The Decentralized Autonomous Organization for academia: Incentivizing a more open scientific process

Abstract

Scientific information should be accessible by all and the scientific process should be transparent. Moreover, scientific data should be stored in a way that it cannot be controlled or manipulated by any centralized entity. Until now, these goals are loosely the consensus among the academic community, but nearly impossible to implement at the global scale. The recent innovations in blockchain technology offer tools that could help decentralize and open science in a comprehensive way. Here, I propose a blockchain-based decentralized autonomous organization for academia. The organization addresses centralization concerns by moving intellectual property onto a decentralized blockchain, but it also opens the possibilities for a lot more. With a decentralized network like this, anyone could publish at any time. The structure of the organization is such that scientists are incentivized to publish early and often, to do thorough peer reviews, and they are recognized for their contributions in the classroom. While blockchain technology is still in its infancy, innovators are now beginning to look for applications beyond the financial sector. Academics and intellectual property will surely benefit from these innovations if we are open to accepting them.

1. Introduction

Science originally developed around the culture of natural philosophy, where a group of thinkers shared ideas with one another. As time progressed and that group of thinkers got bigger, individuals wanted to benchmark ideas as their own. The first official patents were developed in 1474 (Mandich, 1948). Eventually, patenting inventions alone was not enough, thinkers wanted to publish their understanding of the universe, and the first scientific journal was published in 1665 by the Royal Society (Moxham and Fyfe, 2017). This was an exclusive group that initially consisted of only 40 members. Finally, in the late 19th Century, many more subscription and/or for-profit journals started publishing scientific articles. Notably, *Science* and *Nature* magazines were both started at this time. While much science was published as letters between authors, this is also the time where the peer-review process was formalized and the publishing process became what it is today. Unfortunately, publication procedures are still centralized with a few influential publishers controlling most of the scientific information.

In the last decade, we have begun to see an accelerated trend of science moving away from exclusive clubs, subscription journals, etc., and toward more open platforms. Open-access journals allow authors to retain the rights to their work and makes it freely available to the public (Swan and Brown, 2004). Funding organizations like the National Science Foundation require that data be shared. Open version control platforms like Github are a popular way to share software (Thung et al., 2013). Social networks like ResearchGate create a platform for science to be communicated and shared (Thelwall and Kousha, 2017). Orcid and GoogleScholar track scientific work and attach it to an author's identity. Figshare allows scientists to publish and/or cite any form of scientific content including figures, datasets, images, videos, etc. (Thelwall and Kousha, 2016).

These steps toward open science in recent years are an improvement over the previous publishing model in that they open access to more people and encourage transparency of the scientific process. Looking forward though, we should continue to strive for a more wholistic solution that more heavily incentivizes participation and openness. Data products should be accessible to anyone. The publication process should be free to anyone who has an idea that they want to benchmark. The recent innovations in blockchain technology could offer solutions to enhance the decentralization of scientific inquiry. Blockchains are an immutable record; in most cases that record is a set of transactions, but it could just as easily be used as a history of scientific discoveries and to document the scientific process. If science was recorded to a public blockchain it would forever be accessible by anyone.

Here, I outline the organizational structure for a plan to decentralize the scientific process by integrating scientific discovery onto a blockchain. I outline some of the features of the decentralized organization, emphasizing the incentivization scheme that is based on rewarding participants with blockchain-based tokens that represent their work or achievement. I also provide some examples for the flow of work and the tokenomics of the incentivization scheme.

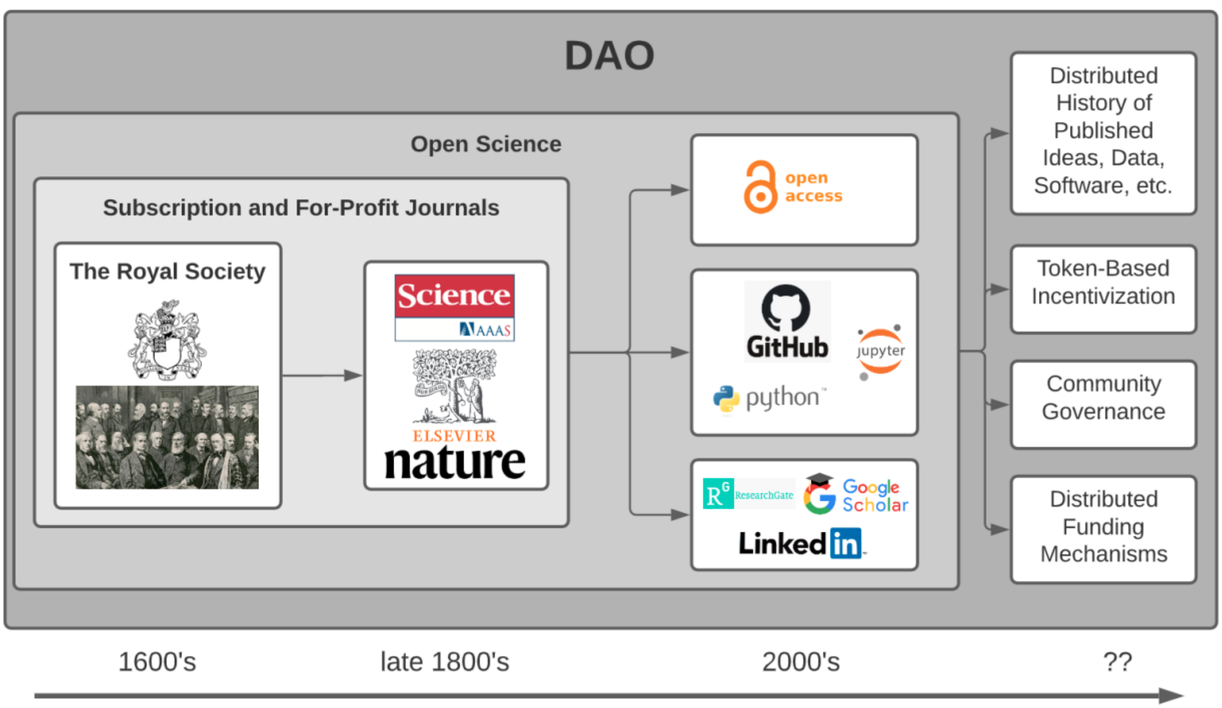

Figure 1: A history and projected future of scientific publishing. The academic *DAO* is meant to include but expand upon all the scientific platforms that we already use today. As open science builds upon the subscription journals that we have used for hundreds of years, so the *DAO*

will take advantage of the open science tools that we have been building in the last decade.

3. A *DAO* for Academia

In this section, I outline the features of a *decentralized autonomous organization (DAO;* see Appendix A*)* for academia. The primary goal of this organization is to open the scientific process, to be more decentralized and public. A secondary goal is to realign the incentives for those who work in academics so that they are aligned with the production of more useful scientific content. I often use language specific to the blockchain community; in those cases I *italicize* the word and provide a definition in Appendix A.

3.1 Decentralization and Openness

Scientific knowledge and the scientific process should be public goods; that is, access should be non-excludable and non-rivalrous. Air is an example of a public good in that every person can breathe it and the act of one individual breathing does not inhibit the same act from another individual. While science is trending toward openness and therefore becoming more of a public good, most journals still institute a paywall, publishing can be inhibitively expensive, and data products are often concealed well after they are published. Design of the academic *DAO* is meant to accelerate open science and enforce sharing of science as a public good.

First, storage and sharing of scientific knowledge including data products, articles, software, etc., will be on a decentralized database (Benisi et al., 2020) or what I will call a '*public ledger*'. In essence, volunteers from the community rent space on their computers which are then used to store the complete archive of scientific information, and those volunteers are rewarded for doing so. Because the network is decentralized, access to this information cannot be restricted from anyone or at any time. Protocols that are currently running decentralized storage networks use rational proof of storage algorithms (Moran and Orlov, 2019) to maintain that the data do in fact exist in the state in which they were sent. When a scientist wants to submit their work to this *public ledger*, they will do so by signing a transaction using their *private key* which is associated with their *digital identity*. Then, the scientific object that was published to the database will forever be attached to their signature and will therefore be provably theirs. That author's *digital identity* with be attributed whenever that data or idea is cited in the future (see Section 3.2).

Of the science stored on this *public ledger*, some will be peer reviewed, but some not. Each scientific object will be assigned a *hash*, much like a doi, and will be recognized as being 'owned' by the authors which submitted this object to the ledger. Peer reviewed science will also show a partial ownership to the reviewer (see incentivized reward mechanisms for both authors and reviewers in section 3.2 below). Scientists often desire to make private contributions, whether that is because they are publishing on a controversial topic or because they are offering a critical review of a colleague's article. Today, those private contributions can be made but only if the author is willing to forego any recognition. Within the *DAO*, recognition can be provided even to private contributions (Halpin and Piekarska, 2017). Anonymous contributions would be made, where the contributor's *digital identity* was encrypted within the scientific object, but future use of that object would still be recognized and cited to the contributor in an anonymous way.

Funds for scientific research are generally distributed through government programs, either at the federal, state, or local level. There are some exceptions including privately funded grants as well as crowd-sourced funding; these are generally looked down upon because these mechanisms can proceed without peer review. An alternative within the academic *DAO* would be to use *smart contracts* to do quadratic funding (Pasquini, 2020). Here, both private and crowd-sourced funds could be matched by public funds within the context of a conventional peer review. Private donors would choose a proposal that suits them and submit funds to a *smart contract*. This smart contract would hold the funds until a peer-review decision is made on that proposal. If the peer review is successful, those funds would be contributed to the funded grant; if not, those funds would be returned to the donor.

Initially, the protocols for this *DAO* will be set up by a group of developers, some from the academic community to outline the structure and some from the blockchain community to implement the protocols as code. Once a model for the *DAO* has been established, any future changes will be subject to community governance. Community members who have made contributions to the *DAO* and have been cited in ways that have earned them tokens (section 3.2) will be allowed to vote on proposals to improve the *DAO*. Since the *DAO* protocols are written as *smart contracts,* they can only be changed by the person or entity which controls that contract; in this case the owner would be the *DAO* itself. In other words, it is impossible to change the protocol in a *DAO* without a passing vote from the members of the *DAO*.

3.2 Token-Based Incentivization

In today's scientific community, there are two general metrics by which an individual is assessed. The first metric is the education level achieved and is assessed by the degree which the individual holds. The main problem with this metric is that it assumes learning happens in a single discrete achievement, upon degree completion. Students who fail to complete their degree are not recognized as having achieved anything, and students who continue to pursue knowledge after their degree are only recognized as having achieved the degree itself. In reality, learning is a continuous process and should be recognized as such.

The second metric is scientific content produced and this is assessed by citation metrics, generally the h-factor. This is a metric based on citations from authored articles that are recognized by Google Scholar. An individual who has earned an h-factor of N has authored at least N articles, each of which has been cited at least N times. This incentivization encourages authorship of manuscripts, again a discrete contribution metric which often fails to recognize significant work that may have been done outside of the manuscript itself. Science is a continuous process and contributions should be recognized in smaller increments. The h-factor-based citation metrics encourage scientists to force every project into a publishable manuscript, when in reality many contributions would be better made in a more continuous way. Software projects, for example, can be contributed primarily by a few individuals but then used by many others. This should be recognized as a significant contribution to the community even if it is not published as a citable manuscript.

Within the *DAO*, academic achievements will be *tokenized*. Tokens will be used to incentivize scientific production as well as overall utility of products, and to recognize those scientific contributions. Tokens are digital objects attached to a person's *digital identity*. These tokens have no monetary value. Instead, a token either represents an achievement or measures the usefulness of a member's scientific output. They will be desirable for the same reason that a high h-index is desirable, because they are proof of one's scientific activity. They will be used in job applications, in proposals, etc. Tokens are distributed by any person or entity whom the *DAO* has credited as a figure of authority, for example universities, journal editors, etc. The signature for token awards will require a *proof of authority* (De Angelis et al., 2018).

There are 5 separate token classifications (each of which is explained in detail below):

- Education
- Teaching
- Publication
- Review
- Citation

Taken together, these tokens are a continuous record of an individual's performance. Hence, they act as an online and ever-evolving CV (Pearce, 2019; Sharples and Domingue, 2016). While the scientist of today is often fixated on their h-index, a measure of their citation count only, the scientist of tomorrow should be incentivized to continue learning to earn education tokens, they should be more rigorously assessed on their teaching effort, they should be adequately compensated for their peer review.

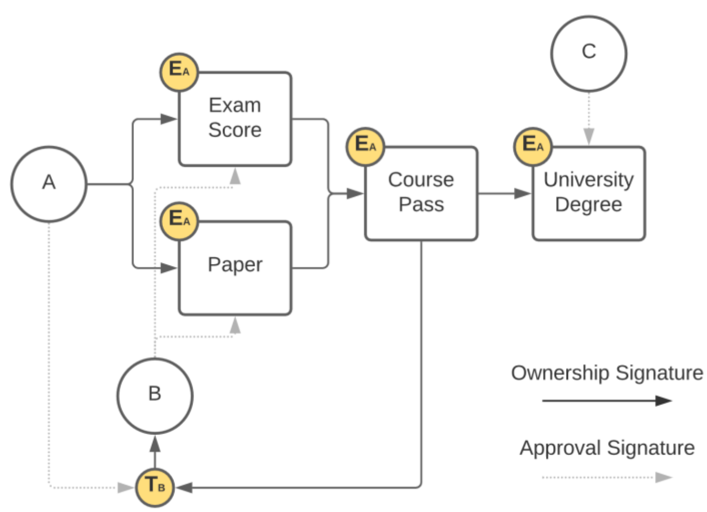

Figure 2: A flow diagram for education and teaching tokens. Large circles represent members of the *DAO*, A, B, and C. Rectangles represent incremental achievements, some of which contain sub-achievements. Each achievement is *tokenized* with an education token (yellow circles with an 'E' for education and subscript letter for the owner) that belongs to the member which

provided the ownership signature. Most achievements are not valid unless signed by some authority with an approval signature. Teaching tokens are rewarded to instructors and are signed by the students.

3.2.1 Education Tokens (*Non-Fungible*)

Education tokens record intellectual effort to the *public ledger*. These tokens are awarded to a student by the instructor of a course. As mentioned above, the current 'reward' for an education is binary, either the student achieves the degree or not. There is almost no recognition given to a student who does not complete their degree, even if they take and pass all but one course. However, the actual learning happens in smaller increments.

Being rewarded through the *DAO*, these education tokens allow for a more continuous record of one's education where learned skills, achievements, even things as small as exam scores are distributed to the *public ledger* (Figure 2). Some tokens are 'sub'-achievements, where many small tokens will be rewritten into one larger encompassing token eventually, but can still be accessed and acknowledged at the smaller level. Exams within a course would be an example of a sub-achievement; a passing score on all exams is required to complete the course, so the course completion token is not awarded if not all exam completion tokens have been achieved. Given enough completed courses, a student could be rewarded a degree, which would be stored as a token on the *public ledger* as well. However, degree completion would not necessarily be required, nor would it be sufficient, these tokens are meant to represent a continuity of learning rather than a discrete achievement.

3.2.2 Teaching Tokens (*Fungible*)

Teacher evaluations will be recorded to the *public ledger* in units of teaching tokens. Each student of a course has the opportunity to reward their instructor teaching tokens alongside their written evaluation of the course. Since these tokens are fungible, there is no uniqueness between courses. The total number of tokens available for an instructor to earn is equal to the number of credit hours of the course. Those tokens are divided between all the students and each has the opportunity to award the instructor with whatever percentage of their token allotment that they see fit. Teaching tokens will come in amounts less than an individual unit.

3.2.3 Publication Tokens (*Non-Fungible*)

Publication tokens identify the author(s) of a scientific product. Much like the education tokens, these are meant to be a more continuous record of scientific achievement than is recognized by today's citation-based metrics. If a data product is usable by others, it should be published with a *hash*, or doi, recorded onto the *public ledger*. The same is true for code, standalone figures, presentations, patents, etc.

Also like the education tokens, these publication tokens can contain 'sub'-achievements where the same authors use a previous piece of work to publish something new. As shown in Figure 3, scientists C and D could work separately on two portions of a project, publishing each item independently, then work together incrementing their progress as they go. Scientist C writes the software for their model and publishes that on their own.

Scientist D, meanwhile, collects data to test their model and publishes that separately as well. To benchmark their progress they then work together to create a figure illustrating their work so far. They publish this with the code from C and the data from D both contained as *hashes* inside of the figure's *hash*. Because the figure directly uses work from both C and D, each is included as an author of the figure as well as the subsequent manuscript. This is true no matter how long the publication chain is. Direct use necessary for authorship is distinct from a citation of another author's publication (as described in section 3.2.5). Published items can always be changed and published as a newer version (even after they are reviewed; see section 3.2.4), but will be considered a distinct publication (and will lose their review status).

As was discussed by Sharples and Domingue (2016), this continuous publication process allows authors to lodge a public record of their ideas and their work. It removes the pressure to hastily write a manuscript in an effort to avoid getting 'scooped'. Continuous publication would slow down the rush to publish a written manuscript, and it would give credibility in publishing investigations that do not result in a peer-reviewed manuscript or those that give an ambiguous result. Today, those efforts generally go uncredited.

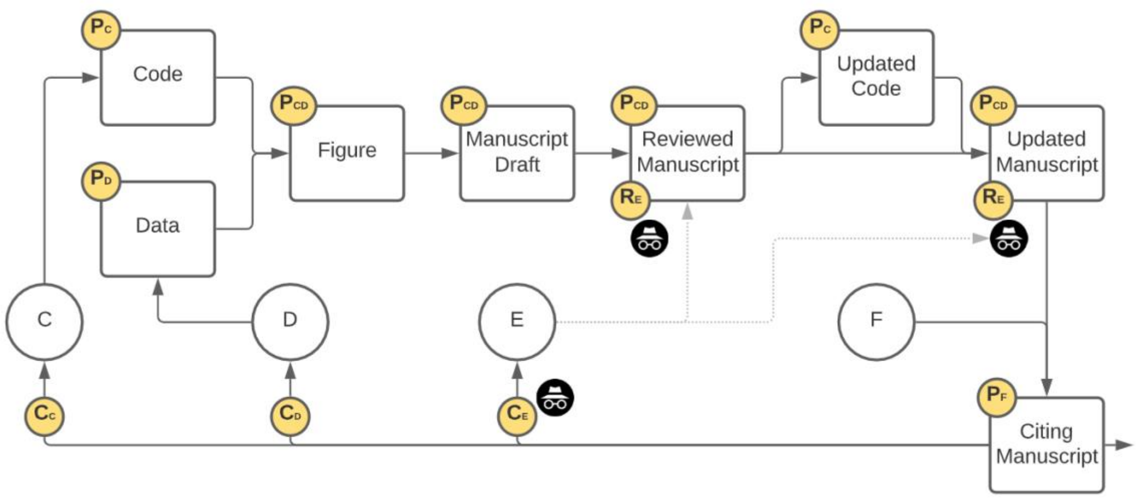

Figure 3: A flow diagram for publication, review, and citation tokens. As in Figure 2, large circles represent members of the *DAO*, rectangles represent milestones or achievements, and small yellow circles represent tokens. The black icon of glasses and a hat represents a private transaction or token. Arrows represent signatures as in Figure 2.

3.2.4 Review Tokens (*Non-Fungible*)

Peer review is an important component of the scientific process. Not only does it weed out illegitimate science, but it enhances the level of science that is published by incouraging a scientist to rethink and refine their own ideas. Sadly though, the act of peer review is expected to be done with no reward and is therefore completely voluntary. Considering that many academics are already overworked with their mandatory obligations, they generally don't have time to do many thorough peer reviews. Within the academic *DAO*, a peer review could be attached to a *digital identity*, either publicly or anonymously if desired, so that the reviewer can do the review transparently through the public ledger (Mackey et al., 2019; Spearpoint, 2017; Tenorio-Fornés et al., 2019) and be recognized (even anonymously) for their contribution.

Continuing with the example in Figure 3, authors C and D publish their manuscript to the *public ledger* and submit for review. The journal editor, who is assigned based on *proof of authority*, chooses at least one reviewer for the manuscript, likely more. When the reviewer(s) are assigned, the review would be initiated as a *smart contract* on the *public ledger* (Mackey et al., 2019; Spearpoint, 2017; Tenorio-Fornés et al., 2019). The contract enforces that upon completion of the review and acceptance by the editor: 1) the reviewed product will be republished with the reviewers private key as a signature proving that the review was done, and 2) the reviewer will be rewarded a token for their efforts. The reviewer token is non-fungible in that it is attached to the reviewed product much like the publication token is.

3.2.5 Citation Tokens (*Fungible*)

Citation tokens are much like Google Scholar citations except that they are recorded and stored on the *public ledger*. Any time that a previously published scientific product is used by another author to publish something new, the cited author is awarded a single citation token. This is true for any scientific product, it does not need to be peer reviewed to be cited, although peer-reviewed products will likely be considered more credible and will therefore be cited more. Peer reviewers also earn credit for articles that they reviewed when that article is cited. This gives the reviewer an incentive to enhance the science as much as they can rather than purely being critical. While I include these peer-reviewer citation tokens in the same category they would actually not be identical to the publisher-citation tokens. Each would be *fungible* on its own, but with each being tracked separately.

In regard to Figure 3, a new *DAO* member, F, publishes a new manuscript which cites the original written by C and D. As soon as the citing manuscript is added to the *public ledger*, citation tokens are awarded to the authors of the cited manuscript, C and D, but a special citation token is also awarded to the reviewer, E. The reviewer chose to be anonymous, so the token is transferred to E in a private way, but the fact that E acted as a reviewer on *some* scientific document that got cited would be public information. If F went on to update their manuscript, no additional citation tokens would be provided, only one set of citation tokens per chain of scientific products.

4. Discussion

With my proposed academic *DAO*, I outline several features which are meant to create a better and more open scientific community. The question remains though, whether this organization needs to be written onto a blockchain and whether it needs to be decentralized at all. Most would agree that in an idealized scientific community, published science should be accessible by anyone at any time. While many academics likely get the feeling of decentralization even today, the reality is that a few journals control the space and that a few critical data centers hold most of the publicly accessible data for a given scientific discipline. Past examples, like book burnings throughout history, have shown us that centralized powers can manipulate access to information in a dangerous way. If scientific discoveries and scientific data were published to a blockchain instead of a centralized server, there would be no single entity in control of the information and scientists could rest assured that a digital 'book burning' would not be possible.

A secondary concern is adoption of the organization. The academic *DAO* is essentially a network of scientists. Like many of these types of social networks, the *DAO* will become exponentially more useful with the more users it has (Metcalfe, 2013). One way to encourage adoption will be to tie the *DAO* into systems that already exist. For example, ORCID (Akers et al., 2016) could move their digital identification system onto a public ledger. Then, ResearchGate (Thelwall and Kousha, 2017) could use the social network that they have initiated to move their users' work onto a public ledger without the users themselves explicitly making the move or even knowing that it was happening.

Significant work has already been done toward this effort. As mentioned above, proposals have already been laid out for smart-contract-based peer review (Mackey et al., 2019; Spearpoint, 2017; Tenorio-Fornés et al., 2019). Going further, organizations like Bloxberg (Kleinfercher et al., 2020) and ResearchCoin (ResearchHub, 2020) are working on implementing similar incentivization mechanisms to that described above. In coming years, we will see the work being done to secure financial infrastructure in the current blockchains being used to secure intellectual property and academic infrastructure in the academic *DAO* or something of the sort.

5. Conclusion

In this article, I outline the structure of a *decentralized autonomous organization* which takes advantage of recent technological developments in the blockchain space to address issues with openness in the scientific process. The record of published science should be a more continuous one, and it should be public. Storing data and intellectual property to a blockchain not only allows anyone to quickly benchmark an idea as their own, it also makes the process of discovery more transparent and hopefully more reproducible.

Appendix A: Some Blockchain Terminology

DAO (Decentralized Autonomous Organization): An ownerless organization controlled by a transparent computer program and governed by the users of that program. A DAO is a group with no leader, a business with no CEO, the controlling program is designed to change only if the members of the DAO vote to change it. Most present-day examples of DAOs are still in development and have not yet achieved complete decentralization.

Public ledger: A ledger is a history of transactions, normally of a financial nature, used to keep a record of accounting. At the smallest scale, a ledger could be a written list among friends to record who bought dinner on a given evening. A public ledger is one that is recorded by all parties, or at least enough of the given parties that it can be trusted by all.

Private Key: An encrypted hash that gives the key-owner control over a wallet and its contents. In the case of this paper, private keys are necessary for ownership and approval signatures (see Figures 2 and 3). The owner of a private key would technically be the individual attached to that digital identity and would be the owner of any associated scientific content.

Digital Identity: An individual identifier that connects a person to an account in the blockchain space. This would signify ownership of tokenized items and be used to interact with smart contracts and DAOs.

Hash: A unique, fixed-length set of characters used to identify a digital object, much like a digital object identifier (doi) but encrypted.

Smart Contract: A self-executing digital agreement that is designed to carry out a given program or protocol under predetermined circumstances. There are a variety of examples of smart contracts, but perhaps the most illustrative is that of insurance. Imagine a farmer who is concerned about low crop yield during a dry season. A smart contract could be designed that collects the farmer's surplus funds on a good year but pays the farmer to cover their expenses on a bad year. In the context of a scientific agreement, smart contracts would be to do with usage of scientific data. An agreement would automatically cite the usage of a data product as soon as it was taken from the repository where it is held.

Tokenization: The representation of some item or concept ownership as a digital object on the blockchain. Tokenized items can be 'fungible', meaning that there are many of the same; for example, US dollars can be represented on a blockchain for means of payment. Tokenized items can also be 'non-fungible' meaning that the token is unique and represents only one real-world item; for example, the deed for a house could have a digital representation on the blockchain and could change ownership in that space. Importantly, ideas such as art and science could be tokenized as well. In this context, the token would not be traded so much as stored on the distributed ledger as an account of who owns the idea. That idea could then be easily and automatically credited when it is used toward the creation of some new idea.

Proof of Authority: A blockchain-based consensus algorithm based on identity as a stake. Much like proof of work [REF] and proof of stake [REF], this algorithm is a means of blockchain security. Here though, instead of computational resources or value at stake, block transactions are validated by approved accounts only. Generally, these accounts would be attached to some member of the community who has a reputation to lose. They are incentivized to make legitimate transactions at the risk of tarnishing their reputation.

References

Akers, K.G., Sarkozy, A., Wu, W., Slyman, A., 2016. ORCID Author Identifiers: A Primer for Librarians. Med. Ref. Serv. Q. 35, 135–144. https://doi.org/10.1080/02763869.2016.1152139

Benisi, N.Z., Aminian, M., Javadi, B., 2020. Blockchain-based decentralized storage networks: A survey. J. Netw. Comput. Appl. 162, 102656. https://doi.org/10.1016/j.jnca.2020.102656

De Angelis, S., Aniello, L., Baldoni, R., Lombardi, F., Margheri, A., Sassone, V., 2018. PBFT vs proof-of-authority: Applying the CAP theorem to permissioned blockchain. CEUR Workshop Proc. 2058, 1–11.

Halpin, H., Piekarska, M., 2017. Introduction to security and privacy on the blockchain. Proc. - 2nd IEEE Eur. Symp. Secur. Priv. Work. EuroS PW 2017 1–3. https://doi.org/10.1109/EuroSPW.2017.43

Kleinfercher, F., Vengadasalam, S., Lawton, J., 2020. bloxberg – The Trusted Research Infrastructure. Whitepaper 1.1 1–26.

Mackey, T.K., Shah, N., Miyachi, K., Short, J., Clauson, K., 2019. A Framework Proposal for Blockchain-Based Scientific Publishing Using Shared Governance. Front. Blockchain 2, 1–12. https://doi.org/10.3389/fbloc.2019.00019

Mandich, Giulio. "Venetian Patents (1450-1550)." *J. Pat. Off. Soc'y* 30 (1948): 166.

Metcalfe, B. "Metcalfe's Law after 40 Years of Ethernet," in *Computer*, vol. 46, no. 12, pp. 26-31, Dec. 2013, doi: 10.1109/MC.2013.374.

Moran, T., Orlov, I., 2019. Simple Proofs of Space-Time and Rational Proofs of Storage. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 11692 LNCS, 381–409. https://doi.org/10.1007/978-3-030-26948-7_14

Moxham, N., Fyfe, A., 2017. THE ROYAL SOCIETY AND THE PREHISTORY OF PEER REVIEW, 1665-1965* 1–35.

Pasquini, R.A., 2020. A Note on Quadratic Funding under Constrained Matching Funds 1–6.

Pearce, J., 2019. Blockchain in Academia : A Literature Review 1–6.

ResearchHub, 2020. ResearchCoin : A Cryptographic Incentive Designed to Accelerate the Pace of Scientific Research 1–5.

Sharples, M., Domingue, J., 2016. The Blockchain and Kudos: A Distributed System for Educational Record, Reputation and Reward. DM Rev. 490–4960. https://doi.org/10.1007/978-3-319-45153-4

Spearpoint, M., 2017. A proposed currency system for academic peer review payments using the blockchain technology. Publications 5, 1–6. https://doi.org/10.3390/publications5030019

Swan, A., Brown, S., 2004. Authors and open access publishing. Learn. Publ. 17, 219–224. https://doi.org/10.1087/095315104323159649

Tenorio-Fornés, A., Jacynycz, V., Llop-Vila, D., Sánchez-Ruiz, A., Hassan, S., 2019. Towards a Decentralized Process for Scientific Publication and Peer Review using Blockchain and IPFS. Proc. 52nd Hawaii Int. Conf. Syst. Sci. 6, 4635–4644. https://doi.org/10.24251/hicss.2019.560

Thelwall, M., Kousha, K., 2017. ResearchGate Articles: Age, Discipline, Audience Size, and Impact. J. Assoc. Inf. Sci. Technol. 68, 468–479. https://doi.org/10.1002/asi

Thelwall, M., Kousha, K., 2016. Figshare: A universal repository for academic resource sharing? Online Inf. Rev. 40, 333–346. https://doi.org/10.1108/OIR-06-2015-0190
Thung, F., Bissyandé, T.F., Lo, D., Jiang, L., 2013. Network structure of social coding in GitHub. Proc. Eur. Conf. Softw. Maint. Reengineering, CSMR 323–326. https://doi.org/10.1109/CSMR.2013.41