel like we can’t foresee the future. So all of that information is tied very, very closely to the point of need and what people are asking. And we customize it for them.

KENNEALLY: Indeed. Well, you know, Bill Trippe, you’re the director of technology, a new position at MIT Press. And yet, moving into the future, you’re having to deal with practices that linger on from the past. I know you were telling me that you just recently put out a digital first – a digital-only product and encountered some real challenges because of the legacy systems that are there from the print days. Talk about that.

TRIPPE: Yeah. We still have a warehouse. And we still have a fulfillment operation.

KENNEALLY: The libraries don’t have any books, but you’ve got a warehouse?

TRIPPE: Right. (laughter) Buy them, please. So we have this new digital-only offering for the first time. And what we found was that we literally couldn’t place the order. There was no way in the warehouse system to say I’m shipping this digital book. It took us a couple of weeks of real sort of head scratching and, you know, really thoughtful analysis on the part of a couple of the programmers to figure out how to make this work. And I’m sure we’re not alone on this. Maybe a bigger publisher probably encountered this five or seven years ago. We encountered this in the last few weeks. So I had brown hair a few weeks ago. (laughter)

KENNEALLY: And I guess that’s why, although you say you’re still a fanboy of MIT and MIT Press, there’s a lot of pain involved. And the other point that was a thread for people was sustainability and, as much as MIT is committed to open access, how are you facing that particular challenge right now? Is it an inevitable future for you as well that we’ll see full open access?

TRIPPE: That’s really hard to tell. I mean I was really intrigued with Judy Luther’s presentation, because open access on the book side is much newer, and we tilt about 80% books to 20% journals. So I think the business models are much more
mature for journals. So I think it’s a very open question how this will play out for books. But I think, at this point, our position is sort of an and-and is, I guess, the way people refer to it. We work with many partners. And we’re doing a lot of different experiments, so.

KENNEALLY: Yeah. Well, one experiment, October Ivins, that some librarians have is regarding copyright – can I put it that way? They have views of copyright that may not jive with what would be for an attorney at an IP law firm. And I wonder how you view all of that, for example, when it comes to fair use. Is there a particularly aggressive stance that’s being taken right now? And is that coming from a conception of copyright that goes back to the past?

IVINS: Fair use is really important to libraries. It’s kind of what allows libraries to exist. And libraries and librarians see themselves as really central to democracy. We buy individual copies that can then be shared with people who couldn’t afford to buy them. That’s what libraries are about. When I say it’s a service profession, everything that librarians do is to get the right content to the right user. So it matters a lot to them. And there is pushback in the library world about licenses overriding law, so people are being encouraged not to sign a content license that makes you give up your fair use rights.

Librarians are being counseled on how to avoid opting into the Digital Millennium Copyright law, that it actually gives you a choice. If you choose to comply, if you want the protections of the DMLA, you have to do these things. If you don’t want those protections, if you’re happy with how you interpret fair use, you don’t have to do that, you can take another course. So I gave you the reference to the ARL best practices for fair use. And I think the library position is quite different.

One of the other hats I wear, I’m a longtime active member of the Society for Scholarly Publishing, which is sort of – we used to say sister organization, now we’re say affiliated organization with NFAIS, which is because it’s also made up of librarians, publishers, vendors, different content providers. And SSP is very careful to stay neutral, whereas other publishing organizations, like PSP and ALPSP, are lobbying organizations and try to get legislation changed, because the librarian members in SSP are very sensitive to the interpretation of fair use that’s put forward by the CCC.

KENNEALLY: Well, I think I would say this about that, that there is a certain need for balance. And so I wanted to turn to Britt Mueller on that, because you see it from really a different perspective. You share a lot of things here, but in the corporate world there are some different concerns and a different approach to that. And you’ve told me that the kinds of licensing negotiations that you get into are really an important part of your job now.
MUELLER: Oh, yeah. It’s critical. So where we license our content – there’s no fair use in a corporate arena, essentially, so let’s just –

KENNEALLY: Let’s just put that on the table.

MUELLER: Yeah, let’s just understand that. And the CCC helps us quite a bit with this process. But essentially, when we want access to that content, we’re buying it. And we are an IP company ourselves. So we are very strict in making sure that we adhere very closely to those license agreements. Our contracts with our vendors are extensive, and they lay out the rights on both sides.

What I was talking about in terms of – what I was mentioning in terms of how we’re licensing going forward is we’re working in those access rights to the contracts so that we can essentially open up that content without having to seek permissions after the fact. We want most of our permissions to be understood in the contract, so that we can go out without having to go back into a lot of back-end work to say do we have permission to do this, do we have permission to do this? And so vendors who work with us to say, yes, let’s distribute this content.

An interesting point for us for e-books – we decided not to go with aggregators. And the reason we don’t want to go with aggregators is that we want to buy directly from publishers. We want that content. We want to own it in perpetuity. We license that for that. We want to be able to bring the content behind our firewall. We want to mine that content. And those are all things we don’t want to do without the people’s permission. We’re talking to you about how we can do that, so we can get value out of it.

I thought it was really interesting, somebody this morning mentioned that books don’t have the same access rights as journals. And yeah, that’s absolutely true. But if you can pull the full content behind your firewall and mine it, you’ve got it. And that’s a really powerful thing to be able to do. So yeah, we’re looking at that. Are we successful across the board? No. Are we successful with some vendors above other vendors? Yes. Are those the vendors that we’re going to be going back to? Yes.

So when we talk about a new model and a software and service on top of the content, the content is not what’s differentiating you. It is commoditized, it’s valuable, but we can get it in other places. The real value is how can we get access to that content and leverage it for what we need to do. And that’s critical, and that’s making a difference in the world. That’s what’s really making an impact in the world.

And that’s why Google is digitizing books – and whether you agree with that or not. That’s why, you know, Facebook is pushing out a lot of content on their site.
These companies are looking at content and how people get access to it and opening it up for people to use as the business model. And that’s what we need to be thinking about. And that’s how we’re trying to leverage our content with you going forward.

KENNEALLY: Maryann?

MARTONE: So just to bring up that point again, when I first started in this whole process of wanting to be able to mine the literature, I was very shocked to find—and I don’t know if this is still the case, but when you see (inaudible) the University of California licenses with publishers, we had no right to mine full text of the articles. It was you’re allowed to download it for your personal consumption. If we wanted to hire armies of students to download every single PDF that we had in our library collections, then we could do full-text mining. But we weren’t allowed to do it electronically. And again it was a very human-centered—oh, one person’s going to read this article—point of view. But that’s really not what people want.

MUELLER: Yeah. And that’s not how we’re using it either.

KENNEALLY: Yeah. And the notion of the human-centered view of the world—you brought it up, and I want to ask you to explore it a little bit further. This whole machine-readable piece of it is very important to you in the academic world and I’m sure in the corporate world too. But talk about that.

MARTONE: So as I’ve gone through this whole process of learning about scholarly communications—and again it’s been a wonderful opportunity through Force 11—I have seen that there’s sort of this duality to almost everything we do now, which is really propelled by the Internet, that you have access to the huge swaths of information just from your little mobile device. But that means that you’ve exposed it in a way that it can be searched.

Now, for data, that’s a gigantic challenge, we know that, for certain types of information. There’s a lots that’s in the hidden Web, and even search engines haven’t cracked that. But I also noticed, looking at different communities and different even working groups and task forces, that those who understood what a human needed to be able to have a certain amount of that content brought to their attention space to evaluate it for whether this is useful for me or not, tended to develop systems that were very human-centric but ignored the machine processability completely.

A really good example of this in biomedical science is the nomenclature to name genetically modified animals and pieces of DNA. It has subscripts, superscripts, slashes, special characters. If you’re a trained curator, it means a lot. You plug that into any sort of average search engine, and it chokes because it can’t deal with the
special characters. You look, on the other hand, at what people were doing in machine processability, and they were coming up with gigantic URIs and all these kinds of, you know, DOIs and other things. And a human looks at that and goes, well, you know, I don’t understand what that means. That’s something for a machine to interpret.

And so I feel that there again needs to be both of these things satisfied in most cases, that you have to be able to expose your content for something else other than yourself to be able to get it to satisfy a lot of what we envision.

At the same time, those people who are working to make this process better for humans and understandable for humans and those who are working to make it better for machines have to understand that, at some point, they are going to come together.

At some point a human is going to have to interact with this, and therefore there are going to need to be conventions and things that actually sit right in the middle that help that translation to happen, something like even a simple accession number of an identifier which is short enough for a human being to say, oh, yes, that is different from that, but not so long that the brain goes, I don’t know, there’s too many things in there. I can’t understand it.

So I do believe that this duality now informs modern scholarship in a way that it never did in the past. It’s no longer just the humans. It is machines and humans that have to come together.

KENNEALLY: Indeed. And Bill Trippe, for MIT Press, how does that resonate for you? You have been embarking on some new efforts to make curation a part of the work that you do, which is really, of course, a human focus. But it must have an underlying machine-readable interest as well.

TRIPPE: Right. So yeah, we do both. CogNet has a machine-driven taxonomy and some tagging that we do. But we’re, one of the open questions on EducationXPress that we would like to answer is about some of the data curation issues. Can that be machine supported? How much of that has to be human supported? And we’re not a huge organization, obviously, so we can’t bring to bear, you know, what we heard about yesterday from Elsevier, for example. But we’re very interested in the questions.

KENNEALLY: All right. Well, we are interested in the questions from the audience right now. If there are any from the floor, we have some microphones set up, as you know. And if you have a question for our panel, approach a microphone and let us know who the question is for, particularly. Do we have any questions at all for –
Well, October Ivins, I have one last question, and then we’ll go to the final evaluations and so forth. In this brave new world, as you describe it, of the library, where the books are disappearing and the spaces are opening up, it reminds me of a dilemma that has been posed around the acquisition of WhatsApp, the online application, or I should say the smartphone application that allows people to text around the world. It’s valued at $18 billion, and it has something like 75 employees. And in the past, a company with that kind of valuation would have had tens of thousands, if not hundreds of thousands, of employees.

As you were describing these new libraries, what I heard was we didn’t hire any new librarians. We cut back on the books. And there seems to be a contraction going on in there. The spaces are opening up, but everything else is contracting.

Is that a concern?

IVINS: Yeah. I think to a degree it is. I mean the thing is, going back to the recession in 2007, 2008, a lot of library budgets haven’t recovered.

KENNEALLY: But is the technology? I guess the point I was making in trying to bring up the WhatsApp case is that technology just enables people to do so many things so powerfully with only a few people.

IVINS: Right. Well, that’s part of it. But, you know, institutions are investing in infrastructure, in facilities, in faculty. And when the library budgets are squeezed, they’re going to the departments and asking them to pay for resources. Like if you need the *Harvard Business Review* and it’s done an exclusive with EBSCO and it means the library has to buy two different databases, maybe the business school will pay for that. So there’s a lot of negotiation going on.

KENNEALLY: So librarians are getting smarter, then, about how they negotiate –

IVINS: Yeah. Well, and the whole budget structure, you know, somebody, I think Nader (sp?) from Proquest was talking about two years ago there was pushback against subscriptions. People wanted to own things outright.

And now that’s sort of shifted. It depends on where you take the money from. You know, the whole budget structure is shifting to what used to be just serials is now all of the continuing resources. You buy something, you buy it outright, but there’s an annual access fee, which makes it more like a subscription. And there’s all the interdisciplinary use, so subject-funded codes don’t really line up anymore.

So that’s another policies and procedures area that has to be shifted to match how the content is vended and available.
KENNEALLY: I want to thank Bill Trippe, October Ivins, Britt Mueller and Maryann Martone. Thank you again.

END OF FILE