Blockchain for Open Science and Knowledge Creation

A technical solution to the reproducibility crisis

PD Dr. med. Sönke Bartling (corresponding author, initiator and maintainer, @soenkeba, soenkebartling@mailbox.org)
Founder of Blockchain For Science &
Associate researcher at the Humboldt Institute for Internet and Society

Many more contributors - please see document history.

Isn´t „really good science not always a break with orthodoxy – and how could the orthodox then fairly assess it?« (Michael Polanyi - potentially wrongly attributed or cited by the maintainer)

BLOCKCHAIN
for Science

Abstract:

Blockchain is a computer protocol involving cryptography, a new way to look at databases and socio-cultural-legal-politcal-economical (r)evolution and knowledge creation will be affected by it.

Blockchain has the capacity to make digital goods immutable, transparent, externally provable, decentralized, and distributed. Besides the initial experiment and data acquisition, all remaining parts of the research cycle could take place within a *blockchain system*. Attribution, data, subject anonymity, data postprocessing (e.g. via smart contracts), publication, research evaluation, incentivisation, and research fund distribution would thereby become time-stamped, comprehensible, open (at will) and provable to the external world. Currently, scientists must be trusted to provide a true and useful representation of their research results in their final publication; *blockchain* would make much larger parts of the research cycle open to scientific self-correction. This bares the potential to be a new approach to the current reproducibility crisis in science, and could 'reduce waste and make more research results true'. Beyond that, this could be used to reduce overhead and accelerate the scientific process and incentivise true innovation.

Introduction

Currently, blockchain is at the peak of its hype cycle (goo.gl/6DbyPC). Many claim that the *blockchain* revolution will affect not only our online lives, but will profoundly change many more aspects of our society [1–4]). Some 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, why is that the case, and what is this all about? More importantly, could knowledge creation benefit from it? Adoption of new technologies is good, however, it should not be an end in itself - there should be problems that can be solved with it. Currently, there is a credibility and reproducibility crisis in science [5–15].

In this article, we will first provide some abstractions and technical points of *blockchain*, then discuss application examples, and finally, identify problems in the research world that might be solved by means of *blockchain*.

Blockchain - the data structure

In a literal sense, blockchain is a computer data structure, a list of data blocks that are linked through a cryptographic function. The earliest description of this data structure dates back to 1991 [16]. If one changes the content of one block, all following blocks need to change as well.

Blockchain became widely known as the data structure (= ledger) that underpins Bitcoin [17,18]. Bitcoin is an online payment processing tool that lacks centrality and trusted third parties such as banks or companies (like Paypal). It is distributed, the blockchain ledger is stored on many computers, and there is no single point of failure. In Bitcoin, long known concepts have been successfully implemented together and found wide use for the first time, as they are:

- Cryptographic tools such as public key cryptography and hashes
- Consensus mechanisms (=ways to settle discrepancies within same data sets that are stored on different computers) [19,20]
- Proof-of-work (=methods that uses laborious computer calculations to prevent a system from being flooded with 'spam' or fake identities) [21]
- Economic incentives (miners are paid with Bitcoins) to agree upon the correct state of the blockchain ledger

Bitcoin continues to function reliably, despite several billion dollars worth of value now within its network. Breaking Bitcoin could potentially make large portions of this money accessible to the attacker.

Blockchain - the (r)evolution

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 context. It describes a system for 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 (≈'append only database')
- Transparent (provable to the external world)

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

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 to do what they are supposed to do. 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 digitized, it can be arbitrarily changed at will without leaving a trace (e.g. by researchers).
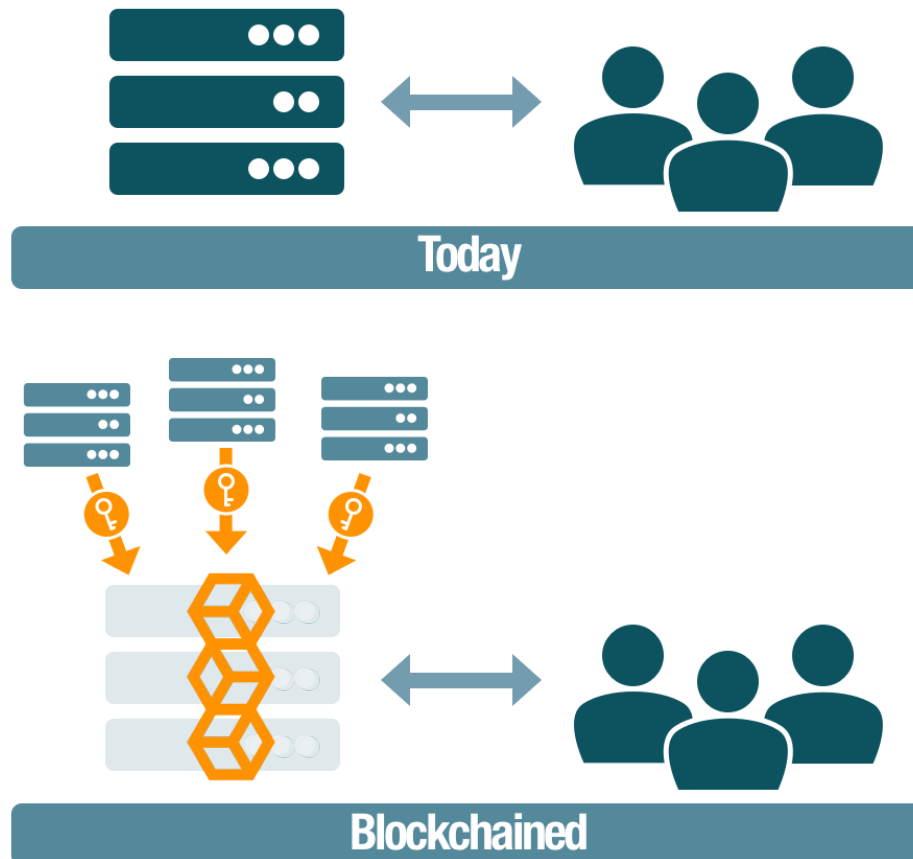
Figure 1: Today the owner (or researcher, academic publisher, data-repository etc.) has full control over their computer, data, and services they run (e.g. a database). After the blockchain revolution, this is no longer the case, as decentralized trust providing systems provide 'cryptographic power' to ensure the integrity of a computer service and authenticity of the underlying database.

After the *blockchain* revolution, this changes fundamentally. The technology has far reaching implications and so it is worthwhile understanding its language - it will be used much more often in the future.

Decentralization means that there is no single point of failure: there is no one single 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 that strictly speaking, data cannot be changed. However, in practice, this 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 of every manipulation is left displayed (Figure 2). Another practical interpretation would be to call a database an 'append only' database. This does not necessarily mean that all data are immutable, e.g. in Blockchain for Healthcare that is an often uttered concern - not the patient data itself immutable, but the access rights to it.

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

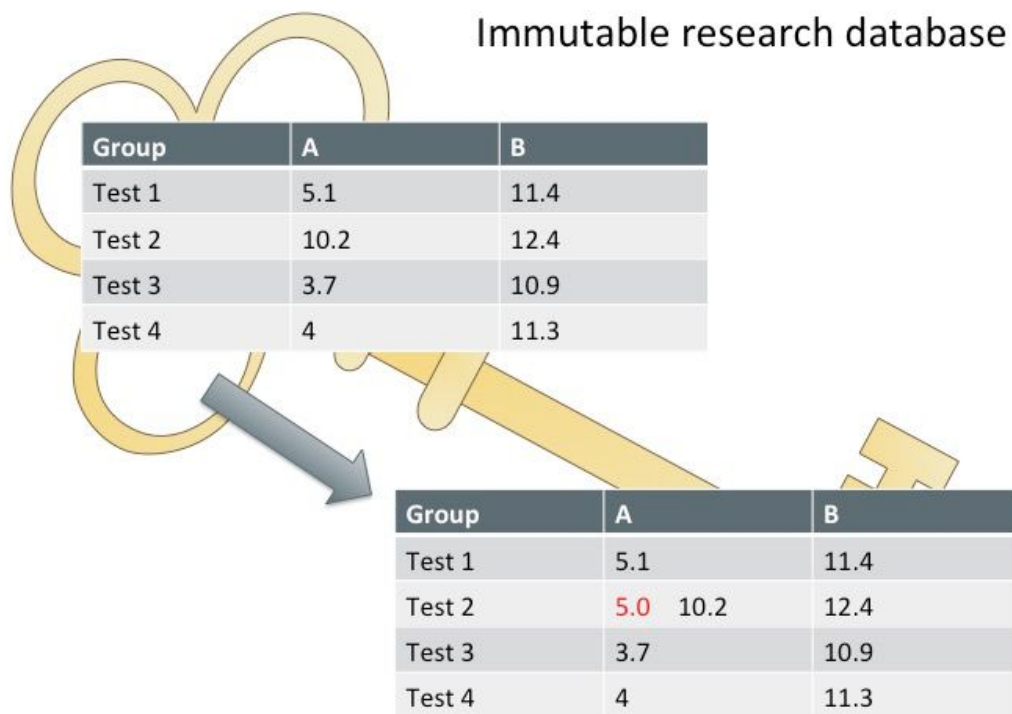| Group	A	B
Test 1	5.1	11.4
Test 2	5.0 10.2	12.4
Test 3	3.7	10.9
Test 4	4	11.3

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

Transparent (provable to the external world) 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 in the manner that they claim; after the blockchain revolution, this will be transparent and provable to peers.

In what follows, blockchain will refer to the data structure and *blockchain* will refer to a system that comprises the above features.

Blockchain - the database view point

Blockchain can be seen as a database with certain characteristics. When compared to current databases interesting correlations can be made (Table 1).

Database	Accessibility to researcher	Decentralisation	Scalability	(Intrinsic) backup	Immutability	Example use case in Research
Spreadsheet (e.g.Excel)	XXX	0	X	0	0	Workhorse in most researchers daily life
Digital Labbook	XX	0	X	X	0	Workhorse in collaborative lab environments
'Databases' (MongoDB, SQL, …)	X	0/X	XXX	XX	0	Backend in data storage, (journal) webpages, libraries, cloud solutions, etc.
Bitcoin-like blockchain	X	XXX	0	XXX	XXX	Notarization functionality
Blockchainified database (e.g. BigchainDB)	X	X	XXX (ongoing debate)		XX	Not yet described, assumed great potential

Table 1: Comparison of research database characteristics

Blockchain revolution - the technical implementations

Blockchain characteristics are being realized through cryptographic methods and consensus protocols. All of these are long since known, and were initially developed to handle hardware failures, e.g. inside big databases [19]. Nowadays, they are used to provide trust among sometimes unknown and distributed entities.

Blockchain systems rely on many discrete computers to secure the *blockchain* system and provide the trust or security that is today provided by administrators (Figure 1). These computers can be anonymous entities (miners) which are incentivized to do so by intrinsic value inherent to the system (e.g. Bitcoin, Ethereum) [22]. They can also be defined by a central authority. For example, the securing computers could be provided by trusted and independent research institutes [23] or governmental organizations. However, in contrast to what trusted third party administrators can do today, the blockchain-securing computers cannot alter data stored in the *blockchain* systems in an undetermined manner, even if someone wanted them to do so. They simply provide 'cryptographic power' so as to secure the *blockchain*. However, if a certain amount of them are compromised, data that is stored in a *blockchain* system becomes completely unreliable and mutable. This is not a bug, but an inherent characteristic of the consensus mechanisms. If they are selected carefully and guarded, such an event would be very unlikely.

Blockchain revolution - beyond Bitcoin

There are many Bitcoin-like blockchain systems. Focusing on their 'coin' aspect, they are called 'altcoins' [24]. Many are just copycats of varying, sometimes questionable legitimacy, some are even scams - but others provide very interesting new features and functionalities that extend far beyond payment processing. A discussion of these is beyond the scope of this article, and would actually be difficult to provide, since innovations and interesting new concepts are being published on almost a daily basis [25]. A list based on current market capitalization can be found here. Here, we will mention some that implement concepts or provide an organizational structure that are especially interesting for research.

One such system, the Ethereum blockchain, goes so far as to provide its own programming language to run distributed, unstoppable, and provable applications [26]. This includes smart contracts [27] which can be used to realize distributed, autonomous applications and organizations [28].

Storj, filecoin, swarm and MaidSAFE are also interesting concepts. They can be seen as *blockchain*-based, distributed cloud services to store data, files (or to provide services…) . Coins are used to incentivise resource providers who provide hard drive space and network bandwidth.

Namecoin is one of the first Bitcoin forks and is purposely built to store key-value pairs, in the foremost case, this is being used to register domain (.bit) names without a central entity like ICANN.

There are several projects out that develop platforms that build an incentivisation/rating/reputation system around providing content (including liking/commenting) using a blockchainified attribution and incentive distribution mechanism (Steemit, userfeeds.io and Synereo). Moreso, they are systems out that `pay` revenue to work at a project (Comakery.com). Certainly, that are very interesting concepts with respect to scientific communication, attribution, work/idea/content sharing incentivisation and have been described as such [29](pevo.science).

Most altcoins work on their own blockchain. However, to make things really confusing, all concepts could technically be implemented in one single blockchain, e.g. the Bitcoin blockchain.

Hyperledger project is a cross-industry collaborative effort, started in December 2015 by the Linux Foundation to support blockchain-based distributed ledgers. The project aims to bring together a number of independent efforts to develop open protocols and standards, by providing a modular framework that supports different components for different uses. This would include a variety of blockchains with their own consensus and storage models, and services for identity, access control, and contracts.

Open Document Repository (ODR) by Kubrik is a global network of document repositories run by public libraries. All repositories share a data storage system based on IPFS where they publish the open data, open access articles and all corresponding meta data. All data updates are tracked on a public permissioned ledger (blockchain) that is run between nodes. All participating research publishing entities will have voting power on this blockchain, so that instead of the energy and cost intensive "proof of work" model, the security of this blockchain will be based upon the trust in all participating public academic institutes. ODR Demo: http://kubrik.io/demos/odr/search login for upload available on request) [30]

Scientific sensemaking itself is much deeper integrated into the protocol itself in dsensor.org and it is designed to evolve peer review to a computational consensus model. Using Dsensor [31] if a scientist creates a thesis and wants to test it the scientist enters the hypothesis in computational form (called a Dmap in Dsensor speak). The Mapping protocol then automates the testing of the science, starting by trawling the Dsensor network for relevant data from other peers. That data is then sampled and 'scored' based on its prediction power to verify or challenge the thesis until a computation consensus is established. Science attaining this status then becomes 'computationally active' in the network meaning any peer has the ability to tap into the collective knowledge and feed in their own unique sensor data get the insights from the science working for them.

Blockchain revolution - and beyond blockchains

In the *blockchain* revolution, other systems that show characteristics of *blockchain* systems, such as being distributed, without a single point of failure, decentralized and immutable, but that are **not** based on a blockchain (the data structure), would exist. Actually, they could play a much larger role in the long term than actual blockchain systems.

IPFS (interplanetary filesystem) "is a peer-to-peer distributed file system that seeks to connect all computing devices with the same system of files. In some ways, IPFS is similar to the World Wide Web, but IPFS could be seen as a single BitTorrent swarm, exchanging objects within one Git repository." Research data or publications that are being stored in IPFS would be available without a centralized server and be very effectively distributed among re-users (See Open Document Repository by Kubrik).

There are database systems that have blockchain characteristics. For example, BigchainDB is a "big data distributed database and then adds blockchain characteristics - decentralized control, immutability and the transfer of digital assets." (https://www.bigchaindb.com/whitepaper/bigchaindb-whitepaper.pdf). Many other companies exist providing similar solutions (e.g. ERIS).

Which blockchain for science and knowledge creation?

Blockchain databases may show different characteristics which can be used to divide them into different groups (Table 2).

First, they can be divided by who secures the blockchain database: Can everyone secure the *blockchain* (permissionless) or only certain entities (permissioned). Permissionless *blockchain* databases use the above described Proof-of-work or Proof-of-Stake approaches together with an incentivisation through an intrinsic value token to prevent attacks to the network. Permissioned *blockchain* databases don´t need this, because there are defined and trusted entities that provide 'cryptographic power' to secure the blockchain database. Furthermore it is very important to mention, that a permissioned blockchain does not mean that the 'cryptographic power' providing trusted third parties have any control over the content that is secured within a permissioned blockchain. They cannot censor or approve beyond the defined protocol in the blockchain system (Not like trusted third parties in the current sense such as service providers (e.g. journal publishers, universities, centralized data repositories, etc.)).

Secondly, they can be divided into public and private blockchains. This differentiation refers to who can actually use the *blockchain* database. Is everybody (public) allowed to use the blockchain database or are only certain parties allowed to use it (private)? However, this differentiation is somewhat coarse, because the access and user rights can be much more differentiated depending on the actual use cases. Furthermore, please note that public/private says nothing about who will be able to read the content. For example, a public *blockchain* can still be used to secure non-public research data.

So, for Science and knowledge creation a *blockchain* that is secured by trusted third parties' computers (permissioned) and to which everybody has access (public) seems to be most suitable in the opinion of Soenke Bartling and other peers (Table 2). There are only very few reasons why it should be permissionless, since trusted third parties exist (research institutes, government agencies). Please notice, that those trusted third parties would have no control over what is actually stored in the Blockchain for Science. It is hard to believe that even under the worst circumstances a government or other entity would try to infringe blockchain securing computers in a Blockchain for Science to censor research results. However, this needs to be discussed carefully by the community.

Table 2: A permissioned, public *blockchain* seems to be most suitable for science and knowledge creation.

	Who can use?	
	Public	Private
Permissioned	Science!	Some banking, supply chain, healthcare, IoT, …
Permissionless	Bitcoin, Ethereum, …	…

(Left axis label: Who secures?)

Blockchain and the research cycle

In this section, we collect and propose applications of *blockchain in science and knowledge creation*. We organize this around the research cycle (Figure 3). Ideas and concepts that are published are marked with a green X, while ideas that are newly introduced (to the best of our knowledge) here by us (see publication date of the living doc or its earlier version) are marked with a yellow Y. Copying of ideas, concepts and text from grey literature (e.g. blog posts) about *blockchain* for science and its unattributed reuse has recently caused controversy [32]. We expect established journals and authors to give appropriate credits in their upcoming articles about blockchain for science that include *all means of current* publication methods.
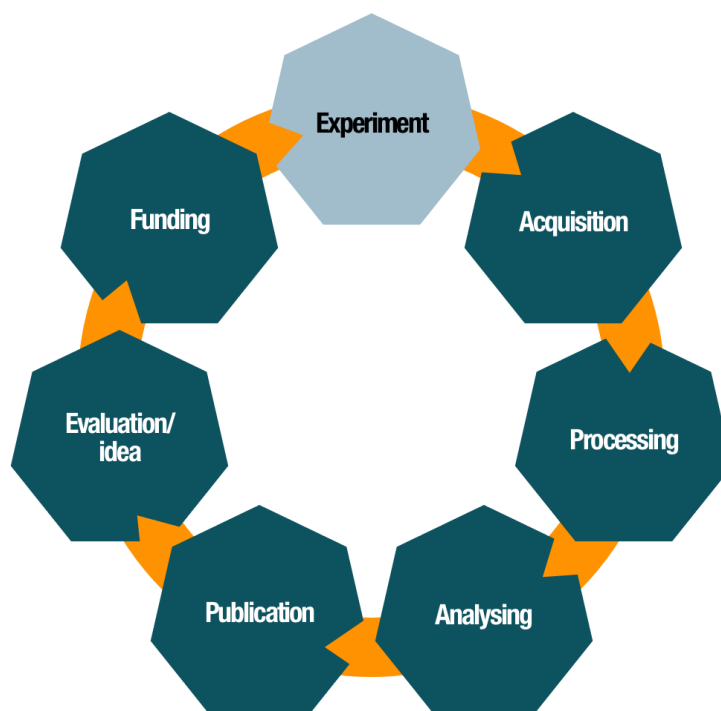

Figure 3: Large parts of the research cycle can make use of *blockchain* (yellow arch); only the experiment itself cannot. From data collection onwards, the rest of the research circle would then become immutable, comprehensible, and externally provable. This would make larger parts of the research cycle open to scientific self-correction and may make more research results reproducible, true, and useful.

Ideas

- X *Blockchains* provide a 'notarization' functionality. Through posting a digest (e.g. cryptographic hash) of a text, data, or general purpose file to a blockchain database, it can be proven that this file or text existed at a certain time point. From this digest,

BLOCKCHAIN
for Science

one cannot conclude on the topic or content of the text or file, but the owner of the text or file can always prove that he or she was in possession of the file/dataset at a certain time point. The time point is defined by the time the block was created in which the digest was posted. This concept is also named 'time-stamping' and 'proof-of-existence' [16]. One easily accessible implementation can be found <u>here</u> [33]. Researchers could post their ideas, research results, or anything else to a *blockchain* system to prove their existence at a certain time point [34][35].

- X For innovations, instead of sending faxes to the patent offices, one could provide a proof-of-existence by posting it to a *blockchain* database [36]. Strong 'prior use' or 'first to invent' claims can be made by the Bitcoin blockchain notarization functionality. <u>Bernstein</u> is working in this space.
- X <u>Lab</u> <u>books</u> could post digests to a *blockchain* system to make them immutable by means of time-stamped entries. A use case is described and potential implications for IP are discussed [37].

Proposal

- X A study design can be pre-registered to a *blockchain*, so that it would prevent the arbitrary alteration of study design after the experiment [38,39]. This can also prevent the arbitrary suppression of research studies from being published in case the results do not meet certain expectations (publication bias) [40]. A registration of studies is recommend to increase the value of research [5,9].

Experiment / data acquisition

- X Using *blockchain* technology, data integrity for approval studies for novel therapy or drugs can be proven to auditors [41,42].
- Y All research data that is acquired could go to a *blockchain* database (Figure X). All data that is acquired during an experiment could then be available first to a certain audience. It could become openly available and could be reused by other researchers. However, this must not necessarily be the case as a researcher could control who may access the data. For example, they could send research data (or representations (e.g. hashes) of it) to a *blockchain* system after initial acquisition, time-stamp it, and still keep it secret up to a certain time point. After this time point (e.g. final publication), they could then release cryptographic codes so as to make the research data publicly available. This could address one issue that is a reason for '<u>Why</u> <u>Most</u> <u>Clinical</u> <u>Research</u> <u>Is</u> <u>Not</u> <u>Useful</u>' and could restore trust in research, which is currently low, because `research is not transparent, when study data, protocols, and other processes are not available for verification or for further use by others [43–45].
- X Clinical trial consent for protocols and their revisions can be made traceable and secured on a blockchain system [46].
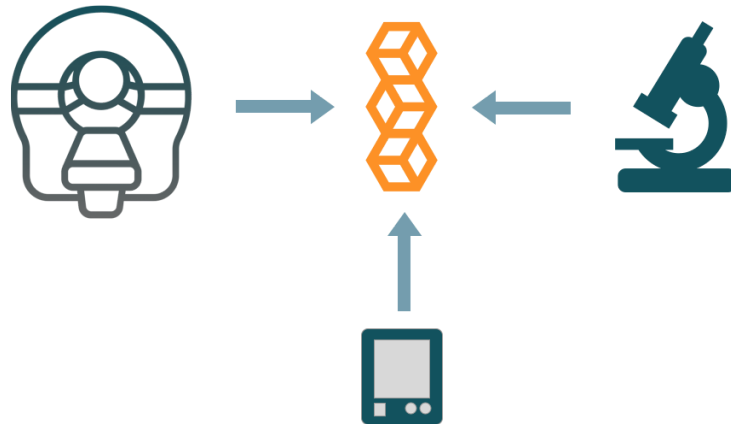
Figure 4: Blockchain to connect Internet of Research things. Lab equipment, microscopes, blots, MRI scanners, digital labbooks (IoRT 'Internet of Research Things) could store the data in a blockchainified database. This would leave an immutable, time-stamped proof of data and its acquisition (limitations: see 'Challenges').

- Y Research data could be acquired by a 'blockchain-ready' sensor (microscope, MRI-scanner, Western-Blot scanner, etc.) in an internet-of-things [47] ('Internet of research things'). Such a sensor would directly encrypt the data (potentially on a hardware level) (Figure 4).
- Y As soon as the data is stored in a *blockchain* database it can be rendered immutable. This means that it cannot be manipulated without leaving a trace (Published at the same time [48]). This can prevent arbitrary data manipulations, be it conscious or inadvertently (e.g. by biased researchers). For example, researchers can prove that they did not drop 'outliers' from the initially acquired datasets, or if so, they would then need to explain as to why they dropped them. Research result manipulations (resulting from whichever motivation it may be) at the level of the initial raw data acquisition would require much more effort than data manipulation in a post-processing sheet - which might only require changing a single digit or image. This could improve scientific reproducibility and may make more research results true.
- Y Blockchainified research data handling significantly extend the ideas and motivations of open data research, since the integrity of the research data can be proven by means of *blockchain*.
- X Blockchainified research data handling could become mandatory for approval studies of novel therapies or drugs, because here truthful data handling, post-processing, and analyses is especially critical, for example, the FDA cooperates with IBM blockchain.

Data management / analysis

- X Bitcoin and many altcoins use large amounts of computational power for the proof-of-work algorithms. The mining incentives could be set in a way so that some of it is also being used for laborious scientific calculations [49].
- X The recommendation to reduce waste in science which reads: 'Public availability of raw data and complete scripts of statistical analysis could be required by journals and funding agencies sponsoring new research' [5,43] could be realized through *blockchain*.
- Y The analysis of the data, postprocessing, and statistics can be protocolized in the *blockchain* database and proven to peers (Figure 5,8). Potentially, statistical analyses and other post-processing steps can run on a *blockchain* system and become provable to the research community. Hashed and time stamped data publication have been suggested.
- Data Management Hub (DaMaHub) is a distributed platform for the scientific data workflow that enable scientists to organise and share research data and outcomes in an easy to use, secure and reputation building way. Data is managed in the normal file system environment and synced between different research partners securely and privately. All users share a data storage system based on IPFS where they publish the open data, open access articles and all corresponding meta data. All data updates are tracked on a public permissioned ledger (blockchain) that is run between nodes. [50]
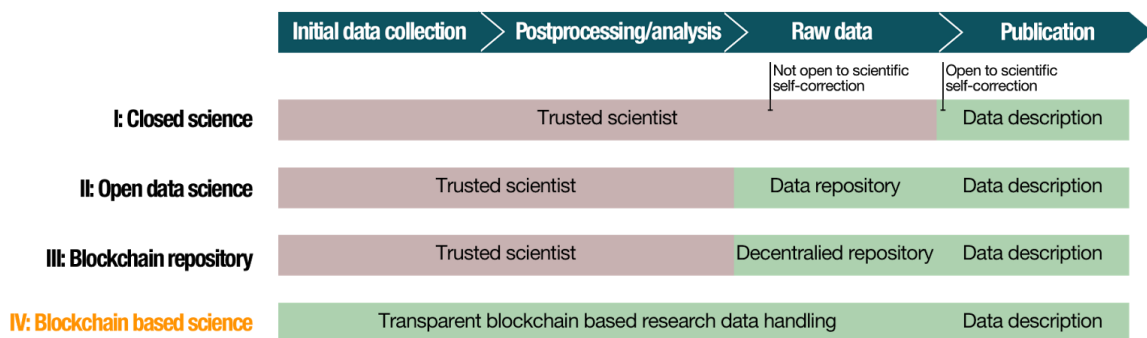

Figure 5: In closed science, scientists just publish a description of their research data and results in their final publication (I). Currently, researchers can publish their research data in data repositories, but that leaves no trace of the data collection or handling process (II). Of course, the final repository could be a blockchain based, decentralized database (III). However, blockchain technology could take the whole process one step further: the whole data handling process could take place in a blockchain system and would therefore be provable and open to scientific self-correction (at will) (IV).

- Y Research data can be post processed and analysed in planned, published and reviewed manner. It can be set in stone and realized as a <u>smart</u> <u>contract</u> (Zach Ramsay, in personal discussion). Ideally a smart contract can result in decision with respect to a research hypothesis (in personal discussions with Zach Ramsay, James Littlejohn). This concept should be called 'Smart Evidence' (Figure 6). It could be a great way to prevent ex-post-facto hypothesizing. Furthermore, it would be a great way for approval studies, e.g. for drugs and novel therapy concepts. The conditions for the approval of a new drug would be set into blockchain stone before the study commences.
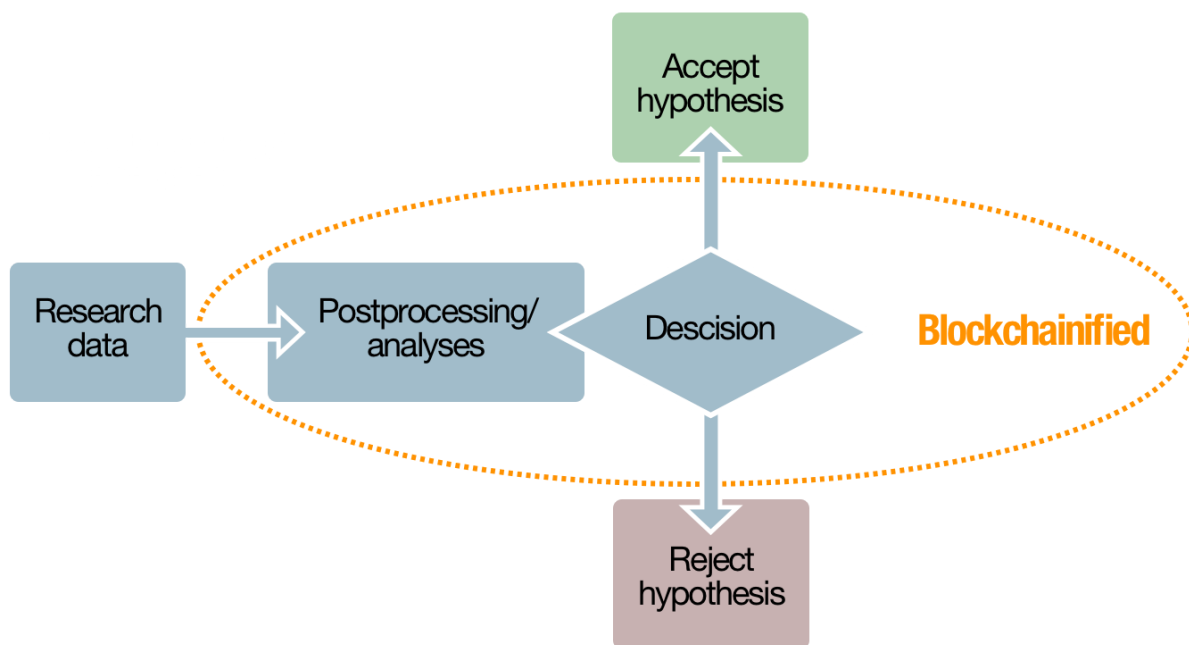

Figure 6: *Smart evidence* - research data postprocessing and analysis are set in 'blockchain stone' before the data is acquired, post-processing and analysis is automated and it may result in acceptance or rejection of the research hypothesis. This would prevent ex-post facto hypothesizing.

- Y Above´s concepts allows anyone to propose (and demonstrate) a different way of doing an analysis. This provides the opportunity for science to act more like a "free-market" where there may be a lab that is really good at producing hypotheses and methodologies, another that has the capacity to run the experiments, and yet another that excels in statistics (Zach Ramsay, personal comment).
- Y Smart contracts can be used to prove that data postprocessing is done in a certain way and only in a certain way, even without revealing the whole transaction process on the blockchain [51]. This opens up novel possibilities to maintain data autonomy and subject privacy in e.g. healthcare or public health research. E.g. Subject data could be sent to smart contract that is openly (/widely) available and that was reviewed by an ethic committee. The smart contract releases data only after a privacy preserving amount of subjects has been reached [52] or only after a certain time period [53], etc. Furthermore, the fundamental problem of identity information

that is being contained in the data (face, genome, etc.) is solved, because the smart contract won´t look for it [54]. The privacy and data autonomy could become so convincing that it might become ethically justifiable that all patient data (even unconsented - under current understanding) could automatically contribute to public health research. Applications are humongous. An example workflow would be: Send all blood-pressure data of all patients to a smart contract, the smart contract averages the patient data with respect to a certain region and time. The smart contract assures that only after reaching a privacy assuring mixing of data the average blood-pressure is made available (Figure 7).

- Y This will shift privacy related questions to: Who do we trust *some* data to do *all* with TO which smart contract do we trust *all* data to do *something* with it.
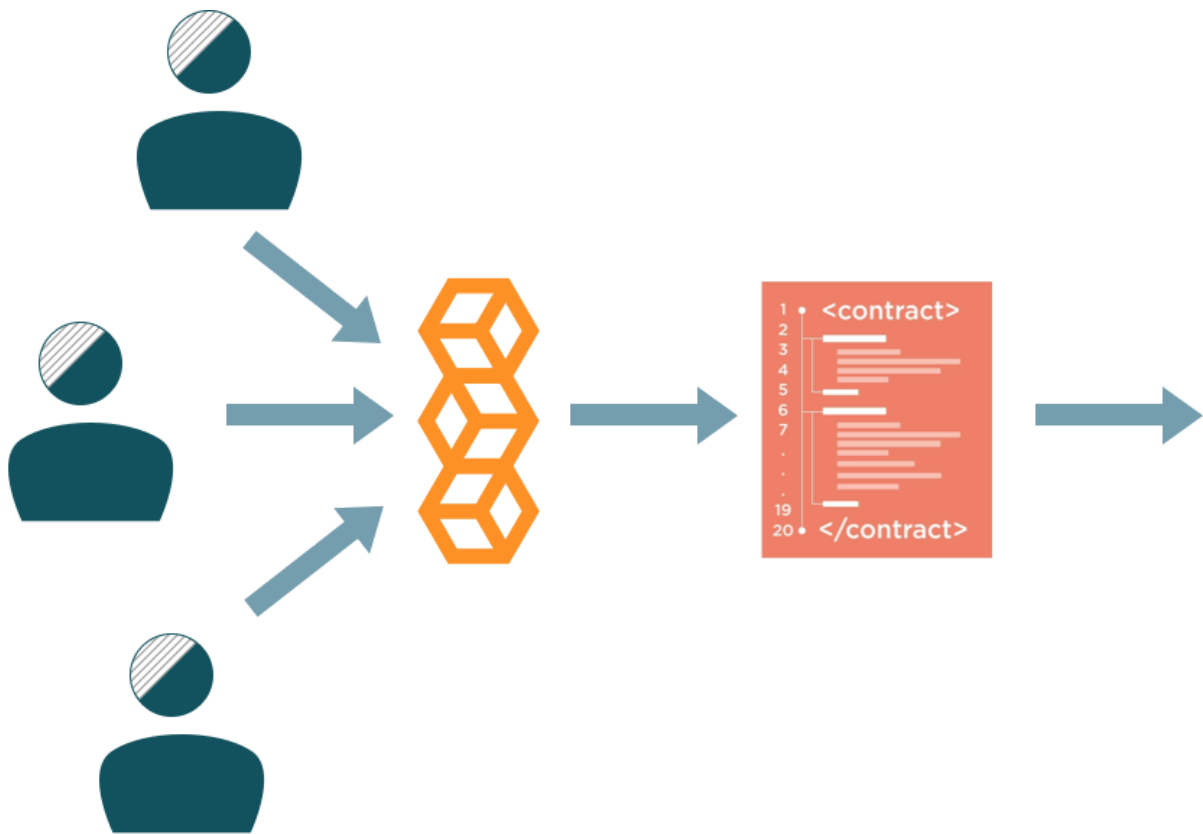

Figure 7: Privacy preserving patient data processing through smart contracts. Patient data is loaded into a blockchain, the subject data (of multiple subjects) sent to a smart contract (Icon from here). The smart contract is reviewed by an commitee and/or public - it is assured that it will only release privacy preserving results, e.g. averages and/or time delayed results. Even if the data itself would reveal the subject´s identity (whole genomes, faces, etc.)- the smart contract won´t look at it. Potential in public health, life science research are humongous.

Data sharing

- X Through *blockchain* databases, data can be stored and shared. Blockchain technologies can provide a redundancy and availability of data, e.g. IPFS. This would be a great way to realize open data research (Figure 4, III).
- X Associated cryptography can assure that the data is only available to certain people, groups and from defined time points onwards. If subject anonymity is of concern, this can be organized by means of using strong cryptography, e.g. in case of healthcare data [55], even without a trustee.
- X Blockchain technology could also be used to 'store' grant money for research and only release it after the publication and/or reproduction of research data/results [39].

Publication/Archiving

- X Publications can be notarized in the blockchain, meaning they can be time-stamped.
- X A decentralized peer-review group (DPG) has been proposed to assure that quality of research [56] or peer-review can be organized using *blockchain [35]*.
- X Ideally *blockchain* systems will be used to timestamp and attribute contributions to dynamic publications and especially low-threshold dynamic publications [57] or granulated publications (e.g. https://www.sciencematters.io/), such as wikis, in which every change (or single scientific observation) can become time-stamped and attributed in *blockchain* (many publications, including personal communication with Lambert Heller).
- X Publications and comments can be shared on a social-media platform and likes, comments, or other interaction can then result in pay-out of coins to incentivise research result sharing [58].
- Y *Blockchain* systems make it possible to publish research anonymously [59] or with a second online identity - and yet one could still get money or other research impact appreciation for it [60,61]. This may make sense if very controversial results are generated and scientists are afraid that this results are 'too disruptive'. Due to the fear of suppression by peers in the complex research social network, they might be afraid to publish such research results or interpretations with their full name [62].
- Y In the form of a 'whistle-blowing' function or anonymous commenting [63], this could also contribute to the internal self-correction of scientific misconduct. If wanted, publications can be claimed later, and the researcher can replace a name placeholder with their real name.
- Y *Blockchain* technology could be used to 'sign' anonymous publications with credibility providing 'signatures'. For example, the publication could be signed with 'An english professor in physics with a Hirsch factor of 15' or 'A German medical doctor'. A research institute could issue cryptographic certificates to do so [64,65].

Research evaluation

- X A *blockchain* (e.g. Namecoin) can be used to register and maintain unique research identifiers like (ORCID) or links to publications or datasets (like DOI) [66].
- X A social network community that incentives content creation and curations can be used to incentives idea, data and results research sharing [29].
- Y The quality of research is currently assessed using impact factor and other altmetrics (like RG score, Altmetric). One has to trust the third parties issuing these to correctly calculate such metrics. With *blockchain* technology and smart contracts, this could change so that the way the metrics are being calculated is externally provable.
- Y A 'research currency' as an incentivization system to 'make more published research results true' as described in [6] could be realized using *blockchain* technology and without a trusted third party, also described as micro-credits [67] .
- Y Science reputation systems can be built using blockchain without a trusted third party.
- X As such, a Decentralized Autonomous Academic Endorsement System has been proposed and interesting implementation ideas and next implementation steps have been disclosed [61].
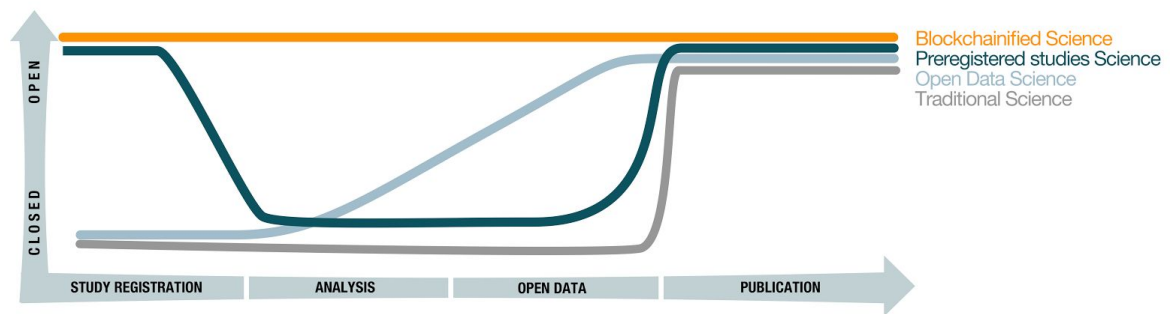
Figure 8: Overview of what parts of the scientific process that are open to scientific self-correction. Blockchainified research may make the whole research process traceable and open - at will.

Research funding

- X Prediction markets [68] to confirm results and to incentivise research could also be used in science [69–72] and could be implemented on *blockchain*.

- X Blockchain could be used to realize a `money-back` functionality for irreproducible research results [48].

- X Blockchain would seem to provide a good mechanism for realising the "credit" systems being proposed for using shared infrastructures like NIH Commons and European Open Science Cloud (Proposed here by Eoghan Ó Carragáin).

- Y New methods of research fund distribution could easily be realized with blockchain technology and smart contracts. For example, a system in which researchers redistribute 50% of their research money among peers [73] can be realized using smart contracts [74]. Research funds could be sent completely anonymously, without trusted third parties.

- Y Similar to a DAO (distributed autonomous organization) [28,75] that could complement functions provided by companies, a DARO (distributed autonomous research organization) can be used to complement research funding agencies.

- Y Concepts similar to colored coins / cryptocurrency tokens could be used to relate research funds to some conditions, even if the distribution mechanism is anonymous and 'black boxed' on *blockchain*. For example, a funding agency could direct the research funds to certain research fields, locations, or institutions. Only researchers that fulfill those requirements would be able to claim those coins.

- Y *Blockchain* could provide many novel ways to distribute research money. For example, research funding provider could pick a combination of characteristics of different kinds of researcher behavior that they want to support. E.g. the amount of patents, citations, tweets, likes, blogs, datasets shared by a researcher, combined with age, location, academic rank, early citations, etc. The problem with the current system is that novel ways of research money distributions aren´t easily employed and system gamers can easily adopt to a constant funding environment. If one asks for patents, there will be patent applications since nothing stops one from writing another rather meaningless application. If funding distribution is under constant and

unforeseeable mutation system gaming will more look like gambling than gaming and researchers might come up with an overall behavior that is best for knowledge creation [76]. At least this might hold true for third party research money. *Blockchain* will also prevent the potential allegation of arbitrariness for research money distribution since the process can be make completely proofable on the blockchain.

- X Blockchain for science and knowledge will not only aid researchers to better conduct and publish their inquiries, but could also engage the public through a more transparent research process. Ultimately this technology could open up the academic process to the public for inquiry and even participation, while simultaneously safeguarding the integrity of their research. This open access could inspire and enable amateur researchers to collaborate with professional researchers in an effort to crowdsource research using the principles of *citizen science*.

Challenges

- One fundamental challenge of *blockchain* is the real-world/blockchain interface problem. How can the blockchain world learn about real-world facts? One instance of this problem is the fact that one has to trust the researcher, sensors, etc. to correctly collect the initial research data. Another example of this problem is the question as to how individual researchers/subjects are recognized within the blockchain world and how their identity is confirmed. This could be done by research institutes (often an institutional email is used to this end, or indeed cryptographic certificates) or other entities that already have a large database of researchers (ORCID, online social networks or publishers).
- The current legislation did not foresee the *blockchain* revolution. Many legal and tax questions remain currently unresolved. These challenges are not specific to blockchain for knowledge creation, but they also exist in other applications of blockchain technology, and are an exciting, evolving field. This is especially intriguing when new funding models (ICOs), anonymous research money distribution, etc. will find widespread use.
- The scalability of most blockchain implementations, e.g. the amount of transactions per unit time, is limited compared to other, centralized technologies, which is kind of obvious, because a status has to flow through a much larger network; the optimization of this scalability is a part of ongoing blockchain research. Sidechains / local blockchains, etc. are one option, many more are discussed on a technical level.
- Implementation! How will a blockchainified research workflow look like? Blockchain needs to be highly integrated into current research workflow and tools.
- Business models! Current business models are arranged around the prospect of creating a single point of failure/container - blockchain changes this - interesting concepts are currently evolving.

Conclusion

The *blockchain revolution* is a game changer and hence chances are that this can be used to break with inappropriate cultures. Indeed, blockchain technology could be used to 'Increase value and reduce waste' [5], by opening the research cycle to scientific self-control

beyond just the final publication and might therefore be a fix to the current reproducibility crisis in science. Furthermore, it could provide new means for the `machine room` of science (e.g. attribution, assessment, research funding, etc.), which could ideally be used to support really innovative research. So blockchain could improve both kind of sciences (Figure 6).

How will blockchain help both kind of sciences?

- Provable, immutable data acquisition, postprocessing & handling
- Smart evidence
- Research subject privacy, crypto-assured study blinding
- Connection to IoRT (Internet of research things)
- Transparent approval studies

More reproducible and true results
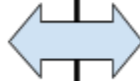
- Unconventional, innovative, but still transparent means of research money distribution
- Early, simple and strong incentivisation of `crazy` concepts / ideas through blockchain tokens
- Researcher anonymity/multiple personae for whistle-blowing / revolutionary stand points

More innovation/discovery

u

Figure 6: How blockchain could help both `kind of sciences'.

1. Swan M. Blockchain: Blueprint for a New Economy [Internet]. O'Reilly Media; 2015. Available: https://books.google.de/books?id=RHJmBgAAQBAJ

2. The trust machine. In: The Economist [Internet]. 2015 [cited 22 Jun 2016]. Available: http://www.economist.com/news/leaders/21677198-technology-behind-bitcoin-could-transform-how-economy-works-trust-machine

3. Meta-Council on Emerging Technologies. In: World Economic Forum [Internet]. [cited 6 Jul 2016]. Available: https://www.weforum.org/communities/meta-council-on-emerging-technologies/

4. Ulieru M. Blockchain Enhances Privacy, Security and Conveyance of Data. Scientific American. Available: http://www.scientificamerican.com/article/blockchain-enhances-privacy-security-and-conveyance-of-data/. Accessed 6 Jul 2016.

5. Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014;383: 166–175. doi:10.1016/S0140-6736(13)62227-8

6. Ioannidis JPA. How to make more published research true. PLoS Med. 2014;11: e1001747. doi:10.1371/journal.pmed.1001747

7. Ioannidis JPA. Why Most Published Research Findings Are False. PLoS Med. 2005;2: e124. doi:10.1371/journal.pmed.0020124

8. Young NS, Ioannidis JPA, Al-Ubaydli O. Why Current Publication Practices May Distort Science. PLoS Med. Oktober 7, 2008;5: e201. doi:10.1371/journal.pmed.0050201

9. Goldacre B. Are clinical trial data shared sufficiently today? No. BMJ. 2013;347: f1880. doi:10.1136/bmj.f1880

10. Engber D. Think Psychology's Replication Crisis Is Bad? Welcome to the One in Medicine. In: Slate Magazine [Internet]. 19 Apr 2016 [cited 25 Jun 2016]. Available: http://www.slate.com/articles/health_and_science/future_tense/2016/04/biomedicine_facing_a_worse_replication_crisis_than_the_one_plaguing_psychology.html

11. Freedman LP, Cockburn IM, Simcoe TS. The Economics of Reproducibility in Preclinical Research. PLoS Biol. 2015;13: e1002165. doi:10.1371/journal.pbio.1002165

12. Angell M. Drug companies & doctors: A story of corruption. New York Rev Books. 2009;56: 8–12. Available: http://www.fondazionedibella.org/cms-web/upl/doc/Documenti-inseriti-dal-2-11-2007/Truth%20About%20The%20Drug%20Companies.pdf

13. Schneider L. Voinnet aftermath: ethical bankruptcy of academic elites. In: For Better Science [Internet]. 7 Apr 2016 [cited 7 Jul 2016]. Available: https://forbetterscience.wordpress.com/2016/04/07/voinnet-aftermath-ethical-bankruptcy-of-academic-elites/

14. pubpeer. A crisis of trust | PubPeer [Internet]. [cited 7 Jul 2016]. Available: http://blog.pubpeer.com/?p=164&utm_source=rss&utm_medium=rss&utm_campaign

=a-crisis-of-trust

15. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, du Sert NP, et al. A manifesto for reproducible science. Nature Human Behaviour. Nature Publishing Group; 2017;1: 0021. doi:10.1038/s41562-016-0021

16. Haber S, Stuart H, W.Scott S. How to time-stamp a digital document. J Cryptology. 1991;3. doi:10.1007/bf00196791

17. Nakamoto S. Bitcoin: A peer-to-peer electronic cash system [Internet]. 2008. Available: http://www.cryptovest.co.uk/resources/Bitcoin%20paper%20Original.pdf

18. Tschorsch F, Scheuermann B. Bitcoin and beyond: A technical survey on decentralized digital currencies. ieeexplore.ieee.org; 2015; Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7423672

19. Lamport L, Shostak R, Pease M. The Byzantine Generals Problem. ACM Trans Program Lang Syst. New York, NY, USA: ACM; 1982;4: 382–401. doi:10.1145/357172.357176

20. Lamport L. The Part-time Parliament. ACM Trans Comput Syst. New York, NY, USA: ACM; 1998;16: 133–169. doi:10.1145/279227.279229

21. Jakobsson M, Juels A. Proofs of work and bread pudding protocols. Secure Information Networks. Springer; 1999. pp. 258–272. Available: http://link.springer.com/chapter/10.1007/978-0-387-35568-9_18

22. Buterin V. On Public and Private Blockchains. Ethereum Blog. 2015;

23. Science B.0 on Twitter. In: Twitter [Internet]. [cited 22 Jun 2016]. Available: https://twitter.com/science_b0/status/712126802064449536

24. Wisniewska A. Altcoins [Internet]. Institute of Economic Research; 2016 May. Report No.: 14/2016. Available: http://ideas.repec.org/p/pes/wpaper/2016no14.html

25. Hurlburt G. Might the Blockchain Outlive Bitcoin? IT Prof. 2016;18: 12–16. doi:10.1109/MITP.2016.21

26. Buterin V. Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform. 2013a. URL $\{$http://ethereum org/ethereum html$\}$.

27. Nick Szabo -- The Idea of Smart Contracts [Internet]. [cited 22 Jun 2016]. Available: http://szabo.best.vwh.net/smart_contracts_idea.html

28. DAOs, DACs, DAs and More: An Incomplete Terminology Guide - Ethereum Blog. In: Ethereum Blog [Internet]. 6 May 2014 [cited 22 Jun 2016]. Available: https://blog.ethereum.org/2014/05/06/daos-dacs-das-and-more-an-incomplete-terminology-guide/

29. (6) T, (7) J, (4) P. A Proposal For An Incentivized Synthetic Biology System Subchain On The Steem Blockchain Platform — Steemit. In: Steemit [Internet]. [cited 6 Aug 2016]. Available: https://steemit.com/science/@transhuman/a-proposal-for-an-incentivized-synthetic-biology-system-subchain-on-the-steem-blockchain-platform

30. Engineering K. Scholarly Publishing & Knowledge Preservation [Internet]. [cited 17 Jan 2017]. Available: http://kubrik.io/projects/odr/

31. NickLambert. Developer Case Study – Dsensor. In: MaidSafe [Internet]. 8 Feb 2017 [cited 20 Mar 2017]. Available: https://blog.maidsafe.net/2017/02/08/developer-case-study-dsensor/

32. Plagiarism concerns raised over popular blockchain paper on catching misconduct - Retraction Watch. In: Retraction Watch [Internet]. 14 Jul 2016 [cited 15 Jul 2016]. Available: http://retractionwatch.com/2016/07/14/plagiarism-concerns-raised-over-popular-blockchain-paper-on-catching-misconduct/

33. Bradley J, Bradley J. Scientific Research Needs a Trustless Blockchain Architecture to Be Trusted - CCN: Financial Bitcoin & Cryptocurrency News. In: CCN: Financial Bitcoin & Cryptocurrency News [Internet]. 13 May 2016 [cited 2 Jul 2016]. Available: https://www.cryptocoinsnews.com/scientific-research-needs-a-trustless-blockchain-architecture-to-be-trusted/

34. Redman J. Clinical Trials Show the Blockchain Can Stop "Fraudulent" Science. In: Bitcoin News [Internet]. 19 May 2016 [cited 2 Jul 2016]. Available: https://news.bitcoin.com/clinical-blockchain-stop-fraud-science/

35. Astroblocks Puts Proofs of Scientific Discoveries on the Bitcoin Blockchain [Internet]. [cited 22 Jun 2016]. Available: http://insidebitcoins.com/news/astroblocks-puts-proofs-of-scientific-discoveries-on-the-bitcoin-blockchain/31153

36. Cawrey D, Wolinsky J, Rampton J, Wolinsky R, Palmer D. Bitcoin's Technology Could Revolutionize Intellectual Property Rights. In: CoinDesk [Internet]. 8 May 2014 [cited 22 Jun 2016]. Available: http://www.coindesk.com/how-block-chain-technology-is-working-to-transform-intellectual-property/

37. Blockchain and Smart-Contracts applied to Evidence Notebook [Internet]. [cited 18 Oct 2016]. Available: https://healthcaresecprivacy.blogspot.de/2016/08/blockchain-and-smart-contracts-applied.html

38. Irving G, Holden J. How blockchain-timestamped protocols could improve the trustworthiness of medical science. F1000Res. 2016;5: 222. doi:10.12688/f1000research.8114.1

39. Carlisle BG. The Grey Literature, apparently — Proof of prespecified endpoints in medical research with the bitcoin blockchain [Internet]. [cited 15 Jul 2016]. Available: http://www.bgcarlisle.com/blog/2014/08/25/proof-of-prespecified-endpoints-in-medical-research-with-the-bitcoin-blockchain/

40. Blockchains For Science: Aligning Research Incentives. In: Doing Distributed Business [Internet]. [cited 22 Jun 2016]. Available: https://db.erisindustries.com/science/2016/03/14/blockchains-and-science/

41. Nugent T, Upton D, Cimpoesu M. Improving data transparency in clinical trials using

blockchain smart contracts. F1000Res. 2016;5: 2541.
doi:10.12688/f1000research.9756.1

42. Repalli J. Blockchain in life sciences [Internet]. [cited 24 Feb 2017]. Available:
https://de.slideshare.net/JayanthiRepalli/blockchain-in-life-sciences

43. Ioannidis JPA. Why Most Clinical Research Is Not Useful. PLoS Med. 2016;13:
e1002049. doi:10.1371/journal.pmed.1002049

44. Schneider L. False priorities at EU2016NL: Mandate Open Data instead of Gold Open
Access! In: For Better Science [Internet]. 28 Apr 2016 [cited 2 Jul 2016]. Available:
https://forbetterscience.wordpress.com/2016/04/28/false-priorities-at-eu2016nl-manda
te-open-data-instead-of-gold-open-access/

45. Dhillon V. Blockchain-enabled open science framework. In: O'Reilly Media [Internet]. 3
Nov 2016 [cited 16 Nov 2016]. Available:
https://www.oreilly.com/ideas/blockchain-enabled-open-science-framework

46. Benchoufi M, Porcher R, Ravaud P. Blockchain protocols in clinical trials: Transparency
and traceability of consent. F1000Res. 2017;6. doi:10.12688/f1000research.10531.1

47. soenkeba on Twitter. In: Twitter [Internet]. [cited 2 Jul 2016]. Available:
https://twitter.com/soenkeba/status/697436268737777664

48. Topol EJ. Money back guarantees for non-reproducible results? BMJ. 2016;353: i2770.
doi:10.1136/bmj.i2770

49. Wagner A. Putting the Blockchain to Work For Science! In: Bitcoin Magazine [Internet].
22 May 2014 [cited 2 Jul 2016]. Available:
https://bitcoinmagazine.com/articles/putting-the-blockchain-to-work-for-science-gridco
in-1400747268

50. Engineering K. Data Management Hub [Internet]. [cited 17 Jan 2017]. Available:
https://kubrik.io/projects/damahub/

51. Kosba A, Miller A, Shi E, Wen Z, Papamanthou C. Hawk: The Blockchain Model of
Cryptography and Privacy-Preserving Smart Contracts. 2016 IEEE Symposium on
Security and Privacy (SP). 2016. pp. 839–858. doi:10.1109/SP.2016.55

52. Twitter [Internet]. [cited 19 Mar 2017]. Available:
https://twitter.com/soenkeba/status/842337875245056000

53. Twitter [Internet]. [cited 19 Mar 2017]. Available:
https://twitter.com/soenkeba/status/842338036553793536

54. Twitter [Internet]. [cited 19 Mar 2017]. Available:
https://twitter.com/science_b0/status/842299809826516992

55. Scott M. The Future of Medical Records: Two Blockchain Experts Weigh In [Internet].
[cited 2 Jul 2016]. Available:
https://btcmanager.com/news/the-future-of-medical-records-two-blockchain-experts-w
eigh-in/

56. Magazine B. Bitcoin Magazine | Bitcoin and Blockchain News [Internet]. [cited 22 Jun

2016]. Available:
https://bitcoinmagazine.com/articles/from-bench-to-bedside-enabling-reproducible-commercial-science-via-blockchain-1464881141)

57. Dynamic Publication Formats and Collaborative Authoring - Springer [Internet]. [cited 22 Jun 2016]. Available:
http://link.springer.com/chapter/10.1007/978-3-319-00026-8_13

58. benjojo, pharesim. Independent scientists could blog their research onto the Steemit blockchain — Steemit. In: Steemit [Internet]. [cited 1 Aug 2016]. Available:
https://steemit.com/research/@benjojo/independent-sientists-could-blog-their-research-onto-the-steemit-blockchain

59. Teif VB. Science 3.0: Corrections to the Science 2.0 paradigm [Internet]. arXiv [cs.DL]. 2013. Available: http://arxiv.org/abs/1301.2522

60. Science B.0 on Twitter. In: Twitter [Internet]. [cited 2 Jul 2016]. Available:
https://twitter.com/science_b0/status/695886036447203328

61. b8d5ad9d974a44e7e2882f986467f4d. Towards Open Science: The Case for a Decentralized Autonomous Academic Endorsement System [Internet]. Zenodo; 2016. doi:10.5281/zenodo.60054

62. Bartling S. Science goes darknet? In: Alexander von Humboldt Institut für Internet und Gesellschaft [Internet]. [cited 29 Nov 2016]. Available:
http://www.hiig.de/blog/science-goes-darknet/

63. PubPeer and Anonymity in Science - Neuroskeptic. In: Neuroskeptic [Internet]. 12 Dec 2015 [cited 13 Mar 2017]. Available:
http://blogs.discovermagazine.com/neuroskeptic/2015/12/12/pubpeer-and-anonymity-in-science/

64. Lab MM. Certificates, Reputation, and the Blockchain — MIT MEDIA LAB. In: Medium [Internet]. 27 Oct 2015 [cited 22 Jun 2016]. Available:
https://medium.com/mit-media-lab/certificates-reputation-and-the-blockchain-aee03622426f

65. The Possibilities of Badges and Blockchain - DML Central. In: DML Central [Internet]. 11 Feb 2016 [cited 22 Jun 2016]. Available:
http://dmlcentral.net/the-possibilities-of-badges-and-blockchain/

66. Lambert Heller on Twitter. In: Twitter [Internet]. [cited 6 Jul 2016]. Available:
https://twitter.com/Lambo/status/192725705556103169

67. Casati R, Origgi G, Simon J. Micro-credits in scientific publishing. Journal of Documentation. emeraldinsight.com; 2011; Available:
http://www.emeraldinsight.com/doi/abs/10.1108/00220411111183546

68. Wolfers J, Zitzewitz E. Prediction Markets [Internet]. National Bureau of Economic Research; 2004. doi:10.3386/w10504

69. Dreber A, Pfeiffer T, Almenberg J, Isaksson S, Wilson B, Chen Y, et al. Using prediction markets to estimate the reproducibility of scientific research. Proc Natl Acad Sci U S A.

2015;112: 15343–15347. doi:**10.1073/pnas.1516179112**

70. Hanson R. Could gambling save science? Encouraging an honest consensus. Social Epistemology. 1995;9: 3–33. doi:**10.1080/02691729508578768**

71. Almenberg J, Kittlitz K, Pfeiffer T. An experiment on prediction markets in science. PLoS One. 2009;4: e8500. doi:**10.1371/journal.pone.0008500**

72. Park I-U, Peacey MW, Munafò MR. Modelling the effects of subjective and objective decision making in scientific peer review. Nature. 2014;506: 93–96. doi:**10.1038/nature12786**

73. Bollen J, Crandall D, Junk D, Ding Y, Börner K. From funding agencies to scientific agency: Collective allocation of science funding as an alternative to peer review. EMBO Rep. 2014;15: 131–133. doi:**10.1002/embr.201338068**

74. Science B.0 on Twitter. In: Twitter [Internet]. [cited 22 Jun 2016]. Available: https://twitter.com/science_b0/status/710412421987508225

75. Popper N. A Venture Fund With Plenty of Virtual Capital, but No Capitalist. The New York Times. 21 May 2016. Available: http://www.nytimes.com/2016/05/22/business/dealbook/crypto-ether-bitcoin-currency.html. Accessed 22 Jun 2016.

76. Bartling S. Research system gamers might become unlucky gamblers through blockchain. In: LinkedIn Pulse [Internet]. 16 Nov 2016 [cited 9 Mar 2017]. Available: https://www.linkedin.com/pulse/research-system-gamers-might-become-unlucky-gamblers-through-s%C3%B6nke

Document status: Living document, comments, suggestions, examples, pointers to wrong attributions and missing references highly appreciated. You disagree with something? Comment anonymously (or pseudonymously by leaving a public key or signature in the comment). The maintainer will never delete or censor any comment, only resolve them!

Published versions here: LSE Impact blog, Zenodo, Researchgate and irights.info (German). This living document (with full history, comments and attributions and hence *post-publication peer-review*) will stay here, open to public since the beginning of July 2016 (publication tweet). A version was time-stamped with proof-of-existence. This publication is an extension of an earlier, pseudonymous, cryptographically signed and hash-time-stamped open publication from February 2015. Attributed 1:1 figure and text reuse most welcome, notification to the maintaining author appreciated.

Reference static versions as:

1. Bartling, Sönke; Fecher, Benedikt. (2016). Blockchain for science and knowledge creation. Zenodo. 10.5281/zenodo.60223
2. Bartling, Sönke; et cont. (2017). Blockchain for Open Science and Knowledge Creation. Zenodo. 10.5281/zenodo.401369