STM US Annual Conference Goes Virtual 28-30 April 2020 – Day 3 Innovation Day

Innovations is now a regular feature of STM conferences. The chair was IJsbrand Jan Aalbersberg of Elsevier, also the Chair of the STM Standards and Technology Executive Committee. He was assisted by a committee of the Future Lab and of course by Eefke Smit

What the user wants: smarter data, smarter machines, smarter science was the strapline from the chair. What researchers should do about data and publisher and funder policies relating to this was central to the programme on this day.

The chair explained the role of the Standards and Technology Committee in STM. Small and large publishers, society and commercial were all represented in it.

The conference for the day began with the **launch of STM Tech Trends 2024 – connecting the dots: it's all about the user.**

The graphic (infograph) can be found at https://www.stm-assoc.org/standards-technology/stm-tech-trends-2024-focus-on-the-user-connect-the-dots It is an open knowledge graph too (said Heather Staines) later and can be added to "verify" assertions.   Eefke Smit, Technology and Standards Director for STM explained her methodology. She stressed the need to take into account the new digital natives with their new demands characterised as go straight there and do not waste time – instant gratification – but part of this is the need for precision information. Trust is very important and so is trust verification. What publishers do must fit in with researcher workflows.
This was followed up by the user if now central: **What these tech trends mean for your business?**

This panel was moderated with characteristic aplomb and interactive skill by Chris Kenneally, Director of Content Marketing for the Copyright Clearance Center, (CCC). They were connecting up the dots on the graph: he said and he referred regularly to it. Heather Staines, who is now CEO of the Knowledge Futures Group, went first. This collaboration of MIT Press and MIT Media Lab has a great line in advocacy for their services: https://www.knowledgefutures.org/about. Why is trust so important asked Kenneally? Researchers have to be able to achieve reproducibility and predatory journals cannot be afforded. Increasing transparency in peer review essential. Younger researchers want to share quickly and they are more open to trying more things. Kenneally also asked about the impact of current disruption. As an industry this is a follow-on from move digital and open access which has given us resilience said Staines.

Renny Guida (IEEE) was asked about data. It has to be more accessible and again the concept of a trusted source. Kenneally riffed on trust and brand. The entire publishing ecosystem has to be trusted. Standards are central. They do a lot of work at IEEE with NISO and STM. As a membership organisation mentorship must be important suggested Kenneally. Maybe the younger researchers want their information in forms other than the article said Guida: publishers have to deliver the solutions to these needs.

Kenneally said we are going to have a lot more than the simple "flat article" (his words). For years the article has been central and it has been trusted. Liz Marchant (Taylor & Francis) made clear that for new researchers the article remains important even with page numbers – her example of conservatism. Now we can look further into bits of the article – a graph or methods for example. She hoped we can bring back serendipity. Authors are customers (it was agreed – a by-product of OA) and publishers have had to respond. We are not yet there with sign posting for the reader. Marchant felt that there was already a change in the way workers in publishing behave – they come to work hoping to do something good.

Sameer Shariff, CEO of Impelsys, was certain that personalising tools are what users need. He has had a revelation from working at home where he is trying to teach Netflix what he likes. He does not want it to get confused if his daughter uses his login. Publishing needs to make it as easy for the researcher as Netflix does – better value. That is what the digi-native expects. Publishers now will all go fully digital – these revenues are writing and now is the good time. Kenneally thinks that voice activation will come in just as mobile is now the technology of choice: this is part of the evolution of the interface. There will be different types of device. It is exciting.

There was a poll of attendees which suggested that what users want a simple interface which was easy to use and not personalised…

US Launch of the STM 2020 Research Data Year: Share – Link – Cite came next. This was a round table discussion, moderated by Joris van Rossum, Research Data Director, STM Association Michael Huerta from National Institutes of Health (NIH) and National Library of Medicine (NLM) lives at this intersection between data and open science. There is a strategic plan for data science for both organisations. There is a draft policy for all those who are funded by NIH. It is being finalised. It is a mandate but not an unfunded one and any costs will be allowed for in the grant. There is another draft policy going through the approval process. NIH wants data put in open domain-specific data repositories with PubMed Center as a default and also (for larger data sets) Mendeley and others. There are stiff criteria for these repositories. For discovery there is another committee working on how to handle this distributed system.

David Mellor of the Center for Open Science is involved in reproducibility projects which are crucial for scientific credibility and understanding of the barriers which do not make it possible. The TOP guidelines are central to FAIR principles (https://www.cos.io/top)

Ian Hrynaszkiewicz of PLOS argued for the policies of the Research Data Alliance as a framework. He was heavy on practicalities. We can see that when publishers implement a mandating policy for data availability statements this works. One angle which came up in questions was the costs of reviewing data. PLOS offered three to four hours as the time involved.

One angle which came up in questions was the costs of reviewing data. PLOS offered three to four hours as the time involved.

Debbie Sweet of Cell Press (Elsevier) explained www.stm-reseachdata.org, and confirmed that mandating data availability does work. Not all data can be shared: it depends on what is normal in their discipline. She has a paragraph in the programme.

It was not clear to me how these various organisations and their policies fit together. There was no obvious disagreement but experts will no doubt see differences in emphasis. It is now clear that publishers are not trying to monetise data so there is no reason why there should be outright disagreement among stakeholders. Van Rossum was positive about sharing FAIR culture but not that many researchers know about the principles.

There was another presentation on **Seamless Access** – see Society Day for links. Ralph Youngen (ACS) argued for a renewed urgency in pressing platform providers.

The last session was entitled **Fairness in data science versus the danger of the autonomous bias** moderated by Chris Graf, director for Research Integrity and Publishing Ethics at Wiley.

The powerful keynote by Julia Stoyanovich of NYU emphasised wider societal issues. A 2019 course piece from her is available at https://cusp.nyu.edu/past-events/julia-stoyanovich/ which covers the some of the same area. Responsible data science is the only sort there should be. Currently a dataset is a mirror of a world which is not as it should be. Is it a faithful image of a broken world or a broken image of an OK world? A human has to intervene and it cannot be left to computers which cannot know. There were a lot of convincing examples from Google ads and Amazon. How is the use of data and automation impacting individuals? How can individuals question an algorithmic decision? Algorithmic transparency is not the same as making the source code open. She is working on systems which might provide an answer. Members of the public have to scrutinise such systems.

Graf quoted Burgelman – all science is now data science. Stoyanovich appeared to agree
The points raised were discussed by a panel with was composed of Merce Crosas of Harvard and Jabe Wilson of Elsevier.

For Crosas see scholar.harvard.edu/mercecrosas. Her main points were as follows. The data and the code have to come together and must be both available in repositories. Openness not always possible and there are ways of transparency being possible for "sensitive" data – what are requirements for data being shared or not? Raw data can be still held if there are privacy guards: they were working on this. These are steps which represent a pragmatic move forward from what the keynote outlines.

Wilson agreed that bias can be introduced at any point in the AI workflow. He gave examples from Elsevier work on this topic.

Kent Anderson moderated and dominated the final session entitled **internet paradigms lost, internet paradigms regained**

For him the original free internet aim has transformed into surveillance capitalism and hopes have soured: this was the thought behind the session.

He likes the subscription model. It works for him now that he no longer works for a large organisation.

You can find his recent ideas at http://www.caldera-publishing.com including some excerpts from his newsletter in case you want to subscribe.

Discussants were Phaedra Cress whose view on predatory publishing are available at https://academic.oup.com/asjopenforum/article/1/1/ojz001/5366231 and Josh Nicholson of Scite.ai who explained about his company.