On the Prospects of Blockchain and Distributed Ledger Technologies for Open Science and Academic Publishing

Editorial

Krzysztof Janowicz [a,*], Blake Regalia [a], Pascal Hitzler [b], Gengchen Mai [a], Stephanie Delbecque [c], Maarten Fröhlich [c], Patrick Martinent [d], and Trevor Lazarus [d]

[a] *University of California, Santa Barbara, CA, USA*
E-mails: janowicz@ucsb.edu, blake@geog.ucsb.edu, gengchen_mai@geog.ucsb.edu
[b] *Wright State University, OH, USA*
E-mail: pascal@pascal-hitzler.de
[c] *IOS Press, NL*
E-mails: S.Delbecque@iospress.nl, M.Frohlich@iospress.nl
[d] *Newgen KnowledgeWorks (P) Ltd, IND*
E-mails: patrick@newgen.co, trevor@newgen.co

Abstract. Distributed ledger technologies such as blockchains and smart contracts have the potential to transform many sectors ranging from the handling of health records to real estate. Here we discuss the value proposition of these technologies and crypto-currencies for science in general and academic publishing in specific. We outline concrete use cases, provide an informal model of how the Semantic Web journal's peer-review workflow could benefit from distributed ledger technologies, and also point out challenges in implementing such a setup.

Keywords: Distributed Ledger Technologies, Blockchain, Crypto-coins, Academic Publishing

1. Introduction and Motivation

Simply put, a distributed ledger is a collaboratively managed database of shared, synchronized, and replicated records that typically does not rely on central governance. The ledger is maintained by a network of nodes that store and verify records, e.g., to prevent double-spending. While the popularity of the block-based Bitcoin often leads to the impression that distributed ledger technologies mainly target the financial market and rely on a (block)chain layout for the ledger, the sector is much more diverse both technically and in terms of the addressed application areas [1]. These range from authentication and rights management, data storage (including handling medical records [2, 3]), credit scoring and risk modeling [4], cloud computing, data provenance [5], e-voting, forecasting, commodity markets [6], and supply chain management [7] to shared business applications [8].

Many ledgers available today are open in the sense that everybody can contribute to them, e.g., by having their transactions included or by casting a vote, as well as in the sense that everybody can run a node. Consequently, the resulting network of nodes is distributed globally thereby spanning across cultures, physical factors such as climate zones, and jurisdictions. This is a key factor in the success of these systems as it increases their resilience, e.g., changes in local law or natural disasters do not immediately impact the entire system.

*Corresponding author. E-mail: janowicz@ucsb.edu.

While distributed ledgers and their underlying technologies are easily confused with what is nowadays called crypto-currencies (or coins), they are not the same. For instance, *blockchain* describes the data structure by which transactions (i.e., messages that alter the state of the ledger) are bundled into blocks of a certain maximum size (for the sake of performance) and then cryptographically linked to a growing list. Bitcoin, in contrast, is a crypto-currency that makes use of this blockchain data structure. Additionally, there are many other aspects that describe and define the workings of a crypto-currency, e.g., protocols, clients, such as wallets used to store coins, smart contracts, consensus measures, and so forth. As a result, there are many ways in which all of these components can be combined to arrive at a final ecosystem. To date, this has resulted in more than 1500 coins, most of which see little to no uptake. Finally, many of these components change during the lineage of a coin, sometimes causing disagreement between supporting parties and ultimately leading to a diverging chain split (called a fork) which creates a new coin.

It follows that the community forming around a particular ecosystem is its greatest asset. After all, hundreds or thousands of people have to trust the system to a degree where they are willing to invest their time, hardware, money, reputation, and so forth, knowing well that only a few of the existing coins will establish themselves down the road.

While the term crypto-coin is misleading in numerous ways and many of the coins are rather securities, a number of joint characteristics distinguish most of them from fiat currencies such as the US Dollar. Coin ecosystems are *decentralized and distributed*, i.e., there is no need for an institution such as the US Federal Reserve System, they are *trustless* in the sense that they do not require users to trust the participating parties,[1] they are *transparent* and *autonomous*, i.e., they are governed by open source algorithms and changes that are not in line with the community can be suppressed or their effects mitigated by a fork, they offer some degree of *anonymity*, and they are *immutable* in the sense that information can be added but not (secretly) edited or removed.

Academic publishing and (open) science more broadly are among the potential application areas for distributed ledger technologies and crypto-coins. Over the past months, this has led to numerous projects, most of them in a very early stage. The visions put forward in these proposals are often bold but also lacking in two critical ways: (1) they typically do not provide the level of detail required to understand their workings and value proposition, and (2) they seem to lack the combination of actors involved in academic publishing that would allow to test drive these visions in a realistic setup.

Here we report on the outcomes of a meeting to explore the potential of distributed ledger technologies for academic publishing that took place in November 2017 in Santa Barbara, California, and brought three parties together: IOS Press as a publishing house, NEWGEN as a software engineering company familiar with journal management systems and academic workflows, as well as researchers from Wright State University and University of California, Santa Barbara, more precisely the editors-in-chief of the Semantic Web journal and some of their team members. Besides conceptualizing how a coin ecosystem may drive academic publishing in general and the Semantic Web journal in particular, the participants sought to understand and anticipate usability and scalability problems for users, i.e., scientists and the general public, but also other parties such as funding agencies and publishers.

2. Distributed Ledger Technologies for Science

We see at least the following areas where distributed ledger technologies such as blockchain could benefit science; see also [9]:

1. Editing, reviewing and publishing academic work, e.g., by making the journal management workflows transparent.
2. Managing, i.e., storing and curating, scientific data to support the reproducibility of results and improve access to scientific data.
3. Connecting researchers to funding sources such as foundations or reversing the process entirely and allowing researchers to bid for existing proposals, e.g., social challenges.
4. Managing intellectual property, establishing identity, and preventing fraud.
5. Democratizing science by making various decisions on the level of funding agencies, journal editorial boards, conference organizers, award and career committees, etc., more transparent and by enabling the research community to vote on important decisions.

[1] As long as malicious actors do not control a substantial share of the nodes or processing power.

6. Opening up the black-boxes resulting from algorithms and closed data sources such as impact factors, citation counts, and so on.

In the following, we will outline use cases for each of these application areas and point out which characteristics of distributed ledger technologies they require.

(1) Open access refers to making research outputs, mainly publications and the utilized data, freely and publicly accessible, often under a Creative Commons license to foster reuse. However, the peer review process as such can also be opened, e.g., by making submissions available during the review process or publishing reviews online as well. In this case, it makes sense to distinguish between openness and transparency. We call a review process open if the submissions and reviews are publicly available and transparent when the entire workflow, i.e., the assignment of editors and reviewers, the decisions taken, potential revisions, author responses, and so forth, are made available as well [10]. The Semantic Web journal follows such a setup and shares all data as Linked Data.[2] Besides making the review process more transparent, this new wealth of data also enables linked scientometrics [11] and more advanced search capabilities for articles, authors, reviewers, and journals by combining vector embeddings computed from the full-text submissions and knowledge graphs generated from the journal management workflow [12].

As long as this setup would be restricted to the Semantic Web journal (or at least the same and trustworthy academic publisher), one would not need distributed ledger technologies as no decentralization is present, the setup is mostly transparent by design, and there is no reason to distrust the involved parties (or, at least, their identity is known). However, all this changes rapidly, when multiple journals and conferences organized by various actors are involved. In such a case, it is unlikely and even undesirable for a few parties to act as central data storage, identity management gateways, and maintainers of services. Hence, such a setup would indeed benefit from distributed ledger technologies [13].

Moreover, these technologies and ledgers could take on some important additional tasks such as managing timestamps, voting on issues that affect the entire ecosystem such as publication fees, and so on. Finally, they could provide a technical solution around incentivization, e.g., by assigning coin rewards to tasks such as editing or reviewing; see also [14, 15]. Put differently, so-called smart contracts could be used to model the agreement between a journal and a reviewer to submit the review by an agreed deadline.

(2) Reproducibility of scientific experiments and reusability of data are major themes in academia but also for the broader public and its credibility crisis. One can envision how all scientific data, scientific procedures and software used for sampling, data preparation, visualization, and so forth [16–20] could be shared on a distributed and immutable file system such as the InterPlanetary File System (IPFS). Similarly, all uses of data, e.g., publications, could be automatically linked to the datasets to generate provenance records. In contrast to today's situation where research teams are often in exclusive control of their data, storing them on a ledger would make all edits permanently visible [21, 22]. Hence, everybody could track what data were used for a scientific publication and whether they have been altered in some way. Most of the currently existing blockchains would not be suitable for such an approach as adding data to them is a slow and very expensive process. However, projects such as Multichain could address these issues and also provide support for private chains. In general, Linked Data and ontologies that have been developed to model scientific workflows, observation data, and provenance records would be well-suited to describe the data to be stored on the chain.

(3) Similar to the examples above, distributed ledger technologies can also be used to handle calls, submissions, reviewing, voting, etc., for research funding. We can envision at least three ways in which this would work. The first case is analogous to the academic publishing use case outlined above but would handle the submission, review, and selection of research proposals. Fairness, in general, seems to be an important issue as the competition for funds is increasing, while researchers that compete for the same resources are reviewing each others' proposals. Second, researchers could vote for or suggest research directions more easily and based on a wider community engagement. Third, one could reverse the funding process and instead of researchers submitting their proposals to a funding agency, interested parties could put forward challenges and rewards for addressing them. Similarly, researchers could put their portfolios online for donors to chose from.

[2] All IOS Press bibliographic data is available as Linked Data at LD Connect. For instance, the SWJ data can be accessed at http://ld.iospress.nl/ios/sw.

(4) With a rapidly growing scientific community and competition, managing the identity of researchers, institutions, funding agencies, publishers, etc., becomes a more pressing issue, e.g., to reduce predatory publishing and other practices. These services could also assist in co-reference resolution, thereby improving information retrieval, knowledge graphs, and so on. Finally, they could also assist in other tasks that relate to trust such as repeated submissions and questions relating to prior work, e.g., whether somebody used ideas and methods from a proposal s/he was asked to review.

(5) Voting has been mentioned in the use cases before because it is a key application area of distributed ledger technologies. In general, these technologies could help to flatten the hierarchies that still dominate the scientific community and enable a broader base to form to arrive at decisions about tenure, the publishing culture, and so forth.

(6) As in so many other areas of everyday life, decisions taken in science rely increasingly on closed data and algorithms. The results returned by these systems can have dramatic consequences for the individual, yet most of these systems are black boxes. The examples above mostly relate to the data storage capability of blockchains, while this use case would also benefit from the (source) 'code is law' culture of crypto ecosystems. Typically, these ecosystems are entirely driven by consensus and open source. For instance, smart contracts are protocols that define and enforce the execution of a (legal) contract, e.g., to handle transactions, without the need for a third party. The code of these smart contracts is public. Using distributed ledger technologies, both the data and the algorithms would be openly available, thereby making the results of measures such as a journal's impact factors or an author's h-index reproducible.

3. Using Technology to Solve Social Problems

Many of the use cases outlined above have a strong social component. After all, distributed ledger technologies were developed to function in a trustless and decentralized environment. Nonetheless, one has to be careful if trying to apply these technologies to solve social problems. For instance, before deciding to make measures such as the h-index reproducible by computing it based on open citation data, we should ask ourselves whether we want research to be governed by such measures in the first place. Similarly, before essentially turning the acquisition of research funding into a bounty hunt, we have to understand the implications, e.g., with respect to the *Matthew effect*.[3] Will voting and bounty-based models lead to an even more unbalanced distribution of funding based on the popularity and marketability of a topic? Along the same lines, if we distrust our colleagues and workflows, is technology the right answer? After all, results can be altered or sensors manipulated before data become available on a blockchain.

Other characteristics of distributed ledger technologies may have problematic side effects as well. For instance, we discussed the immutability of the InterPlanetary File System and blockchains in general as an advantage above, but one can also take a different perspective. As editors-in-chief of the Semantic Web journal, we are frequently asked by authors to depublish rejected manuscripts and we have a policy for doing so: papers can be depublished after a minimum of 4 weeks after the decision letter has been announced. The need for such a compromise highlights why strict immutability is problematic. Many journals and conferences have fixed rules about publishing overlapping work and their systems will search the Web for existing similar publications or plagiarism. In the early days of the Semantic Web journal, we frequently had to explain to editors of other outlets that a paper available on our webpage has not been published and can safely be resubmitted to another journal. We also had numerous other cases that required the ability to change records permanently, e.g., where authors asked to resubmit a paper (before reviewers had been assigned) because they forgot to remove a potentially embarrassing comment from the submission. Similarly, we assume that most conference organizers are familiar with reviewers accidentally submitting a review for another paper or putting the confidential part of a review into the textbox that will become visible to authors.

Next, there are cases where authors need to submit their work without identifying themselves such as William Sealy Gosset having to publish his famous t-distribution under the pseudonym *Student*. While permanent identifiers such as ORCID make this increasingly difficult, distributed ledger technologies have a potential to do both [23]: either make anonymity almost impossible by linking all scientific activities to profiles/addresses, or enable strong anonymity – a feature promoted by some crypto-currency ecosystems.

[3] Named after Matthew 25:29: 'For to every one who has will more be given, and he will have abundance; but from him who has not, even what he has will be taken away.'

The discussion whether we want pseudonymous contributions needs to be a social question before it is a technological one.

Finally, crypto-currencies have also been proposed as an incentive and reward model for reviewers and editors. The Semantic Web journal publishes the names of the reviewers and editors in the header of every paper not only to increase transparency but also to give them credit for their time and energy. This show of appreciation could be complemented with a coin reward, however financial incentives may lead to unintended consequences [24].

4. Modeling Journal Management Workflows using Distributed Ledger Technologies and Crypto-Coins

In the following, we will introduce an informal model for use case **(1)** to highlight how distributed ledger technologies and crypto-coins could be integrated into the Semantic Web journal's workflows and academic publishing in general. This shall serve as a demonstration of the kinds of decisions that would be involved, which problems may arise, and what the value proposition of such a setup would be.

4.1. A Publishing Ecosystem

As described above, the potential for distributed ledger technologies is best utilized by taking the publishing ecosystem into account and by not merely focusing on a single journal or conference. Hence, we will assume that multiple outlets such as journals and conference proceedings from various publishers are involved. These publishers, outlets, funding agencies, and researchers (in various roles such as editors, reviewers, and authors) form a publishing ecosystem that will be driven by a crypto-coin (called HypatiaCoin [HYC] here) and will use a blockchain as shared data storage for metadata relevant to academic publishing. The coin will be used to vote, as an incentive for reviewers and editors, to cover the costs of (open access) publishing such as developing and maintaining journal management systems, typesetting, printing, administrative overhead, and so on. Most of the functionality offered by the ecosystem will be implemented in the form of smart contracts.

4.2. Minimal Smart Contract Functionality

Without going into technical details, we assume that the coin will be a so-called *ERC20 token*[4] that utilizes the existing Ethereum (ETH) blockchain and ecosystem.[5] There are many advantages and disadvantages to doing so, most of which play no role in this early design phase. We will discuss some of them below to give the reader an impression of the details (and their surprising consequences) that would have to be worked out before any serious coin ecosystem for science could go into production mode. As far as the review process (as a subpart of the larger setup) is concerned, these smart contracts would have to handle at least the following functionalities, explained in more detail further below, where each function has its own set of permissions and preconditions that govern who can invoke it and when. We will assume an open and transparent review process as defined by the Semantic Web journal.

- instantiate % *A journal or conference*
- receivePaper % *Submit manuscript & coins to the journal*
- assignEditor % *Assign an editor to a paper*
- changeEditor % *Calls assignEditor*
- assignReviewer % *Set reviewer, deadline, and reward*
- deleteReviewer
- acceptInvitation % *Reviewer accepts contract*
- submitReview % *This will trigger the reward as long as it was called before the deadline and after approveRecommendation has been called*
- submitRecommendation % *Editor submits decision*
- approveRecommendation % *EiC approves decision or returns it to the editor; in the first case this will trigger the reward to reviewers and the editor*
- publishDecision % *Decision becomes official*
- receiveHistory % *Returns a URI of a previous paper page*
- submitOpenReview % *For non-solicited open reviews*
- approveOpenReview % *Editor decises whether to include the review*
- withdrawSubmission % *One can only trigger withdrawal if a public key has been submitted together with the submission*

[4] https://github.com/ethereum/EIPs/blob/master/EIPS/eip-20.md

[5] Other ecosystems such as NEM (https://nem.io/) or even a consortium, i.e., non-public, blockchain may be suitable as well.

- increaseActivity % *A score of all contributions such as reviewing, editing, authoring, etc.*
- sendToLottery % *Send unused coins to the lottery*
- releaseFunds % *Releases all assigned coins*

4.3. Workflow Integration

Here we briefly discuss how the smart contract above would be executed to model the Semantic Web journal's review workflows; details such as the lottery will be motivated and explained below.

An author team submits their paper (*receivePaper*) together with a pre-defined amount of HYC. These tokens will be used up during the review and publication process or will be sent to a lottery. The unique paper URI – generated through hashing of linked metadata and full text – is created and written to the blockchain. Hence, the submission time is known and the content cannot be altered without leading to a new hash.

An editor is assigned (*assignEditor*) and begins to invite reviewers (*assignReviewer*). The editor and those reviewers who accepted the invitation (*acceptInvitation*) are added to the Linked Data that describes the submission and written to the blockchain. Reviewers are essentially modeled via their public keys, i.e., (wallet) addresses, thereby allowing for anonymity.

The reviewers are expected to submit their reviews on time (*submitReview*). Otherwise, the contract is not fulfilled and no reward is paid out. This may require additional functionalities not modeled here, such as the editor granting a deadline extension. After all reviews are submitted and their hashed URIs have been added to the blockchain, the editor recommends (*submitRecommendation*) a specific decision, say *minor revisions*.

The editors-in-chief either approve or disapprove the recommendation (*approveRecommendation*) and publish the final decision (*publishDecision*). Once the decision has been written to the blockchain, *increaseActivity* and *releaseFunds* will be used to determine the distribution of coins. Of course, submissions often take multiple iterations before finally getting accepted. Coins that remain after the entire process finished by either finally accepting or rejecting the manuscript will be transferred to a lottery (*sendToLottery*).

4.4. Understanding Coin-based Incentives

The model above seems straightforward and while the community forming around the use of distributed ledger technologies for science is evolving rapidly, we believe that it is a representative example of what has been proposed to date. One of the goals of the November 2017 meeting was to go beyond the surface and explore the consequences of such a coin-driven model.

Starting with the positive, the presented workflow increases transparency in multiple ways, rewards editors and reviewers for their work, allows authors that would not be able to pay for (open access) publications to participate by using coins received from reviewing, manages timestamps and deadlines, and so on.

However, creating such an incentive-based model, specifically one that largely relies on the automatic execution of source code while minimizing human interaction, can lead to unforeseen consequences. Consider, for example, the case where a paper submission costs 100 HYC. Let us also assume that the editor receives 10 coins, and each reviewer receives 20. An additional 20 coins go to the journal and/or publisher. Given that a paper should receive at least 3 reviews, this process would spend 90 coins, leaving 10 HYC for open reviews or the lottery.

Such a setup will most likely encourage overly positive or negative reviews as the reviewers get paid a single time, and, thus, will try to minimize the revisions that a paper has to go through. The same can be said for the editors. There is also no room for additional open reviews once the budget is used up. The alternative would be that each round of revisions has to be paid by the authors. In this case, however, the incentive for reviews and editors will be to protract the process even if one would implement a model for diminishing returns. In such a case, reviewers could merely contribute to those rounds that seem worth their time and then either accept or reject the manuscript, or become unavailable. Moreover, authors would be discouraged from submitting their work due to uncertain costs.

Finally, as the coin has to cover costs in the physical world, such as the time of reviewers, software development, backups, storage costs, and so forth, the coin will have some value as measured against other coins or fiat currency, e.g., US Dollars. Today's coins, such as Bitcoin, Ethereum, or ADA, are very volatile, often increasing by hundreds of percent within months just to lose 90% of these gains over a brief period. If the HypatiaCoin turned out to be relatively unstable, authors would submit their papers during low prices as measured against the US Dollar (USD) and act as reviewers during higher prices. This would create the unfortunate situation that most papers would become available at the same time where the least researchers are willing to review the new submissions and the other way around. Keeping the coin stable, however, may prove to be difficult. If the coin would indeed be an ERC20 token– as many existing proposals suggest – each transaction of HYC would result in small transaction fees determined by a so-called GAS price, the reward received by ETH nodes for executing smart contracts. Hence, changing ETH-USD rates will impact HYC prices. To give a concrete example, the price per ETH was at about \$10 in January 2017, reached nearly \$1400 in January 2018, and dropped below \$400 in April of the same year. Without going into technical details, the GAS price is also proportional to the likelihood at which nodes decide to include a transaction and, thus, authors may see the need to pay more to get their papers submitted close to a deadline or during increased network activity.

4.5. *Winning the Lottery*

The issues described above will require a consensus-based solution that combines social and technical aspects. Here, we will outline how the technical component could be handled to avoid introducing false incentives.

Simply put, preventing actors from gaming the system can be achieved by introducing an element of randomness. To do so, we propose to assign an *activity* score to all possible interactions that can occur during academic publishing such as submitting a manuscript, reviewing, editing, commenting, organizing an event, publishing an article, and so forth. Every time an actor performs such an activity her score is increased by a certain amount (*increaseActivity*) and written to the chain. For instance, submitting a manuscript may yield 10 points, while writing a review may increase one's activity score by 5 points. Such an activity score models how much an actor contributes to the ecosystem. Coins that have not been used up during the review process are transferred to a lottery that regularly distributes coins among members of the community. The likelihood of this distribution is proportional to the activity score. After receiving coins from the lottery, the activity score of the winning actors would be reduced by a certain amount to mitigate the Matthew effect, while others would continue to accumulate activity. This makes it unattractive to unnecessarily drag out the review process (as returns are uncertain) and still leaves sufficient incentive for multiple rounds of revisions. It also mitigates the negative effects of value fluctuation as submitting a manuscript increases the activity score, thereby making a (partial) return of the invested coins likely. The exact setup of the system, the number of coins distributed to a participant, e.g., 100 HYC, and the relation to the fixed rewards (if any), have to be fine-tuned and will require experiments and adjustments. Algorithm 1 illustrates how the lottery would function.

Algorithm 1 HYC Lottery Outline

```
procedure SELECTACTORFORPAYOUT
    p                 ▷ Fraction of actors that get a payout
    a                 ▷ Number of actors with activity > 0
    t                 ▷ Sum of all activity scores of all actors
    s                 ▷ Payout penalty
    i ← a ∗ p
    ral ← random.shuffle(actor_list)
    while i ≠ 0 do                    ▷ Find i actors for payout
        r ← rand(1, t)
        while ral.hasNext() do        ▷ Iterate actors
            ra ← ral.next()
            r ← r − ral.getActivityScore()
            if r <= 0 then
                ra.reduceActivityScore(s)
                return ra             ▷ Actor selected
        i − −;
    ...
```

The following parameters can be used to adjust the lottery to balance the relative importance of contributions and prevent an overly strong Matthew effect. First, each activity, such as reviewing, has a fixed score associated to it. Second, p determines the fraction of actors that will receive a payout during any given round, say 10%. We assume everybody gets the same payout but other models could be considered as well. Third, s determines the reduction in activity score after winning the lottery. Finally, instead of a linear relation between accumulated activity and payout, one would likely use $\lceil log_b(activity\_score) \rceil$ instead. As the lottery payout is driven by an actor's activity budget and not HYC coins, the setup rewards contribution instead of holding coins.[6]

[6] Some readers may notice that this differs from many so-called (delegated) proof-of-stake coins and is more analogous to NEM's proof-of-importance which is also rewarding contribution over mere holding. Note, however, that in our examples we assume that HYC is a ERC20 token running on the Ethereum blockchain which uses a proof-of-work consensus mechanism (and will move to proof-of-stake in the future). Hence, strictly speaking, HYC does not have a consensus mechanism and the model described here is more accurately described as a so-called airdrop.

In order to generate random numbers and achieve fairness in the lottery, one cannot simply implement a traditional pseudorandom number generator, e.g., the Mersenne Twister, in the smart contract code since the exact state that determines the next output is public knowledge and thus susceptible to attack. Instead, generating random numbers for lottery services on the blockchain remains an ongoing challenge. Currently, many decentralized applications (dapps) rely on external *oracle services* which are trusted third parties that are capable of providing unpredictable random numbers to a smart contract. However, these are centralized services that lack the trustless nature of decentralized apps. A more promising candidate, based on the BLS signature scheme by Boneh, Lynn and Shacham [25], is a blockchain protocol capable of generating random numbers with verifiable results and is aptly dubbed "Proof-of-Randomness"[26].

4.6. Deflationary Coin Model

We assume that there will be a fixed amount of HypatiaCoin. This implies that after all coins are in circulation, the price at which the coin will be traded, and, thus, the costs and rewards for activities such as submitting a paper, will be adjusted over time. Since new coins will not be added and existing coins may be held instead of being circulated[7] or even lost, e.g., by losing a private key, the value (as compared to fiat currencies) of HypatiaCoin will likely increase over time by virtue of limited supply.

At the beginning, however, the coin has to be distributed and the ecosystem kick-started at a time when the lottery does not hold sufficient coins for the incentive model to work. Hence, we propose to only distribute a fraction of the total amount of HYC at the beginning. This could be done by releasing the amount of coins required for the submission of one manuscript to all researchers that have been involved in the participating journals. The remaining coins will be distributed in two ways, one portion will be set aside and will be released whenever a new journal joins the ecosystem; we will discuss this case in the next section. The other portion will be stepwise included in the lottery until all coins are used up. By the time all coins are in circulation, the system will be self-sustainable by the constant flow of coins between actors and the trust in the ecosystem will be sufficient for new actors to acquire coins on the market, e.g., in exchange for USD. These assumptions are in line with existing crypto-coins and their ecosystems. Whether they will hold in general remains to be seen.

[7] Although, we hope that the lottery will discourage mere holding.

Naively, the distribution of coins towards the lottery can be implemented by simply reducing the amount of distributed coins by some percentage per step (Eq. 1). Here, p_r represents the coin payout added to the lottery in a given round r on top of the payouts left from the review process and other activities, t_r is the amount of coins left after the payout at $r-1$, d is the default payout percentage, e.g., 0.1.

$$p_r = t_r * d \quad (1)$$

However, such an approach would over-proportionally favor early adopters, and, thus, may not be suitable for a flatter adoption curve. Hence, similar to so-called difficulty adjustment algorithms, one could relate the payout to the overall activity in such a way that a lower than average activity increases payout, while an increasing activity reduces the payout (Eq. 2). Simply put, one could realize such a model as follows, where p_r is the payout at round r, t_r is the amount of coins left after the payout at $r-1$, d is the default payout percentage, e.g., 0.1, $\overline{a}_x$ is the average activity over a moving window of x rounds, say 3, and a_r is the activity accumulated during lottery round r.

$$p_r = t_r * d * (\overline{a}_x / a_r) \quad (2)$$

4.7. Beyond a Single Journal

Handling academic publishing will require substantially more functionality than the smart contract and its incentive models outlined above. For instance, voting procedures for adding new journals, conferences, their publishers, and so forth have to be designed. Each of these outlets may have their own fees and rewards structure, may follow a slightly different review workflow, and will require a complex setup to grant and revoke rights, e.g., for guest editors. There may even be reasons to exclude actors such as predatory publishers. To truly realize the potential of a distributed ledger technologies based academic publishing ecosystem would likely require a so-called decentralized autonomous organization (DAO), i.e., an organization that is governed by smart contracts for all of its tasks. As the DAO would also be in control of the lottery and changes to smart contracts that govern the journal management process, errors in its setup may have severe consequences. One of the first organizations of this kind, simply called 'The DAO', was hacked briefly after its launch leading to a loss of $50 million.

We believe that the creation of a DAO for academic publishing or even open science more broadly may be a multi-year project and would have to be staged in the sense that one would start with a loosely organized and consensus-driven group of pioneering journals to which more structure is added over time. Thereby, researchers would already benefit from distributed ledger technologies for academic publishing without the final system having to be in place. This would also enable each journal to run a different setup, and, thus, speed up the process of finding the right incentive models and their parameterizations. Whether these journals could already share a common coin remains to be seen. This would not necessarily be a drawback as a new ledger can be created once the DAO is in place that imports the balances from each journal and its actors. In fact, this is a common operation in crypto-coin ecosystems.

In contrast, the idea that a DAO could be developed early on and by a single project outside of the realm of academic publishing seems naive.

4.8. Distributed Versus Decentralized

So far, we did not explicitly distinguish between a distributed and a decentralized ecosystem. For example, a system can be distributed in the sense that it runs on nodes spread out geographically across continents but still be centralized in that all nodes are run by the same agent. Ethereum co-founder Buterin distinguishes between three kinds of decentralization: architectural, political, and logical centralization. Blockchains are typically decentralized architecturally and politically as they neither have an infrastructural single point of failure nor a central governing agent. However, they are logically centralized as they share a common state and act as a single computational unit.

We believe that a similar distinction has to be made on the level of blockchain-driven applications. To illustrate this point, we will discuss four cases where usability is improved by increasing centralization; thereby weakening some of the value propositions of the distributed ledger paradigm.

First, consider the minimal smart contract functionality introduced in subsection 4.2. Editors decide and approve decisions, trigger the publication of reviews, and so forth. We believe that such an intermediate step is important for a variety of reasons. For instance, it helps to prevent cases where reviewers have accidentally submitted a review for a different paper or have posted confidential notes to the editor in the field accessible to authors. Similarly, in contrast to the standing editorial board, guest editors are less familiar with the general quality expectations of a journal, its internal workflow, and review criteria for different paper types. Without discussing and approving their decisions, a journal would not be able to establish a consistent policy and profile. Hence, pushing manuscripts, reviews, and decisions to the blockchain immediately, may have negative consequences for authors, reviewers, editors, and possibly the entire journal. Nonetheless, introducing an intermediate step reduces transparency and introduces central architectural and political steering elements to an otherwise decentralized system.

Second, the possibility to submit reviews anonymously is an important instrument. However, it is not straightforward to implement using smart contracts on an unencrypted blockchain such as Ethereum, as opposed to the anonymity that comes for free with an encrypted blockchain such as ZCASH.[8] This is for multiple reasons. First, if the editors should invite reviewers through a function call, they would need to collect their public addresses. Second, the review rewards have to be sent to those addresses. Making the reviewers responsible for their own anonymity, e.g., by using new addresses for every transaction, puts an unreasonable burden on them. An alternative would be not to pay out rewards directly after each review but only if enough payouts have accumulated across the entire ecosystem so that the periodic payouts cannot easily be traced back to a reviewer. However, this essentially introduces a trustee – the very actor that distributed ledger technologies are trying to supersede.

Third, if authors should directly interact with the smart contract, they will have to pay GAS. There are three ways to handle this. Either the authors hold a small amount (less than 1 US Cent worth) of ETH themselves, the journal triggers the contract and incurs the cost, or the journal sends ETH to the authors for them to interact with the contract. The issue here is not the amount of ETH needed but the introduced complexity and technical expertise required by the authors.

[8] https://z.cash

We believe that the HYC ecosystem should provide simple interfaces to hide the complexity and all technical details from its users but this will come at the cost of increased centralization.

Fourth, for sake of simplicity, we assumed that only hashed URIs are stored on the blockchain which give access to all journal management workflow metadata. Other solutions could include storing more information or even the full papers, reviews, figures, used data, and so on. However, what data will be stored off-chain is not just a question of reducing costs[9] but also one of decentralization and distribution. Finally, if data are automatically written to an immutable chain without human curation, will malicious actors be able do compromise the entire system by uploading illegal content thereby distributing it across nodes that may face legal consequences for storing such data?

5. Conclusions and Future Steps

In this work, we outlined the potential of distributed ledger technologies for science and more specifically for academic publishing. Our goal was to showcase how the journal management workflow of the Semantic Web journal could be modeled using a crypto-coin ecosystem. The journal is particularly suited for such an experiment as it already follows an open and transparent review process and stores most of the data generated in the process as publicly available Linked Data using a SPARQL endpoint.

Distributed ledger technologies and crypto-coin ecosystems have the potential to transform academic publishing in a number of ways. For instance, they could store scientific data and results of the peer-review process transparently in a distributed, permanent, fail-safe fashion, they could break apart the strong relation between journals and publishers and instead give publishing houses a new role as digital service providers, they could reduce the risk of least publishable units, would enable the creation of a marketplace for reviewers and reward their contributions, they would make the manipulation of scientific data, experiments, and the review process more difficult, they handle deadlines more efficiently, would enable contributions from actors that would otherwise not be able to pay for publication costs, support community-wide voting on major issues, and so on.

Setting up such an ecosystem and modeling the interactions which may occur using technologies such as smart contracts is far from trivial. Hence, the goal of this work was to present an outline of the potential benefits, demonstrate how such an ecosystem could be set up, illustrate the difficulties that may arise from incentivisation and potential solutions to them, and give the reader an impression of the many decisions that would have to be made as well as their consequences. While we believe that distributed ledger technologies (and to some degree crypto-coins based on them) hold great potential and we are enthusiastic about the role they will play in fostering open science, the so-called whitepapers put forward by many of the projects that surfaced over the past months do not contain any of the details discussed on the previous pages. We believe that most or even all of the issues raised can be overcome, but this will require a consensus process developed together with a broad base of researchers, publishers, and funding agencies. In the meantime, experiments with individual journals or conferences may lead the way.

6. Appendix: Existing Projects

As mentioned above, there are more than a dozen projects that aim at incorporating distributed ledger technologies into various stages of science. Most of these projects are in a very early stage and it is often not clear what their value proposition is and whether they have a team and roadmap to follow through. Moreover, the whitepapers published so far do not contain sufficient details on how to archive their goals. Below, we list some representative projects that seem to be moving forward as of June 2018.

- **Blockchain for Open Science**: A 'living' document[10] outlining the potential of blockchain for various stages of the research cycle ranging from data collection to publication.
- **ScienceRoot** [11] tackles similar problems to the one described in our work and the blockchain for open science document, e.g., incentivizing reviewing and creating an immutable archive of scientific publications.

[9] as the ecosystem would have to be able to add terabytes of new or versioned data per day to an immutable blockchain.

[10] Available at https://tinyurl.com/y87og5tz

[11] Whitepaper available at: https://www.scienceroot.com/resources/whitepaper.pdf.

– **Frankl**,[12] a distributed ledger based toolkit to foster open science with a focus on data archiving.
– **Pluto Network**[13] aims at creating a decentralized scholarly publication framework together with a reputation mechanism.

References

[1] M. Walport, Distributed ledger technology: Beyond blockchain, *UK Government Office for Science* (2016).
[2] S. Angraal, H.M. Krumholz and W.L. Schulz, Blockchain technology: applications in health care, *Circulation: Cardiovascular Quality and Outcomes* **10**(9) (2017), 003800. doi:https://doi.org/10.1161/CIRCOUTCOMES.117.003800.
[3] M.B. Hoy, An introduction to the Blockchain and its implications for libraries and medicine, *Medical reference services quarterly* **36**(3) (2017), 273–279.
[4] H. Byström et al., Blockchains, real-time accounting and the future of credit risk modeling, *Lund University, Department of Economics* (2016).
[5] X. Liang, S. Shetty, D. Tosh, C. Kamhoua, K. Kwiat and L. Njilla, Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability, in: *Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, IEEE Press, 2017, pp. 468–477.
[6] J.J. Sikorski, J. Haughton and M. Kraft, Blockchain technology in the chemical industry: Machine-to-machine electricity market, *Applied Energy* **195** (2017), 234–246. doi:https://doi.org/10.1016/j.apenergy.2017.03.039.
[7] H.M. Kim and M. Laskowski, Toward an ontology-driven blockchain design for supply-chain provenance, *Intelligent Systems in Accounting, Finance and Management* **25**(1) (2018), 18–27.
[8] M. Pilkington, 11 Blockchain technology: principles and applications, *Research handbook on digital transformations* (2016), 225.
[9] J. van Rossum, Blockchain for research, *Digital Science, November* (2017). doi:https://doi.org/10.6084/m9.figshare.5607778.
[10] K. Janowicz and P. Hitzler, Open and transparent: the review process of the Semantic Web journal, *Learned Publishing* **25**(1) (2012), 48–55. doi:https://doi.org/10.1087/20120107.
[11] Y. Hu, K. Janowicz, G. McKenzie, K. Sengupta and P. Hitzler, A linked-data-driven and semantically-enabled journal portal for scientometrics, in: *International Semantic Web Conference*, 2013, pp. 114–129. doi:https://doi.org/10.1007/978-3-642-41338-4_8.
[12] G. Mai, K. Janowicz and B. Yan, Combining Text Embedding and Knowledge Graph Embedding Techniques for Academic Search Engines, in: *4th Workshop on Semantic Deep Learning (SemDeep-4)*.
[13] J.P. Tennant, J.M. Dugan, D. Graziotin, D.C. Jacques, F. Waldner, D. Mietchen, Y. Elkhatib, L.B. Collister, C.K. Pikas, T. Crick et al., A multi-disciplinary perspective on emergent and future innovations in peer review, *F1000Research* **6** (2017). doi:10.12688/f1000research.12037.2.
[14] Z. Jan, A. Third, L.-D. Ibanez, M. Bachler, E. Simperl and J. Domingue, ScienceMiles: Digital Currency for Researchers, in: *Companion of the The Web Conference 2018 on The Web Conference 2018*, International World Wide Web Conferences Steering Committee, 2018, pp. 1183–1186.
[15] M. Hauser and E. Fehr, An incentive solution to the peer review problem, *PLoS Biology* **5**(4) (2007), 107.
[16] T. Lebo, S. Sahoo, D. McGuinness, K. Belhajjame, J. Cheney, D. Corsar, D. Garijo, S. Soiland-Reyes, S. Zednik and J. Zhao, PROV-O: The PROV Ontology, *W3C recommendation* **30** (2013).
[17] A. Haller, K. Janowicz, S.J. Cox, M. Lefrançois, K. Taylor, D. Le Phuoc, J. Lieberman, R. García-Castro, R. Atkinson and C. Stadler, The Modular SSN Ontology: A Joint W3C and OGC Standard Specifying the Semantics of Sensors, Observations, Sampling, and Actuation.
[18] B. Ludäscher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E.A. Lee, J. Tao and Y. Zhao, Scientific workflow management and the Kepler system, *Concurrency and Computation: Practice and Experience* **18**(10) (2006), 1039–1065. doi:https://doi.org/10.1002/cpe.994.
[19] Y. Gil, E. Deelman, M. Ellisman, T. Fahringer, G. Fox, D. Gannon, C. Goble, M. Livny, L. Moreau and J. Myers, Examining the challenges of scientific workflows, *Computer* **40**(12) (2007). doi:10.1109/MC.2007.421.
[20] C.A. Goble and D.C. De Roure, myExperiment: social networking for workflow-using e-scientists, in: *Proceedings of the 2nd workshop on Workflows in support of large-scale science*, ACM, 2007.
[21] A. Extance, Could Bitcoin technology help science?, *Nature* **552**(7685) (2017), 301. doi:10.1038/d41586-017-08589-4.
[22] S. Bartling and B. Fecher, Could Blockchain provide the technical fix to solve science's reproducibility crisis?, *Impact of Social Sciences Blog* (2016).
[23] J. DuPont and A.C. Squicciarini, Toward de-anonymizing Bitcoin by mapping users location, in: *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*, ACM, 2015, pp. 139–141. doi:10.1145/2699026.2699128.
[24] U. Gneezy and A. Rustichini, Incentives, punishment and behavior, *Advances in behavioral economics* (2004), 572–89.
[25] D. Boneh, B. Lynn and H. Shacham, Short signatures from the Weil pairing, in: *International Conference on the Theory and Application of Cryptology and Information Security*, Springer, 2001, pp. 514–532.
[26] UltraYOLO Team, UltraYOLO: Decentralized Lottery Protocol and DAO using Proof-of-Randomness, Technical Report, UltraYOLO, 2017. https://ultrayolo.com/whitepaper.pdf.

[12] Whitepaper available at: https://docsend.com/view/gn8t7k9.
[13] Whitepaper available at: https://assets.pluto.network/Pluto_white_paper_v04_180719_1355_BSH.pdf.