

Scholarly Social Machines

David De Roure^{1, 2}

¹Oxford e-Research Centre, University of Oxford, Oxford, UK

²david.deroure@oerc.ox.ac.uk

Identifier: <http://www.oerc.ox.ac.uk/sites/default/files/users/user384/scholarly-social-machines.html>

In Reply To: <https://linkedresearch.org/calls>

Despite many attempts to perturb a scholarly publishing system that is over 350 years old, it feels pretty much like business as usual.^[1] Here I question whether we have become trapped inside the machine, and argue that if we want to change anything in an informed way then we need to step outside and take a look. How do we do this? First I describe what I mean by a social machine, and the “scholarly social machines ecosystem”. The article closes with a list of questions that I believe we need to be asking.

The evolutionary growth of new social engines

Once upon a time, interacting with digital content was an option, as was turning to social networking sites to communicate with friends and colleagues. Today our lives are mandatorily mediated by technology that enables academic, social, economic and cultural interactions at scale. Our widespread adoption of Web, laptop and smartphone, with many more devices still to come, means we find ourselves living in interleaved physical and virtual worlds.

The design and analysis of these socio-technical systems has attracted much academic attention, exploring both social science and computer science perspectives. Here we focus on one model in particular, because it is an abstraction that underpins the Web—it is the *Social Machine*.

Tim Berners-Lee provides a definition of Social Machines in his book *Weaving the Web*.^[2] “processes in which the people do the creative work and the machine does the administration”. A less quoted but more complete definition follows in the same passage: “The stage is set for an evolutionary growth of new social engines. The ability to create new forms of social process would be given to the world at large, and development would be rapid.” Written in 1999, Berners-Lee was already anticipating the “social engines” like Wikipedia and twitter that were to follow over several years.

SOCIAM

In 2012 a consortium of UK universities embarked on the SOCIAM project, its five year mission to explore *The Theory and Practice of Social Machines*. The SOCIAM team started their journey by identifying individual social machines to study, and – like true explorers – endeavouring to identify and categorise the social machines that were out there in the jungle. For example, Wikipedia became a popular embodiment of the notion of the Social Machine – an open platform, operating at scale, widely known and observed, clearly socially constituted, and complete with a crowd and automation.

The SOCIAM project went on to study many others, and especially Zooniverse, which evolved from the Galaxy Zoo citizen science site into a kind of social machine factory. In its latest incarnation it's a platform that empowers citizens to create their own social machines — recalling that empowerment underpins that original definition.

Significantly, Zooniverse represents a new way of conducting scholarship, exploiting the new affordances of the digital, especially scale, automation and empowerment. Can our knowledge infrastructure^[3] cope with this shift in scholarship? We return to this question later.

The Scholarly Social Machines Ecosystem

Studying individual machines is clearly important in order to understand how to build them. But over time citizens typically engage with more than one social machine, and really we have a socially-coupled ecosystem of social machines. We need to understand this ecosystem: as designers we are not really creating standalone social machines, but making an intervention in the ecosystem with an intended outcome in mind (and what happens might be completely different).

Interested in the ecosystem angle from the outset, I observed the auto-ethnographic opportunity: we are all engaged in a “scholarly social machines” ecosystem. We author, review and publish; we generate born-digital content and repositories to put it in; we discover and read and recommend; we crowd-source our research and we engage the public. Also we use software for our research, and this lives in an adjacent region of the ecosystem, the land with github, stack overflow and other social machines coupled by developers and research software engineers.

In scholarly communications, traditional centralised monolithic processes looked set to give way to a vibrant ecosystem of new intermediaries throughout the research lifecycle and for every aspect of communication — and significantly they are available for us (and our service providers) to select and to assemble, joined up by DOIs, APIs and ORCIDs. For me it is this very ability to assemble, reconfigure and repurpose social machines that makes them distinctive in the landscape of sociotechnical models. This is not to say that

scholarly social machines need to be mediated by IT: we have also looked at a historical perspective, not just pre-web but early modern.^[4]

So I made an early slide with the logos of various tools, websites, platforms and publishers that were being promoted at events like the FORCE conferences. Since then I've spotted many similar slides – but the logos change quickly, because this ecosystem is quite dynamic, and natural selection is at play. We see disintermediation and new intermediaries, at various granularities. And we see historical intermediaries, like publishers, acting to avoid disintermediation—I once called this phenomenon ‘antidisintermediationism’.

Trapped inside the machine

Social machines give us a lens and an opportunity for academic insight into a vital ecosystem, but in practice this ecosystem hasn't attracted the attention I hoped. Reflecting on this, I think it might be precisely because we are inside the machine and find it hard to step outside and take a look at ourselves.

We still talk in traditional terms. We talk about data but forget about software. We don't discuss how citizen science doesn't fit very well. When time or resource for change is limited, everything looks like open access (yet again) and the parallel world of open data that has been invented to mimic it. We forget about cultural publishing differences and look for one size fits all. And we write yet more reports that say pretty much the same things.

And then there's the Catch-22: the way we try to tackle the problem is to use traditional publishing, to use the very machines that we believe are flawed. For example, you are (probably) reading this article in the existing social machine (and if you're not then congratulations, you escaped! And I offer you this piece as a historical artifact from an uncertain time, with an uncertain archive, and congratulate you on a miracle of preservation and discovery).

And then there are the antidisintermediationists, who would rather everyone stayed inside publishing as we know it, seducing us with the familiarity of a revamped status quo instead of a radical rethink.

The view from outside

But I believe we must step outside, because as long as we are inside we are not asking important and hard bigger questions. Here are some examples:

1. We know that the *real-time* data supply to our research is going to increase dramatically, but have we really thought about what this will do to the ecosystem? Is our knowledge infrastructure ready? Have we rehearsed the methods, and if so where?
2. Can we achieve the full potential of shifts in scholarship, such as citizen science and its augmentation through machine learning, and facilitate rather than constrain further innovation?

3. Are our teams ready? Can you tell me what research team sizes we will be working with in the future? What specialists do we need? How restricted are we by disciplinary silos?
4. How much will be automated? What percentage of academic content will be produced by machine? Consumed by machine?
5. Which components and processes will become obsolete? Are we ready to replace rather than revamp? Will policy interventions be effective, and will they have unexpected side effects? e.g. What percentage of publications need to comply with FAIR or data citation principles to have a useful effect?
6. And once we figure out what we need to do, how do we figure out the best interventions to achieve it?
7. How do we use social machines as an abstraction that helps describe, understand, analyse, and model the scholarly social machines ecosystem?
8. And finally, how do we evidence the optimum granularities in our scholarly communications ecosystem on the spectrum between extreme decentralisation, which aims to empower the individual and community, and massive monolithic social platforms which harness collective energies to benefit a smaller constituency?

Perhaps it will help if we look at a different ecosystem and then turn round and look back at ours. The software ecosystem is an excellent exemplar but still quite close. So I offer the social machines of music: downloads, streaming, music recognition, music publishing, uploads, fandom. Why is this relevant? Well for one thing, the music industry has “gone digital” end to (nearly) end, in a way that science still aspires to.^[5] It’s about planning, performance, recording, production, distribution, discovery, delivery, consumption and reuse. It’s about creativity, and fundamentally it’s about people.

Acknowledgements

I am grateful to many colleagues who have engaged in discussions about scholarly social machines, including Dave Murray-Rust, Ségolène Tarte, Pip Willcox, and the participants in the *Social Humanities* workshop at the Digital Humanities Oxford Summer School 2016.

This article is a response to the Call for Linked Research.

References

1. De Roure, D., (2014). *The future of scholarly communications*. Insights. 27(3), pp.233–238. doi: 10.1629/2048-7754.171
2. Tim Berners-Lee, Mark Fischetti. 1999. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor* (1st ed.). Harper San Francisco.
3. Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman,

C. L., Ribes, D., Burton, M., & Calvert, S. (2013) *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Ann Arbor: Deep Blue. <http://hdl.handle.net/2027.42/97552>.

4. David De Roure and Pip Willcox (2015). *Coniunction, with the participation of Society: Citizens, Scale, and Scholarly Social Machines*. Scholarly Communications Workshop, Boston, MA. April 2015. Available on <http://www.academia.edu/12103878/>

5. David De Roure, Graham Klyne, Kevin R. Page, John Pybus, David M. Weigl, Matthew Wilcoxson, Pip Willcox (2016). *Plans and performances: Parallels in the production of science and music*. 2016 IEEE 12th International Conference on e-Science, Baltimore, MD, 2016, pp. 185-192. doi: 10.1109/eScience.2016.7870899