Reenvisioning a future in scholarly communication

Chris H.J. Hartgerink

2017-05-09

**For the 2017 IFLA conference, theme: The Open Revolution**

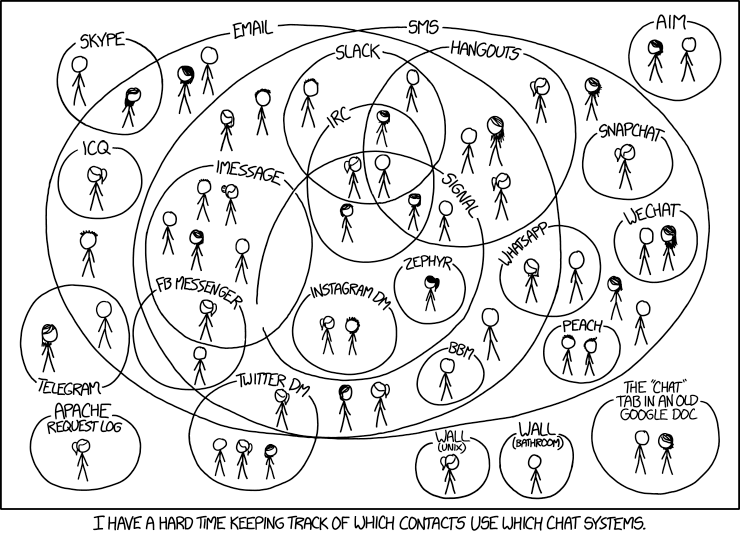
## Abstract

## Body

Scholarly communication is in (dire) need of disruption. Subscription costs to scholarly journals are untenable already (e.g., Sample 2012) and will become even more untenable in the long run (i.e., serials crisis; Panith and Michalak 2005). The current scholarly incentive system detracts validity instead of adding it (e.g., **???**; **???**) by promoting behavior that is antithetical to the norms of science (Merton 1942; Mitroff 1974; Anderson, Martinson, and De Vries 2007) and seek legitimacy instead of accuracy. Results cannot systematically be verified because data are rarely shared (e.g., **???**; **???**; **???**) or preserved (Vines et al. 2014). Widely publicized results regularly fail reproducibility tests (e.g., **???**; Begley and Ellis 2012; Baker and Dolgin 2017, but see also Schweinsberg et al. (2016)), in part because published results are highly pre-selected for unscientific reasons (e.g., **???**; Nuzzo 2015). These problems are not new but they are also not absolute --- scholars want to behave in accordance with the scientific norms (Anderson, Martinson, and De Vries 2007) but the system seems to prevent them from doing so.

In this perspective piece, I try to radically reimagine the scholarly communications infrastructure based on the digital tools that are available to us now, instead of the legacies of the analog era. I thoroughly believe that the Mertonian norms in science (Merton 1942) align perfectly with transparency in science (see also Hartgerink and Wicherts 2016) and that a scholarly communications system can be built on this. I invite everyone to criticize these ideas, for only with diversity of ideas can we truly make progress. Currently the decisions are driven by a few oligopolies (**???**) or the highly pre-selected sample of tenured professors and policymakers, who mostly stem from the generation that has created this broken system. As a result, they are likely to underestimate the problems we now face and cannot be expected to intervene in the ways that are necessary.

In the last decades, interpersonal communication has changed tremendously. The past decades have brought much diversity in communication, such that it can become confusing (see Figure 1). At the same time, the diversity has liberated us by allowing innovative ways to communicate (e.g., social media) or making available previously difficult to use safety measures easily accessible to the masses (e.g., encryption). Diversity has thus spawned new ways of communicating.



*Figure 1.* A comical example of the diversity in communication channels.

However, the way scholars communicate their findings has barely changed; we are still stuck in a paper paradigm, albeit writing digital pages. Such digital pages radically restrict the way we imagine how we can communicate results. In recent years, options to break free from this have arisen with tools such as [Rmarkdown](http://rmarkdown.rstudio.com/) or [Jupyter notebooks](https://jupyter.org/). These options allow the user to include code to dynamically generate figures, tables, or values in the text. By extension, how data are cleaned and analyzed can also be included as part of a manuscript for co-authors to verify or readers to discuss. However, the scholarly communications system currently still requires submitting to a journal that fits all of that back into a page-by-page, static, and paper format without all of these options. This not only prevents new ways of communicating findings, it also prevents further encouragement of discussion amongst readers about the findings. As a result, the current system encourages scholars to passively consume findings instead of critically evaluating them.

The scholarly process is much more complicated than can be captured in a single static and retrospective publication. For example, a quantitative scholar needs to develop methods to investigate a research question. Feedback from peers would be valuable at that point. With so called Registered Reports (**???**) a two-staged submission process is introduced to move away from a single static publication by allowing selected peers to discuss the contents at an interim stage. Additionally, this ensures chronology of predictions such as pre-registrations aim to do (**???**). However, these behaviors are for the proactive, hence, remain the exception rather than the norm (Kidwell et al. 2016). Additionally, researchers create materials, protocols, and other outputs during the research process that now are not (aptly) sharable in a static publication.

Additionally, the scholarly process has grown more complicated due to increased variation in the research tools available. Point-and-click software has increased the number of statistical analyses a researcher can do within minutes, increasing researcher degrees of freedom to present results that are easier to process (Simmons, Nelson, and Simonsohn 2011). The development of technology has created new types of data output that greatly exceed human comprehension and require algorithms before being able to make sense of the data (e.g., fMRI data). How a researcher applies all these tools is essential, but the current scholarly communication system does not require the details. Lack of direct reproducibility of results (e.g., Beaulieu-Jones and Greene 2016) therefore is no surprise when we do not know how data are processed and analyzed.

As such, the scholarly process has become more of a black box than it was before because communication has not adapted proportionate to the increasing complexity. Increasing complexity and lack of transparency makes it increasingly difficult to reproduce steps taken, simply due to an increasing garden of forking paths (Gelman and Loken 2014). Regardless, telling peers what path was taken could simply be required, if applicable. Moreover, much of the software is proprietary and it is simply impossible to verify results or understand how they originated (Ince, Hatton, and Graham-Cumming 2012). Trying to reverse engineer these pieces of software is actually often illegal based on the terms of service agreement that you have to accept before using the software. As such, it is practically impossible to understand how software development affects how results are computed, while testing for this is easy with open-source software.

Due to the fact that these complexities cannot and are not currently reported, breaking the reign of the paper paradigm is a necessity to better represent and understand how scholarly results come into existence. Of course paper is a valid medium to communicate nonetheless, which also is not the case when scholarly communications become truly digitally born and continuous. PDF versions of outputs can still be produced if a paper format is suitable or desirable. However, it is easy to imagine videos being embedded for protocols to be easily clarified (having "Video X" alongside "Figure X" and "Table X", for example).

Moreover, given the way we currently organize scholarly communication, we cannot break free from the legacies of the paper paradigm. Paper-based communication includes the legacy of printing research outputs, which by extension includes the idea of subscriptions, purchasing content, and restricting the flow of information. Copyright is the legacy of the paper-based age that seems to limit the digital potential of the outputs. As such, the legacy of a paper-based paradigm is that of knowledge commodification.

For scholarly communications, commodifying knowledge as it is currently done can be seen as a human rights violation. The Universal Declaration of Human Rights (UDHR), Article 27 section 1, reads as follows > Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits. When knowledge is commodified, participation, enjoyment, and sharing of scholarly progress is detracted and would therefore transgress this internation human right. I am not a law scholar, but there is no precedent that disconfirms this notion. Moreover, this is the main legal motivation proposed by Alexandra Elbakyan as to why [Sci-Hub](http://sci-hub.bz) is not illegal from her point of view (**???**).

Commodifying knowledge is not the only way to make the scholarly ecosystem sustainable. If it was the only option, it would potentially justify the current system. However, an alternative is the commodification of how information is consumed. greatly encourages society to innovate how people can discover and reuse information.

By focusing on the commodification of consumption instead of commodification of the resource, the problem of access to knowledge can be resolved in a sustainable manner. When the consumption is commodified, all service providers benefit from creating a large pool of resources that can be consumed, because having only a selection to consume is less attractive to potential customers. However, the largest resource is freely available for reuse if everyone participates and shares, which creates a market that stimulates access instead of limiting it (i.e., when knowledge is commodified).

Additionally, commodification of consumption removes several perverse incentives from the scholarly system that now produces unreliable knowledge.

The main tenet underlying the themes of Open Access, Open Data, Open Science, and replication initiatives in scholarly communication is sustainability.

In this perspective piece, I propose a redesign of the scholarly communication system that is flexible, decentralized, and distributed. Such a system would allow for more sustainability and diversity in the way we consume, produce, and access knowledge, which would ultimately benefit everyone who reaps the benefits of scholarly results.

## Sustainable scholarly process

Single points of failures make a system vulnerable to outages, (malicious) adjustment of content, or even complete removal of content.

The scholarly process is ridden with single or few points of failures at the moment.

These points of failures occur at each stage of the scholarly process.

In order to remove these failure points, a sustainable scholarly communications system would have to be both decentralized and widely distributed.

Decentralization is achievable with new peer-to-peer (p2p) Internet protocols such as dat and ipfs.

Such p2p networks also allow for ready redistribution when copyright is leveraged to work for distribution.

Alongside decentralization and distribution, integrity of the scholarly record can also improve. The chronology of the scholarly process can be ensured by splitting up the paper article as we know it into nodes of information.

These nodes can easily be logged on a blockchain based ledger (e.g., [Hyperledger](https://github.com/hyperledger/fabric)-based) of which everyone can have a copy. Although explaining the entire operating principle of the blockchain is beyond the scope of this piece, the essence is very simple. Each entry on the blockchain needs to satisfy a predefined mathematical rule [@], which is dependent on the contents of the previous entry in the ledger, the contents of the new entry, and a constant to offset the current entry. Solving this offset for the new entry to satisfy a specific mathematical rule is called 'mining'. However, because the proof of record is dependent on the contents of the previous entry (i.e., block), the chain is broken when the contents of that previous entry are changed. Given that each user has a copy of the entire blockchain, those who have a different chain from the majority of other users are simply rejected. As such, in order to change the ledger, each change would have to be re-verified by the entire network to become verified (for an introduction to the blockchain, see <https://www.youtube.com/watch?v=_160oMzblY8>).

{  
 "parent": null,  
 "self": "dat://92b9b1b95e18c1dd1546c25ac49609da6fb0e3f2abc94ecf17609c565888c16b",  
 "time-deposit": "2017-04-10T18:38:03+02:00",  
 "time-available": "2017-04-10T18:38:03+02:00",  
 "authors": "b601f276870ebf4920b3c78c4af51d3ed3bf4db7ab2b9c4faa74e0b457aa13ed",  
 "type": "hypotheses"  
}  
  
{  
 "parent": "dat://92b9b1b95e18c1dd1546c25ac49609da6fb0e3f2abc94ecf17609c565888c16b",  
 "self": "/ipfs/QmVLDAhCY3X9P2uRudKAryuQFPM5zqA3Yij1dY8FpGbL7",  
 "time-deposit": "2017-04-12T12:07:06+02:00",  
 "time-available": "2017-08-12T12:07:06+02:00",  
 "authors": "b601f276870ebf4920b3c78c4af51d3ed3bf4db7ab2b9c4faa74e0b457aa13ed",  
 "type": "data"  
}

Each entry in the ledger needs to be self-containing and refer to all relevant files. The ledger can be platform agnostic.

A parent node can never occur after a child node,

The ledger can also allow for time-locked sharing but with a clear chronology of creation. Time-locked sharing would require a third-party that can be trusted

Nodes cannot be made available

Combinations of ORCIDs can be hashed into unique author combinations to provide identification over time, but remove gender and status bias during initial assessment (amongst other things).

What types of nodes are contained in the ledger, is something that requires discussion and can change over time. Nonetheless, the fact remains that if someone presents results without

For a sustainable distribution system, persistent nodes need to be identified: libraries.

Nonetheless, other stakeholders benefiting from maintaining the distribution of the decentralized network are (not-)for-profit companies that build services on top of the resources.

## Locking the system open

When Brewster Kahle spoke at OpenCon2016, he mentioned that we need to lock the system open. At first, it would seem only the open advocates win in that case, but I think the for-profit companies can benefit from this as well by opening up a whole new market where more return can be made and in a more sustainable manner.

## References

Anderson, Melissa S, Brian C Martinson, and Raymond De Vries. 2007. “Normative Dissonance in Science: Results from a National Survey of U.S. Scientists.” *Journal of Empirical Research on Human Research Ethics* 2 (4): 3–14. doi:[10.1525/jer.2007.2.4.3](https://doi.org/10.1525/jer.2007.2.4.3).

Baker, Monya, and Elie Dolgin. 2017. “Cancer Reproducibility Project Releases First Results.” *Nature* 541 (7637). Springer Nature: 269–70. doi:[10.1038/541269a](https://doi.org/10.1038/541269a).

Beaulieu-Jones, Brett K, and Casey S Greene. 2016. “Reproducible Computational Workflows with Continuous Analysis.” *BioRxiv*. Cold Spring Harbor Labs Journals. doi:[10.1101/056473](https://doi.org/10.1101/056473).

Begley, C. Glenn, and Lee M. Ellis. 2012. “Drug Development: Raise Standards for Preclinical Cancer Research.” *Nature* 483 (7391). Springer Nature: 531–33. doi:[10.1038/483531a](https://doi.org/10.1038/483531a).

Gelman, Andrew, and Eric Loken. 2014. “The Statistical Crisis in Science.” *American Scientist* 102 (6). Sigma Xi: 460. doi:[10.1511/2014.111.460](https://doi.org/10.1511/2014.111.460).

Hartgerink, Chris H. J., and Jelte M. Wicherts. 2016. “Research Practices and Assessment of Research Misconduct.” *ScienceOpen Research*, September. ScienceOpen. doi:[10.14293/s2199-1006.1.sor-socsci.arysbi.v1](https://doi.org/10.14293/s2199-1006.1.sor-socsci.arysbi.v1).

Ince, Darrel C., Leslie Hatton, and John Graham-Cumming. 2012. “The Case for Open Computer Programs.” *Nature* 482 (7386). Springer Nature: 485–88. doi:[10.1038/nature10836](https://doi.org/10.1038/nature10836).

Kidwell, Mallory C., Ljiljana B. Lazarević, Erica Baranski, Tom E. Hardwicke, Sarah Piechowski, Lina-Sophia Falkenberg, Curtis Kennett, et al. 2016. “Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency.” Edited by Malcolm R Macleod. *PLOS Biology* 14 (5). Public Library of Science (PLoS): e1002456. doi:[10.1371/journal.pbio.1002456](https://doi.org/10.1371/journal.pbio.1002456).

Merton, Robert K. 1942. “A Note on Science and Democracy.” *J. Legal & Pol. Soc.* 1. HeinOnline: 115.

Mitroff, Ian I. 1974. “Norms and Counter-Norms in a Select Group of the Apollo Moon Scientists: A Case Study of the Ambivalence of Scientists.” *American Sociological Review* 39 (4). American Sociological Association: 579–95. doi:[10.2307/2094423](https://doi.org/10.2307/2094423).

Nuzzo, Regina. 2015. “How Scientists Fool Themselves – and How They Can Stop.” *Nature* 526 (7572). Springer Nature: 182–85. doi:[10.1038/526182a](https://doi.org/10.1038/526182a).

Panith, Judith M., and Sarah Michalak. 2005. “The Serials Crisis: A White Paper for the Unc-Chapel Hill Scholarly Communications Convocation.” <http://wayback.archive.org/web/20170509082718/http://www.unc.edu/scholcomdig/whitepapers/panitch-michalak.html>.

Sample, Ian. 2012. “Harvard University Says It Can’t Afford Journal Publishers’ Prices.” <http://wayback.archive.org/web/20170509082722/https://www.theguardian.com/science/2012/apr/24/harvard-university-journal-publishers-prices>.

Schweinsberg, Martin, Nikhil Madan, Michelangelo Vianello, S. Amy Sommer, Jennifer Jordan, Warren Tierney, Eli Awtrey, et al. 2016. “The Pipeline Project: Pre-Publication Independent Replications of a Single Laboratory’s Research Pipeline.” *Journal of Experimental Social Psychology* 66 (September). Elsevier BV: 55–67. doi:[10.1016/j.jesp.2015.10.001](https://doi.org/10.1016/j.jesp.2015.10.001).

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. “False-Positive Psychology.” *Psychological Science* 22 (11). SAGE Publications: 1359–66. doi:[10.1177/0956797611417632](https://doi.org/10.1177/0956797611417632).

Vines, Timothy H., Arianne Y.K. Albert, Rose L. Andrew, Florence Débarre, Dan G. Bock, Michelle T. Franklin, Kimberly J. Gilbert, Jean-Sébastien Moore, Sébastien Renaut, and Diana J. Rennison. 2014. “The Availability of Research Data Declines Rapidly with Article Age.” *Current Biology* 24 (1). Elsevier BV: 94–97. doi:[10.1016/j.cub.2013.11.014](https://doi.org/10.1016/j.cub.2013.11.014).