A glimpse into the scientific paper of the future: fully computational and interactive chemistry

Alberto Pepe[1] and Matteo Cavalleri[2]

[1]Authorea Team
[2]Wiley

October 7, 2020

Why are scientific ideas disseminated via "papers"? Is a paper the best way to share and publish research results today? The format and function of research communication 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. The two papers may differ in page layout, color, and typeface, but they are essentially identical in format - a collection of text and figures. Indeed, the fact that we refer to the mainstream outlet of research communication as "paper" speaks volume of its boundness to print.

While the published *format* has not changed in the last 400 years, the change in published *content* is astronomical: a proclamation of the success of science. The discovery of molecular structure of DNA (WATSON & CRICK, 1953), penicillin (Fleming, 1980), and the formulation of general relativity (Einstein, 1916) 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 a traditional paper format (Abbott et al., 2016). LIGO's groundbreaking was certainly not analyzed on a 2D piece of paper.

So, how is it possible that scientists produce and write cutting-edge "21st-century research" and still publish it in a "17th-century format"? (Somers, 2019; Pepe, 2017)

Obviously, the paper format, being so enduring and persistent, has served science well. But things have changed in the last three decades. The recent explosion of content digitalization, growing internet speed and connectivity, and reliance on data, code, and computational power are leading to an unprecedented and irreversible path to changing the way we publish and disseminate research ideas. A Gutenberg-style revolution in scholarly communication is upon us, and we believe it is being pioneered by the Open Science movement.

The Open Science initiative 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 (the availability of preprints in subject-based repositories has moved beyond the realm of physics), for many computational domains Open Science translates into making code and data available to everyone, and into practicing "open notebook" science.

In other words: 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 data and code are provided by authors, and published, they are oftentimes relegated to

Supplementary Information or to entirely separate platforms, disconnected from the published "full text". 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. 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, say, changed parameters.

The idea of a multimedia, multi-dimensional, scholarly publication that defies the limitations of the 2-dimensional paper format is not new. The publication history of the first detection of gravitational waves by the LIGO collaboration is an example of how much this is needed in scientific publishing. The discovery was reported in a series of traditional articles (Abbott et al., 2016)(Abbott et al., 2016) but with an associated and externally hosted supplemental Jupyter notebook (LIGO Binder, 2016). The notebook allows readers to run and tweak the code, change parameters to alter the analysis, and, in its section dedicated to the signal processing of the gravitational waves into sound, it even allows readers to play the bloop of two black holes colliding. Yet, the notebook and the multimedia elements had to reside outside the article. Why?

Rich media available at `https://youtu.be/QyDcTbR-kEA`

In chemistry, the intrinsic limitation to 2-dimensional publishing 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.

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 two decades at least and would have but all disappeared if it weren't for scientific journals. Now the same tools that are at the disposal of researchers when analyzing their data can be finally offered to readers, speeding up the impact of the research.

Figure 1: Example of an interactive 3D visualization from (Hanwell et al., 2020). Readers can choose visualization style and parameters for the data, enhancing their understanding of the results.

Modern science produces massive amounts of data whose visualization translate poorly in a static format. Interactive visualization that allows readers to hover over values and zoom in/out graphs help identifying outliers among a large number of data points or detecting anomalies in time series plots. Since every modern scientist is essentially a data scientist, a tool like Plotly becomes an invaluable tool for research visualization that can be incorporated into a HTML based research article.

It is apparent that the two-dimensional file formats dominating scientific publishing (the PDF "paper") compound such issues, and that journals need to enable more interactive publications to truly support scientists in presenting and publishing the reported results. It is time to bring scientific content and format to the same speed and level of progress.

Today we're making a little step in this direction. Wiley's International Journal of Quantum Chemistry (IJQC) is pleased to announce a collaboration with Atypon and Authorea which culminated in the publication of the first set of research articles featuring interactive visualizations and integrated source data and code: (Peverati, 2020; Griego et al., 2020; Folmsbee & Hutchison, 2020; Pablo-García et al., 2020)

These papers provide a glimpse in this vision of the scientific article of the future, offering 3D interactive graphs, data and code integration (via stand-alone Binder containers) and open peer-review details.

They are the start of our journey towards the scientific "paper" of the future, with more to come....

References

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.

Die Grundlage der allgemeinen Relativitätstheorie. (1916). *Annalen Der Physik*, *354*(7), 769–822. https://doi.org/10.1002/andp.19163540702

Observation of Gravitational Waves from a Binary Black Hole Merger. (2016). *Physical Review Letters*, *116*(6). https://doi.org/10.1103/physrevlett.116.061102

The Scientific Paper Is Obsolete. (2019). 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/

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

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

ASTROPHYSICAL IMPLICATIONS OF THE BINARY BLACK HOLE MERGER GW150914. (2016). *The Astrophysical Journal*, *818*(2), L22. https://doi.org/10.3847/2041-8205/818/2/l22

LIGO Binder. (2016). 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

Open Chemistry, JupyterLab REST and Quantum Chemistry. (2020). https://doi.org/10.22541/au.158687268.81852407/v2

Fitting elephants in the density functionals zoo: Statistical criteria for the evaluation of density functional theory methods as a suitable replacement for counting parameters. (2020). *International Journal of Quantum Chemistry*. https://doi.org/10.1002/qua.26379

Acceleration of catalyst discovery with easy fast, and reproducible computational alchemy. (2020). *International Journal of Quantum Chemistry*. https://doi.org/10.1002/qua.26380

Assessing conformer energies using electronic structure and machine learning methods. (2020). *International Journal of Quantum Chemistry*. https://doi.org/10.1002/qua.26381

Turning chemistry into information for heterogeneous catalysis. (2020). *International Journal of Quantum Chemistry*. https://doi.org/10.1002/qua.26382