ALPSP 2017, Noordwijk, The Netherlands

Day 1 Amsterdam: met up with colleagues and Adrian Stanley, a VP of Digital Science. Main topic was the forthcoming ReadCube Discovery initiative, which CALJ is initiating. Also talked about the CALJ Readership Analytics project.

Day 2 Train and Bus to Noordwijk. Lunch included Toby Green publisher of OECD materials, who has just written an article on the failure of OA. Toby has expanded their successful "Freemium" operational model (read for free, anything more requires payment). And he is aggregating NGO publishing through the OECD. This is a very positive development since most NGOs have unsustainable publishing models.

Chat with Phill Jones of Digital Science. I agreed to send him a copy of the CALJ Readership Analytics study for comment and an expression of interest by Digital Science to provide CALJ with analytical services. Looks promising, one would think they could manage the work we need done and might be able to pay the foundations for adding value. AI

Keynote speakers: Trust, Truth, and Scholarly Publishing
Together they provided a broad view of the status of journals in the context of science. The first paper given by a health researcher saw that the largest problem facing researchers and research communication was not plagiarism or predatory journals but research integrity. The main factor he identified as crucial was the general climate in the research community which has been brought about in part by rewards for publication, and more specifically, the number of publications with no nuanced rewards for telling the whole truth inclusive of failures and negative results. He noted that researchers inevitably are selective in what they publish, a behaviour that biases the research record. His notion was that the matter might be based in a commitment to "research integrity" and increased transparency through the publication of accessible data. In short he called for a redesign of the rewards system and the involvement of the whole range of actors in the research and research communication process.

The second keynote was delivered by a former director of the US Office of Science and Technology Policy, Kei Koizumi. His basic message was how the US was entering an era not unlike that experienced in recent years in Canada. Research data was being suppressed in large part through funding cuts of science and spending in such areas as Homeland Security. Thus policy is ceasing to be driven by information. He used the term "science free policy." He noted that there has been a growing distrust of scientific, including social scientific information but that it has been much heightened in the Trump administration where there has been a direct suppression of uncomfortable and inconvenient information. This is damaging to the public trust of research and the distrust is taking a general form and is liable to reduce the value of journals. His overall assessment is that the research infrastructure is battered but still standing largely bulwarked by the public access plans that were instituted during the Obama years. In other words, public access works against the erosion of trust.

Plenary 1: Learned Societies, Navigating a Sea of Change
This first plenary addressed how learned societies are navigating changes in the generation and exploitation of information. The general point made by both speakers was how learned societies were changing their roles to regain a significant space in the public mind. Hetan Shah, executive director of the World Statistical Society made two points. First was that the traditional three-legged stool of support, membership, meetings and publication, still governed much of what the WSS does but that the approach they were taking is considerably more innovative. Starting with the notion that learned societies have no inherent right to exist, he has developed and implemented a strategic plan that aims to strengthen the discipline and the profession, increases

statistical literacy in the general public, and puts forward statistics in the public interest. He has also professionalized the society rather than running it largely with volunteer labour. With respect to journals he has required journal service providers to bid on a contract for services. He utilizes a commercial mindset to cross-fund membership, uses members as statistical ambassadors, and generally involves the membership in the cause of the discipline.

The second speaker, Alix Vance of GeoScience World – an amalgam of geology oriented societies – also emphasized a commercial outlook as part and parcel of being a not-for-profit publisher underlining the benefits of scale that arise from collaboration. The use a shared risk/reward structure for their publication of journals, books and data; are involved in innovation of geoscience tools, and look for both horizontal and vertical growth: A good combination of DIY and commercial publishing.

The short list for the Innovation Awards on ALPSP website
Kudos
Escalex
Publons
Delta Think
INASP Author Aid

SECOND DAY
Plenary 2: Rethinking Publication: The drivers, technologies and financing behind bold evolutions
The panel members addressed "the article economy and have we achieved it." The meaning of this phrase was unclear but in general all the speakers agreed that while lots of changes are taking place, the publication of articles remains central because scholars are rewarded for the number of articles publish and little else, not negative results, not tool development, not data reanalysis, bot replication. Brand, Impact, non-article-shaped activity is not getting out, old time peer review is still in place, OA growth has flattened, and while the west is changing developing countries are still trying to catch up to the traditional system.

One speaker noted that the pirates (including predatory publishers) are taking the OA road to create different business and subvert the whole. It was noted that a version of PLOS exists behind SciHub as well as on its own site but it does not seem to be draining traffic away from PLOS. It was noted that SciHub is a *cul de sac* in that the built-in outlinks to other articles disappear – they must be copied and pasted. While reader feedback has been introduced there does not seem to be a lot of uptake. Preprint servers are emerging everywhere i.e., socRxiv, bioRxiv but it is not clear what their eventual role will be. Later in the conference it was noted that there really isn't an emerged business model: there are costs with no income.

Just as in Open Access, everyone agrees that Open Science is a great desideratum but if built on the traditional model of journal publishing (i.e., earning money in the marketplace) it is unsustainable. Again here, later in the program a rep from the Gates Foundation noted that not only can the people funded by Gates take advantage of the publishing of Gates Open Science which has a very quick peer review process and immediate OA publication, but also, projects funded by Gates can get enough funds to pay APCs.

Implicit in much of the discussion was a call for a restructuring of incentives so that scholars are recognized to things other than article publication. In science, at least, there is a great deal of critically important information not being published: data, tools, protocols because of inadequate incentives. Also called for here and later again and again was the funding to be available to pay

the costs of publishing as the Gates Foundation does. It was refreshing for that discussion not to be dominated by the consumer-based demands of librarians for free goods.

Note was taken of the need for transparency by the availability of data. Two derivative points emerged in this regard. One was the need for data analytics. Of far greater use is analyzed data unless one is oriented to the issue of reproducibility given that society and researchers are already suffering from information overload. In a presentation by the New England Journal of Medicine it was noted that over 200 health researcher groups responded to a contest to try to reanalyze data. Some managed to generate some interesting hypotheses for further research.

John Inglis talked about the establishment of bioRxiv by the Cold Spring Harbor Lab and the establishment of an easy transfer of articles from that preprint server to various journals. In general, hiring and tenure committees in science as well as funders are treating lodging articles in preprint servers as necessary but not sufficient behaviour. Generally speaking, the opinion was that preprint servers do play a value-enhancing role in science. Articles do get feedback and changes are often made before journal submission. Plus, researchers use them daily to keep up with what is happening in their field. Other points:
- preprint: circ prior to peer review,
- server vs service;
- servers as a stand-alone mechanism
- no business model for preprint servers
- preprints are being recognized for grant getting and tenure in science
- will preprint servers be gamed by fraudulent postings (in order to undermine the influence of science)?

This talk caused me to wonder whether Compute Canada would be interested in providing a pre-print server for Canadian SSH research.

Liz Allen of F1000, an organization that suggests articles relevant to life science researchers, provides research writing and management tools, and a platform for open science publishing explored how researchers can share more of what they are doing noting that metrics are built around articles – not research, e.g., data, notes, guidelines, etc. She mentioned the development of an open research platform by F1000 for the Montreal Neurological Institute.

Hannah from Wellcome noted that Wellcome is intervening in the research communication process to try to reshape research communication to meet its goals and needs which, generally are to make health research publicly available and to increase the recognition of other info/research activity using Wellcome brand – i.e., building legitimacy.

Kristan Ratan from the Collaborative Knowledge Foundation CoKo talked in terms of an emerging research ecosystem composed of a constellation of research outputs saying that in general we are far from achieving such a state of play. To help this along, CoKo has developed an open source code library to help researchers report research activities other than the distillation of research into a "dumbed down" (sic) article format. One such tool is WAX, an html editor. The problem here is that researchers are focused on data not communication tools.

In a session labelled Industry Updates the audience was introduced to a monograph publishing services provider called GlassTree. Daniel Berse outlined the services they provide which parallel those provided by OJS for journals. He overemphasized author earnings but also mentioned cost reductions for students. Their motto: ROI or perish, not publish or perish.

SAM BRUINSMA of RA21 described a NISO/STM initiative aimed at resolving issues that create barriers to moving to federated identity in place of IP address authentication (which, presumably, make theft possible. Evidently, some products and services are available in the identity discovery space today.

Tasha Mellins-Cohen outlined the activities of Project Cupcake which is oriented to the development of metrics for the assessment of the quality of scholarly (not science) journals. It aims to create a database of information about pub services and purchase capturing and conveying the quality of service provided by scholarly publishers to their customers - authors, funders, and subscribers.

▪▪▪

Andrew Pitts reviewed security issues with emphasis on SciHub

Editorial: Jo Adetunji
The Conversation: "matches journalists with researchers in the UK." It is funded by UK universities and is an attempt to increase public engagement aimed at building public trust in media and research. The speaker claimed that there was a branch in Canada however, when I looked up Canada, the vast major of what was written was by UK journalists. On the other hand, Canadians wrote about non-Canadian issues. The Conversation is akin to a digital wire service generating Creative Commons content and contains code that allows republication to be tracked. It also provides writing tools to help academics to write for the general public. They also run a column called Fact Check to debunk lies and misrepresentations.
theconversation.com/uk

Heather Staines reviewed the annotation technology, Hypothes.is. Basically it is software that allows for layers of annotation e.g., for the public, for an expert community, etc. and they are working towards an annotation standard. They have an overall concern with the ability of students to evaluate digital sources and to promote the understanding of digital polarization algorithms.

AI:
Sam Herbert from 67 Bricks gave a simple introduction to AI. His examples of applications to journal publishing were a bit weak (identifying reviewers, improving marketing communications (targeting)) but he certainly got the basics of AI across. He talked about machine learning and how it can lead to, for example, computers identifying breast cancer with greater reliability than trained human being.

Ruth Pickering of Yewno described how they created searching software to query a database composed of the digitized content of the Stanford library. It was very impressive. A student could type in The Faerie Queen and be presented with of screen left the context of the work and then all the literature that the library holds on the book. A student might then create a new node (e.g., Queen Elizabeth I) and then a third and see information further afield that would add to the richness of his/her inquiry. Apparently, the search process can take the user to the exact pages relevant in the initial query. She explained that what they were trying to do is model a neural network. One was left with the feeling that a Stanford student would outperform students from other universities purely on the availability of this tool.

AI-2 Two Publishing Case Studies

INSPEC is an indexer of engineering content encompassing a considerable collection of journal content (maybe an exhaustive set). David Smith told how INSPEC had digitized its database and then subjected the content to queries. HE described the results as spectacular in that they were exhaustive and captured all the relevant content in addressing a query. He noted that they did not use the whole article but did not say explicitly whether it was just the abstract or whether it was an INSPEC-created abstract.

MARCEL KARNSTEDT-HULPUS described some database work being done on Springer Nature content. He claimed the ability to produce a structured world of knowledge built upon an ocean of semantically enriched and integrated facts. Said differently, he offered an intelligent search that goes beyond keyword search and addressed users' intents and mapping their queries to entities of different types. He saw the future as facilitating interactive knowledge exploration, smart browsing and serendipitous discovery. He talked of (semi-)automated entity recognition in natural-language texts, over taxonomies, and advanced machine-learning techniques and data analytics like graph & network mining and the resulting benefits. The proof will be in the pudding.

I skipped the session on the Belgian RRO's loss to HP on trying to get tariffs applied to copying machinery. Luckily Canada has largely avoided such levies except on audio recording matter.

Day 3
Plenary 4 Maximizing the Value of Research Findings and Data: Cross community innovation

Carly Strasser from the Moore Foundation (Moore's Law) described the range of projects that they supported. She emphasized data driven discovery and not surprisingly, given Moore's illustrious applied career, the thrust of their support is aimed at giving recognition to other research activities beyond article publication, e.g., tool development, code, etc. She introduced their project development software, the Jupyter interactive lab workshop which they have placed on GitHub (so it can be downloaded); they are monitoring developments on "open access" and working at estimating the costs of implementing OA (costs of bioRxiv).

Ashley Farley from the Gates Foundation talked about its activities and its commitment to the open sharing of information, note not Open Access per se. All research funded by Gates is published CC BY and is made immediately available (including the data). Gates covers APCs and insists on no exceptions so it generates high compliance. They are launching this fall what sounds like a new publishing operation that they probably world not call a journal. Its name is Gates Open Access. I'm not sure if this is actually Gates Open Research which was announced in March 2017. I suspect so. They expect to publish quickly turning around a submitted article within 10 days with open review (not clear what review incentives are in place if any). They intend to measure impact and address the challenges in open science and data management head on.

Pam Miller from the New England Journal of medicine described the project summarized earlier in this report. I think it was she who talked about scientific ecosystems and responsibility.

Marcus Munafo, prof of biological psychology, reviewed research practices quite thoroughly noting that a general practice was really unacceptable, specifically the generation of post-data collection hypotheses. Many see this as absolutely acceptable. It is statistically unsound. It generates hypotheses rather than testing them. 90 of psychological articles found what they hypothesized. He suggested prior registration of hypothesis and methodology and a journal commitment to publication prior to data gathering. It won't happen. He suggested the creation of quality control badges on journals; he argued that openness in review and through the publication

of data helps to control quality; he suggested the alignment of journals and funders prior to experimentation; and that because replication is rarely published, science is not very self correcting.

Plenary 5 AI The robots are coming Chair: Isabel Thompson

Joanna Bryson, a prof. from Bath, argued that sustainability and inequality are the key issues not robots. She also agued for human responsibility in anything that robots or, for that matter, any AI or algorithm did.

Sebastian Huempfer from Echobox described his company as providing intelligent automation. It distills data and designs strategy for identified audiences. They provide the headlines for various online sites such as *The Guardian*, which drives readers to the main content. They are also social media consultants.

Venture capitalist, Volker Hirsch ran through a set of alarming developments in the replacement of human workers by robots (and similar disruptive scenarios). E.g., China replaced 60,000 workers who were earning $20 per day with robots. He noted that data vastly accelerates AI because it provides the foundation for AI. Robots are now teaching robots. The program, Deepface performs better on face recognition than human beings. Machines can invent new languages that work for them but not humans. Lawyers are being replaced by AI, anything involving pattern recognition is probably best handled by a machine.

The presentations were followed by a lively discussion.