A glimpse into the scholarly article of the future: fully computational and interactive research

Alberto Pepe[1] and Matteo Cavalleri[2]

[1]Authorea Team
[2]Wiley

August 18, 2020

Why are scientific ideas disseminated via "papers"? Is a paper the best way to share and publish research results? For the last decade or so, we have complained that scientists write cutting-edge "21st-century research" and publish it in a "17th-century format" (Pepe, 2017)(*The Scientific Paper Is Obsolete*, n.d.)

In sum: the format of research papers has not changed much in the last 400 years. Take any paper published this week, download it, and compare it to a digitized version of a paper from the 1600s. They may differ in page layout, color, and typeface, but they are essentially identical in format - a collection of text and figures. While the publishing format has not changed in the last 400 years, the change in content is a proclamation of the success of science. The discovery of DNA (WATSON & CRICK, 1953), penicillin (Fleming, 1980), and the formulation of general relativity (Einstein, 1905) are some of the biggest and most splendid scientific discoveries of all time. They were all published in a two-dimensional paper format. Even more recently, the groundbreaking discovery of gravitational waves, which earned the 2017 Nobel Prize in Physics to the leads of the LIGO collaboration, was published with the traditional format.

Obviously, the paper format, being so enduring and persistent, has served science well so far. But things have changed in the last three decades. Digitalization, internet, etc plus the volume of data, code, and cyber infrastructure necessary to run even the simplest simulations and computations.... We believe that a Gutenberg-style revolution is needed to bring scientific content and format to the same speed and level of progress.

The raise of the open science movement aims to make scientific research and its dissemination accessible, reproducible, and transparent. In addition to encouraging publication of research as open access as early as possible (preprints), in computational sciences this translates into making code and data available to everyone, and into practicing open notebook science. There are little doubts that readers and reviewers must be able to understand how the authors produced the computational results, which parameters were used for the analysis, and how manipulations to these parameters affect the results. Increasingly, journals and funding agencies are mandating that researchers share their code and data when reporting on computational results based on code and data.

However, even when these are published, they are provided in separate platforms from the main text that is based on traditional and static formats such as PDF. Since code, data, and text are not linked on a deep level, readers and reviewers are faced with barriers that hinder their ability to understand and retrace how the authors achieved a specific result reported in the article, whether in a figure, table, or the main text. In addition, while data and code may be available in repositories external to the corresponding article(Antoniol et al., 2002), it takes readers and reviewers considerable effort to verify the software and re-run analyses with e.g. changed parameters.

It is apparent that the two-dimensional file formats dominating scientific publishing (PDF) compound such

issues, and that journals need to enable more interactive publications to truly support scientists in more deeply presenting the reported results. *code integrationin with figures*

The publication history of the first detection of gravitational waves by the LIGO collaboration illustrates this aspect perfectly. The discovery was reported in a series of tranditional articles but with an associated and externally hosted supplemental Jupyter notebook(*GitHub: losc-tutorial/LOSC$_E$vent$_t$utorial/master, n.d.*). The notebook supports the reader in walking them through the work that generated every figure in the paper. Anyone is able run tweak the code, playing with the analysis to get a better handle on how each one works. The notebook, in its section dedicated to the signal processing of the gravitational waves into sound, allow to play in your browser the bloop of two black holes colliding.

Today we're making a little step in the right direction. The International Journal of Quantum Chemistry (QUA) is pleased to announce a collaboration between Wiley, Atypon, and Authorea which resulted in the publication of the first set of research articles featuring fully computational and interactive visualizations.

The idea of an innovative 21st century scholarly article is not new. eLife, for example, has pioneered blah blah through a pilot. What's new is that these newly published articles are real scientific papers, not just test articles.

In chemistry, the intrinsic limitation to two-dimensions in displaying molecular structures of the PDF format limits the understandability of the research(Fatemah et al., 2020). The three dimensional shape or configuration of a molecule is an important characteristic and it is of paramount importance to understand the compound's reactivity, properties, and characteristics, including toxicity. *Something something on solution to display molecules in 2D, with FIG* 3D interactive visualization of molecular structures have been employed for decades in chemistry to analyze and display research results (Ihlenfeldt & Engel, 1998), but this web-ready application did not make it into mainstream academic publications due to their print-first focus. In fact, one can argue that 2D representation of molecules is not part of the research toolkit in chemistry for 2 decades at least and would have but all disappeared if it weren't for scientific journals. *3D visualization example*

something something Plotly

References

Science was always meant to be open. (2017). Authorea. https://doi.org/10.22541/au.148777700.03352906

https://www.theatlantic.com/science/archive/2018/04/the-scientific-paper-is-obsolete/556676/. https://www.theatlantic.com/science/archive/2018/04/the-scientific-paper-is-obsolete/556676/

Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. (1953). *Nature*, *171*(4356), 737–738. https://doi.org/10.1038/171737a0

Classics in infectious diseases: on the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae by Alexander Fleming, Reprinted from the British Journal of Experimental Pathology 10:226-236, 1929.. (1980). *Rev Infect Dis*, *2*, 129–139.

Recovering traceability links between code and documentation. (2002). *IEEE Transactions on Software Engineering*, *28*(10), 970–983. https://doi.org/10.1109/tse.2002.1041053

https://mybinder.org/v2/gh/losc-tutorial/LOSC_Event_tutorial/master. https://mybinder.org/v2/gh/losc-tutorial/LOSC_Event_tutorial/master

Interactive 3D Visualization of Chemical Structure Diagrams Embedded in Text to Aid Spatial Learning Process of Students. (2020). *Journal of Chemical Education*, *97*(4), 992–1000. https://doi.org/10.1021/acs.jchemed.9b00690

Visualizing chemical data in the internet—data-driven and interactive graphics. (1998). *Computers & Graphics*, *22*(6), 703–714. https://doi.org/10.1016/s0097-8493(98)00091-0