Could Blockchain provide the technical fix to solve science's reproducibility crisis?

blogs.lse.ac.uk/impactofsocialsciences/2016/07/21/could-blockchain-provide-the-technical-fix-to-solve-sciences-reproducibility-crisi

Blockchain technology has the capacity to make digital goods immutable, transparent, and provable. **Sönke Bartling** *and* **Benedikt Fecher** *look at the technical aspects of blockchain and also discuss its application in the research world. Blockchain could strengthen science's verification process, helping to make more research results reproducible, true, and useful.*

Currently **blockchain** is being hyped. Many claim that the blockchain revolution will affect not only our online life, but will profoundly change many more aspects of our society. Many foresee these changes as potentially being more far-reaching than those brought by the internet in the last two decades. If this holds true, it is certain that research and knowledge creation will also be affected by this. So, what is blockchain all about? More importantly, could knowledge creation benefit from it? One potential area it could be useful is in addressing the credibility and reproducibility crisis in science.

Blockchain – the data structure

But first, let's look at what blockchain is technically, In a literal sense, blockchain is a computer data structure. A row of data blocks that are connected through a cryptographic function, the earliest description of this data structure dates back to 1998. If one changes the the content of one block, all following blocks need to change as well.

Blockchain became widely known by being the data structure (= ledger) in Bitcoin. Bitcoin is an online payment processing tool that lacks centrality and trusted third parties such as banks or companies (like Paypal). A defining feature is that the structure is distributed – the blockchain ledger is stored on many computers, and there is no single point of failure.
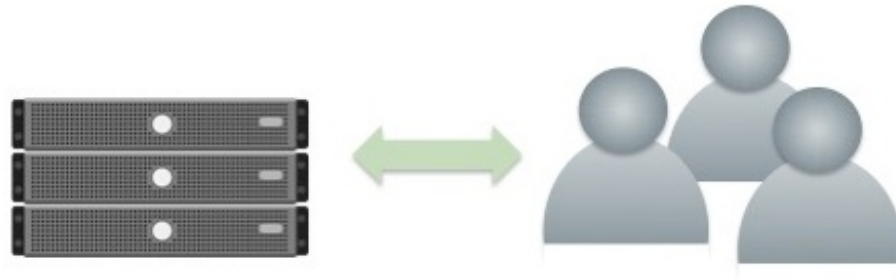
Blockchain – the (r)evolution

However, payment processing is just one application of blockchain systems. To differentiate the characteristics of the upcoming online (r)evolution from the payment processing tool and implementation of Bitcoin itself, the term 'blockchain' is nowadays used in a much wider meaning. It stands for a system of organizing all kinds of digital things, be it files, databases, or assets, in ways that were first widely perceived in Bitcoin. Attributes of this system include:

- **Decentralized**
- **Distributed**
- **Immutable**
- **transparent (provable to the outside)**

Before we explain more in detail what this means, let us first take a look at how we use computer services today:

For us nowadays, it is clear that whoever provides online services, be it a cloud storage service, a bank, an email provider, or a scientific publisher, needs to be trusted. We know that the provider could technically alter our accounts, change scientific results, or indeed our emails and files at will. We rely on those trusted third parties not to do so (Figure 1). Furthermore, we know that once data is digitalized it can be arbitrarily changed at will without leaving a trace (e.g. by researchers).
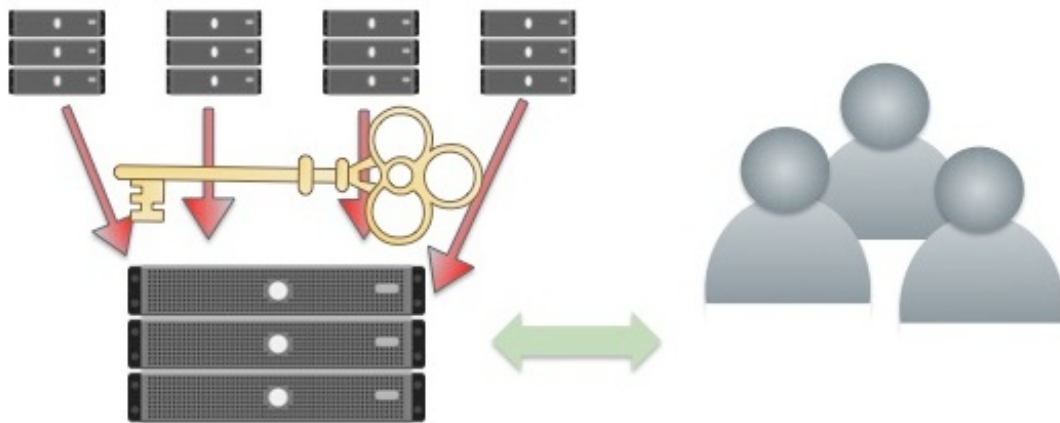
Figure 1: Today the owner (or researcher, academic publisher, etc.) has full control over their computer, data and services they run (e.g. a database). After the blockchain revolution, this is no longer the case as system is decentralized.

After the *blockchain* revolution, this changes fundamentally. The implications are far-reaching and it is worthwhile to understand the following terms:

Decentralization means that there is no single point of failure: there is no computer system that can be switched off, censored, or otherwise blocked in order to stop a service.

Distributed means that there is no single hardware infrastructure holding the service. Often, this means that a copy of a database exists on several computers, however, it may also be the case that a database is split between many computers.

Immutability means strictly speaking that data cannot be changed. However, in practice that means that data cannot be changed without leaving a trace. Most of the time, this means that old versions can be recovered and that any changes will be protocolled in a system. It is like comparing an excel sheet in which values can be changed at will to a piece of paper. On paper a trace that shows every manipulation is left (Figure 2).

Immutable research database

Group	A	B
Test 1	5.1	11.4
Test 2	10.2	12.4
Test 3	3.7	10.9
Test 4	4	11.3

Group	A	B
Test 1	5.1	11.4
Test 2	5.0 10.2	12.4
Test 3	3.7	10.9
Test 4	4	11.3

Figure 2: ***Blockchain* can make research databases immutable, meaning that they cannot be changed without leaving a trace.**

Transparent (provable to the outside) means that a computer program is really running as is publicised (advertised). At the moment, we must rely upon others to calculate things (e.g. impact factor) or to apply post-processing tools to research data as they claim that they are doing it; after blockchain revolution, this will be transparent and provable to peers.
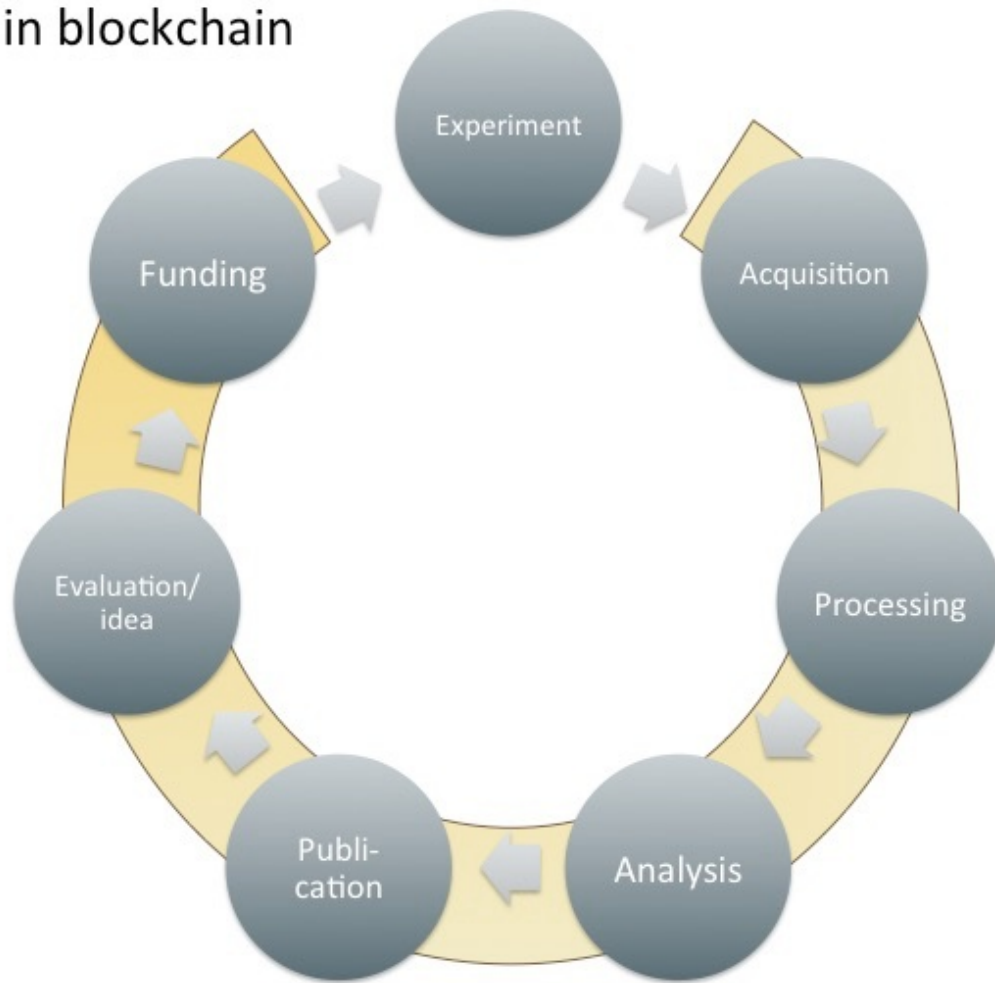
Figure 3: Large parts of the research cycle can make use of blockchain (yellow arch). Only the experiment (or initial data collection) phase cannot. Through blockchain technologies, the rest of the research circle would then become immutable, comprehensible, and provable to the outside. This would make larger parts of the research cycle open to scientific self-correction and may make more research results reproducible, true, and useful.

Blockchain and the research cycle: A Literature Review

For a full literature review of Blockchain and the research cycle, see this living doc by the authors which covers the applications of blockchain in science and knowledge creation in further detail. The authors have collected and organised published and newly introduced ideas for blockchain's relevance across the research cycle.

This blogpost is an extension of an earlier, pseudonymous, cryptographically signed and hash-time-stamped publication.

Note: This article gives the views of the author, and not the position of the LSE Impact blog, nor of the London School of Economics. Please review our Comments Policy if you have any concerns on posting a comment below.

About the Authors

Sönke Bartling *is an affiliated researcher at the Humboldt Institute for Internet and Society, board-certified radiologist and scientist in basic medical imaging sciences. He is an Open Science researcher.*

Benedikt Fecher *of the German Institute for Economic Research (DIW Berlin), is a doctoral researcher at the Humboldt Institute for Internet and Society. The focus of his dissertation is the participation in Open Science Projects.*