

Reproducibility Crisis with Computer Science

CANDIDATE NUMBER – 2, Universitetet i Tromsø, The Arctic University of Norway.

1 INTRODUCTION

Before the advent of scientific journals, science was more oriented towards a closed approach, so that only a restricted number of people could have access to it. Many scientists, including Galileo, Newton, Kepler, and Hooke, made their discoveries into something they could profit, and often papers were even encoded in anagrams. Following this closed science logic though, it was difficult to identify ownership and priority of scientific discoveries, leading even to a debate on whom first discovered calculus between Leibniz and Newton.

In modern times, scientific research is moving towards openness and free knowledge, and ownership and priority issues are not the main concern anymore. However, there are still research data, journals, and paper which are not accessible to all. Data is not open due privacy and secrecy by for-profit organizations, and papers are not open because they are published in pay-walled journals. Merton [4] arguments in favor of open science are that, knowledge-creation is more efficient if scientists work together, and it is morally binding on the professional scientist.

Besides knowledge and discoveries, science needs also to constantly proof itself to be right, hence it needs to be reproducible. If scientists such as Newton, Kepler or Galileo, could guarantee the reproducibility of their discoveries only by mathematical proofs, in modern science, and especially after the advent of computers and information technology, it is not possible anymore. The complexity and amount of data used to formulate theorems and proofs has drastically changed the way of doing research, and having an open knowledge approach is not good enough to perform good science.

Being our field of research computer science, and in particular big data on blockchain technology and cryptocurrencies, we want to enhance the main challenges of doing good

scientific research when research data and source code are not openly accessible. Excessively closed research data can lead to reproducibility issues, especially when machine learning and big data analysis are involved. We will show in this manuscript the importance of open source code and open research data when it comes to the scope of making good science.

2 OPEN SCIENCE

Open science is the movement to make scientific research (i.e. data, software, publications) accessible to all levels of an inquiring society, amateur or professional [6]. However, the term open science encapsulates many different concepts, such as: open educational resources, open access, open peer review, open data, open source, and open methodology [3]. Also, according to Fecher et al. (2014) [2], there are different open science schools. They can be for instance, oriented towards a pragmatic, public, or democratic approach, but the purpose is commonly identifiable in opening up the process of knowledge creation, making knowledge freely available for everyone, and enabling science to be accessible for citizens. We focus in particular on the open access, and open data concepts, enhancing difficulties for scientists, amateur and professionals to contribute in scientific development and innovation, if research data are not openly accessible.

Our research topic involves blockchain technology and cryptocurrencies. By definition of blockchain, research data are all public and always available. If we then perform searches in a topic of our interest, using public and open source web applications to explore research data, such as Dataverse [1] or Google Dataset Search¹, we can easily find the information we need about real-world implementation of blockchain technology, such as Bitcoin [5]. However, not all data in different fields of computer science are easily accessible. If we perform research about privacy, or eHealth for instance, we will not get the same amount of results, and most of the research data are not public.

3 PRIVACY PRESERVING

REFERENCES

- [1] CROSAS, M. The dataverse network®: an open-source application for sharing, discovering and preserving data. *D-lib Magazine* 17, 1 (2011), 2.

¹<https://toolbox.google.com/datasetsearch>

- [2] FECHER, B., AND FRIESIKE, S. *Open Science: One Term, Five Schools of Thought*. Springer International Publishing, Cham, 2014, pp. 17–47.
- [3] KNOTH, P., AND PONTIKA, N. Open science taxonomy. *FOSTER open science* (9 2015).
- [4] MERTON, R. K. Science and technology in a democratic order. *Journal of legal and political sociology* 1, 1 (1942), 115–126.
- [5] NAKAMOTO, S., ET AL. Bitcoin: A peer-to-peer electronic cash system. *Working Paper* (2008).
- [6] WOELFLE, M., OLLIARO, P., AND TODD, M. H. Open science is a research accelerator. *Nature Chemistry* 3, 10 (2011), 745–748.