

The future of scholarly communications

Based on a paper presented at the 37th UKSG Conference, Harrogate, April 2014

The academic publishing industry is set to celebrate 350 years of peer-reviewed scientific journals. However, there are significant shifts in the practice of scholarship, as scholars and citizens alike participate in an increasingly digital world. Is the scholarly article still fit for its purpose in this data-driven world, with new interdisciplinary methodologies and increasing automation? How might it be enhanced or replaced with new kinds of digital *research objects*, so as not to restrict innovation but rather create a flourishing sense-making network of humans and machines? The emerging paradigm of social machines provides a lens onto future developments in scholarship and scholarly collaboration, as we live and study in a hybrid physical-digital sociotechnical system of enormous and growing scale.

Introduction

In 1665 there was a revolution. The Royal Society of London launched its *Philosophical Transactions of the Royal Society*, the world's first journal of science. This mode of scholarly communication, so natural to us now, was created to further the Royal Society's mission to promote knowledge. It invited and encouraged them, and us, to stand on the shoulders of giants. The story of open science had begun.

Scholarly practice has changed since then, most evidently with the adoption of digital and computational methods, yet the principles and presentation of *Philosophical Transactions* and all its successors have remained largely the same. After 350 years of successful and continuous publication of peer-reviewed science, academic publishing has become an international industry. The question might reasonably be asked: is this model of scholarly communication still fit for purpose? Yet it has been robust in the face of change. While the web was invented at CERN to enhance science communication – incidentally affecting so much of our lives as a consequence – scientific publication still looks remarkably as it did in years past. Hence we might also observe that it must be doing something right.

This article looks at this dilemma: at the shifts in scholarship due to the increasing scale of digital research and societal engagement; the fitness of today's scholarly article for its purpose; and at how we might enhance the idea of the article so as not to restrict innovation. It also provides a lens onto future developments in scholarship and scholarly collaboration, through the emerging paradigm of *social machines*.

Scholarship at scale

Citizens and scholars alike interact with the digital world every day, explicitly through apps or websites, and implicitly through the digital transactions of our daily lives as we interact with the devices around us. We live in a hybrid physical-digital sociotechnical system of enormous and growing scale, characterized in Figure 1.

The technology axis, forecast by the electronics industry, is well understood – it is essentially Moore's Law. This means we see growing computational capacity that affords real-time data analysis, and an increasing number of devices in our environments (at home, in the laboratory, or the field) through the 'internet of things'. Meanwhile progress along the social axis has been rapid and has given us social computing, Science 2.0¹ and the collective wisdom of the crowd.



DAVID DE ROURE

Professor of
e-Research
University of Oxford

"... Is this model of scholarly communication still fit for purpose?"

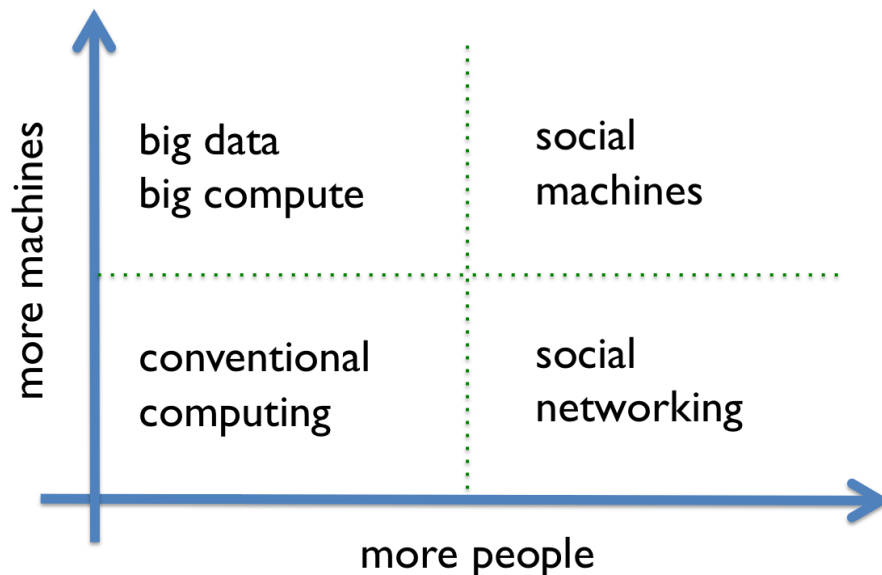


Figure 1. Society and technology unite at scale in the social machines quadrant

Here we focus on the fourth, top right quadrant, where the crowd meets the digital world in our scaled-up information society. This is where digital transactions and interactions give rise to the new forms of 'big data' that we study, from supermarket loyalty cards and social media to citizens collecting data. It is also where we see new methodologies for study and innovation, such as citizen science, collaborative design and social editing. While this growing societal engagement characterizes today's fourth quadrant research, we can anticipate an increasingly automated future – we will run out of humans and yet the technology axis goes on.

The fourth quadrant is complex and hugely important, and we are beginning to understand its knowledge infrastructure through the notion of *social machines*. (We return to these later.)

Shifts in scholarship

The trends in scholarship toward data-driven or data-intensive science are well documented, for example through the collection of essays in Microsoft's *Fourth Paradigm*². New capabilities provide researchers with additional tools, though some commentators take the stronger – and controversial – view that new methods represent a more fundamental shift in the scientific method itself³. Certainly we can answer old questions in new ways, and faster, which is important in competitive scientific endeavours. These methods are more cost effective too. But most importantly we can answer entirely new questions, and obtain research results that simply were not possible before, by starting from the data rather than with the hypothesis that led to collection of data in the first place.

The August 2011 riots in several London boroughs and other cities across England led to a compelling example of this new scholarship. Citizens communicated and reported actions via Twitter, and the Twitter records of the event have been a subject of study – indeed, social media was implicated in the riots. *The Guardian* newspaper's 'Reading the Riots' project⁴ set out to "investigate England's summer of disorder", in which the social processes had no central coordination or control and false rumours were spread. This excellent case study in social media analytics leads to understanding how new social processes were created at the scale of the population involved, and in real time. This is the new, fourth quadrant scholarship about fourth quadrant society, and with it we witness the emergence of the new disciplines of computational social science and web science.

"– we will run out of humans and yet the technology axis goes on."

235 This example also illustrates something about the changing mode of knowledge production in contemporary society, increasingly recognized by researchers using phrases like 'living lab' and 'research in the wild'. Significantly, this is also recognized by funders, for example in the Research Councils UK cross-council e-Science programme and now especially the interdisciplinary Digital Economy Theme⁵. Characteristically, their goals are not pure scientific knowledge, but also *applicability*; there is an emphasis on co-production, and researchers conducting research in the context of its application are said to be 'in it' rather than 'on it'.

In the spirit of co-production, we also see the success of citizen science. Galaxy Zoo⁶, for example, engaged 165,000 people in the visual characterization of galaxies. It also generated scientific discoveries as the citizens interacted through discussion forums and engaged with the science itself. Its successor platform, Zooniverse⁷, has already engaged over a million users in projects ranging from nature to the humanities, while ClimatePrediction.net⁸ is generating scientific discoveries through people volunteering their computers in the world's largest climate modelling experiment. These fourth quadrant methods are leading to new outcomes, incorporating new means of research collaboration and challenging the traditional scholarly record.

"... goals are not pure scientific knowledge, but also *applicability* ..."

End of the article

Where traditional scholarship might be described as a sense-making network of humans exchanging scholarly writing, today it is a sense-making network of humans and machines, with the communications produced and consumed by both. A 'future history' thought experiment blogpost, looking back from 2065 on the four centuries of *Philosophical Transactions*, anticipates the demise of the traditional article in about 2030 for eight reasons⁹:

- it was no longer possible to include the evidence in the paper
- it was no longer possible to reconstruct a scientific experiment based on a paper alone
- writing for increasingly specialist audiences restricted essential multidisciplinary reuse
- research records needed to be readable by computer to support automation and digital curation
- single authorship gave way to casts of thousands of collaborators and citizen scientists, leading to failure of the authorship incentive model
- quality control models scaled poorly with the increasing volume and open access movement, obscuring innovation
- alternative reporting was found necessary for compliance with increasingly stringent scientific and industrial regulations
- frustrated by inefficiencies in scholarly communication that stifled progress, research funders demanded change.

But we should not dismiss the article so easily, as we can usefully ask what is right about it that has enabled its success and longevity. My suggestion is that articles are *social objects*: we share them, cite them and discuss them – they enable us to cross boundaries of time, place and discipline. Significantly, they are the subject of discourse around which social networks form, we measure our reputation by them, and they enable us to collaborate (in order to compete). It is not just about what is inside the chunk of knowledge represented by an article, but the fact that it is encapsulated into a social object with its own social life.

"... articles are *social objects* ... they enable us to cross boundaries of time, place and discipline."

Research objects

The notion of the *research object* is that a researcher can bundle together all the aspects of a piece of research that make up its record, in one sharable and citable object; for example, gathering the evidence for a research outcome or a decision¹⁰. By aggregating the multiple digital pieces into one object with one identifier we achieve a new sharable, citable social object that drops into the tools of digital research. Crucially, the components might be exchanged with computers as well as humans. Research objects need not contain software or executable parts, but this is likely to be the case in comprehensive records of digital research.

An early example of a research object is the *pack* in myExperiment, a social website for sharing computational workflows¹¹. The workflow, typically a data analysis pipeline, was seen as the social object in this Web 2.0 site, by analogy with photos on Flickr and movies on YouTube. But myExperiment users soon requested the ability to attach data, logs, papers and presentations to their workflows – to record, share and publish their experiments. This led to the notion of the *pack*, essentially a bundle of URLs gathering together content. Packs are represented using the Open Archives Initiative Object Reuse and Exchange (OAI-ORE) standard and are also available as linked data, hence semantically described for ease of discovery and reuse, while persistent URIs and digital object identifiers (DOIs) facilitate citation. Subsequent work has more fully developed the notion of workflow-centric research objects¹², and systems which capture software, scripts, protocols and models can be viewed similarly in that they have components which describe computations.

“... an ongoing reflection on the nature and purpose of research objects ...”

Analysis of the growing corpus of myExperiment packs has led to an ongoing reflection on the nature and purpose of research objects, an exercise known informally as ‘the R Dimensions’¹³. This teases apart aspects of the role of research objects: one is that they should be *reusable*, another is *repurposable*. More generally, they enable people or machines to *reconstruct* a piece of research. This helps with *reproducibility*, but note that a research object is not ‘reproducible research’ by itself: reproducibility means reusing a research object with a change to some circumstances, inputs, resources or components in order to see if the same results are achieved independent of those changes. The 21 ‘R words’ to date are grouped under six categories:

- scientific method – reproducible, repeatable, replicable, reusable
- access – referenceable, retrievable, reviewable
- understanding – replayable, reinterpretable, reprocessible
- new use – recomposable, reconstructable, repurposable
- social – reliable, respectful, reputable, revealable
- curation – recoverable, restorable, reparable, refreshable.

Today research objects are typically shared by humans and, when new data is available, people might choose to rerun experiments to achieve new results. However, there is no need for a human to press the button, as the objects can be executed automatically. We see some of this already in the automatic execution of workflows on arrival of new data, and also for curation, to validate, maintain and even repair them. With increasing automation, we anticipate the evolution of the notion of research object into the *computational research object*, a model that enables machines to assemble and execute systems of research objects. These thoughts apply equally well to documents, and the idea of executable documents has been explored in the community, for example in Elsevier’s *Executable Paper Grand Challenge*¹⁴ in 2011. Beyond scholarly communication, the same ideas can be applied in other digital end-to-end systems, such as music production¹⁵.

237 We see then that machines are users too, but this brings a set of concerns. Who gains credit and who owns the intellectual property generated when research runs automatically? Who is liable for any damage that arises? What are the implications of unintended or accidental assembly of research methods and outcomes? What are the consequences of automated research that occurs at very high speed, possibly speculatively, without human intervention? Where is the empowered, critical, creative, subversive human in the loop?

“...machines are users too, but this brings a set of concerns.”

Social machines

Tim Berners-Lee provides a popular definition of social machines in *Weaving the Web*¹⁶:

“Real life is and must be full of all kinds of social constraint – the very processes from which society arises. Computers can help if we use them to create abstract social machines on the Web: processes in which the people do the creative work and the machine does the administration.”

This final phrase establishes a crucial principle for sociotechnical systems at scale and under automation, requiring that computers empower rather than replace humans. But the less quoted and 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, this definition anticipates the development of social media but significantly talks of new forms of social process and that these are given to the world at large. Twitter is an excellent example, because it illustrates new social processes in the hands of citizens – it is not necessary to fill out forms to register a new hashtag (as if it were a domain address, for example) but anyone can create one: this came about because people did. The Twitter infrastructure provides a communications mechanism, and the protocol and etiquette built on it are socially constituted – in other words, the behaviour of the Twitter social machine is the result of programming by citizens. It follows that the study of social machines is both inherently social- and machine-oriented¹⁷.

In fact we are seeing the co-constitution of a new scholarly communications system. We have encyclopaedia co-production in the shape of Wikipedia, and as libraries and publishers reinvent themselves for digitized and 'born digital' content, we see a plethora of new websites emerging, from repositories for data sharing to new models of peer review. Through the social machines lens these form an evolving ecosystem of *scholarly social machines*, vital as new experiments are conducted and sites come and go. We all are participants, authors and readers alike, and many of us are designers too. Scholarship itself is becoming an in-the-wild experiment in the co-production of social machines.

“Scholarship itself is becoming an in-the-wild experiment in the co-production of social machines.”

Conclusion

Three hundred and fifty years after *Philosophical Transactions* was launched, we seem to be stuck in our ways. Today's scholarly communications infrastructure is already a constraint on innovation – and research funders should care about this. Today we are obsessed with data itself, but what we do with the data – our method – matters at least as much. We seem to be trying to retrofit digital scholarship into historical practices and disciplinary divisions, but we need to learn from past practice and look ahead to new paradigms. Whether we can move on by evolutionary change in scholarly communication, or whether we need another revolution, is an important question. Some changes cannot just evolve: changing the side of the road we drive on calls for an overnight revolution. Scholarly social machines may be a useful lens through which to view, understand and design these changes as we move ahead.

The irony is not lost that this article too is part of the established scholarly publication system – can we use a flawed scholarly communications system to fix a flawed scholarly communications system? To recognize this quandary requires us to defamiliarize the article, the monograph and the book, and to focus on future practice. This article is a call to action.

“This article is a call to action.”

Acknowledgement

SOCIAM: The Theory and Practice of Social Machines, is funded by the UK Engineering and Physical Sciences Research Council (EPSRC), under grant number EP/J017728/1.

References

1. Shneiderman, B, Science 2.0, *Science*, 2008, 319(5868), 1349-1350.
2. Hey, T, Tansley, S and Tolle, K, *The Fourth Paradigm: Data-Intensive Scientific Discovery*, 2009, Microsoft Research.
3. Anderson, C, The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, *Wired Magazine*, 2008, 16(7).
4. Vis, F, Twitter as a reporting tool for breaking news, *Digital Journalism*, 2013, 1(1).
5. Digital Economy Research in the Wild:
<http://www.ahrc.ac.uk/Funding-Opportunities/Pages/Digital-Economy-Research-in-the-Wild.aspx> (accessed 1 September 2014).
6. Galaxy Zoo:
<http://www.galaxyzoo.org/> (accessed 1 September 2014).
7. Zooniverse:
<http://www.zooniverse.org/> (accessed 1 September 2014).
8. climateprediction.net:
<http://www.climateprediction.net/> (accessed 1 September 2014).
9. De Roure, D, 17 March 2013, Pages of History, SciLogs blog:
<http://www.scilogs.com/eresearch/pages-of-history/> (accessed 1 September 2014).
10. Bechhofer, S, Buchan, I, De Roure, D et al, Why Linked Data is Not Enough for Scientists, *Future Generation Computer Systems*, 2013, 29(2), 599–611.
11. De Roure, D, Goble, C and Stevens, R, The design and realisation of the myExperiment Virtual Research Environment for social sharing of workflows, *Future Generation Computer Systems*, 2009, 25(5), 561–567.
12. Belhajjame, K, Corcho, O, Garijo, D, Zhao, J et al, Workflow-Centric Research Objects: A First Class Citizen in the Scholarly Discourse. In: *ESWC2012 Workshop on the Future of Scholarly Communication in the Semantic Web (SePublica 2012)*, 2012, Heraklion, Greece.
13. De Roure, D, 27 November 2010, Replacing the Paper: The Twelve Rs of the e-Research Record, SciLogs blog:
<http://www.scilogs.com/eresearch/replacing-the-paper-the-twelve-rs-of-the-e-research-record/> (accessed 1 September 2014).
14. Executable paper grand challenge:
<http://www.executablepapers.com/> (accessed 1 September 2014).
15. De Roure, D, Executable Music Documents, *Digital Libraries for Musicology (DLfM '14)*, 2014, London, UK, ACM.
16. Berners-Lee, T, *Weaving the Web: the Original Design and Ultimate Destiny of the World Wide Web*, 1999, San Francisco, Harper.
17. SOCIAM – The Theory and Practice of Social Machines:
<http://sociam.org/> (accessed 1 September 2014).

Article copyright: © 2014 David De Roure. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and distribution provided the original author and source are credited.



David De Roure

Professor of e-Research, University of Oxford, Oxford e-Research Centre, 7 Keble Road, Oxford OX1 3QG, UK
E-mail: david.deroure@oerc.ox.ac.uk

ORCID iD: <http://orcid.org/0000-0001-9074-3016>

To cite this article:

De Roure, D, The future of scholarly communications, *Insights*, 2014, 27(3), 233–238; DOI:
<http://dx.doi.org/10.1629/2048-7754.171>