



ELEPHANT

IN THE LAB

OPINION

Decentralizing Science

Short title	Decentralizing Science
Long title	Advancing science through incentivizing collaboration, not competition
Authors	Dr. Martin Etzrodt ¹
Author affiliation	¹ ETH Zürich, Department of Biosystems Science and Engineering
Author bios	Dr. Martin Etzrodt is currently a postdoctoral researcher at ETH Zurich studying stem cell differentiation. He is interested exploring blockchain technologies as a means to reward scientific contributions of interdisciplinary teams and make scientific progress more efficient. He claims that decentralized, trusted platforms for the dissemination scientific information and attribution might be the solution.
Author social links	E-Mail: LinkedIn - Twitter
Date published	29 January 2018
DOI	https://doi.org/10.5281/zenodo.1156360
Cite as (APA)	Etzrodt, Martin (2018). Advancing science through incentivizing collaboration, not competition. <i>Elephant in the lab</i> . DOI: https://doi.org/10.5281/zenodo.1156360



ELEPHANT

IN THE LAB

OPINION

Introduction

Today a system that sets wrong the incentives for the scientific community is prevailing that is relying on an outdated system for the communication of science through centralized publisher cartels. As a result, science in some fields is suffering from increasingly poor reproducibility. This bears the risk of loss of credibility among the public and increased scarcity of public funding (Ioannidis, [2005](#); Siebert, [2015](#); Reardon, [2017](#); Nature, [n.d.](#)).

Because publishing in scientific journals today is the only way to achieve attribution and reputation for scientific work and is the core requirement to secure future funds scientist are incentivized to thrive for their publications impact as the major goal of the academic enterprise (Siebert, [2015](#)).

The negative consequences of relying on scientific journals as a trusted third party thus go beyond the already well identified problems of paywalls and the need for open access that are being addressed by movements aiming for open science (UNESCO, [n.d.](#); TIB, [n.d.](#)) and institutional initiatives against publisher cartels (Project Deal, [n.d.](#)).

More importantly, we need to establish an entirely new paradigm of “trustless” permanent publication, attribution and interoperability of scientific data. This opinion is written from my perspective of 15 years’ experience in biomedical research, however the proposed solutions may be applicable to other scientific fields alike.

In 1945 Vannevar Bush, head of the US office of Scientific Research and Development during WWII anticipated the creation a futuristic device he called “Memex”, a “*device in which, an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory*” (Bush, [1945](#)). Further he proposed that “Memex” would lead to “*wholly new forms of encyclopedias [...], ready made with a mesh of associative trails running through them, ready to be dropped into the Memex and there amplified*” (Bush, [1945](#)).

Arguably today we hold such devices in our hands and the internet serves as an underlying infrastructure, yet still 73 years later, past the invention of the personal computer and the internet we can agree with Dr. Bush’s position that “*professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose*” (Bush, [1945](#)).

Indeed, the formats of scientific communication remain the same as during the time of Dr. Bush’s essay in 1945. Running a website and issuing the print-version of articles as downloadable PDFs, that does not really make one a “digital publisher” anymore in the year 2018.

Dr. Bush continues that, “*if the aggregate time spent in writing scholarly works and in reading them could be evaluated, the ratio between these amounts of time might well be startling. Those who conscientiously attempt to keep abreast of current thought, even in restricted fields, by close and continuous reading might well shy away from an examination calculated to show how much of the previous month's efforts could be produced on call. Mendel's concept of the laws of genetics was lost*



ELEPHANT

IN THE LAB

OPINION

to the world for a generation because his publication did not reach the few who were capable of grasping and extending it; and this sort of catastrophe is undoubtedly being repeated all about us, as truly significant attainments become lost in the mass of the inconsequential” (Bush, [1945](#)).

Indeed, today the amount of scholarly output has even more dramatically increased (Ioannidis et al., [2014](#); Boon, [2016](#)). The risk of drowning out key findings in this vast amount of published work is pervasive and small-scale projects aiming for low hanging fruits instead for more risky, long-term and collaborative projects are prevailing, especially in the life sciences. The publisher’s and thus often the scientist’s main focus remains on “story telling” and favouring novelty over diligence of reporting and reproducibility. This has concerning effects on the credibility of science and our technological progress, which to date has been best documented and might be most apparently leading to harmful consequences in the biomedical sciences (Ioannidis, [2007](#); Sarewitz, [2016](#)).

The role of scientific publishing companies for communication, attribution and reputation building

Scientific publishers have traditionally served two important roles for science. First, they have guaranteed the efficient collection and distribution of scientific information. This included the distribution of physically printed versions of the scientific articles to subscribers, such as academic libraries around the world. Second, publishers serve as a trusted third party. As such, they are filtering content, handle a peer review process and serve as a solicitor to attribute the scientific findings to a single or groups of individuals. While today most scientist access scientific literature through the internet, and in many disciplines articles are circulated prior to publication on pre-print servers (e.g. arXiv.org, [n.d.](#)), the role of journals as trusted third party remains. Indeed, peer review seems to be an inevitable mechanism to guarantee scientific quality. Unfortunately, despite that notion, the quality of the scientific output appears to be concerning in decline (Ioannidis, [2005](#); Ioannidis, [2007](#); Freedman, [2010](#)) and review processes with scientific publishers are often excessively long leading to delays in communicating new discoveries, while adding apparently nothing against stopping the decrease in quality. Especially in experimental fields it is hard to assess more than just the plausibility of the presented work through peer review, but not the more important aspect of reproducibility. It has been also shown that the amount of data demanded by editors and reviewers alike for a single publication in these fields has steadily increased, again adding nothing to securing reproducibility (Sarewitz, [2016](#)).

A major bottleneck of disseminating new data is not only the review process, though. Troublingly, key discoveries are withheld by the scientist themselves because of the requirement of publishing “a full story” instead of single, validated observations, with *Science Matters* being a notable new type of journal that aims to change that (Science Matters, [n.d.](#)).



ELEPHANT

IN THE LAB

OPINION

The requirements of storytelling, exclusivity and novelty often degrades trust among scientist within the community to talk about or share new findings rapidly (a debate among scientists on that in a forum of “Science careers” that makes you want to pull your hair out can be found here (Science Careers, [2016](#)). This stifles cooperation and efficient allocation of awarded funds by unwittingly encouraging labs to reinvent the wheel instead of cooperating. On the other hand, the requirement of novelty for most scientific publishers, disincentives to report reproducibility or irreproducibility of other scientist’s results.

Scientists need to embrace more collaborative actions and faster reporting of their results. In the following I outline technological advances that are emerging to achieve this goal.

Open access data infrastructures

Efforts to create platforms with the aim for a better dissemination of information, sharing and commenting on scientific data are emerging (OECD, [2016](#); OpenAire, [n.d.](#); European Commission, [2017](#)) and open data repositories already exist (Nature Website Repositories, [n.d.](#)). The existing platforms today are however based on a centralized infrastructure requiring different formatting and annotation requirements, suffering from a lack of interoperability, even if they may adhere overall to such principles as FAIRsharing ([n.d.](#)). For most disciplines data and software tools are only dumped in repositories following their publication as a peer reviewed paper, thus at the end of a research cycle and will likely often have little future use.

A new, interoperable standard of communication for scientific information, in essence “peer-to-peer science” in which data is stored in a global, decentralized database which can be openly accessed would be desirable (European Commission, [2017](#); Heller, [2017](#)).

Decentralization means that within the databases addressing of information would occur by its type, not the location. Data packages would carry cryptographic hashes as “fingerprints” and data would be stored in multiple locations making the data immutable, censorship resistant and essentially undeletable (IPFS, [n.d.A](#); DAT, [n.d.](#)). The Interplanetary File System (IPFS) proposed by Juan Benet (IPFS, [n.d.](#)) or the Distributed Data Community (DAT) present to date the most advanced solutions for this goal. Notably, IPFS has already been instrumental to secure uncensored access to Wikipedia (IPFS, [n.d.B](#)) and the US climate data archives (GitHub, [n.d.](#)). It would be advisable for initiatives such as the European Open Science Cloud (European Commission, [2017](#)) to adopt IPFS/ DAT and support a network of academic libraries to oversee network nodes.

Using these standardized, interoperable protocols all data should be stored as well annotated research objects including the experimental design, raw data, analysis scripts and final analysis reports, all hyperlinked and most importantly immutably attributable to the publishing scientist (Littlejohn, [n.d.](#)).



ELEPHANT

IN THE LAB

OPINION

In essence, the scientist's lab book could, if logged publicly in such a system permit the option of "instant" publication. This should be the first and foremost interest of the scientist to secure the attribution of his findings and may be followed by adequate reporting in a research or review article for a wider non-specialist audience.

Logging the entire research cycle in this way would also prevent practices of "ex-post-facto" reporting, thus retroactively changing the hypothesis in a research study. This infamous "spinning the story" often distorts the results during interpretation in an attempt to make them appear more favourably for publication (Chiu et al., [2017](#)).

In a decentralized infrastructure for science documentation and reporting dedicated "science browsers" (Denis, [n.d.](#)) probably implemented as desktop clients would permit direct access to data streams in specialized "channels", representing a specific scientific field. Anyone could subscribe to and – under certain criteria – contribute to such "channels", but likely only specialists of the field may be able to make sense of the primary information presented there. Of note, all contributions and references toward other contributions would be logged and immutably traceable within the system.

The role of publishers would need to adapt, requiring the industry to provide new services that justify charging fees for their content. Making primary data streams accessible to a wider audience could be a new business model, for instance. Further, providing the best "browser" software, indexing tools and reviews by scouting for interesting current research from the decentralized data streams and presenting them to a wider, non-expert audience could be a future role for scientific publishers.

In all, scientists would spend less time with story-telling and trying to "sell" their data in a publication but would need to focus more on ensuring their data suffice to address proposed hypotheses, are reproducible and diligently annotated to be correctly submitted to a decentralized "web of knowledge". Their reputation would directly depend on the reproducibility and utility of their contributions for their peers. While duplication would inevitably occur, this should be easier traceable than in the current system relying on full articles and citations. Work building on top of a previous finding would normally also include replication of this finding and serve as further evidence of replicability. A larger number of future work emerging from a submitted scientific object would increase its relevance.

In the next section I describe in more detail the governance requirements for open access data infrastructures.



ELEPHANT

IN THE LAB

OPINION

The vision of a collaborative culture of “Open source science”

It is important to stress, that “Open Science” is destined to fail, if we do not ensure the creation of the correct incentives for the scientific community to seriously embrace it. If the adverse mechanisms linked to traditional publication and funding remain, the “European Open Science Cloud” is dead on arrival. Instead of “Open Science” I propose to implement “Open Source Science”, much like open source software projects (The Open Source Way, [2016](#)).

A decentralized open access data infrastructure would require very strict rules of how new research data objects are generated within the system. In “Open Source Science” one would first register ideas and hypotheses and ideally also propose an experimental design for a given problem (“a whitepaper”; Ioannidis, [2005](#); COS, [n.d.](#)). Some expert community of the field that is following the proposal would have their own idea about how to improve the experimental design at this point. They could participate and comment on the proposal through “Open Source Science Platforms” (e.g. Denis, [n.d.](#)) and improve the design before a costly experiment is conducted or, as perhaps more often the case in small scale experimental efforts, decide to execute the experiment independently. Others would already be aware of the experiments conducted, even by competing teams and, as soon as the data are online, crunch the numbers and provide feedback on the experimental outcomes and independently verify the results.

Conferences (alike publications) that today are often a loose collection of reports on experiments that “already happened”, mostly in secrecy and without engagement of the community beforehand, could turn into “DevCons” or “Idea conferences”. Such developer conferences would permit agreement on standards, for instance on the exact execution of sequencing experiments thus ensuring reproducibility.

While there is often agreement about the nature of an exciting area of research, the path to approach the new problems may differ and it would be a healthy process to set up challenges and have different teams follow up on the same question with slightly different strategies. Even if the discovery does not turn out according to expectation, useful discoveries may be made along the way. Given studies are preserved throughout in the open access data infrastructures no data or finding is lost, including negative results.

Today, approaches and technologies already exist that enable teams to engage in large scale collaborative efforts, for instance in high-energy physics or observational astronomy. These include new ways of data storage, processing and dissemination, but also tools for managing groups of people to work productively in collaborative, (mostly) well annotated projects (e.g. Github). An important technology that would ensure that the above mentioned visions become a reality and can be implemented along open science platforms is blockchain technology (Bartling & Fecher, [2016](#)).



ELEPHANT

IN THE LAB

OPINION

Blockchains are globally distributed open and immutable transaction ledgers that provide a new means of addressing key issues such as transparent and censorship-resistant

- attribution of data to the operator or machine that generated them
- timestamping of data, methods, results and interpretations □
- financial transactions & incentives

Immutable ledgers replace a third party and create trust among different participants of an open science platform. For instance, the report of a single observation would be time-stamped and therefore could be irrevocably attributed to a single scientist. Tokens can be issued that serve as a medium to execute transactions of data or as incentive of replication studies, if required. It should be even possible to establish a tokenized funding model which includes the possibility for tracing the performance of allocated funds by sponsors (Bartling & Fecher, [2016](#)). Bad players, even pseudonymous ones, would be penalized in the long-term as their irreproducible results would naturally become obsolete.

Concluding remarks

Given the complexity of problems we face today we need to step away from the winner-takes-it-all mentality in science and embrace new ways to solve the challenges we face as a society building on a solid base of knowledge everyone can access and contribute to.

Public funding agencies will have a far better argument for justifying scientific funding when the gained results are accessible by everyone. Getting data out to the public through open science platforms as fast as possible also should be more advantageous than hoarding them in secrecy as new, unimaginable associations become possible.

Backed by trustless blockchain mechanisms of attribution and validation this all may now become a reality.

After all, we do not have to wait or even attempt to “convince” the current players to adopt such new approaches. There is enough momentum to simply try it out. Let’s build the trustless and blockchain backed decentralized infrastructure for science and knowledge creation now. Everyone may be free to join and try it out. It is an experiment where no-one will be harmed if it fails, but there is much to gain, if succeeds!



ELEPHANT

IN THE LAB

OPINION

Acknowledgements

I thank Gotthold Fläschner and Thomas Lee Russel (both ETH Zurich) for critical reading of this manuscript. The ideas presented here are based on extended conversations in the fall of 2017 during Blockchain for Science related hackathons in Berlin and London. For this I especially thank Soenke Bartling (Alexander von Humboldt Institute for Internet and Society), James Littlejohn (Living Knowledge Network), Dennis Parfenov (Data Management Hub), Eoghan Ó Carragáin (University College Cork, Ireland), Roman Gonitel and other members of the Blockchain for Science (www.blockchainforscience.com) Think Tank.



ELEPHANT

IN THE LAB

OPINION

References

- Bartling, S., & Fecher, B. (2016). Blockchain for science and knowledge creation. Zenodo. <https://doi.org/10.5281/zenodo.60223>
- Boon, Sarah. (2016). 21st Century Science Overload. 2016. Retrieved from <http://www.cdnsciencepub.com/blog/21st-century-science-overload.aspx>
- Bush V. (1945). As we may think. *The Atlantic Monthly*. Retrieved from <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>
- Chiu, K., Grundy, Q., & Bero, L. (2017). ‘Spin’ in published biomedical literature: A methodological systematic review. *PLoS Biology*, 15(9), e2002173. <https://doi.org/10.1371/journal.pbio.2002173>
- COS. Preregistration Challenge [Internet]. (n.d.). Retrieved from <https://cos.io/prereg/>
- DAT. Distributed Data Community [Internet]. (n.d.). Retrieved from <https://datproject.org/>
- Denis, P. (n.d.). A decentralised system for scientific data management. [Internet]. Retrieved from <https://github.com/kubrik-engineering/damahub>
- European Commission. (2017). European Open Science Cloud [Internet]. Retrieved from <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>
- FAIRsharing. [Internet]. (n.d.). Retrieved from <https://fairsharing.org/standards/>
- Freedman, David H. (2010). Lies, Damned Lies, and Medical Science. *The Atlantic*. Retrieved from <https://www.theatlantic.com/magazine/archive/2010/11/lies-damned-lies-and-medical-science/308269/>
- GitHub. IPFS Climate.[Internet]. (n.d.). Retrieved from <https://github.com/ipfs/notes/issues/210>
- Heller L. (2017). How P2P and blockchains make it easier to work with scientific objects – three hypotheses. Retrieved from <https://blogs.tib.eu/wp/tib/2017/05/09/how-p2p-and-blockchains-make-it-easier-to-work-with-scientific-objects-three-hypotheses/>
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Ioannidis, J. P. A. (2007). Why Most Published Research Findings Are False: Author’s Reply to Goodman and Greenland. *PLoS Medicine*, 4(6), e215. <https://doi.org/10.1371/journal.pmed.0040215>
- Ioannidis, J. P. A., Boyack, K. W., & Klavans, R. (2014). Estimates of the Continuously Publishing Core in the Scientific Workforce. *PLoS ONE*, 9(7), e101698. <https://doi.org/10.1371/journal.pone.0101698>
- IPFS. Interplanetary File system [Internet]. (n.d. A). Retrieved from <https://ipfs.io/>



ELEPHANT

IN THE LAB

OPINION

IPFS. *Uncensorable Wikipedia on IPFS* [Internet]. (n.d. B). Retrieved from <https://ipfs.io/blog/24-uncensorable-wikipedia/>

Littlejohn, J. (n.d.). LIVING KNOWLEDGE NETWORK. Retrieved from <http://lkn.foundation/>

Nature. (n.d.). Nature Brexit Special [Internet]. Retrieved from <https://www.nature.com/news/brexit-1.20225>

Nature Website Repositories [Internet]. (n.d.). Retrieved from <https://www.nature.com/sdata/policies/repositories>

OECD. (2016). *Technology and Innovation Outlook 2016*. Retrieved from <http://www.oecd.org/sti/oecd-science-technology-and-innovation-outlook-25186167.htm>

OpenAire [Internet]. (n.d.). Retrieved from <https://www.openaire.eu/>

Project Deal [Internet]. (n.d.). Retrieved from <https://www.projekt-deal.de/about-deal/>

Reardon, B.Y.S., Tollefson, J., Witze, A., & Ross, E. (2017). Science under fire in Trump spending plan. *Nature*, (543), 471–472.

Sarewitz, D. (2016). Saving Science. *The New Atlantis*, 4–40.

Science Careers. [Internet]. (n.d.). Retrieved from <http://scforum.sciencecareers.org/viewtopic.php?f=1&t=10587>

Science Matters [Internet]. (n.d.). Retrieved from <https://www.sciencematters.io/>

Siebert, S., Machesky, L. M., & Insall, R. H. (2015). Overflow in science and its implications for trust. *eLife*, 4. <https://doi.org/10.7554/eLife.10825>

The Open Source Way [Internet]. (2016). Retrieved from https://www.redhat.com/en/The_Open_Source_Way_Red_Hat_Forum_2016#description

TIB. Leibniz Information Centre for Science and Technology [Internet]. (n.d.). Retrieved from <https://www.tib.eu/en/>

UNESCO. (n.d.). Open Science Movement [Internet]. Retrieved from <http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/open-science-movement/>