University of Southampton

Web and Internet Science

# Towards Decentralised Open Science with Blockchains

by

**M. R. Hoffman**

Thesis for the degree of Doctor of Philosophy

June 2023

UNIVERSITY OF SOUTHAMPTON

# Abstract

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE
WEB AND INTERNET SCIENCE GROUP

Doctor of Philosophy

**Towards Decentralised Open Science with Blockchains**

M. R. Hoffman

This work proposes and evaluates a sociotechnical system decentralising scholarly communication (SC), i.e., how researchers *(1)* disseminate and *(2)* evaluate their findings. The current set-up of SC relies on *(1)* academic publishing (AP) for the dissemination and *(2)* publisher-driven peer review (PR) and scientometrics for the evaluation of research. Currently, SC relies on resources that are computed and stored centrally in private, corporate databases. This thesis not only outlines the tensions between the oligopolistic nature of corporatised SC vs. the collaborative public-domain mindset of science, but it also documents the '*wicked problem*' of large market players failing to curate public domain work, often privately appropriating it. Also described are shortcomings of PR and scientometrics in their current, ossified, and non-transparent formulations. Against that backdrop, using engineering and social approaches, the thesis presents an all-encompassing study into what is expected of modern SC and how to envision its future. One novel finding is that given the problems of corporate centralisation pertaining to SC, the tenets of Open Access, Open Science and Open-Source Software apply synergistically, and when implemented using the features of blockchain technology, this combination provides a valid platform to decentralise & *'de-corporatise'* SC, to ensure its fit within the public domain. It is documented how, as part of the emerging *Web3* paradigm, blockchain-based sociotechnical systems facilitate interactions among parties, reducing the number of intermediaries required and providing new efficient modes of coordination. The thesis addresses the tension between the singular ideal of truth in research juxtaposed against the pluralistic nature of research production. On the one hand, it is the use of blockchain *smart contracts,* whose execution is validated globally by the network, that helps with providing a *single version of the truth* to all scholars, thus organising research information and value exchange with fairness, transparency, and efficiency in mind. On the other hand, blueprints for SC *decentralised autonomous organisations* acting as federated support networks are proposed, including the governance smart contracts containing the logic for disintermediated AP (*Smart Papers*), decoupled PR (*DAO4PR*) and transparent scientometrics computations. Accordingly, novel blockchain SC software is created and evaluated computationally and sociologically. The results suggest SC needs to be modernised to stay fit for purpose and that blockchain solutions are useful towards that end, especially the proposed solution that incorporates open-source smart contracts running atop a Proof-of-Stake blockchain, Ethereum.

# Table of Contents

# Declaration of Authorship

*I, M.R. Hoffman, declare that this thesis and the work presented in it are my own and have been generated by me as the result of my own original research. I confirm that:*

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

(Hoffman, Ibanez, & Simperl, 2020),
(Hoffman, Ibanez, & Simperl, 2019),
(Hoffman M. R., Ibanez, Fryer, & Simperl, 2018),
(Ibanez, Hoffman, & Choudhry, 2019),
(Hoffman M. R., 2018).

Signed:

Date: 21 June 2023

# Acknowledgements

*SOCIAM (The Theory and Practice of Social Machines) project is acknowledged for funding the initial part of this research.*[1]

> *"Do the things that interest you and do them with all your heart… If you fail the first time, then you'll just have to try harder the second time. After all, there's no real reason why you should fail."* - ***Eleanor Roosevelt***

The supervisors of this thesis have, somehow organically, fluctuated in their percentage responsibility, so the order in which I list them may not necessarily be a key indicator of how immensely important they have been. My gratitude goes to Dr Luis-Daniel Ibañez, Prof. Elena Simperl and Prof. Les Carr for their invaluable guidance and advice.

The interim progress examiners, Assoc. Prof. Kieron O'Hara and Dr George Konstantinidis have provided helpful feedback to the later direction of this research. I must also mention Dr Manuel Leon-Urrutia and Prof. Age Chapman for providing ample opportunities to teach exciting and relevant subjects to students. Furthermore, I am grateful to the staff of my Department and the Postgraduate School here at the University of Southampton (Megan, Lynn, Rachel), as you have always been helpful and kind. Finally, thank you to Prof. Thanassis Tiropanis and Prof. John Domingue for their valuable comments that shaped key amendments.

Outside of Southampton, I want to acknowledge Annette Baldenegro for hiring me into the Big Data industry years ago, which really opened my eyes to what the new era of computing looked like; Dr Virginia Eatough of Birkbeck College for teaching me qualitative research skills, and the University of Winchester team that hired me as Lecturer in Computer Science and Cyber Security, in particular, thanks go to Dr Ben Sanders and Dr Adriana Wilde (who is at Southampton again).

I also want to thank my parents, Urszula and Mariusz, for their emotional support throughout this formidable endeavour. This paragraph cannot be finished without acknowledging Ed, my Auntie, Grandma and my brother. I do want to dedicate this thesis to my nephew and my niece to inspire them in the future.

Even though not all the names can be listed here, I am so thankful for having in my life Jacqui, Joanna, Kasia, Sara, Ania *(1&2)*, Hania, Ola, Piotr, Maciej, Chrissy, Hannah, Tomek *(1&2),* Marianna, Clive, Jakub, Ben, Billy, Chris-Jae, my friends and my tutor from my professional diploma course in counselling, and my very close friends Cameron, Euan and last but not least, my love, Adam –

**for being my support and my inspiration.**

[1] www.sociam.org *{Accessed August 2022}*

# Abbreviations

**ACL** ................................. Access Control List (*information security*)

**AOS** ................................. Abe-Ohkubo-Suzuki Ring Signatures (*cryptography*)

**AP** ................................. Academic Publishing

**API** ................................. Application Programming Interface (*IT*)

**CLI** ................................. Command Line Interface *(IT)*

**DAO**................................. Decentralised Autonomous Organisation *(crypto)*

**DApp** ............................... Distributed Application (*crypto*)

**DAO4PR**........................ DAO for Peer Review

**DeFi** ................................. Decentralised Finance (*ecosystem*)

**DeSci**................................ Decentralised Science *(emerging ecosystem)*

**DI** ................................. Decentralised Identity

**DID** ................................. Decentralised Identifier

**DLT** ................................. Distributed Ledger Technologies

**DOI** ................................. Digital Object Identifier

**DON**................................. Decentralised Oracle Network

**EES** ................................. Elsevier Editorial System

**EIP** ................................. Ethereum Improvement Proposal (*standard*)

**ERC** ................................. Ethereum Request for Comments (*standard*)

**ETH** ................................. Ether (cryptocurrency)

**EVM** ................................ Ethereum Virtual Machine

**FSN** ................................. Federated Support Network (see: *Zuboff & Maxmin, 2004*)

**GDPR**.............................. General Data Protection Regulation

**HCI** ................................. Human-computer Interaction

**I4OC**................................. Initiative for Open Citations

**IPFS** ................................. Interplanetary File System (*protocol and network*)

**LSAG** .............................. Linkable Spontaneous Anonymous Group (*cryptography*)

**NFT** ................................. Non-Fungible Token (*crypto*)

**OA** ................................ Open Access (*publishing model*)

**OJS** ................................ Open Journal Systems

**OOP** ................................ Object-Oriented Programming (*computer science*)

**ORCID** ............................ Open Researcher and Contributor Identifier

**OS** ................................ Open Science (*movement*)

**OSF** ................................ Open Science Foundation

**OSS** ................................ Open-Source Software

**P2P** ................................ Peer-to-Peer (application architecture)

**PGP** ................................ Pretty Good Privacy (*encryption scheme*)

**PKI** ................................ Public Key Infrastructure

**PoS** ................................ Proof-of-Stake (*algorithm*)

**PoW** ................................ Proof-of-Work (*algorithm*)

**PPPR** .............................. Post-Publication Peer Review

**PR** ................................ Peer Review

**SAG** ................................ Spontaneous Anonymous Group (*cryptography*)

**SC** ................................ Scholarly Communication

**SCs** ................................ Smart Contracts

**SLR** ................................ Systematic Literature Review

**SMs** ................................ Scientometrics

**SoC** ................................ Separation of Concerns (*design principle*)

**SSoT** ................................ Single Source of Truth *(information science)*

**SVoD** .............................. Single Version of Data *(enterprise data management)*

**SVoT** .............................. Single Version of the Truth

**SP** ................................ Smart Paper (*living document*)

**TA** ................................ Thematic Analysis (*research*)

**TCR** ................................ Token-Curated Registries

**Tx** ................................ Transaction (*on a blockchain*)

**UI** ................................ User Interface

**UML** ................................ Unified Modelling Language (*software design*)

**URS** ................................ Unique Ring Signature (*cryptography*)

**UX** ................................ User Experience

**W3C** ................................ World Wide Web Consortium

**ZKP** ................................ Zero-Knowledge Proof (*cryptography*)

# Definitions of Key Terms

**Bibliometrics**........................ statistical evaluation of recorded information (publications), focusing on metrics such as analysing the number of citations of an article or a publication

**Blockchain**........................... a type of distributed ledger (made up of time-stamped blocks containing hashed transactions) that is usually public and requires consensus across all nodes globally

**Crypto** ....................... the culture and community associated with **C*ryptocurrency***

**Cryptocurrency**.................. digital currency exchanged online, secured by cryptographic keys, and normally built on its own specially designed blockchain

**Cryptographic key**............ a string of characters used to encode or decode data, see *Public key* and *Private key*

**Cryptographic token**........ digital asset exchanged online, secured by cryptographic keys, and normally designed for specific use on top of existing blockchains

**DAO** ........................... distributed autonomous organisation, a non-hierarchical organisation represented by rules encoded in smart contracts, self-organising around projects and outcomes

**DeFi Ecosystem**................ blockchain-based financial system containing decentralised exchanges, marketplaces, prediction markets, credit and lending, insurance, as well as asset management and trade automation

**Distributed capitalism**..... according to Zuboff, this is a new iteration of capitalism recognising the shift from products and services to tools and relationships enabled by interactive technologies, rescuing assets such as knowledge, skills and people from corporate control

**Distributed economy**........ according to Johansson, this is a regional approach to promote innovation through fostering small and medium-size enterprises

**Decentralised Identifier**... (DID), a globally unique and cryptographically verifiable identifier that uses distributed infrastructure for its protocols

**Distributed Ledger** ........... a database that can follow block or graph structure for transactions, maintained at multiple nodes instead of central location

**Ether** ........................... the main cryptocurrency operating on the Ethereum blockchain

**Ethereum** .......................... a blockchain designed by Vitalik Buterin, capable of running smart contracts

**Federated Support Network-**according to Zuboff and Maxmin, an FSN links enterprises following the metaprinciples of distributed capitalism, where all value is traceable to the individual realising that value

**Fork** see ***Hard Fork***

**h-Index** .......................... scientometric proposed by Hirsch that measures "relevant productivity" (either of a particular author or sometimes of a particular journal), where $h$ is the threshold of minimum citations accounted for; often said to be correlated with the "impact" of the entity

**Hard fork** .......................... a way to upgrade a blockchain, where new endorsed records are created using a new blockchain protocol, following new rules

**Hash** output of a ***Hash (hashing) function***

**Hash (hashing) function** . one-way function mapping data of arbitrary size into fixed-size values

**Impact factor** ..................... scientometric devised by Garfield that measures the ratio of new citations to recent publications; often said to be correlated with the "quality" of the published research

**Informetrics**........................ the study of quantitative aspects of information, including ***Bibliometrics*** and ***Scientometrics***

**Interoperability**................. ability for multiple applications or systems to exchange data

**Oligopoly** .......................... a market where pricing is controlled by a small number of producers

**Oligopsony**.......................... a market with multiple producers yet only a small pool of consumers

**Oracle** .......................... a service that updates a blockchain/distributed ledger using data from the outside world

**Proof-of-Work** ................... distributed consensus algorithm deterring manipulation of data by introducing burden of mathematical proof that a certain amount of computational effort ("work") has been expended; used for confirming transactions, adding valid blocks to the blockchain and rewarding participation; expends much more energy than ***Proof-of-Stake***

**Proof-of-Stake**.................... distributed consensus algorithm deterring manipulation of data by allowing participants to stake their interest in the network; used for confirming transactions, adding valid blocks to the blockchain and rewarding participation; expends much less energy than ***Proof-of-Work***

**Private key** .......................... a cryptographic key stored on the intended receiver's device and used to decrypt data

**Public key** .......................... an openly accessible cryptographic key, associated with the intended receiver, and used to encrypt data

**Public key certificate** ....... a digital document used to prove the validity of a public key

**Public key infrastructure** a set of authorities and policies that manage public key certificates

**Ring signature** .................... a type of digital signature that can be performed by any one member of a particular group using the group's public keys and the signer's private key in a way that cannot trace the signer but confirms that he or she is a member of that particular group, e.g. ***AOS signatures***

**Scientometrics** .................... a field of ***Informetrics*** focusing on scientific metrics such as impact factors and h-Indices (compare with bibliometrics)

**Smart contract** .................. computer programs stored on a blockchain that are validated by multiple parties globally, and run automatically when certain conditions are met, without the need for a third-party to initiate them; they can move blockchain funds between accounts

**Social machine** ................... a Web-based environment comprising human and technological constituents interacting to produce and curate outputs

**Socio-technical system** ..... a goal-directed organisation optimising the interaction of social and technical elements in work design and task execution

**Solidity** .......................... a programming language designed for Ethereum smart contracts

**Token** ......................... see ***Cryptographic token***

**Trustlessness** ...................... a trustless system means that its participants do not need to know or trust each other or a third party for the system to function

**Zero-Knowledge Proof** .... a method by which the prover can prove to the verifier that they know that a given statement is true whilst not revealing any additional information or indeed withholding the statement in full; e.g. *"proving that you know the value of x, without revealing x"*

# Chapter 1

# Introduction

## 1.1 The Tensions of Scholarly Communication

**Scholarly communication (SC)** is the set-up through which research findings are published, evaluated for quality, disseminated to the public, and preserved for future access. Also described as a family of cultural processes enabling scholars to communicate their findings (Kling & McKim, 1999), this *loosely-coupled system* (Schwartz, 1994) includes formal means, such as publications in peer-reviewed journals, and increasingly, informal channels, such as micro-blogging via social media (ACRL, 2003) even though most peer-reviewed papers are still disseminated traditionally, through publishing houses.

This thesis explores the motivations and the possibilities involved in reducing the dependence of SC on corporate publishers, outlining how the recent advancements of the Web can empower scholarly researchers and ordinary citizens alike to decentralise the operations of science. There are opportunities to utilise the advances of Web technologies, focusing on blockchain technology, to improve the processes through which science reaches its audiences, and the ways in which scholars are recognised. However, these opportunities do not come without their challenges. Since there are no agreed-upon or well-established open data/process models for scholarly communication practices (Sompel, Payette, Erickson, Lagoze, & Warner, 2004), any efforts to 'upgrade' or 'project manage' SC appear tricky. Moreover, it is difficult to even imagine a fully public system that is not run for profit, especially because today's way of doing SC borrows heavily from the internal processes of Brobdingnagian commercial publishing houses which are relied upon by most scholars. These mega-publishers do indeed make a lot of money from our research, for example Elsevier's profit margin tends to be larger than Google's year-on-year. [2]

However, SC is still evolving. For example, in recent decades, electronically mediated scholarly communication has been witnessing quantitative and qualitative developments in how researchers collaborate utilising various communication styles and methods over the Web (Borgman & Furner, 2002). But despite its changing

[2] https://libraries.mit.edu/scholarly/publishing/elsevier-fact-sheet/ *{accessed April 2022}*

nature, scholarly communication system of the new millennium is still being described as *badly flawed* (Odlyzko, 2002), in that it fails at providing the qualities that scholars and the general public desire, such as openness, trustworthiness and a fair approach to disseminating and accessing new scientific information. Because any improvements to SC are generally slow due to the system's inertia (Hill, 2016), bringing it up to speed with what modern technology has to offer would require an active, sustained and globally coordinated intervention.

In light of the above, the purpose of this PhD is not to revolutionise the world of SC, but rather to focus on presenting one innovative approach to improving scholarly communication, in line with Open Access (Suber, 2007) and Open Science (Fecher & Friesike, 2014) principles, through the use of blockchain-based *smart contract* technology (Cong & He, 2019). My goal in this work is to offer a well-defined set of alternatives that could help science get unstuck from the sub-optimal ways in which publishing, reviewing and evaluating scientific findings is carried out at present. A compelling collection of reasons[3] necessitates such a reform of the scholarly communication processes and structures that are currently in use around the globe.

## 1.2 Public Science?

Philosophers of science have raised the question of values that people hold around science, for instance, whether it should be public or private (Bridgman, 1940). Even though science may oftentimes happen in what appears as private settings, I argue that it is an inherently public endeavour.

Saliently, one of the fundamental assumptions legitimising scholarly research is that most of it gets communicated to the world (ACRL, 2003) to build common knowledge and facilitate human inquiry. A substantial portion of scientific research is, thus, assumed to be publicly funded and should be held to benefit the public[4]. For example, in 2017/18 (when I started working on my thesis), the majority of UK research was publicly funded, as outlined in **Figure 1** - 2017/2018 UK research funding sources - data from Jisc.ac.uk and parliament.uk.

[3] these reasons are exhaustively listed in the subsection 1.5 – Motivations.

[4] I review these claims as part Chapter 2 – Background and Related Work.

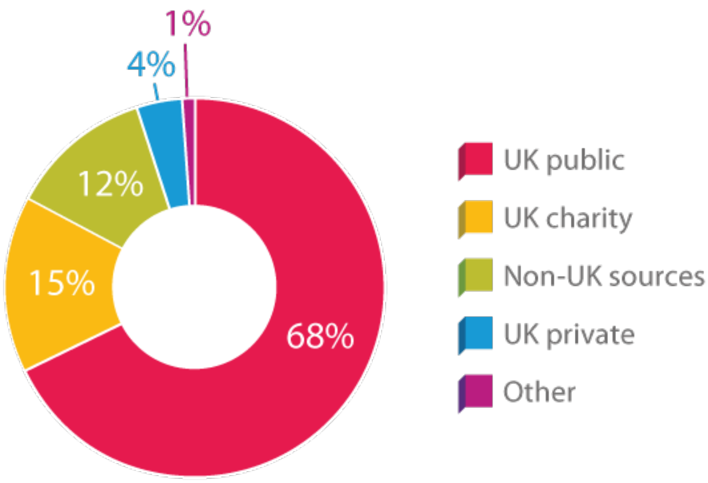



**Figure 1** - 2017/2018 UK research funding sources - data from Jisc.ac.uk and parliament.uk

The vast majority of scholars working for publicly funded universities develop and disseminate their research *pro bono publico*, without direct financial rewards, unless they commercialise their research outside of traditional scholarly structures - and even then, it is typically just the final deliverable, or the implementation, that is commercialised. The profit motive may sometimes clash with the public nature of science. Some researchers are said to be motivated by developing their reputation in a particular domain[5] while others may value the freedom to work on problems that they find interesting (Deci & Ryan, 2013) but this generates tension against the for-profit, commercial nature of private publishing companies that financially benefit from restricting access to scholarly research. It has been suggested that should better incentives be created for scholars to publish in non-for-profit outlets (outside of the incentive structure that focuses on impact factors and university rankings that tend to favour *megapublishers*), the research community could 'regain control' of SC (Larivière, Haustein, & Mongeon, 2015 b).

Moreover, there is growing evidence that the general public have less than satisfactory access to findings that are seen as important, and researchers claim that this lack of access is a hindrance to scientific progress (Maxwell, 2012). There is also growing dissatisfaction with how the established scholarly communication practices do not fully address the needs of the public (Sompel, Payette, Erickson, Lagoze, & Warner, 2004) and those of the scholars (Lagoze, Edwards, Sandvig, & Plantin, 2015).

[5] this claim is also evaluated in Chapter 2.

One of the big issues raised is the situation, whereby private enterprises sell papers (journal articles, conference proceedings, monographs; (Bjork, Roos, & Lauri, 2009)) in return for hugely inflated fees, constituting an oligopolistic market which acts like a black hole sucking up public money (Larivière, Haustein, & Mongeon, 2015 a). Academic publishing, positioned at the very core of scholarly communication, thus creates a very central tension. It is useful to understand the origins of this tension, which is why it is expounded in the next Chapter of the thesis (*Background and Related Work*). When envisioning change to the status quo, it is also useful to understand why blockchains have been suggested as a possible solution to the coordination problems posed by SC, which is what the next sub-section illuminates.

## 1.3 The Significance of Web3, Blockchain & Smart Contracts

Challenging the social, technological, and economic status quo, a new family of Internet applications has emerged that allows its users to transact (exchange value and information) in a transparent and decentralised manner. **Blockchains** are based on the idea of distributed consensus, and provide immutable, permanent records of data (Wattenhofer, 2016). They have been successfully used to implement cryptocurrencies such as Bitcoin (BTC) and Ether (ETH), allowing users to exchange value without relying on traditional central bank-issued fiat currencies (Nakamoto, 2019). Blockchains come with their own set of incentives and rules (Kiayias, Koutsoupias, Kyropoulou, & Tselekounis, 2016) that aim to guarantee the chain's propagation and survival. The singular term '**blockchain'** is usually used to describe an idealised version of blockchain technology that transcends any particular implementation, and now became common parlance. Nevertheless, this thesis, rather than referring to abstract blockchain, focuses on individual blockchain implementations, or blockchains, predominately Ethereum.

Newer generation blockchains like Ethereum allow their users to run so-called **smart contracts**, an idea first introduced by Nick Szabo in the 1990s (Szabo, 1997) but only recently successfully implemented. A smart contract is a set of promises, specified in software form, including protocols within which the parties perform on these promises, and stored on the blockchain. The blockchain's notarisation and consensus protocols will guarantee the execution of the contract, as well as the authenticity of the contract's results, even in an environment where the transacting parties do not fully trust each other. Smart contracts have been shown to be useful in developing distributed applications (dApps), and in building so-called decentralised autonomous organisations (DAOs) (Raval, 2016), where socio-economic rules are encoded as computer code, thus reshaping the ways in which people organise their work and transfer value and information (Luu, Chu, Olickel, Saxena, & Hobor, 2016). Therefore, they are seen as belonging to the next generation of the World Wide Web, the Web3[6], which decentralises transactional and communications architectures on open protocols and platforms, built on P2P (peer-to-peer) nodes, and running on decentralised storage systems. More information about blockchains and smart contracts is provided in the Background and Related Work chapter, which follows in this thesis.

The relevance of blockchains to the problem at hand, i.e., decentralising scholarly communications to make science more open, is that smart contracts can be used as blueprints that define scholarly workflows, disseminate papers, coordinate peer review and compute scientometrics, running on top of Web3 infrastructure that does not require a central third-party to coordinate. Given the novel nature of this sort of endeavour, the next subsection presents the motivations for this thesis, leading to problem formulation and research questions associated with such an undertaking.

---

[6] Speaking of the direction towards which the Web is heading, there is slight confusion regarding the terms Web 3.0 which is older, and has been connected to semantic technologies, and Web3 as advocated by the Web3 Foundation, created by Gavin Wood, the co-founder of Ethereum and the Solidity language. Web3 is the vision of the serverless internet or the decentralised web. It echoes many of the goals of Web 3.0, one understanding of which is that it is created by 'networked digital technologies that support human cooperation' (Fuchs, 2007). While Web 3.0 technologies like Wikipedia are not only communicative, but also cooperative, forming the *cooperation Internet* enabling participation and self-organisation, they do not use blockchain. Web3, however, is increasingly adopting blockchain for many of its co-operative use cases.

## 1.4 Motivations for this Thesis

### 1.4.1 Fostering Research Trustworthiness

MAPS TO RQ1 (P.11) AND CHAPTERS 5 & 6

Even though the narrative of "science in crisis" remains heavily debated (Fanelli, 2018), the majority of academics believe in there being a significant **reproducibility crisis** (Baker, 2016). To combat the problems associated with failed study result reproductions, multiple solutions have been proposed, and a combined approach may be necessary. Firstly, a single-version-of-truth ($SVoT$) approach, in which studies are pre-registered in a durable, tamper-proof form, can mean that each piece of research can be traced back to its original hypotheses. Secondly, the relevant data and other connected artifacts need to remain robustly linked and transparent, in a way that illuminates the provenance of each paper. This underlines the importance of designing or re-using appropriate data structures, schemas and formats, to foster the trustworthiness of researchers and their outputs. Moreover, the issue of distinguishing valuable data from noise is important here, as fighting spam and disinformation becomes more and more relevant with each passing year. Thus, can we design a system that boosts public trust in research? The relevant research question is developed as **RQ1** (see the next section, **Problem Formulation and Research Questions**) The design of the proposed system is presented in Chapter 5 and evaluated in Chapter 6.

### 1.4.2 Fostering Research Openness

MAPS TO ***RQ3*** (*P.12-13*) *AND* ***CHAPTER 5 SECTION 5***

Open Access, Open Science and the use of Open-Source Software in science are often cited as the converging openness initiatives that promote the vision of intellectual properties as public goods, representing academe's commitment to the larger public sphere (Willinsky, 2005). For example, Open Systems Pharmacology Community[7] promotes the mix of OA, OSS and OS to build models and simulations in Pharmaceutical Sciences to foster reproducibility and collaboration. There is a need to back Open Science initiatives to improve the circulation of knowledge, by increasing faculty awareness and interest, and providing new tools and frameworks. The main contribution answering this need is that I provide blueprints for the use of relevant Open Source technologies in Chapter 5 Section 5 "Implementation". Secondary contribution related to fostering research openness is my analysis of the Open Science movement, and relevant open and decentralised technologies in the Background (Chapter 2) of the thesis.

### 1.4.3 Decoupling Peer Review from Corporate Interests

MAPS TO ***RQ1*** (*P.11*) AND ***RQ2*** (*P.12*) *AND* ***CHAPTER 5 SECTION 7***

A compelling set of moral, ethical and economical reasons necessitates the reform of the publisher-driven peer review processes that are currently in use around the globe. New approaches leveraging applied cryptography offer help in envisioning dependable and adaptable peer review platforms that incorporate flexible features such as post-

[7] https://www.open-systems-pharmacology.org [Accessed August 2022]

publication peer review, different combinations of blinding and a tamper-proof approach to storing reviews and annotations. The reader will be able to familiarise themselves with the ailments of peer review in the next Chapter (*Background and Related Work*). Can we design a system that does peer review better? The main contribution answering this question is Chapter 5 Section 7 “Decentralising Peer Review with Smart Papers”.

#### 1.4.4 The Need for Reviewer Recognition

MAPS TO ***RQ3 (P.12-13)*** *AND* ***CHAPTERS 6.5*** *AND* ***7***

At the time of writing this, findings have been published from a *Nature* experiment that indicates academics want more recognition for reviewers’ valuable contributions and to introduce more transparency into the process (Editorial, 2019). So far, 91% of *Nature* authors and 55% of reviewers have opted into this trial. Not only are reviewers not getting paid, which has got its pros and cons, but there is also no standard, transparent way of tracking reviewers’ hours, leading to the invisibility of their efforts. There is no compensation involved except the kudos of the editor (who in gratitude will simply send you more papers to review in the future) and "*a sense of obligation that you've done your duty and service to the larger scientific enterprise*", as Paul M. Sutter, astrophysicist and science educator put it. Having all that in mind, could we design a system that, at the minimum, recognises and, at best, rewards the reviewers adequately and fairly for their underappreciated efforts? This maps to the contribution of the survey in Chapter 6.5 (Findings) outlining how much importance scholars attribute to such a feature, which informed the proposed future research directions in Chapter 7.

### 1.4.5 The Struggle for Optimal Incentives for Open Science

MAPS TO ***RQ3 (P.12-13)*** *AND* ***CHAPTER 2.10*** *AND* ***CHAPTER 7***

Higginson and Munafo (2016) wrote that current incentive structures of SC (such as incentivising novelty and the sheer number of published items) stand in direct conflict with maximising the '*scientific value of research'*. Optimising the incentive structures for scholarly authors may involve recognising their effort in conducting exploratory studies and publishing negative results, as well as taking into consideration a wider chunk of scholarly outputs regardless of how novel or exciting they are. As Frankenhuis and Nettle (2018) argue, Open Science re-incentivises researchers to embrace doubt and uncertainty, something that the current SC system de-incentivises. The reader will be introduced to the ideals of Open Science in the next Chapter (*Background and Related Work)*. Could we envision a system informed by Open Science principles that operates with better incentives for scholars? The contribution of the thesis is reviewing the existing related work on the blockchain that deals with incentives, such as ScienceMiles (Chapter 2.10). Due to time and scope limitation no original work could be carried out on incentives but it informed the proposed future research directions in Chapter 7.

### 1.4.6 Scientometrics as a Computational Problem

MAPS TO ***RQ2*** *(P.12) AND* ***CHAPTER 5.3.1***

Scientometrics are often based on bibliometrics such as citation counts, and it has been accepted that generating citation-based scientific metrics is a computational problem (Krapivin, Marchese, & Casati, 2010). The algorithms and the data stores used to compute modern scientometrics come in various shapes and forms but are not openly accessible as they belong to private companies, as studied by Falagas et al. (2008). The authors suggest that there is a need for a "*systematic reconstitution of the pros and cons of each database and the development of a formula for free access*".

Can we develop such a formula? The addition of Bibliometrics to the Smart Paper model in Chapter 5.3.1 is a contribution towards that direction. Also see the relevant Section in the Background Chapter (2).

### 1.4.7 Avoiding the Private Appropriation of Public Work

MAPS TO ***RQ1, RQ2, RQ3*** *(P.11-13)* AND ***CHAPTERS 2*** AND ***4.***

The unit of scholarly communication is a scholarly paper. Scholars are not directly reimbursed per paper because this is neither common practice nor is it possible to know in advance what the price tag may be. In any case, the incentives of scholars are not usually profit-driven, but rather reputation-driven (Fecher, Friesike, Hebing, & Linek, 2017) and there are other human values at play such as curiosity, a desire to pursue knowledge and a desire to communicate with other scholars. On the other hand, papers are priced by academic publishers[8]. They can be bought individually or in bulk. The money paid per published paper does not in any way make its way to the author of the work. This can be seen as private appropriation of public work because publishers operate as private corporations whilst academics perform a **public** role, the idea that was first substantiated in the 18th century by the Kantian notion of *public reason*[9], then cultivated by the Victorian culture of *public science* throughout the next century (Turner, 1980), subsequently restated in the 1940s as the American presidential credo that scientific progress is essential to *public welfare* (Pielke, 2010). Informed by these ideas, can we design a system that does not unnecessarily privatise scientific knowledge creation? This informs the exploration of different narratives in Chapter 2 (Background) and the proposal of a decentralisation framework in Chapter 4.

---

[8] https://onlinelibrary.wiley.com/library-info/products/price-lists (Accessed April 2022)

[9] Kant is concerned with enlightenment as something that we must work towards together. He recommends "nothing is required but… freedom: namely, freedom to make *public use* of one's reason in all matters". This encompasses the scholar's public rights and responsibilities as a 'writer of reason'—in Kant's words, the application of reason "*as a scholar* before the entire *public* of the *world of readers*" (Williams, 2018 ).

## 1.5 Problem Formulation and Research Questions

We have now seen that there are multiple concerns about the fairness and accountability of modern scientific processes, however, there is also a possibility for evaluating new technology to create a fairer and more accountable system. Therefore, there is a need to design and evaluate such a system, in which a decentralised approach for disseminating scientific knowledge would be implemented. As part of this PhD project, a blockchain-based system is proposed. For its success, the following research questions must be answered:

### 1.5.1 Research Question 1

**RQ1.** Within the context of scholarly communications (SC), *(a)* what is the meaning of centralisation/decentralisation, *(2)* what are those components of SC that exhibit centralised tendencies and can be re-designed, re-imagined or re-formulated in a decentralised manner?

*Discussion*: The main burden of answering RQ1 was qualitative in nature and required exposing the weaknesses of the centralised approaches to SC. Through applying mixed methods including Systematic Literature Review, Stakeholder Analysis, Actor Network Theory and Activity Theory, I identified the importance of decentralising paper publishing and dissemination, peer review and research evaluation relying on scientometrics. Furthermore, I identified the importance of the following guiding themes: openness, fairness, transparency, career progression and impact evaluation. I show that publishing, research governance, peer review, research evaluation and even grant management can happen in a decentralised environment using blockchain without relying on many existing intermediaries. Answering this question also tells us more about the limits and drawbacks of centralisation and provides a lot of material for the Background section (Chapter 2).

### 1.5.2 Research Question 2

**RQ2.** What are the requirements for a decentralised academic scholarly communications platform - including (a) what features are relevant? (b) how can they be implemented? (c) how can they be evaluated?

*Discussion:* To answer RQ2, I applied the interdisciplinary methods described in subsections on Action Research, Software Development Life Cycle, Agile Development, Blockchain Software Design. I developed a blockchain-mediated decentralisation framework described in Chapter 4.

I used this framework to design blueprints for a decentralised paper publishing platform described in Chapter 5, as the main question was also narrowed down to the following sub-questions that are addressed in Chapter 5.

**RQ2.1.** *How to manage releases of scholarly papers and their attribution agreements in a trusted way?*

**RQ2.2.** *How to avoid malicious/accidental modifications in remote data stores affecting the computation of bibliometrics for scholarly work?*

**RQ2.3** W*hen decentralising peer review, what is the viability of using ring signatures for ensuring double-blindness of the process?*

The limits of decentralisation were also explored this way and form part of the Conclusions.

### 1.5.3 Research Question 3

**RQ3.** What are the attitudes of scholars towards decentralised Open Science software, and, towards decentralising science with blockchains?

*Discussion:* The question of perceptions and attitudes towards decentralised open science with blockchain is deserving of focus, mostly because there has been little systematic inquiry into Open Science software - a cursory Google Scholar search yields fewer than a couple of dozens results across literature. Firstly, this lack of momentum is surprising and motivated me to probe into scholarly perceptions of the state of the art of Open Science software and the attitudes towards possible innovation in this space. Secondly, I wanted to know the differences between the perceptions and attitudes of decentralising Open Science in general, versus decentralising it with blockchains, to see what the differentiating variables are.

Also, because in this thesis, I present a vision for open science that is no longer rigidly organised into Journals and Conferences, but works as a more loosely-coupled setup that dynamically organises knowledge based on search-and-sort keywords that are linked to scholarly disciplines and topics, I wanted to gauge not only the demand for a system like Smart Papers (developed in this project) but also illuminate the perceived problems associated with moving onto this new software system. Of further interest is building the understanding into whether the blockchain-specific features are useful to the wider academic community and perceived as clearly understandable and value-adding.

The qualitative evaluation of the features and their implementation is related to the subjective satisfaction of the users (scholars). The theoretical underpinnings of "usability" evaluation can be traced to activity theory, a framework that facilitates productive cooperation between social scientists and software designers (Kaptelinin & Nardi, 2006). Surveys were launched to evaluate how scholars perceive the Smart Papers features and how efficiently they judge it solves the problems affecting scholarly communication, and this was documented in Chapter 6.

To provide more context and understanding of how the above research questions were answered in my published peer-reviewed work, I present my summary of publications from this thesis in the next Section of this Introduction.

## 1.6 Summary of Publications from this Thesis

The present work has led to peer-reviewed publications as listed in

**Table 1** below. This data has been sourced from Google Scholar[10].

| TITLE | CITED BY | YEAR |
|---|---|---|
| Can blockchains and linked data advance taxation<br>MR Hoffman<br>Companion Proceedings of the The Web Conference 2018, 1179-1182 | 25 | 2018 |
| Toward a Formal Scholarly Understanding of Blockchain-Mediated Decentralization: A Systematic Review and a Framework<br>MR Hoffman, LD Ibáñez, E Simperl<br>Front. Blockchain | 16 | 2020 |
| Smart papers: Dynamic publications on the blockchain<br>MR Hoffman, LD Ibáñez, H Fryer, E Simperl<br>European Semantic Web Conference, 304-318 | 9 | 2018 |
| Scholarly publishing on the blockchain - from smart papers to smart informetrics<br>MR Hoffman, LD Ibáñez, E Simperl<br>Data Science 2 (1-2), 291-310 | 8 | 2019 |
| Blockchains and Digital Assets<br>LD Ibáñez, MR Hoffman, T Choudhry<br>https://www.eublockchainforum.eu/sites/default/files/research-paper ... | 5 | 2019 |

**Table 1** - Peer-reviewed publications resulting from this research.

[10] https://scholar.google.com/citations?hl=en&user=GtKm0gUAAAAJ *{Accessed April 2023}*

## 1.7 Author's Background

I received an MEng in Computing from Imperial College London and an MSc in Psychology from Birkbeck University of London. I have worked across many industries from investment banking to online music streaming and have always been fascinated with the crafty applications of computer software to solve diverse problems of human beings. I taught Big Data engineering for a couple of years, working for a Hadoop-driven Silicon Valley start-up that later became a big corporation. When travelling to customer sites across the world, I realised that people begin to struggle with Big Data when it becomes too chaotic and the single version of the truth (SVoT) is lost. In the world of limitless data, ordering that data becomes important so that value can be distinguished from noise. I was aware of many techniques that aimed to bring order and meaning to data at scale, from old-fashioned database schemas to newer semantic Web technologies, but I was also wary that the idealistic, community-oriented rather than financial/market-based nature of the incentives[11] driving the adoption of these techniques meant that their full value is challenging to realise.

At around the same time, blockchain, together with cryptographic tokens, became a buzzword, and I became intrigued by whether or not it can actually help with all the use-cases that it was being hyped for. It was clear to me that blockchains provide SVoT for their respective ecosystems, but they are decentralised, so there is no single source of data (SSoD) and hence no single source of truth (SSoT), instead relying on coordination mechanisms to piece "the truth" together whilst providing economic incentives to facilitate that. Having been accepted into my candidature in Web Science and having started talking to academics who did *real* scholarly research, I realised that the application of SVoT principles to scholarly communications is urgently needed, and that these approaches have been successfully studied via ontologies and linked data. Yet suddenly, blockchains emerged, offering hope that they could boost those existing approaches to de-cluttering the world of Big Data, and especially big scholarly data. Seeing blockchains as a kind of incentivised hyper-trust databases, I started probing into them further. Meanwhile, I remained active in the teaching field and focused on Data Science and Machine Learning, whilst also focusing on exploring the world of Open Science, which inspired this thesis.

With my background as a Software Engineer, I can hope that the blueprints, arguments and schemata provided by my work are useful in propelling Open Science more and more towards decentralisation. With my background as a Psychologist, I can hope that researchers can feel psychologically safer knowing that the system in which they work can be made fairer and more transparent, and that they can trust it not only to reflect their values but also to represent their economic interests.

---

[11] See, for example, (Simperl, Cuel, & Stein, 2013)

## 1.8 How to Navigate this Thesis

The first three Chapters contain the necessary context to understand the problem and the outcomes of this research. Further chapters explain the artifacts produced by this research. The thesis is concluded by a summary of findings and recommendations for future work. The Chapters are designed to build on one another, but for those who may want to read the Chapters non-linearly, here follows a synopsis of each separate Chapter:

**Chapter 2** – this Chapter describes **Background and Related Work**, starting out with the problems with academic publishing, explaining the meaning of decentralisation, and outlining the literature and research on blockchains, smart contracts and decentralised autonomous organisations. Also explored is the relevant literature on peer review, bibliometrics and the meta-principles of distributed capitalism.

**Chapter 3** – describes the overview of the **Methodology** that this thesis follows, and is split into engineering and social sciences sections. I explain the interdisciplinary approach, action research, activity theory, actor-network theory, thematic analysis, as well as smart contracts (Solidity) programming and the software design and engineering approach.

**Chapter 4** – outlines the novel **Framework** for decentralising with blockchains.

**Chapter 5** – presents original Smart Papers, the **blockchain implementation** of SC including Publication flow, Peer Review and Scientometrics calculation.

**Chapter 6** – presents new **survey findings** from a questionnaire administered to senior scholars, asking them to evaluate decentralised scholarly communication approaches proposed by the author.

**Chapter 7** – finalises the thesis with an overview of **contributions** and **limitations** of this thesis, a summary of **findings**, **conclusions** and **recommendations**, including the next steps for **future work**.

**Appendixes pt 1** – contain the overview of the modelling for the implementations, diagrams outlining the software artifacts, workflows and the security audit

**Appendixes pt 2** – contain the link to the survey responses from senior scholars, the questionnaire used for collecting this data, ethics approval

**Appendixes pt 3** – contain the outcomes of the SLR into blockchain-mediated decentralisation.

# Chapter 2

# Background and Related Work

## 2.1 The Value Expropriation of the Scholarly Economy

Since the creation of first scientific journals almost 400 years ago, large commercial publishers have increased their control of the scientific ecosystem (Larivière, Haustein, & Mongeon, 2015 a) giving rise to an oligopolistic market environment in which scientific research is disseminated. This peculiar market operates within the constraints of a concentrated supply side of a handful of publishing houses controlling a huge share of publishing activity, whilst facing a rather fragmented "buy" side of academic institutions (Ponte, Mierzejewska, & Klein, 2017). In fact, more than half of all research is accounted for by only 5 publishers (Larivière, Haustein, & Mongeon, 2015 a).

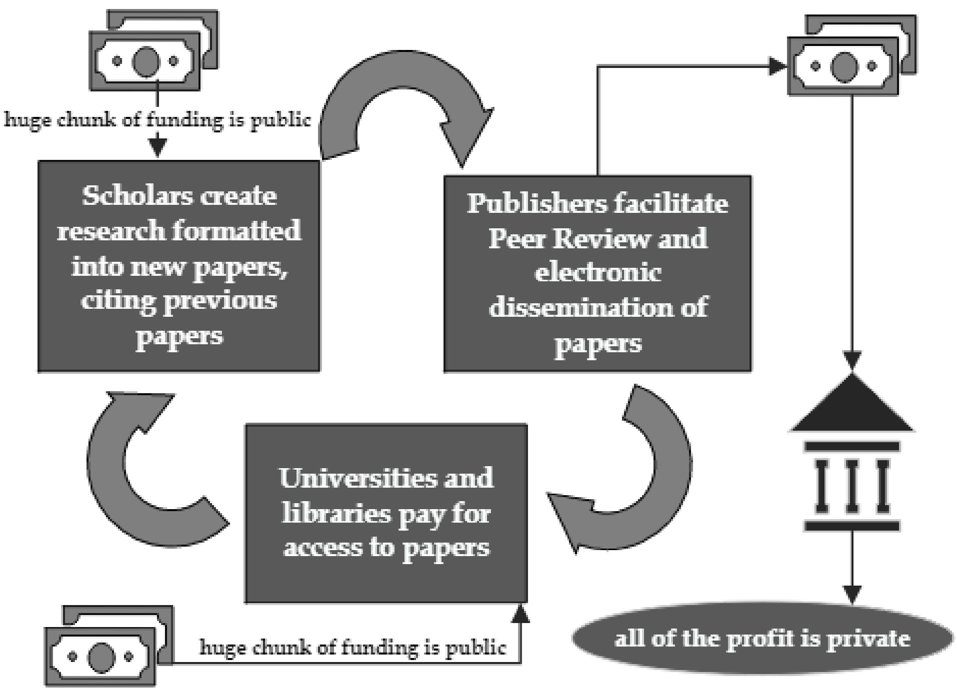



**Figure 2** – a model of SC adapted & evolved from McGuigan & Russell (2008)

There are significant concerns associated with this peculiar balance of power. In their analysis, McGuigan & Russell (2008) point towards the weak bargaining power of faculty and scholars as suppliers of intellectual property in today's publishing economy, observing that prestigious publishers are able to *"expropriate the value added by authors"* whilst increasing article prices that they are able to charge due to their oligopoly status. The authors quote a Deutsche Bank analyst stating that *"the value added to the publication process by the academic publishers is not high enough to explain the margins that are earned"* (Braley, 2005). The overall schematic of the system described by the authors is modelled in **Figure 2.** This set-up leads to many an apprehension among researchers about the fairness and accountability of the mechanisms at play. The main sub-problem here is double appropriation, where academics "*produce work not for profit, only to have publishers earn significant profit margins by selling the work back to academia"* (Tötösy de Zepetnek & Jia, 2014). It is not difficult to notice a moral hazard (and an incentive misalignment) inherent to an economy that exploits both scholarly labour as well as universities' budgets. Other problems include "*rapidly rising subscription prices, concerns about copyright, latency between results and their actual publication, and restrictions on what can be published and how it can be disseminated"* (Sompel, Payette, Erickson, Lagoze, & Warner, 2004). To make matters worse, "*loyalty to the scholarly publishing infrastructure is both enforced and self-reinforced by the role that the scholarly publisher plays in the career path of scholars.*'' (Lagoze, Edwards, Sandvig, & Plantin, 2015)

The current configuration of SC mechanisms also restricts public access to science. In 2018, one of the biggest publishing houses, Elsevier, stopped thousands of scientists from Germany and Sweden from accessing its recent journal articles, because the stakeholders failed to negotiate a satisfactory deal[12]. In 2019, the University of California, in a bold move stopped its purchases from Elsevier and walked away from contract negotiations - the motive was to save money and "*stop supporting the production of research that is inaccessible to the public*". Since, the University managed to re-gain access to Elsevier journals through a 2021 deal with a caveat that UC researchers get to publish OA versions of their articles without any fees payable to Elsevier[13]. These are just a few examples from the harrowing saga of universities across the world failing to strike satisfactory deals with publishing houses, which raises the question whether a system in which universities, libraries and governments cannot negotiate fair deals with publishing behemoths is the best one that we can envision? To come up with a better sociotechnical set-up to support global scientific enquiry and human progress through a more open and accessible scholarly communication that empowers both the scholars and the public requires us to understand better exactly what is broken.

---

[12] *"Dutch publishing giant cuts off researchers in Germany and Sweden"*, Nature (News) 19/07/2018, https://www.nature.com/articles/d41586-018-05754-1 *(Accessed April 2022)*

[13] *"Big Deal for Open Access"*, Inside Higher Ed (News) 17/03/2021 https://www.insidehighered.com/news/2021/03/17/university-california-reaches-new-open-access-agreement-elsevier *(Accessed April 2022)*

## 2.2 Why the System Prevails in an Inadequate State

One of the most pressing obstacles to change lies in the incentive structures of the academia, which are frequently inadequate in recognising and rewarding the efforts to open up the scientific process, possibly endangering advancement of researchers if they embrace non-traditional ways of working and publishing (Wilsdon, et al., 2017). The recent scholarly publishing market failures can be suitably framed as an *inadequate equilibrium* according to Yudkowsky (2017), a relatively stable state of a market system that could be improved upon, but for reasons of change being difficult to implement and coordinate, prevails in its inadequate state. As the decision theorist explicates, inadequate equilibria happen when unfortunate combinations of information failures and misaligned incentives lead to suboptimal match-making - this includes principal-agent problems[14], asymmetric information[15], and collective action problems[16]. In the case of scholarly communication, Yudkovsky argues that the SC system optimises for 'wrong' variables, such as "prestige per dollar" - from a grant-maker's point of view, any funding should ideally get their organisation's name in the headlines, and the research proposal promising a certain number of citations and an opportunity to be associated with a famous name or a famous topic will get cherry-picked by the funders. This cherry-picking does not quite work for research that could provide 'altruistic benefits' such as neglected health intervention studies that may not be immediately cited by others, minimising the opportunity for altruists to use this system. Therefore, for any scholar trying to do the 'right' thing from an altruistic standpoint, that thing is to ignore the current system's incentives.[17] However, from a game-theoretical point of view, those scholars that deflect from the current system will lose out if this system continues to prevail in the long run. Carrying out the proposed inadequacy correction is thus an example of a *wicked problem* defined by Churchman's principle that:

> *"whoever attempts to tame a part of a wicked problem, but not the whole, is morally wrong"* (Churchman, 1967),

---

[14] in political sciences and economics, a principal-agent problem occurs when the agent is acting on behalf of the principal but the agent's best interests are contrary to those of the principal

[15] in a transaction, information asymmetry occurs when one party has better information than the other

[16] in governance, this problem occurs when one agent acts on behalf of multiple principals, and the collective interest of the principals may become unclear, leading to high autonomy of the agent

[17] https://equilibriabook.com/an-equilibrium-of-no-free-energy/ *(Accessed April 2022)*

where many simultaneous changes need to be carried out for the system to reach another preferable equilibrium, but no individual benefits if they alone change their behaviour. Furthermore, solving one sub-problem can reveal the existence of new, previously unknown problems. In science, this can be illustrated by the lengthy delays, the limited impact and the local scope of the existing, mostly local, initiatives that aimed at changing the status quo of scholarly publishing (Rabesandratana, 2019). Taming the whole of the problem is a lofty ideal but it cannot be operationalised in a thesis like this one. Thus (at the risk of being morally wrong) to break down the wicked problem into more manageable subproblems, we must define them first. Scholarly publishing being an oligopoly would be one such subproblem, as it comes with the risk of unethical collusion by the big players and a high barrier to entry for new players (Stigler, 1964). Furthermore, oligopoly is a result of market centralisation, where the most bargaining power becomes concentrated in the hands of select few decision-makers. Power concentrations in profit-driven corporates may be motivated by greed and materialistic growth which may be parasitical if it involves exploitation, according to Schumacher (1973). To re-humanise economies of production, the author recommends a baby-step approach that is the opposite of Churchman's slogan, i.e.:

> *"the stupid man who says 'something is better than nothing' is much more intelligent than the clever chap who will not touch anything unless it is optimal" (Schumacher, 1973).*

I do argue hereby that certain aspects of the scholarly culture make the scientific community a good testing ground for "something that is better than nothing", i.e., a more decentralised approach, which would enable us to experiment with new incentives, new mechanisms and new infrastructures that could encourage new, arguably healthier behaviours among publishing scholars. But to be able to design a blueprint for a decentralised scholarly communications set-up, we must first understand the meanings of 'decentralisation' that relate to technology and society.

## 2.3 What is Decentralisation?

Decentralisation (American spelling decentralization - used frequently in computer science) means the dispersion, delegation, or distribution of power away from a central authority[18]. To successfully propose a means of decentralising scholarly communication, this thesis will incorporate an analysis of the unique meanings of decentralisation from the perspective of scholars, blockchain users, developers, and researchers. This will require a coherent framework, yielding a novel definition that encapsulates those meanings. The rationale for focusing on decentralisation is based on the fact that in the blockchain sphere, claims about decentralization have always been part and parcel of the debate on the value that this technology promised to offer – however - the idiosyncrasies with which the term "decentralization" is used within the blockchain space have generated confusion, especially since computer science already had its long-standing discipline of distributed systems research, equipped with its own vocabulary that focuses on "distributed," rather than decentralized, technical set-ups. For instance, in the late 1990s, distributed computing enabled the push toward a peer-to-peer (P2P) mode of communication over Internet networks, also known as server-less communication, as it did not require a central server to facilitate most of its communication flows. With that in mind, a chiefly technological classification would categorize blockchains as a type of peer-to-peer and a serverless distributed system, without calling it a decentralized one, suggesting that the sources of meaning of "decentralization" in blockchain contexts are not necessarily technological in nature, thus motivating the need for the present effort.

Nevertheless, the debate in recent years, mostly on Medium, Twitter and other social media, has centred around the notion of decentralization as the raison d'être for blockchains. Whilst the whitepaper specification for Bitcoin, the first-generation blockchain (Nakamoto, 2019) does not directly mention decentralization, it does still make the point that in this set-up, there would be "no central authority." The term "decentralization" is explicitly mentioned in the description of the Ethereum blockchain, whose official description talks about "a global, decentralized platform for money and new kinds of applications." But then again, what is a decentralized platform exactly, and how is it different from a peer-to-peer one? And fundamentally, what sort of "authority" might such a decentralized

---

[18] Merriam-Webster.com Dictionary, s.v. "**decentralization**,"' accessed April 2022 https://www.merriam-webster.com/dictionary/decentralization

platform be setting out to avoid, and what type of power is it dispersing? To answer these questions, and to clarify the meanings associated with "decentralization" in blockchain discourses, the thesis first chronologically reviews the historical meanings of decentralization, and subsequently contributes a survey of the modern usage of the term, comparing the scholarly usage with the emerging use of the term in the blockchain sphere, also including grey literature. Academic rigor will be injected into scholarly blockchain efforts, as rigorous approaches to blockchain research are notoriously scarce (Treiblmaier, 2020). Toward this end, a systematic review will be used to map out cross-domain meanings of decentralization, then subsequently, a framework will be proposed that enables the mapping of blockchain-mediated decentralization efforts into their more specific meanings. Finally, a resulting definition for "blockchain-mediated decentralization" is put forth. I will argue that groups, communities and societies, in order to function, need to constantly coordinate complex match-making and decision-making processes that are networked (and increasingly online), and which can be coordinated through an optimal allocation of authority, on a continuum between "centralized" and "decentralized" power, with the aid of blockchain-mediated decentralization as one technique for improving existing set-ups within the scholarly communication context.

To the best of the author's knowledge, this thesis embodies the first attempt at providing a framework for blockchain-mediated decentralization that would be rooted in the historical understanding of this concept. It should be noted that taxonomies do exist for classifying elements of decentralized systems, such as the one developed by Glaser & Bezzenberger (2015) which classifies cryptographic tokens, cryptocurrencies, decentralized organizations and decentralized applications; or the more recent framework developed by Tan (2019) which formalizes cryptographic token economics. However, these sorts of technocentric frameworks would on balance tend to ignore the sociocultural reality from which blockchains have emerged, usually either focusing on the technical aspect of their implementation, or examining just their design, or just how they are applied within a narrow context. In contrast, the main scope of the instant section stresses the archetypical nature of decentralization and how that archetype has informed and shaped the emergence of blockchain-mediated decentralization. Moreover, this work addresses the conclusions outlined in recent research by (Schneider, 2019) that urge for more specificity when discussing decentralization.

### 2.3.1 Historical Meanings of Decentralisation

Despite plenty of evidence in the literature of the various meanings assigned to the term "decentralization" (Kim, 2008, pp. 5-7), it is safe to say that this term has always had its origins in political science. As early as the mid-1800s, Tocqueville already distinguished governance centralization from decentralization, the latter being a prerequisite for healthy **democracy** (Janara, 1998, p. 208). Tocqueville also claimed that whilst centralized governance brings about efficiency, it is *decentralization* that **empowers** the individual (p. 210). This notion has been echoed throughout centuries; for example, Tiebout (1956) believed that decentralization improves the provision of public goods, "increases variety," and addresses the needs of local populations; Seabright (1996) observed that decentralization increases **accountability**. Since the nineteenth century, competing definitions for decentralization have proliferated – from their survey of over forty uses of this term, Dubois and Fattore (2009) found that decentralization usually focuses on themes of authority, responsibility, power and ownership, and that it often emphasizes the role of regional and local governments in politics and administration. Indeed, many countries have witnessed the rise of decentralization as a reaction to government failure, in an attempt to make their governments more accountable to individuals (Kim, 2008, p. 8). Thus, an initial glance at the literature points to the political science understanding of the term which focuses on addressing individual needs and preferences and individual empowerment. However, the role of technology in political decentralization requires a critical lens.

Almost in parallel to the political meaning of the word, in the last couple of decades, the term "decentralization" has made inroads into the world of computers and technological advancement. Curiously, a 1958 forecast in Harvard Business Review (Whisler & Leavitt) predicted that by the 1980s, the newly emerging field of information technology would *re-centralize* the techniques of organizational management, leading to the increased concentration of power at the top, and the disappearance of the middle management. The authors predicted that the only remaining reasons for attempting managerial decentralization would be psychological in nature (as the authors put it, to make better use of "the whole man" by encouraging "active participation" rather than mechanistic work), whilst any major economic reasons for decentralization would have become obviated by emerging technology facilitating unprecedented efficiencies through concentrated power. These efficiencies were assumed to inevitably lead to the concentration of business insight in the hands of the select few top executives. It is particularly interesting how the psychological needs and preferences of the working population were downplayed by that narrative. The authors minimized the role of the human in the loop, as his/her agency increasingly becomes thwarted by prcryptography proved to justesses and automated mechanisms.

However, by the end of the 1980s, and coinciding with the popularization of computer networks in the enterprise, the technological tide seemed to be turning against centralization. In (1988), Phil S. Ensor criticized the insular, inflexible top-down information flows in organizations and coined the term "functional silo syndrome" to mean a dysfunctional and unbending organizational structure; soon enough the talk of "**silo mentality**" commenced, and management consultancies began warning corporations against informational silos, encouraging individual departments to share insights. In the context of increasing demands for mechanisms and techniques promoting "variety and diversity" in manufacturing and services provision, it was Eugene B. Skolnikoff (1994) who concluded that "*technological change will tend, on balance, to favor decentralization of political power within societies over time.*" How much of this was wishful thinking is difficult to say due to the challenges in objectively measuring decentralization and political power, which will be discussed later in the article.

Meanwhile in counterculture, the late 1980s and early 1990s witnessed the development of the *cypherpunk* movement (Assange, Applebaum, Muller-Maughn, & Zimmermann, 2012) spreading the gospel of the reduction of governments with the help of technology (in particular, cryptographic technologies). The cypherpunk vision was defined in the *Crypto Anarchist Manifesto*, where Tim May (1992) famously stated that "*just as the technology of printing altered and reduced the power of medieval guilds and the social power structure, so too will cryptologic methods fundamentally alter the nature of corporations and of government interference in economic transactions.*" Hence, technology-mediated decentralization was no longer just limited to business and academia, it found its embodiment as a political movement. Many proponents of cypherpunk thought the cypherpunk ideal of a **free society** to be ultimately achievable through **cryptography**, the branch of mathematics and computing that deals with keeping one's individual preferences secret and enabling the private exchanging of communications in a way that could not be intercepted by governments and corporations. Cypherpunks, in particular believed that **individual freedom** requires active opposition to an emerging authoritarian and technocratic order (Beltramin, 2020), and this set of values and attitudes can, nowadays, also be observed in **communities** centered around **blockchain** and **cryptocurrency** development (the so-called "**crypto**" enthusiasts). However, despite the value attached to cryptography by cypherpunks and the "crypto" crowd, for many years, cryptography proved to just be one small facet of the efforts to counter authoritarianism, whilst there also emerged other non-cryptographic technologies attempting to empower the individual.

### 2.3.2 Modern Technology-Mediated Decentralisation

Cryptographic methods enabled some small and medium-scale end-user solutions to the problem of preserving individual autonomy in cyberspace in the 1990s, such as PGP-encrypted emails and SSL-secured financial transactions in Web browsers, but they did not seem to facilitate any major form of political change. At the time, a tectonic political shift was about to be achieved by peer-to-peer technologies that did not heavily rely on encryption, but rather focused on the efficient routing of data between nodes in a network, without having to rely on a central server to coordinate the flows of information. A famous example of a peer-to-peer technology upsetting the centralized status quo was Napster at the turn of the millennium, a P2P application that took on the major labels, allowing anyone to freely download any music (and other types of content) that they wanted from their peers, albeit not legally. Inevitably, the big recording shops and artist rights' groups launched a stream of legal proceedings against Napster's creators, which led to the downfall of not just the Napster application but also other alternative P2P solutions, with a social consequence of widespread stigma becoming attached to P2P usage.

What the P2P saga illustrated was that regulation had to play catch-up with technology and that pre-existing entrenched interests re-asserted their dominance through the political state apparatus. At the same time, the dramatic rise to power of large, consolidated technology providers including Google, Facebook and Amazon (dubbed the "**Big Tech**") placed Skolnikoff's hypothesis that "technology favours decentralization" under question, as well as putting to shame cypherpunk visions of the "collapse of governments." As recently as in the 2010s, researchers increasingly criticized the monopolistic and oligopolistic powers wielded by the centralized technology intermediaries, especially on the **World Wide Web**, which was originally designed to be universal and distributed. Some have called for the "**re-decentralization**" of the online space (Ibáñez et al., 2017) to try to rebalance the Web back to its Golden Age, if there ever was one. Blockchains/distributed ledgers have been embraced as the building block that would directly enable this re-decentralization. Fast forward to the 2020's, the contradictory claims about blockchain-mediated decentralization have generated a lot of confusion in the online, social media space, with widespread disagreement as to whether blockchains really render decentralization feasible.

To better understand **blockchains** (which are a subcategory of distributed ledger technologies, or DLTs, where records are shared by multiple participants), one should appreciate that these were initially designed with the following three goals in mind:

> ***(1)*** *To move away from centralized control of a ledger of transactions;*
> ***(2)*** *To provide a tamper-proof synchronization mechanism for the above;*
> ***(3)*** *To do the above among peers that do not necessarily trust each other or know each other.*

Blockchain platforms, such as Bitcoin and **Ethereum**, keep permanent and unchangeable records of transactions (such as fund transfers) between multiple parties, whilst also enabling those parties to run smart contracts, which can be thought of as self-executing agreements that do not need a centralized third party to verify (Underwood, 2016). Within this setting, blockchains can provide a single version of the truth to everyone involved, making all sorts of agreements arguably easier to manage whilst producing a secure audit trail. Blockchains can also be used to model state machines, which makes them ideal for implementing control flows (workflows) consisting of multiple stages that need to be completed in a particular transaction (for example, check clearing or crowdfunding). Consequently, one can use a blockchain as a building block to provide decentralized services, such as a decentralized autonomous organization (DAO), a decentralized identity (DI), or decentralized finance (DeFi). On top of those building blocks, we see new decentralized ways of coordinating social undertakings. For a better appreciation of the details of blockchain-enabled decentralization, I refer to my systematic review results in the Appendices.

### 2.3.3 Scholarly Communication as a Shared Resource / Commons

A growing number of researchers propose that scholarly resources be understood as common pool resources (CPR) in the style of the "commons" ("a resource shared by a group of people that is subject to social dilemmas"; (Hess & Ostrom, 2007 a)) as defined and described by the 2009 Nobel Prize winner political scientist Elinor Ostrom. In 2007, Ostrom applied the Institutional Analysis and Development (IAD) framework to further the recent commons debate to delineate the so-called "knowledge commons" relying on knowledge as a shared resource (Hess & Ostrom, 2007 b). The link between knowledge and commons can be made by acknowledging the problems traditionally associated with commons usage - such as resources abuse, over-extraction, pollution, and inequalities in resources access and distribution. Solutions to these problems are also inspired by the "natural" commons (e.g., fisheries), and include community-based non-private mechanisms that rely on social rules and management structures to preserve fair and sustainable resource accessibility. Potentially reframing scholarly communication as a commons-based system draws attention to the need for adaptive rules, collective action and self-governance in order to manage, disseminate and preserve the ever-growing scholarly record (Hess & Ostrom, 2004). Embracing commons-based thought could also mean that scholars would need to re-evaluate their “co-dependent” relationship with multi-billion-dollar-profit corporations. Blockchain could facilitate this transition according to available research on blockchain and commons - within the world of commons-based thought, blockchain technologies generated some enthusiasm, introducing two types of discourses: (1) those characterised by **techno-determinist**[19] and **market-driven**[20] values, tending to ignore the complexity of emerging social dynamics, and (2) critical accounts contributing to identifying limitations of traditional central institutions and whether or not they are necessary to enable democratic governance (Rozas, Tenorio-Fornés, Díaz-Molina, & Hassan, 2021). Blockchain enriches commons thinking by providing the following affordances: tokenisation of value and incentives, self-enforcement of rules,

---

[19] a reductionist theory that assumes that a society's technology progresses by following its own internal logic of efficiency

[20] the primacy of `free' markets in commons-based decision-making ensures that power remains in the hands of those who direct and control financial resources

formalisation of rules, automatization of processes, decentralisation of power over the infrastructure, increasing transparency, and codification of trust (Rozas, Tenorio-Fornés, Díaz-Molina, & Hassan, 2021).

## 2.4 The "Open" Initiatives

### 2.4.1 Open Science

Most research outcomes are hidden behind the paywalls and only accessible to financially strong institutions. Findings are published preferentially, so that important information remains inaccessible in private data stores. Methodologies and tools are inadequately documented and impossible to understand by the wider public. Against this concerning landscape, Open Science is a relatively new and promising approach to democratising the scientific process based on cooperation and digital collaboration. Open Science entails a move from the traditional practice of publishing research findings in journals, towards sharing available knowledge as early as possible in the research process (Wilsdon, et al., 2017). Open Science has democratic goals, as it exposes the alleged hypocrisy of science, which portrays itself as working for the common good, but still some argue that it exists in its own ivory tower, too isolated from society (Holbrook, 2019). The authors Fecher and Friesike (2014) propose five OS schools of thought:

- infrastructure school (concerned with technologies and systems)
- public school (concerned with the accessibility of knowledge creation),
- measurement school (concerned with impact measurement)
- democratic school (concerned with access to knowledge)
- pragmatic school (concerned with collaborative research).

The assumptions, goals and keywords of each of the five OS schools are drawn in **Figure 3**.

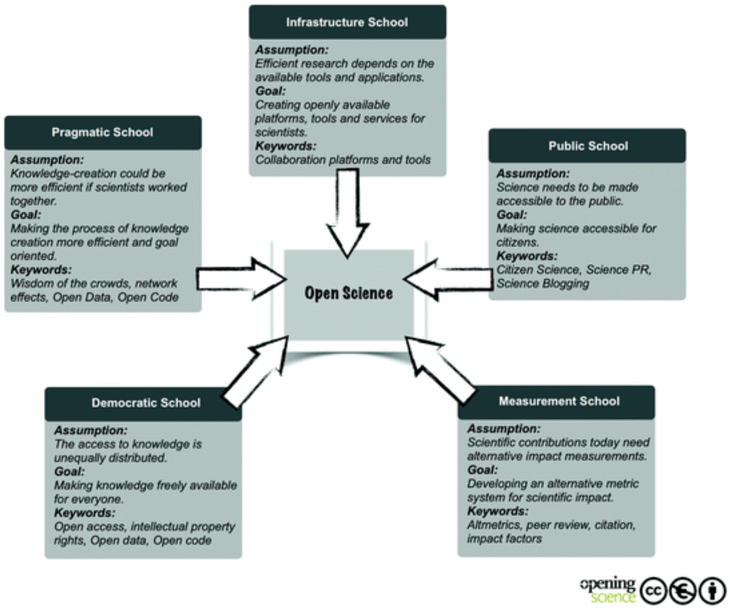



**Figure 3** - The five schools of Open Science from Fecher and Friesike (2014)

The five-schools proponents recognise that novel methods of knowledge creation and dissemination go hand in hand with systemic changes (e.g., new ways of measuring scientific impact), changes in the daily practices of researchers (e.g., using new tools and methods), as well as changes in the practices of academic publishing (e.g., embracing alternative publication formats). The authors urge for holistic approaches and futuristic thinking, advocating the introduction of new incentives: for sharing Open Data, for working collaboratively online, for exploring alternative research communication media (like blogging). (Fecher & Friesike, Open science: one term, five schools of thought, 2014).

Apart from the five-schools model, The FOSTER Taxonomy of Open Science is often quote as an authoritative source on defining the various branches of the OS movement. FOSTER[21] (Fostering the practical implementation of Open Science) is an EU-funded project, carried out by 11 institutions from across 6 countries. The Taxonomy is given in **Figure 4** - The FOSTER Taxonomy of OS.

[21] https://www.fosteropenscience.eu

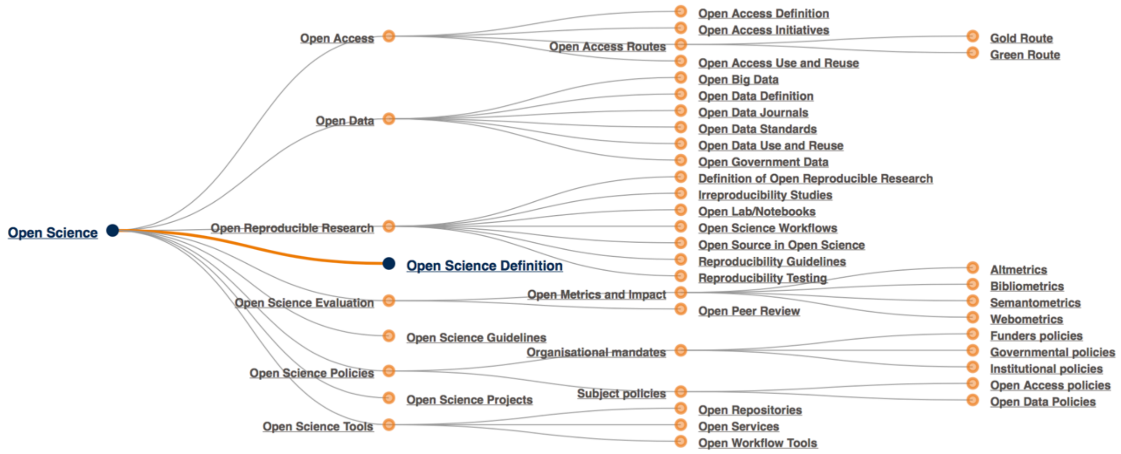



**Figure 4** - The FOSTER Taxonomy of OS

### 2.4.2 The FAIR Data Stewardship Principles

The FAIR acronym is defined to provide researchers with a list of principles for improving data management in research (Wilkinson, 2016), positing that all research data be:

- **F**indable (assigned a globally unique identifier, indexed)
- **A**ccessible (data and metadata are retrievable, access is free and universal)
- **I**nteroperable (shared vocabularies are used)
- **R**eusable (properly licensed and associated with detailed provenance)

The authors Wilkinson et al. (2016) propose that FAIR Data principles. when implemented, will promote data stewardship to the benefit of the entire academia, as a "pre-condition supporting knowledge discovery and innovation". The full benefits are formalised as the Enabling FAIR Data Commitment Statement, containing recommendations for the following stakeholders: repository managers, publishers, societies, communities, institutions, funding agencies and researchers.[22]

### 2.4.3 Open Access

An initiative called Open Access (OA) (Laakso, et al., 2011) has grown in popularity, promoting free online access to scientific knowledge. Advocates of OA

[22] https://www.nature.com/articles/d41586-019-01720-7

claim that commercial publishers using traditional business models are making unreasonably large profits, whilst the critics are dubious of OA's economical sustainability noting that information "wants to be expensive".[23] Regardless of the ongoing critique of the Open movements, there currently exist active transnational efforts to mandate OA through policy interventions. Most importantly, *Plan S*[24], an initiative for open-access science publishing that was launched by Science Europe on 4 September 2018. This effort is coordinated by the *cOAlition S*, a consortium launched by major national research agencies and founders from twelve European countries. Plan S requires that, from 2021, scholarly publications that result from publicly-funded research be published in compliant Open Access journals or Open Access platforms.

There seem to be some strong benefits of adopting and promoting OA, and the biggest one seems to be the popularisation of science - articles are generally viewed and downloaded more often than non-OA articles (Wang, Liu, Mao, & Fang, 2015) and generate more sustained interest (Davis, 2011). Readership is also higher in previously excluded demographics, such as patient groups, non-profit sector workers and independent researchers (ElSabry , 2017). OA articles are also cited more, and are referenced more on social media (Shema, Bar-Ilan, & Thelwall, 2014).

The criticism of Open Access is related to its lower barrier to entry and alleged correlation with predatory publishing, e.g. low-quality journal publishing with lax peer review and deceptive marketing practices, damaging the reputation of OA. To combat predatory publishing, reputable OA journals ensure their auditing by means of community curation, through websites such as Scientific Electronic Library Online (SciELO)[25] and the Directory of Open-Access Journals (DOAJ)[26].

Open Access journals should not be confused with preprint servers, an informal mechanism for disseminating papers, with only rudimentary quality checks. Many big funders have chosen to recognise preprints, e.g., The U.K. Medical

---

[23] *The pros and cons of Open Access*, Nature (Web Focus), by Kate Worlock: https://www.nature.com/nature/focus/accessdebate/34.html *accessible only through: https://web.archive.org/web/20170112192537/https://www.nature.com/nature/focus/accessdebate/34.html*

[24] https://www.coalition-s.org/ (Accessed April 2022)

[25] https://scielo.org/en/

[26] https://doaj.org

Research Council, Wellcome Trust, NIH encourage applicants to cite preprints (not necessarily peer-reviewed) in their grant proposals. One criticism of pre-print culture is their potential to attract plagiarism due to lower quality standards than those expected by journals. This is increasingly causing calls for the regulation of preprints and introduction of stricter preprint standards. In any way, it is believed that preprints do not diminish the need for reputable peer-reviewed journals (Sarabipour, et al., 2019). It is important to note that Open Access and preprints are just technical solutions, whose success is evaluated quantitatively, to the social issues highlighted by Open Science that should be addressed in qualitative ways. The next section illustrates the qualitative problems of the complicated economy of science that may be hindering the adoption of the open movements.

### 2.4.4 Open Source Software in Research

In their report, (Hasselbring, Carr, Hettrick, Packer, & Tiropanis, 2020), the authors evangelise the need to make research software open source for "good scientific practice", including reproducibility and reusability purposes. They motivate their recommendation through the example of computer sciences, where Open Source research software is a "central asset for development activities". It is observed that given the benefits of open-source licensing, it should be "the rare exception to keep research software closed".

The proponents of OSS (also FOSS and FLOSS[27]) in science note that software engineering methods may be beneficial to research in other disciplines. The essential freedoms associated with Open-Source mean that the software can be ran without limitation as to its purpose as long as it is ethical to do so, it can be studied as to how it works, it can be studied as to how it was developed, it can be redistributed to help others, and it can be modified by others where due credit is given.

---

[27] OSS – Open Source Software,
FOSS – Free and Open Source Software,
FLOSS – Free and Open Source Software.
The use of libre denotes the idea that the software is "free as in free speech, not free beer"

## 2.5 The Convoluted Economy of Science

The purpose of this subsection is to highlight the complexities associated with studying how "Science" works as an "economy". This is because at many points of this Thesis, we will need to be discussing the details of how "Science" (i.e. scholars, scientists, academics) produces, consumes and exchanges value, and how it allocates resources within its ecosystem. As most research productivity can be attributed to the "Global West", the analysis will be heavily skewed towards narratives associated with capitalist economies.

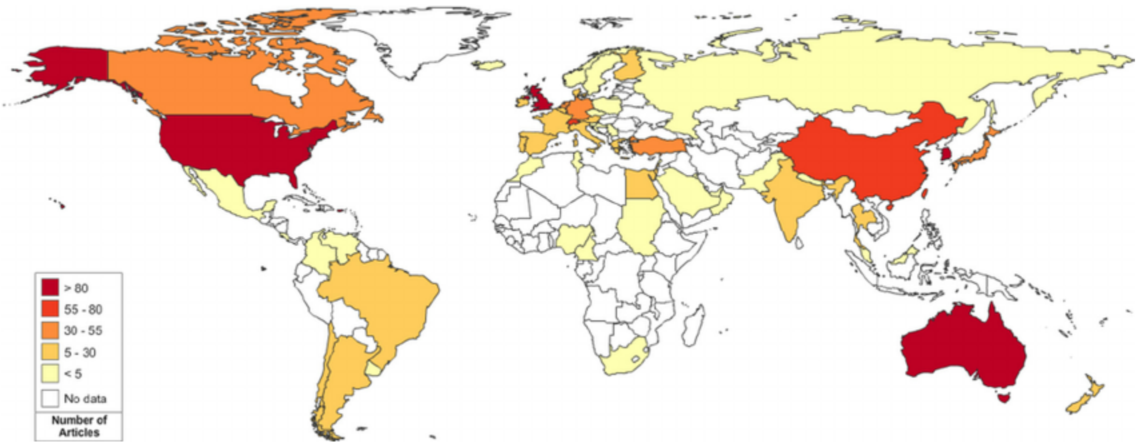



**Figure 5** The world map of worldwide research productivity from Luo et al., 2015

### 2.5.1 Monopolies of Knowledge

Economic historian Harold Innis developed his ideas of monopolies and oligopolies of knowledge when writing about communications theory (Innis, 2007). He argued that whenever new media emerge, so do monopolies of knowledge, reinforcing the power of political elites in direct relation "to the demands of force". Innis focuses his concerns on the United States of America, where he observes how privately owned media networks undermine independent thought and increase passivity in their audiences by creating monopolies of communication. He also observes the role of the monopolies of knowledge in suppressing scientific ideas. In order to study monopolies of knowledge, Innis highlights the importance of studying economics, communications and technology.

The researchers (Lightfoot & Wisniewski, 2014) note that academia is not immune to pressures from corporate and financial elites, and that the education system is a "key component in creating and maintaining power and informational imbalances within societies". To counter the stifling effects of the oligopolies of power, they recommend that knowledge is spread more equally and transparently. They predict that this change can never come from corporate spheres and highlight the role of the academe in progressing the cause. When considering the issues of knowledge democratisation, they highlight the importance of considering political theory, sociology, and psychology.

Monopolies of knowledge can play a significant role in exacerbating the current crisis in science. These monopolies can take various forms, including the dominance of a few large publishers, the concentration of research funding in specific institutions or regions, and the influence of gatekeeping mechanisms in determining what research gets recognised and disseminated. This limits the availability of scientific knowledge to a select few who can afford access, hindering scientific progress and hampering the dissemination of research to broader communities. When it comes to researchers from underrepresented or economically disadvantaged backgrounds, they can be left out of the *monopoly* sphere, which can impede scientific advancement and limit diversity in scientific perspectives and breakthroughs.

### 2.5.2 Publishing in Light of the Economic Nature of the Firm

At this point, a certain peculiarity must be explored as regards publishing organisations. All human labour has been subject to economic scrutiny since economics started looking at it through the production-consumption lens. However, scholarly publishing does not seem to conform neatly to major economic theories. One such theory is Coase's theory of firms. Ronald Coase proposed in 1937 that "firms" exist for two main reasons. Firstly, to reduce transaction costs by optimising internal transactions and minimising the reliance on external (market) ones. And secondly, to reduce production costs by increasing productivity through specialisation and teamwork. The main drawback of firm organisation, according to his theory, is then the difficulty of assessing contributions of individual employees (Coase, 1937). For his work, Coase received a Nobel Prize in Economics.

What can be learned from his findings? It can be observed that scholarly communications may possibly (weakly) conform to his ideas about productivity (researchers work in teams and increasingly specialise) but it is very difficult, if not outright impossible, to talk about research items as products, because their creators are not just driven by the desire to sell their research on the market. Some could even argue there is no market - you cannot, for example, negotiate the price for your scholarly research paper, unless you explicitly commercialise your research outside academia. Under Coase's' very early economic model *"the price mechanism should give the most efficient result"* for selling the product. In modern-day academia, where this is not the case, what is then the purpose of charging article fees by publishing houses? It is arguable whether it enriches Science in any tangible way. Coase, in 2012 and shortly before his death, wrote in the Harvard Business Review that *"the degree to which economics is isolated from the ordinary business of life is extraordinary and unfortunate… The tools used by economists to analyze business firms are too abstract and speculative to offer any guidance to entrepreneurs and managers in their constant struggle to bring novel products to consumers at low cost."*[28] He went on to state that:

> *"it requires an intricate web of social institutions to coordinate the working of markets and firms across various boundaries. At a time when the modern economy is becoming increasingly institutions-intensive, the reduction of economics to price theory is troubling enough."*

Coase's' use of the Web imagery points to the decentralised and interconnected nature of successful markets that resemble networks in which value is produced, exchanged and consumed. It is worth looking at Web-inspired new economic models that have recently emerged under the umbrella of sharing economies.

### 2.5.3 Sharing Economies

Having introduced a concept of marketplaces with incentives that are not strictly monetary, I must mention *sharing economies* - a type of technology-mediated exchange system that creates new ways of provisioning products and services that are capable of reallocating wealth away from middlemen and towards small producers and consumers (Schor, 2016). In the context of academia, its *reputation economy* can be seen as having similarities in common with a "sharing economy", as both value peer recognition and tokenised trust (for example, in the form of reviews or ratings). By cutting out middlemen, a sharing economy requires individual users to be trustworthy and to trust each other. In sharing economy websites, trust is anchored by identity checks performed by the platform and mediated by a flow of user reviews that result in user ratings based on multiple

---

[28] https://hbr.org/2012/12/saving-economics-from-the-economists

scales. The platform can still be seen as a trusted third party, but its role and complexity is reduced, whilst more power is placed in the hands of platform users. In case of AirBnB, a sharing economy platform for room rentals, individual messaging between peers is a key element that can make or break a successful transaction. Thus, a lot of information is exchanged prior to value exchange. It is an48example of a meaningfully (but not highly) decentralised system with the benefits of economic flexibility that empowers everyday small-scale property owner rather than major centralised players such as hotel chains.

The main criticism of platforms like AirBnB, eBay and Uber is that, although they allow users to transact directly with each other, it is the centralised platform remains in control of search logic, result presentation and even arbitration in case of conflict. This wider problematic has been studied by Srnicek (2017) and dubbed "platform capitalism". The only solution to this problem is to free up the core source code and open up the platform's protocols. Some analysts go as far as to suggest that the sharing economy term is 'vague and may be a marketing strategy' and the Economic Policy Institute[29] refers to these firms as internet-based service firms (IBSFs). There is a threat that sharing could put jobs in the regular economy under pressure with unintended negative consequences to workers' rights, but also consumer protection and privacy. Finally, because of the lack of any fractional ownership or governance mechanisms in core sharing economy firms, the end users have no voice and no stake in the platform, and therefore no control over the platform's future. One thing that could solve this conundrum is changing the legal and operational structure of the platform so that it allows its users to have a real stake in it, which I will discuss later in the DAO (decentralised autonomous organisation) section of this thesis. DAOs represent the collaborative ideals of the early Web and place emphasis on openness and transparencies.

### 2.5.4 The Collaborative Web vs the Competitive Free Market

Open Science and Open Access share attitudes with the Open Source[30] initiative which emphasises collaboration, a form of cooperation. This focus on cooperation can be seen as a return to the core values of science (Chan & Loizides, 2017), p.

[29] https://www.epi.org
[30] https://opensource.org

154. Whilst competition is essential to capitalist economic development, so is cooperation – the balance between the two is seen as necessary. However, modern economics does not discuss cooperative forces in much detail, implicitly assuming that the price mechanism could possibly effect whatever coordination is needed for the successful functioning of the economic system (Teece, 1992), a thought that may be traced to Austrian economists, chiefly Hayek. Austrian economists tend to favour the idea of "free markets" and apply this idea to all sorts of problems – this notion had been famously promoted throughout Hayek's work, who believed that "we are able to bring about an ordering of the unknown only by causing it to order itself", thus, yielding most manual interventions into economic systems harmful or at best ineffectual (von Hayek, 1991). Despite my background not being academically in Economics, it is my understanding from studiously reviewing Hayek's works, that his arguments make plenty of assumptions about there being a positive correlation between economic and personal liberty and the importance (or even inevitability) of the competitive price mechanism as a balancing force for good. However, those assumptions seem to clash with coercive aspects of oligopolies and monopolies that do not neatly respond to competitive pricing. At least for a century, we have been hearing social critiques of centralised silos in areas such as business, technology and even music (Baym, 2010) which is dominated by major record labels. The critiques of the "big shops" usually converge around how they unfairly get away with their misdeeds (market manipulations) due to their overwhelming size, resulting in a loss of accountability (Hindriks & Lockwood, 2004), (Schminke, Ambrose, & Cropanzano, 2000), (Arnold & Gillenkirch, 2011). In (Mookherjee, 2006), the author identifies a moral hazard for intermediaries (eg. Publishing houses) with incentives that differ to those of their principal's incentives (eg. Academia) in monopsony conditions (a situation where there is only one buyer, e.g. the government), something that Hayek would recognise (as a "negative externality"), but without offering a solution to the question of how to prevent it from happening.

Even though Hayek's influential school of thought, having dominated Western neoliberal policy space of the last couple of decades, has been trusted to spontaneously bring order to markets, for many years, and culminating in the 2008 financial crisis, we have been witnessing multiple large-scale market failures and the creation of entities that are "too big to fail". Thus, this belief in the supremacy of free markets and the price mechanism, simultaneously downplaying the importance of decision-oriented nature of human action, may have been used to formulate a rather unfortunate basis for modern policy making, as (Fuchs, 2007) argues, without intentional human action, there cannot be economy, and there cannot be society. Hayek was somewhat aware of the possibility of market failures due to those negative externalities. In particular, he understood the danger of "all exclusive, privileged monopolistic organizations" (Hayek F. A., 1960). He later also acknowledged that public goods cannot be provided by the market mechanisms. He did not, however, give much thought to alternative mechanisms to the provision public goods in a manner that avoids the emergence of monopolies and oligopolies, and he wasn't much interested in the corporate operations of scholarly communication, which were not as centralised in his times as they are nowadays.

Whilst Hayek recommends that we trust the well-established institutions that have brought us prosperity, we may now be evolving past needing those institutions. Thanks to technological advances, the information society promises a new transcendental space, a cooperative society (Fuchs, 2007), forming a type of participatory democratic economy in which human agency can thrive. The cooperative aspect of this new paradigm is crucial. A key feature of cooperation as a form of *doing good* is that it is directed at alleviating structural problems with society (e.g., caring for the sick) rather than incidental problems (like helping a friend with her broken leg) (Van Vugt, Biel, Snyder, & Tyler, 2000). Cooperation is thus seen as the very essence of a healthy and helpful society, yet its value is often antagonised by competitive market forces, for example when governments decide to privatise their health systems. This is perhaps partially caused by some other trends in economic thought – Fehr et al. ( (2000), (2002)) have shown economists at large to be systematically biased in favouring the self-interest hypothesis whilst ruling out heterogeneity in the realm of social preferences (including interpersonal altruism, fairness, reciprocity, and inequity aversion), essentially showing them to favour corporate interests whilst downplaying the role of non-corporate cooperation-based motivations.

Nevertheless, this thesis takes an altruistic value-driven stance favouring collaboration, and it is therefore assumed that following the free-market order is not optimal for coordinating the economy of science. I argue that it is only through active centrally coordinated interventions that small- and large-scale scholarly cooperative and collaborative initiatives can be incentivised, fostered, and sustained. Thus, we arrive at a sort of paradox, where central coordination of a project/programme can lead towards a transition to a more decentralised system. Thus, decentralisation brings us back to Hayek, who observed that:

*“Decentralization has become necessary because nobody can consciously balance all the considerations bearing on the decisions of so many individuals . . . because all the details of the changes constantly affecting the conditions of demand and supply of the different commodities can never be fully known, or quickly enough be collected and disseminated, by any one center, what is required is some apparatus of registration which automatically records all the relevant effects of individual actions.”* (Hayek F. , 1944)

Whilst in Hayek’s conception, that *apparatus of registration* is assumed to be the free-market price system. But Mirowski (2011, pp. 317-319) argues that *‘markets are not only limited and intermittently unreliable information processes; they can equally well be deployed to produce ignorance’,* for instance, to manufacture junk science by corporations, as an effective delay tactics to avoid regulation, compliance or compensation, such as in the case of tobacco industry (Proctor & Schiebinger, 2008). Corrupting science thus becomes “yet another creative way for corporations to engage in politics”[31]. Nonetheless, there emerges a parallel in Hayek’s language between the functioning of the market and the functioning of the Web, including blockchains (which would have been unknown to him at the time). Hayek believes the market to be a *communication system* and attributes its success to the *dispersion of knowledge,* where both functions can be attributed to the Internet (Castro, 2016) and Web3, the new iteration of the Web that introduces blockchains. However, price discovery is only but a tiny functionality achievable by the use of the modern Web. The value of information and knowledge can still be measured by the price of the relevant article in a competitive online marketplace, but it can also be measured by a myriad of collaborative informetrics available thanks to the Web, such as social media likes, downloads, shares and retweets, and formal scientometrics that could be calculated by a blockchain smart contract. These will be discussed in the next subsection.

[31] https://www.societyandspace.org/articles/science-mart-by-philip-mirowski, Review by James Porter (13/06/2013)

## 2.6 Scientometrics

Informetrics (a wider study of how information moves through social systems, not just science-oriented ones (Bar-Ilan, 2008)) is the study of the quantitative aspects of information, and it includes scientometrics, which in turn deals with statistical evaluation of recorded scientific information (publications), focusing on statistics such as numbers of citations of a scholarly paper (sometimes called bibliometrics). This is a very self-aware paragraph of this thesis, since at the time of writing the present work, there is a shake-up happening at the Elsevier-owned *Journal of Informetrics*. The original editorial board has left the big publisher, justifying their move by the need to make bibliographic references freely available for analysis, and by an aspiration towards being truly Open Access and community-owned. The journal's board members wanted Elsevier to lower its article-publishing charges and to take part in the *Initiative for Open Citations (I4OC)* [32], so as to free up citation data for research. When Elsevier declined to meet the requests, they launched a free-to-read journal called Quantitative Science Studies (QSS) published by MIT Press.[33] The journal studies issues related to citation analysis and its use in academia.

### 2.6.1 Citation Analysis

Davenport and Cronin (2000) view citations as "links" between papers that form tokens representing trust in that when, for example, Alice cites Bob, Alice assumes that Bob's claims are trustworthy enough to support her claims. In this general case, citation statistics may be used in recommender systems to prioritise highly

[32] a collaboration between scholarly publishers, researchers, and other interested parties to promote the unrestricted availability of scholarly citation data, see: https://i4oc.org/

[33] https://www.mitpressjournals.org/loi/qss

cited items. However, many authors take issue with this simplistic positive view of citations, instead pointing out the relevance of various meanings, contexts and sentiments of citations (Abu-Jbara, Ezra, & Radev, 2013). Hence, under a more nuanced model, we would have to pay attention to the polarity of the citation, either positive, representing trust, or negative, representing mistrust in someone else's findings. Even more complex NLP-driven scientometrics-based citation analysis methods have been proposed recently by (Jha, Jbara, Qazvinian, & Radev, 2017) but they remain out of scope of this thesis. Also out of scope is the discussion of the modern standards by which academics are judged and evaluated in their careers, and the widespread criticism of basing complex grant and hiring decisions on rather simplistic metrics like citation counts (Lawrence, 2007).

Coming from a different angle, citations can be explained as symbolic payments of intellectual debt (Small, 2004), making citation analysis a useful tool for sociological analysis into how credit (or discredit) flows first to the publication, and then to the associated authors, institutions, countries, and journals that are its attributes. Despite the perceived great significance of citation counts as a scientometric resource, alarmingly, currently available citation counts are not free from miscalculation (Adam, 2002). For example, one of the biggest platforms nowadays to provide citation counts for scientific articles, Google Scholar, contains errors that are reported in literature (Harzing & Van der Wal, 2008). As an author of more than one published paper, I am even able to confirm that the Google Scholar count for one of my papers is incorrectly reported, as one source has been counted twice. Similar significant informetrics errors like phantom citation counts, missing citations and duplicate references have also been reported for other scholarly data providers such as Scopus (Franceschini, Maisano, & Mastrogiacomo, 2016). For the curious reader, a good overview of the limitations of scientometrics is provided in (Abramo, 2018).

### 2.6.2 The problems with h-Index and Impact Factor values

Most traditional scientometrics, quoted by research portals such as Google Scholar and ResearchGate, are based on citation analysis, including h-Indices and Impact Factors. The h-Index, proposed by Hirsch, measures "relevant productivity" (either of a particular author or sometimes of a particular journal), where $h$ is the threshold of minimum citations accounted for; often said to be correlated with the "impact" of the entity but this is a common misconception, as it does not consider any actual outcomes in the real world, outside of the realm of citations. Now, the Impact Factor, proposed by Garfield, measures the ratio of new citations to recent publications; and is often said to be correlated with the "quality" of the published research, but this seems to be another misconception, since as any citation-based system can be gamed (for example by means of manipulating editorial policies), the objective interpretation of its meaning becomes challenging. Furthermore, citation distributions within journals are highly skewed, the properties of citations are field-specific, and the data used to calculate them is currently not available to the public. The final limitation can be addressed by a public blockchain which will be explained in Chapter 5, but this will not remove the remaining limitations of these scientometrics.

The San Francisco Declaration on Research Assessment (DORA), led by Stephen Curry of Imperial College London (my alma mater), must be mentioned here as it aims to spread awareness of the poor correlation of the Impact Factor (Journal Impact Factor, IF, JIF) and the impact or merit of an individual's scholarly work. The declaration has been signed by 2628 research institutions worldwide[34]. One signatory, the Utrecht University announced that it discontinues the use of impact

[34] https://sfdora.org/signers/ [Accessed August 2022]

factor in hiring and promotion[35]. For the purpose of evaluating research, DORA recommends that researchers consider the value and impact of all research outputs (including datasets and software) in addition to publications, and consider a "*broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice*" [36]. Nonetheless, it is a known fact in academia that citation-based metrics are still being used for many purposes to hep scholars "get an intuitive feeling" of the weight carried by a particular paper, author, or journal [37], [38]. Critics of this approach argue that over-reliance on metrics has led to perverse incentives enabling a *publish-or-perish* environment that is detrimental to research quality (Weingart, 2005).

### 2.6.3 Altmetrics

Altmetrics are a broad group of metrics, capturing various parts of impact a paper or work can have, mostly by using the functionality of the World Wide Web, including open APIs and social media. Commonly found altmetrics report on the numbers of times an article has been Read or Viewed, Discussed, Blogged about, Tweeted about, Shared via a link, Saved as a bookmark, or Upvoted/Recommended. Researchers suggest that altmetrics can be gamed as easily as traditional scientometrics, and recommend that traditional and alternative metrics should complement (and not compete with) each other (Bornmann L. , 2014).

---

[35] Woolston, Chris (2021-06-25). "Impact factor abandoned by Dutch university in hiring and promotion decisions". *Nature Career News*. doi:10.1038/d41586-021-01759-5

[36] https://sfdora.org/read/ [Accessed August 2022]

[37] https://explore.researchgate.net/display/support/h-index [Accessed August 2022]

[38] https://scholar.google.co.uk/intl/en/scholar/metrics.html#metrics [Accessed August 2022]

## 2.7 Peer Review

Peer review is situated at the very core of science, its structure and operations. As a family of mechanisms whose role is to power a deeply social enterprise of scholarly communications, peer review has got to be supported by a social network of “values, beliefs and myths" as observed by (Chubin, Hackett, & Hackett, 1990). Its purpose, at a very basic level, seems clear enough - to improve the quality, even trustworthiness, of findings that are disseminated with a hope of them becoming scientific knowledge. However, peer review processes have frequently been described as *flawed* (Smith R. , 2006), hard to define and difficult study in a purely objective manner. Critics of peer review go as far as to ask the question whether peer review should “continue to operate on trust”. Ultimately, if citizens perceive peer review to be secretive and elusive, then is this really the best way to organise science in democratic societies? Can transparency and accountability be improved?

Peer review receives consistent support from academics, who agreed in a recent survey (Ware, 2008) that it plays an important role in science. The survey outcomes suggested that peer review is believed to “help scientific communication” thus “improving the quality of published papers.” Nevertheless, if peer review is so essential to science, it is somewhat ironic that science has so very little to say about whether it works. A systematic review of the evidence related to peer review effects concluded that "the practice of peer review is based on faith in its effects, rather than on facts" (Jefferson, Alderson, Wager, & Davidoff, 2002). The peer review process can also introduce bias. Significant bias and parochialism have been identified in the peer review system (Goldbeck-Wood, 1999). Can we have a better approach that addresses these problems with a view to improving the quality of science?

It is often said that peer review has a gate-keeping role. This gatekeeping may be done in a selfish way (“this paper contradicts me – reject”, “cite my paper, pay homage to my idea - conditional accept”, etc.) or altruistically (regardless of the outcome - honest evaluation of the pros and cons, and advice on how to improve). Regardless of how objective the reviewer may strive to be, research from 2006 shows how 150 proposals submitted to the National Science Foundation were evaluated independently by a new set of reviewers, only to prove that getting a research grant depends significantly on random chance (Jennings, 2006). Rather than denying the chance factor, it would be interesting to look retrospectively at past citations that led to accepted/ rejected papers, and to examine how editors vary in their ability to pick the winners, by measuring the review-outcome across subjective/objective dimensions. Clearly, a decentralised peer review data analytics platform is needed to free the peer review data for such complex scientometrics. This would enable us to better understand how true quality of scientific outputs is reflected in a variety of transparently calculated metrics.

PEERE[39] is a large European Cooperation in Science and Technology (COST) initiative that teams up with major publishers to devise a protocol for sharing information on all aspects of peer review. Experience shows that sharing peer-review data can foster transparency and accountability in publishing (Squazzoni, Grimaldo, & Marušić, 2017). Nevertheless, as of the time of finalising this thesis, the details of the PEERE protocol remain elusive and unpublished.

[39] http://www.peere.org/

### 2.7.1 Deconstructing the Scholarly Journal

Models known as the “deconstructed journal" (Smith J. W., 1999) and the “decoupled journal" (Priem & Hemminger, Decoupling the scholarly journal., 2012) have been proposed in recent years. The authors of both concepts argue that due to the tight coupling of the journals function’s, i.e. their essential services of research registration, quality control, dissemination, and notarisation, it is difficult to change any single aspect of the journal system Decoupling, closely related to the concept of separation of concerns, is a cornerstone computing concept that has informed many modern Web architectures from microservices to service-oriented architecture (SOA). For example, PLoS[40] treats copy-editing as a service which can be run independently from its other services, an approach that has proven both successful and profitable. Deconstructed/decoupled journals have the potential to use the full power of the modern Web, where individual Web services evolve flexibly according to their users’ needs.

It has been increasingly argued that peer review has the potential to be decoupled from the journal system and evolve into a standalone process, and possibly be combined with other quality control mechanisms. Nevertheless, with each Web-based journal and conference following their own and oft-closed specifications for their various services, there is a lack of open peer review standards. The existing Web-based journal systems also rely predominantly on closed software, except for select few new crypto-oriented publications like Ledger[41] which, although it is based on a PHP and MySQL back-end, additionally employs a public blockchain to timestamp published manuscripts; also noteworthy is the Editorial Workflow in PKP Software Open Journal Systems (OJS) which employs an open-source Editorial Workflow with well-documented steps for editors, reviewers, copy-editors and proof-readers[42].

---

40 https://journals.plos.org/plosone/

41 https://ledgerjournal.org/

42 http://pkpschool.sfu.ca/courses/editorial-workflow-in-ojs-3/

## 2.8 Modelling the Research Metadata

### 2.8.1 Semantic Web / Linked Data approaches

Conceptual models like Liquid Publications (Casati, Giunchiglia, & Marchese, 2007) and Dynamic Publication Formats (Heller, The, & Bartling, 2014) have been proposed to leverage Semantic Web technologies to transform research objects from static to evolutionary entities. In these models, authors collaborate on a living version of the research object that, upon the authors' agreement, has periodical snapshots or releases published on the Web. Releases can be open for comments and reviews from the members of the public or submitted to Calls for Contributions of conferences or journals. Authoring tools like *Dokie.li* (Capadisli, et al., 2017) go one step further and provide decentralised implementations of living research objects that allow authors to retain the ownership of, and sovereignty over their data. This supplies an alternative to the current state of play, where scholarly publication processes are centralised in publishing houses and large technology providers.

However, an under-explored aspect in these models is how to manage the interactions between authors and contributors of a research object in a trusted way, which is of utmost importance for computing bibliometrics transparently. Examples of these interactions *are (i) Agreement between authors on which snapshot of a working version should be released (ii) Agreement between authors on the attribution due to each of them for each release of a living research object (iii) Public comments and reviews of public releases, both as a mean to complement bibliometrics - often overlooked, yet crucial labour in academia.*

From the point of view of a single scholar that co-authors several papers with different teams, receives reviews and comments from peers, and reviews and comments research made by others, data produced by these interactions, used to measure their performance, is not only controlled by her, or a single third party, but also by many other scholars (or their trustees). Any accidental or malicious change in a data store that is out of her control might have catastrophic impact on her performance measures. This thesis advances several Semantic Web research areas, including trust management for the Semantic Web and decentralised scholarly publication. By proposing a system that uses distributed ledgers and smart contracts to manage trust in a scenario which has been long understood as a critical showcase of semantic technologies, we provide a timely contribution to an ongoing discourse on the role and future of the Web as a (re-decentralised) platform for progress and social good.

### 2.8.2 Microsoft Academic Graph and OpenAlex

Microsoft Academic Graph (MAG) provided a taxonomy of more than 100,000 research topics and was the biggest graph of its kind before its discontinuation by the parent company in 2021[43], a snapshot if its schema is given in **Figure 6** - Microsoft Academic Graph Schema - example of main entities and fields. Unfortunately, as MAG has been discontinued as of the time of writing this thesis, its alternatives have not been properly documented.

Nonetheless, in response to MAG's discontinuation, non-profit grant-funded start-up OurResearch created OpenAlex (https://openalex.org), meant as an alternative for MAG and named after the ancient Library of Alexandria. Open Alex provides a public API that can be used to carry out complex searches[44]. It is supposed to contain metadata for more than 200 million works from across thousands of institutions (Priem, Piwowar, & Orr, 2022). Unfortunately, current research on OpenAlex usefulness and comprehensiveness is scarce – the author could not find enough data to suggest that the project could be widely adopted in its current form, as there is work to be done on important API characteristics such as normalisation and disambiguation of entities, mostly authors and institutions (Priem et al., 2022).

[43] https://www.natureindex.com/news-blog/microsoft-academic-graph-discontinued-whats-next

[44] https://www.nature.com/articles/d41586-022-00138-y

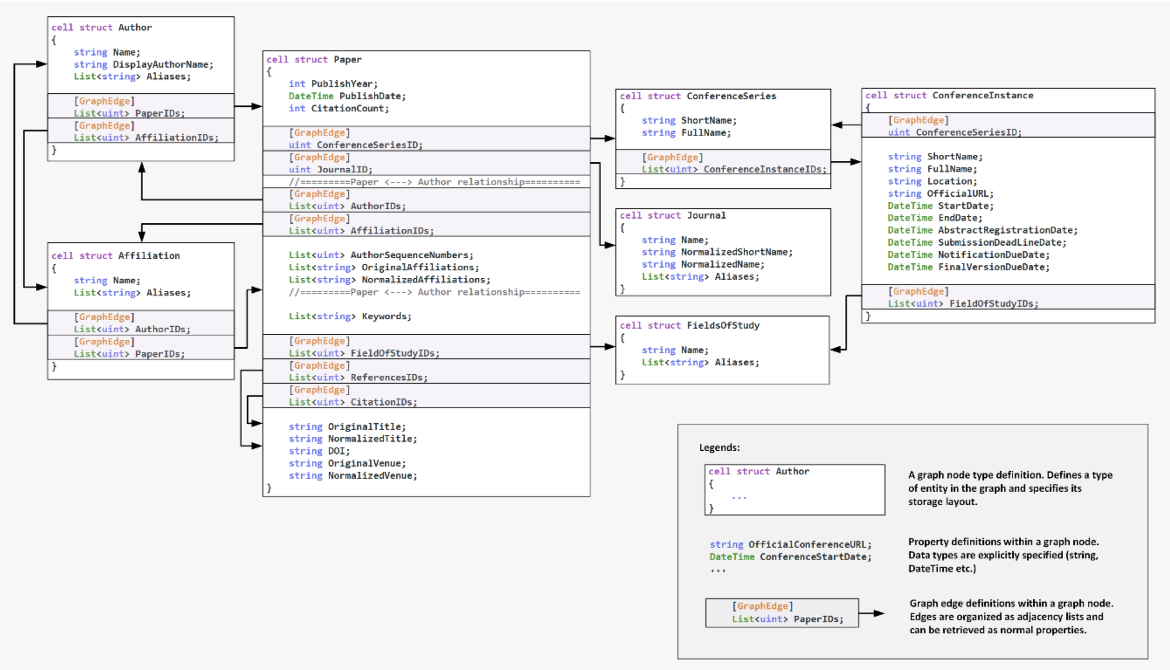



**Figure 6** - Microsoft Academic Graph Schema - example of main entities and fields

## 2.9 Blockchains

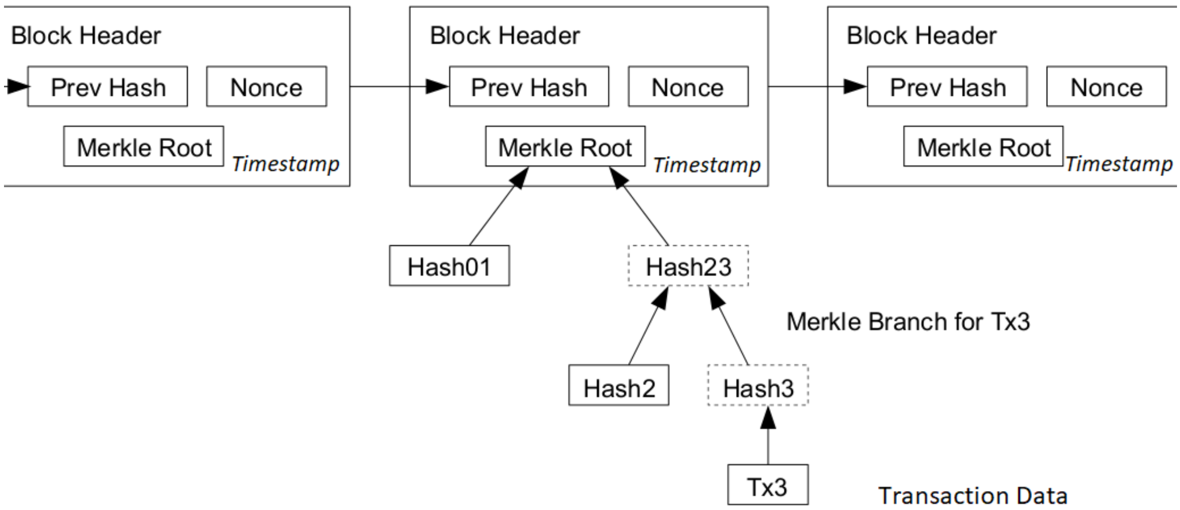



**Figure 7** - Simplified Blockchain Data Structure

Recently, distributed ledger technologies, also known as **blockchains** (Wattenhofer, 2016), have emerged as a novel decentralised solution to the problem of managing transactions of digital assets among parties that do not necessarily trust each other (or even a common third party) whilst guaranteeing the immutability and verifiability of records. Their record-keeping capabilities have been extended to user-defined programs that specify rules governing transactions, a concept known as **smart contracts**, which offer guarantees of security, tamper-resistance and absence of a central control. One of the core advantages of a distributed ledger over a database is that it helps avoid replication errors (Swanson, 2015) and delays when dealing with a huge number of parties. Blockchains are also superior for preventing forgeries and disabling unauthorised reversals of transactions. Moreover, on top of existing, "more conservative" blockchains, we now see the emergence of so-called "Layer 2" state-channel based technologies and federated chains. These new blockchain-based solutions offer more liberal models of governance and are designed for smaller, trust-minimised environments that are not as hostile as the public *mainchain* environment. The underlying data structure has been mapped in **Figure 7**, whereas **Table 2** illustrates the differences between the two currently most popular blockchains.

| | **Ethereum** | **Bitcoin** |
|---|---|---|
| **Launched** | 2015 | 2009 |
| **Consensus** | Proof of Work (merging with Proof of Stake in 2022) | Proof of Work |
| **Smart contracts** | Yes | No |
| **Block time** | Between 12 to 17 seconds | 10 minutes |
| **Main crypto** | ETH | BTC |
| **Max crypto supply** | Uncapped | 21,000,000 |
| **Number of full nodes** | 5,000 | 10,000 – 15,000 |

**Table 2** - Key differences between Bitcoin and Ethereum blockchains

In a blockchain like Bitcoin or Ethereum, the validators, who are anonymous and bound by the distributed consensus algorithm (Proof of Work, although other protocols have been explored like Proof of Stake), are part a decentralised governance mechanism. This governance mechanism relies on the exercising of the rights by blockchain developers, miners, investors and nodes, and it entails the possibility of forking the core code in case of major disagreements. These decentralised features create the *trustless* aspect in a blockchain network and without those elements, a blockchain would become an untrustworthy architecture[45]. Blockchain-generated trust is meaningful as long as the incentives of keeping the system running are larger than the one-shot value of attacking it (Budish, 2018).

[45] https://hackernoon.com/decentralized-data-why-blockchain-is-meaningless-and-trustless-is-everything-318fd14d3827

As blockchains consist of various modules, for simplicity I have grouped them into three layered classes – *Fundamentals*, *Features* and *Interfaces*, as depicted in **Figure 8** - Layered architecture of blockchains. The Fundamentals are those primitives on the system that act as building blocks on top of which more abstract Features can be built. The fundamental primitives are obviously blocks (datasets of timestamped transactions), Merkle trees (data structures for linking transactions and blocks), P2P functionalities (network discovery and network connectivity), *addressing* functionality (enabling participants to identify various accounts and contracts by means of their address), transaction functionality (enabling value to be transferred between addresses), fee-charging functionality that supports the operations of the network, and the consensus algorithm that ensures that everyone has the same copy of the data and that it is not viable to tamper with that data.

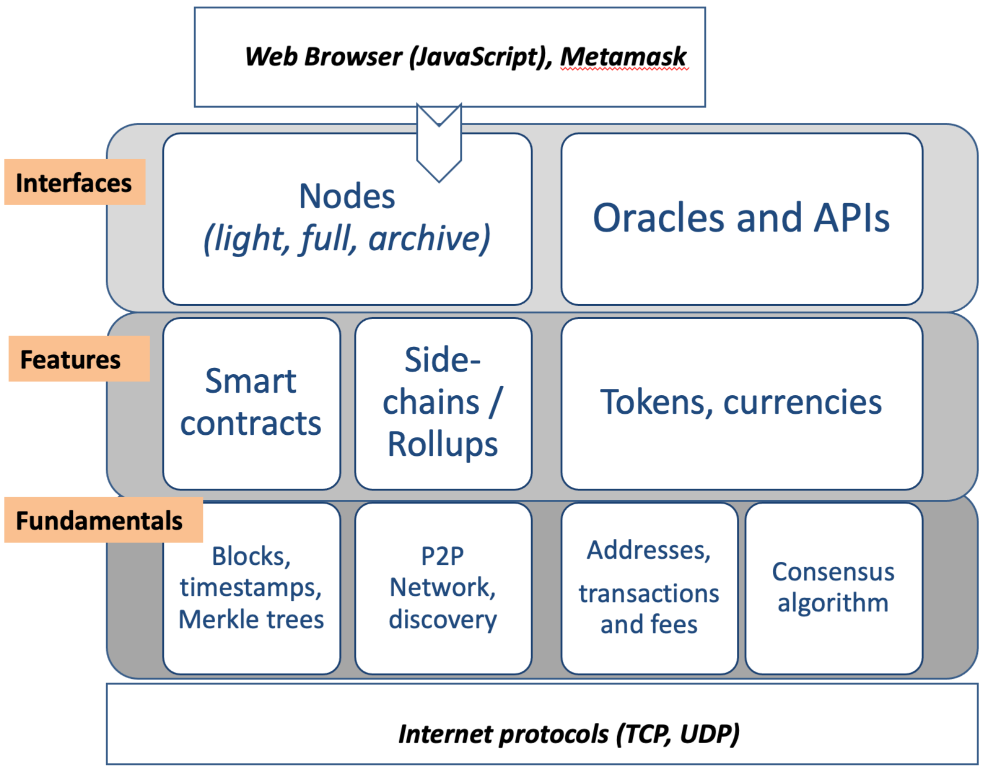



**Figure 8** - Layered architecture of blockchains

The Features built on top of those primitives include smart contracts, automated programs that can alter the state of the blockchain; side-chains and rollups which are mini-blockchains speeding up certain types of transactions (for example zk-Rollups which speed up privacy-related operations), tokens which form the so-called incentive layer, as they can be used to reward participants, and currencies which can be used to store and exchange value. At the very top of this architecture we have interfaces that enable clients to use the blockchain and its features, these interfaces include different types of nodes (light, full and archive nodes), Oracles and APIs. Some nodes can be accessed from a Web browser using browser extensions such as Metamask, which enables Web access to the blockchain.

### 2.9.1 Decentralized Oracle Networks (DONs)

As blockchain smart contracts are deterministic, and hence restricted to access data from inside the blockchain, oracles are data feeds that provide a two-way bridge between the blockchain and the outside world. Oracles post data on the blockchain, and secure the quality of that data using on-chain and off-chain components.

To avoid the problem of relying on a SSOD, oracles are typically decentralized, forming decentralized oracle networks (DONs) and a quorum is used to agree on the SSOT. An overview of the four distinctive Oracle patterns (pull-based inbound, push-based inbound, pull-based outbound and push-based outbound) is given in (Mühlberger, 2020) and shown in Figure 9.

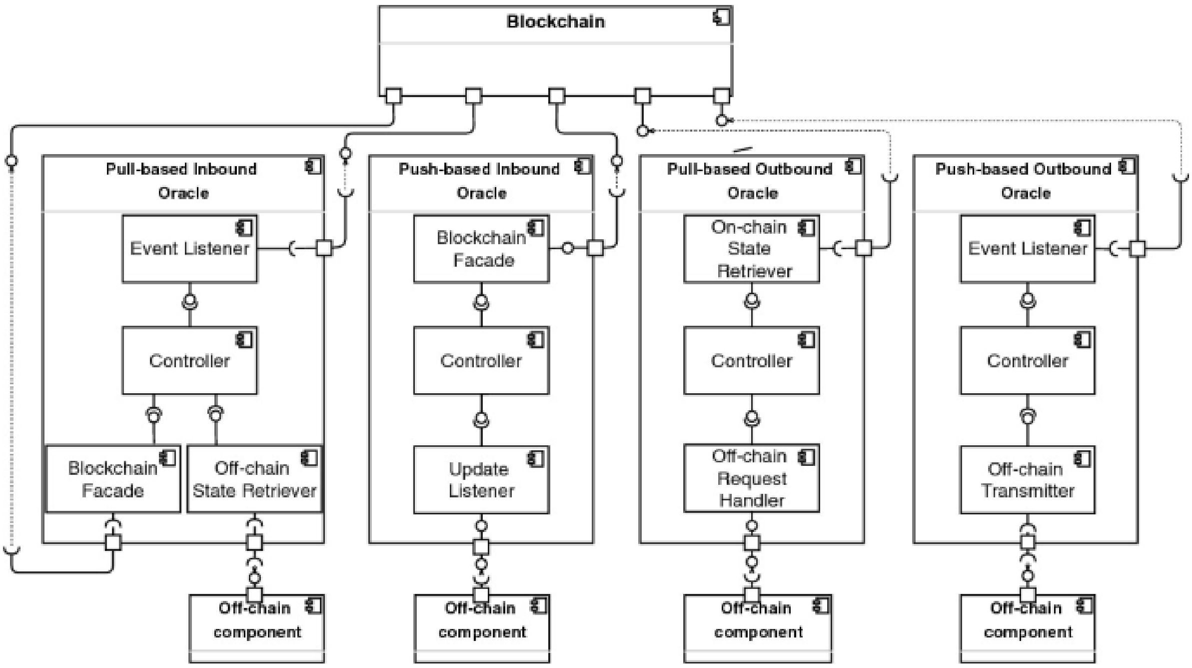



**Figure 9** - the four distinctive Oracle patterns

One of the biggest oracle networks in the market today, Chainlink[46], is composed of nodes capable of integrating with most third-party APIs. Witnet[47] is anoter popular DON, from the creators of the Aragon framework, and it provides complete verifiable proofs of data integrity (de Pedro, Levi, & Cuende, 2017).

### 2.9.2 Blockchain as a Trusted Third Party

The smart contracts (in tandem with DON's) take the place of a trusted third party in keeping records, with the critical difference being that of data and execution not being controlled by a single entity, but rather inheriting all the guarantees of the host blockchain platform. The basic unit of application-level work on a blockchain is a transaction, which is initiated when a participant sends a service call to an identifiable (addressable) smart contract that typically had been already deployed on that blockchain. If the call is not referencing an already deployed smart contract, it may deploy a new contract, that can later be re-used. The call is always signed

[46] https://chain.link
[47] https://witnet.io

by the requesting party – they can store their signature in the contract and become its owner.

Trust (or at least one particular type of "trust", cryptographic trust) in Bitcoin and Ethereum blockchains is guaranteed by the Proof of Work scheme, which solves the Byzantine Generals Problem, a situation in which, in order to avoid the collapse of the system, the system's participants must agree on a common strategy (consensus), even when some of these participants are unreliable. Bitcoin miners run a program which gives each miner a chance to discover the solution to a puzzle which involves guessing the hash of the newly mined block that is pending addition to the existing chain, even though these miners are competing with each other. The first miner to create a working solution will broadcast this solution to the whole network, and it is then verified by other miners. The 'consensus' here means that for every new block, a majority is incentivised to reach an agreement on who the winner was and tampering with historical records is not economically viable. The IT consultancy Gartner[48] further explains that "trust" in blockchain is generated by the actions of: authenticating users, validating transactions and recording the information on a tamper-proof ledger; and that all of this functionality relies on the "five key elements" of blockchain that Gartner lists as Distribution, Encryption, Immutability, Decentralization and Tokenization (see **Figure 10**).

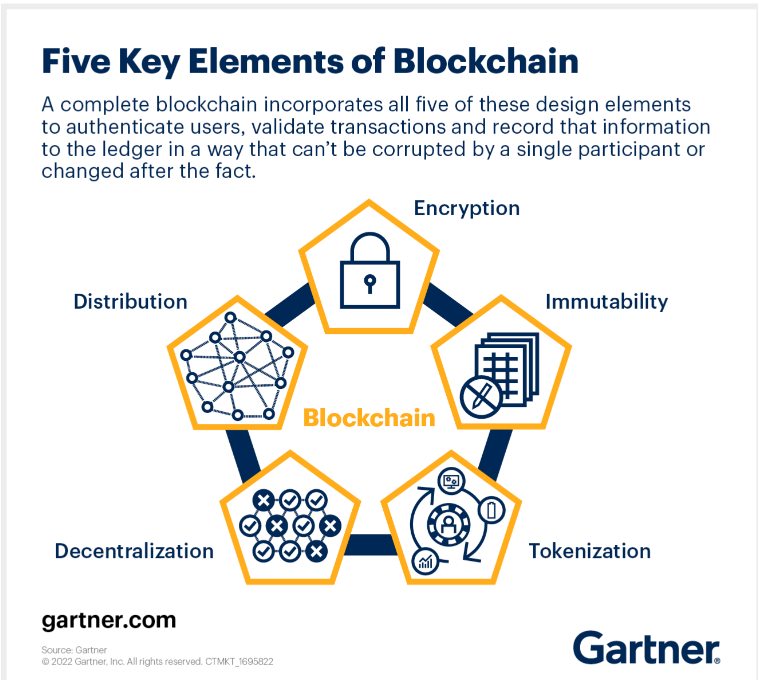



**Figure 10 -** Gartner's Five Key Elements of Blockchain (2022)

[48] https://gartner.com

### 2.9.3 Blockchains as an Evolving Technology

Blockchains are still evolving. It can be assumed that without interoperability and standards, it will be difficult for any blockchain-based solutions to truly gain traction. The rapid establishment of multiple consortiums to examine governance and standards in Blockchain development are evidence that efforts towards this goal are underway. ISO/TC 307[49] is the official ISO code for the standardisation of distributed ledger technologies. There are currently 6 standards being discussed to normalise terminology, privacy, security, identity, reference architecture and ontology. While some of these efforts have now finished, others are in moderately advanced stages. Meanwhile, a number of open-source frameworks for blockchain interoperability have emerged including Cardano[50], Aion[51], Cosmos[52], Polkadot[53] and Blocknet[54].

In the context of European Union, the General Data Protection Regulation (GDPR) issued in 2016, requires an in-depth investigation by legal experts on how it should apply to distributed ledgers (Berberich & Steiner, 2016). A better understanding is needed on what constitutes personal data and personally identifiable information on a blockchain, and what legal and technological mechanisms are required to protect it. Another practical consideration is that all transactions on blockchains are final, which is a desirable design principle, guaranteeing trust through non-repudiation. Nevertheless, mistakes are bound to happen in everyday transactions leading to disputes and reversals. A layer needs to

---

[49] https://www.iso.org/committee/6266604.html

[50] https://whycardano.com/interoperability/

[51] https://aion.network/

[52] https://cosmos.network/

[53] https://polkadot.network/

[54] https://www.blocknet.co/

be provided on top of distributed ledger data to manage conflicting edits, or when a party challenges an established record.

Finally, current blockchain implementations rely on digital signatures, which have been found vulnerable to attacks by means of quantum computers so parties with access to quantum computation would have unfair advantage in procuring mining rewards. Recently, approaches have been described for mitigating the post-quantum blockchain challenge (Kiktenko, et al., 2018), (Aggarwal, Brennen, Lee, Santha, & Tomamichel, 2017). Within this rapidly evolving environment, the practical part of this thesis builds on top of the Ethereum framework, which is becoming a de-facto blockchain standard, together with its established tooling environment. (Leydesdorff,, 2021)

### 2.9.4 Permissioned vs Permissionless Blockchains

In a **permissionless** distributed ledger such a the Ethereum platform, or the original blockchain underpinning Bitcoin, anyone can operate a node and participate in the chain through investing their CPU cycles. In some use cases, the use of heavy computation to create trust is not just redundant (e.g. when dealing with existing authorities, like tax authorities, which inherently possess trust which has been created in a legal, albeit non-computational manner), but also prohibitively expensive. As the use cases become more sophisticated, the advantage of the permissioned mode in terms of cost and throughput becomes tangible. In the permissioned model, the platform controls who is allowed to participate in the validation processes and in the protocol itself. Permissioned blockchains have a stronger notion of **identity** that can be managed by use of certificates. Furthermore, they allow grouping users according to a particular consortium, where trust is "transferred" from real-world trusted entities, such as government agencies.

### 2.9.5 Ethereum – a Permisionless Blockchain

Ethereum is an open-source, public, blockchain-based distributed computing platform featuring smart contract functionality (Wood, 2014). Ethereum blockchain was designed to be entirely deterministic. This means, that everyone should always end up with the same, correct state, if they try to replay the history of Ethereum transactions. In Ethereum, the code execution layer is provided by the Ethereum Virtual Machine (EVM), a Turing complete virtual machine that allows anyone to execute code that references and stores blockchain data. Every contract on the Ethereum blockchain has its own storage which only it can write to. When deployed, Ethereum contracts get an **address**, that can be considered like an URI in Ethereum's namespace. Using this address, a client can call functions defined in a smart contract, in a similar fashion to a web service.

Ethereum allows user to deploy smart contracts. A smart contract is a piece of computer code intended to digitally facilitate, verify, or enforce the negotiation or performance of a contract. Smart contracts allow the performance of credible transactions without third parties. These transactions are transparent and irreversible. **Figure 11** depicts a snippet of smart contract code written in Etherum's native Solidity[55] language which is a transaction-oriented (with the elements of object-oriented) high-level programming language. Solidity is statically typed, supports inheritance, libraries and complex user-defined types among other features. Early Solidity use cases involved creating contracts for voting, crowdfunding, auctions, and transactions requiring multiple signatures. We have seen how blockchain technology can coordinate participants to maintain a ledger of transactions maintaining cryptocurrency accounts. I believe that looking at the evolution of blockchain use cases, it appears that now it is time to extend the use of this technology to coordinate organisations in how they manage their resources,

[55] https://solidity.readthedocs.io/en/v0.5.7/

and this includes experimenting with blockchain for the purpose of improving scholarly communications.

#### 2.9.5.1 Privacy on a Permisionless Blockchain

As all transactions on a permissionless blockchain are public, so are the details of transacting parties (their addresses, values transferred, dates, and smart contracts used). To implement a level of privacy on top of a permissionless blockchain, custom solutions are required involving specially crafted smart contracts. For example, JP Morgan contributed an open-source library called Anonymous Zether[56], which uses “a fully decentralised zero-knowledge proof (ZKP) system” that allows users to conceal transaction amounts on Ethereum or similar account-based smart contract blockchains, allowing users to “obfuscate their identities in addition to transacted sums”. A Zether confidential transaction costs ca. 0.014 ETH or around $30 (as of April 2022).

JP Morgan is not the only company that released a privacy solution for the Ethereum mainnet, legitimising the public blockchain for business use, and driving focus away from private blockchains and Ethereum forks. EY also released Nightfall[57], another ZKP protocol, with a set of smart contracts and microservices to enable standard ERC-20 and ERC-721 tokens to be privately transacted on the Ethereum mainnet. EY believe that “public blockchains are much more secure than private ones” and reported a cost of around $10 per private transaction[58].

---

[56] https://github.com/jpmorganchase/anonymous-zether/

[57] https://github.com/EYblockchain

[58] https://www.linkedin.com/pulse/say-hello-nightfall-paul-brody-1f/

#### 2.9.5.2 Criticism of Proof-of-Work and the move to Proof-of-Stake

Proof-of-Work, currently used by Ethereum, is a distributed consensus algorithm deterring manipulation of data by introducing burden of mathematical proof that a certain amount of computational effort ("work") has been expended; used for confirming transactions, adding valid blocks to the blockchain and rewarding participation; and it expends much more energy than Proof-of-Stake, which is also a distributed consensus algorithm, that aims to deter manipulation of data by allowing participants to stake their interest in the network.

High energy consumption of the Proof-of-Work (PoW) algorithm remains an issue in broadening the applications of blockchain technology. To avoid environmental side effects of blockchain, policymakers are urged to develop standards that move away from PoW, as it cannot be made environmentally friendly even when renewables are introduced (Schinckus, 2021).

Ethereum is planning to move (as of late 2022) to a consensus mechanism called proof-of-stake (PoS) which exhibits better energy efficiency, as participants do not require a lot of electricity to be expended for mining blocks. It is also claimed to have stronger immunity to centralisation, as it should lead to more nodes in the network.[59]

### 2.9.6 Hyperledger: Fabric and Sawtooth – Permissioned Blockchains

Hyperledger Fabric[60] is an implementation of a permissioned distributed ledger platform, running smart contracts, on top of a modular pluggable architecture. The distributed ledger of the fabric is run as a peer-to-peer protocol. The blockchain supports two kinds of peers: a validating peer responsible for running consensus

---

[59] https://ethereum.org/en/developers/docs/consensus-mechanisms/pos/

[60] http://github.com/hyperledger/fabric

and maintaining the ledger; and a non-validating peer that functions as a proxy to connect clients. Some key features of Fabric are:

- it is a permissioned blockchain with immediate finality;
- runs arbitrary smart contracts (called chaincode);
- user code becomes encapsulated in Docker containers;
- pluggable consensus protocol;
- supports certificate authorities (CAs) for identity management;
- supports transaction certificates;
- persistent state using a key-value store interface;
- an event framework that supports pre-defined and custom events;
- a client SDK (Node.js) to interface with the fabric;
- support for REST APIs and CLIs.

Hyperledger Sawtooth[61] is much newer and was released at the time of writing this paper. It shares many capabilities with its Fabric predecessor, whilst also implementing the following:

- On-chain governance: smart contracts can be used to vote on blockchain configuration settings such as roles and permissions;
- Ethereum compatibility: solidity code can be deployed and executed on Sawtooth;
- Smart contracts can be programmed in Go, JavaScript and Python Smart Contracts

The idea of using permissioned blockchains relies on fine-grained access control that they provide, where not everything is public, and users can belong to different roles with different privileges, similar to traditional databases. Permissioned blockchains are easier for corporations to embrace because they can be readily integrated with corporate compliance structures. They are also more cost-effective, as validation can be done cheaper. However, they do not offer the true

[61] http://hyperledger.org/projects/sawtooth

decentralisation and trustlessness that public permissionless blockchains provide, i.e. they could be easily switched off by a corporation.

```
     */
    function decrementCitations(uint256 step) auth(DECREMENT_CITATIONS_ROLE) external {
        value = value.sub(step);
        emit DecrementCitations(msg.sender, step);
    }

    /**
     * @notice Create a new SmartPaper with an initial Version
     * @param _metadata The metadata identifying this paper
     */
    function newPaper(string _metadata) external auth(CREATE_PAPERS_ROLE) returns (uint64 paperId) {
        paperId = numPapers++;
        uint256 versionId = numVersions++;
        papers[paperId] = SmartPaper(paperId, msg.sender, new address[](0), getTimestamp64(),
                                     50*10e16, 1, _metadata, new uint256[](0), false, 0, 0);
        versions[versionId] = Version(versionId, msg.sender, new address[](0),
                                 getTimestamp64(), "InitialVersion", IpfsMultihash(0,0,0), false);
            papers[paperId].versionIds.push(versionId);
        emit CreatePaper(msg.sender, paperId);
    }

    /**
     * @notice Add a new Version to an existing Smart Paper
     * @param paperId The Smart Papers ID identifying this paper
     */
    function addVersion(uint64 paperId) returns (uint256 versionId){
```

**Figure 11** - a snippet of smart contract code written in Solidity

### 2.9.7 Identity on a Blockchain

#### 2.9.7.1 DIF and DIDs

The Decentralized Identity Foundation (DIF) was created by Microsoft, uPort Blockstack, Sovrin, and others to establish a basic common framework for making claims about identities using different kinds of decentralised protocols. Any identity is represented by it's DID (Decentralized IDentifier) consisting of

did:[method]:[method specific identifier]

e.g., *did:sovrin:3nafiQG6Cgm1GYTBaaKAgr77uY7iSerUkqX*.

Each DID has a *DID document* associated with it. containing its public keys, any public claims made by this Identity, and optional endpoints that can be used for interacting with it. A special resolver is used to lookup the DID document for a given DID. For example, uPort uses hashes in its IPFS claims registry to resolve DID documents[62].

#### 2.9.7.2 W3C Verifiable Credentials

Whereas the DIF is working on a high-level definition of Decentralised Identity, the W3C is working on standards for different kinds of claims about an Identity as part of its Verifiable Claims Working Group[63]. On 28 March 2019, a W3C Candidate Recmmendation was released for the Verifiable Credentials Data Model[64].

---

[62] https://medium.com/uport/different-approaches-to-ethereum-identity-standards-a09488347c87

[63] https://www.w3.org/2017/vc/WG/

[64] https://www.w3.org/TR/verifiable-claims-data-model

#### 2.9.7.3 ERC-725, ERC-735 and ERC-780

The ERC-725 Ethereum Identity Standard was authored by Fabian Vogelsteller, creator of the ERC 20 standard and Web3.js. ERC 725 describes proxy smart contracts that can be managed by multiple keys and other smart contracts. Self-sovereign Identity Keys (ERC725) is essentially an access control list for Ethereum addresses with four defined purposes: Management (setting up and modifying identity), Action (calling other contracts), Claim Signing and Encryption.

ERC 735 is an associated standard to add and remove claims to an ERC 725 identity smart contract. While ERC 725 does support different kinds of signature schemes, the burden of verification of these signatures is placed on 'claim consumers'. The vision for ERC725 + ERC735 is to become an open, portable standard for identities that will enable decentralized reputation, governance, and more. Users will be able to migrate their identities across different Dapps and platforms.

A competing standard, the uPort approach as represented by ERC 780, is called the Ethereum Claims Registry. It uses shared common contracts that are not controlled by anyone and can thus be thought of as a form of common good. Technically, it's an immutable key-value store where the identity of the writer is public, and claims have subjects. A claim has 4 fields (Issuer, Subject, Key, Value). There is a strong off-chain component in uPort for privately sharing claims that should not be public.

### 2.9.9 Decentralised Autonomous Organisations

A decentralized autonomous organization (DAO) is a form and technique for organising financial and decision-making processes in a group of people bound by rules that are encoded as open-source smart contracts, where decisions are made democratically by stakeholders without relying on governmental or corporate coordination mechanisms (McGregor-Lowndes, 2019). A DAO's budget, membership and voting rules are thus maintained on a blockchain. Decentralized autonomous organizations aim to be open platforms where individuals control their identities and their personal data, and are incentivised to take interest in the DAO's activity and able to influence its future. The vision for DAOs as opposed to traditional top-down power structures is shown in **Figure 12**. This figure illustrates the importance of peer-to-peer value transfers between users, miners and exchanges in a DAO model.

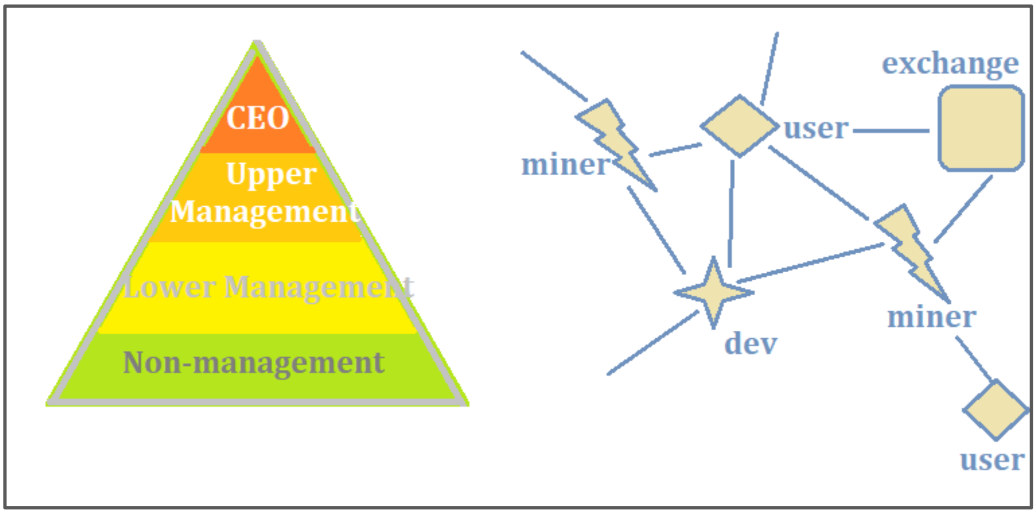



**Figure 12** - Traditional top-down organisation compared to a DAO

Daniel Larimer first proposed the concept of a "Decentralized Organized Company" in an article published on September 7, 2013. Shortly thereafter, Vitalik Buterin proposed that DAO's can be created on top of his Ethereum blockchain, and that after a DAO was launched, it might be organized to run without human managerial oversight, provided that the DAO's smart contracts contain sufficient business and managerial logic. However, DAOs suffer from a unique combination of governance issues due to a low barrier to entry and the difficulty in obtaining legal recourse in traditional courts. One major consideration is that it is not unreasonable for a minority stakeholder to be concerned that a majority might decide to liquidate the DAO and exclude minority stakeholders. To avoid this scenario, a DAO needs constraint-imposing mechanisms, and ideally, a whole new dispute-resolution ecosystem that allows on-chain arbitration by default but also enables such a minority stakeholder to raise a dispute off-chain (e.g., in a traditional court of law) if necessary. To be able to implement this, DAOs must be defined in human-readable terms, and not just as computer code.

Another point requiring attention is that if everyone in the DAO is following just their own personal incentives, the system risks degenerating. This is why some Web researchers have proposed that DAOs still need some form of top-down control in the form of "unincentivized incentivizers" to strengthen their governance. The difficulty of achieving strong DAO governance has been illustrated by how DAOs have consistently received a lot of bad press due to abject governance failures, starting with the DAO Hack in June 2016 which saw millions of dollars stolen from Ethereum users by hackers exploiting smart contract vulnerabilities in the original Ethereum DAO. Even more recently, MakerDAO, which is also an Ethereum-based DAO, has been facing liquidation due to under-collateralization triggered by Ether volatility related to wider market volatility during the early 2020 financial crisis. At the time of writing this thesis, MakerDAO was facing complete shutdown due to a lack of "crisis governance" processes which were never built into its governance design in the first place (Barrera, 2020) and this failure resulted in a long period of chaos among the participants of MakerDAO.

Nevertheless, DAOs present tremendous scope for innovation and experimentation. They can be used in tandem with decentralized identities (DI), which are currently under active research, but some proof-of-concept work is already being promoted by the Decentralized Identity Foundation (created by Microsoft, uPort Blockstack, Sovrin, and others), the goal of which is to establish a basic common framework for making claims about identities without the need for a central verifying authority. In such an approach, a decentralized identity is represented by its DID (Decentralized Identifier) that is linked to the public keys that are associated with it, and each such identity can be strongly associated with any public claims made by that identity online. DAOs and DIDs are examples of blockchain-based building blocks that can be used in decentralization initiatives, such as decentralized finance ("DeFi"). A robust online financial ecosystem based on the blockchain has now evolved, in which one can find blockchain-based financial intermediaries ranging from decentralized exchanges to "algorithmic central banks." Given the variety of technological propositions on offer that claim to be decentralized, from identity to finance, is blockchain really a viable vehicle for decentralization, and if so, then what sort of decentralization is it? Knowing that it was the loss of trust in centralized entities that gave birth to blockchain in the first place, it is to be expected that blockchain decentralization will not be purely technological in scope, but rather human-centric, transforming psycho-social aspects related to trust, identity management and user experience (Gaggioli et al., 2019).

#### 2.9.9.1 Aragon DAO Framework

Aragon DAOs manifest themselves as dual Web-blockchain entities with a secure connection linking the browser to the blockchain. Aragon DAOs specify which addresses have access to perform actions on behalf of the organisation in a permission registry called the Access Control List (ACL). Blockchain addresses that belong to this registry can be externally owned accounts or smart contracts. Some contracts are intended to forward actions based on pre-defined criteria, for example a voting app will forward action only after a successful approval vote was proposed and executed. By chaining multiple contracts together we can define complex criteria which direct how actions are performed within Aragon organisations. For example, a DAO may want to allow royalties to be transferred, but only if the transfer is 1) proposed by a member of the DAO, 2) approved by a majority of members, and 3) within a pre-determined budget. This can be accomplished by configuring a chain of permissions with each link imposing logical constraints on the final action, and implemented through individual "Apps" (Token Manager, Voting App, Finance App, Vault) that are nothing but smart contracts tightly coupled with tailored JavaScript and CSS/Web functionality. The Vault App, responsible for storing the organisation's assets, may grant the transfer role only to the Finance application, which implements budgeting logic, etc.

*N.B.* It's not unreasonable for a minority stakeholder to be concerned that a majority might decide to liquidate the DAO and exclude minority stakeholders. To avoid this scenario, a DAO needs a mechanism to impose a constraint that can be enforced by the actions of any individual rather than a majority. Proposal Agreements are designed towards this end. They enable a DAO to define human-readable terms that proposals must conform to and require collateral before proposal can be forwarded to a voting app. These agreements can be coupled with a Voting app by assigning the create vote, pause vote, and cancel vote roles. If a minority stakeholder feels that a submitted proposal has violated the terms of the proposal agreement then they can raise a dispute.

#### 2.9.9.2 Moloch DAO, UkraineDAO and other DAOs

Moloch DAO[65] was a new type of organisation, developed by Ameen Soleimani, Arjun Bhuptani, James Young, Layne Haber & Rahul Sethuram[66] that focuses on grant-funding and redefines the concept of a firm. Unlike traditional organisations, the shares are not used to continuously control the capital of an overgrown organisation. Instead, the only way to get access to the capital is through destroying the shares along with any appertaining rights (e.g., liquidating 5% of the stock will yield 5% of the capital, and there is no other way of redistributing the underlying resources). The name Moloch, a biblical god associated with sacrifice, was chosen as it can be loosely understood as personifying societal failure due to coordination problems[67] (individuals optimizing things, resulting in a grossly suboptimal overall outcome), and was previously used in this context by Alan Ginsberg[68] and more recently in Scott Alexander's "Meditations on Moloch"[69], where Alexander notes:

> *no individual journal has an incentive to unilaterally switch to early registration and publishing negative results, since it would just mean their results are less interesting than that other journal who only publishes ground-breaking discoveries. From within the system, everyone is following their own incentives and will continue to do so.*

Alexander argues that coordination through an unincentivized incentiviser the one thing that can prevent a complex social system from degenerating into chaos.

In technical terms, Moloch is based on just 2 Solidity smart contracts[70] and operates through the submission, voting on, and processing of a series of membership proposals. To combat spam, new membership proposals can only be submitted by existing members and require a deposit. It is proposed that Moloch DAO incentivises coordination, through synthesising traditionally separate parts of a company into one workflow, and by creating additional incentives for disgruntled members to exit (thus strenghtening core commitment). The incentives can be managed in terms of any ERC-20 tokens (these are explained in the next section).

---

65 https://molochdao.com/
66 https://medium.com/@simondlr/the-moloch-dao-collapsing-the-firm-2a800b3aa2e7
67 https://twitter.com/ameensol/status/1084652405285736449
68 https://www.poetryfoundation.org/poems/49303/howl
69 https://slatestarcodex.com/2014/07/30/meditations-on-moloch/
70 https://github.com/molochventures/moloch

The UkraineDAO[71] raises money for Ukrainian defence in the 2022 military conflict with Russia and has been described as “the largest crypto contribution to Ukraine’s war efforts to date,”[72] created by activists, hackers, musicians and artists including PleasrDAO[73] with a goal to help Ukrainians immediately. A snapshot on March 3rd at 6 pm of the initial donations (around $6m in 72 hours) was converted into the $LOVE token, commemorating the donors’ contributions. The creators motivate the new NFT with the following claim:

> *Seeing these tokens in one’s wallet reminds people of the bigger picture behind Web3 building and decentralized organizations[74].*

The DAO presently plans to distribute 1000 $LOVE tokens (or a fraction thereof depending on donation size) for each ETH donated and encourage people to hold $LOVE in their crypto wallets “as a reminder of our world’s ongoing humanitarian needs”. One of the founders of the DAO, of Trippy Labs fame, also states:

> *“It is apparent a revolution cannot be started with fiat as there are too many ways for traditional funds to be intercepted or halted by traditional institutions, governments, and other factions with intent to control, harm, or simply shift funds without transparency. Decentralized Autonomous Organizations are a model for the world to witness how people with a common purpose can join and work together quickly to distribute funds to present humanitarian causes.”*
> ~Trippy[75]

Other noteworthy DAOs that are currently active include:

> MakerDAO, PolkaDAO, dxDAO, DigixDAO, metacartel, humanityDAO, DAOstack, BlankDAO, MyBit, Dash DAO, Identity DAO, Decred DAO, Bisq DAO, SikkaDAO, MovementDAO and Humanity DAO[76] which has an ambitious goal of providing every registered user with universal basic income.

---

[71] https://www.ukrainedao.love/ukraine-dao
[72] https://edition.cnn.com/style/article/ukrainedao-pussy-riot-nft-flag-war-fundraising/index.html
[73] https://pleasr.org/
[74] https://app.newscrypto.io/ru/markets/ukrainedao-flag-nft/overview
[75] https://mobile.twitter.com/trippylabs
[76] https://www.humanitydao.org/

### 2.9.10 On When Blockchains Are Needed

The applicability of blockchains to relevant use cases has become a hot topic recently among business consultancies and technology firms. There has been growing expertise based on real-world experience among blockchain practitioners, and informative frameworks have emerged that can guide users towards the right solution for their use case, answering the question on whether blockchains are indeed needed to solve the problem at hand. Among alternatives to blockchains are other types of data stores, such as email servers, shared drives, and databases (including managed and encrypted databases). The framework (**Figure 13**) developed by the US Department of Homeland Security and popularised by the Center for Global Development provides a flowchart that can help the user decide on whether blockchains fit their requirements.

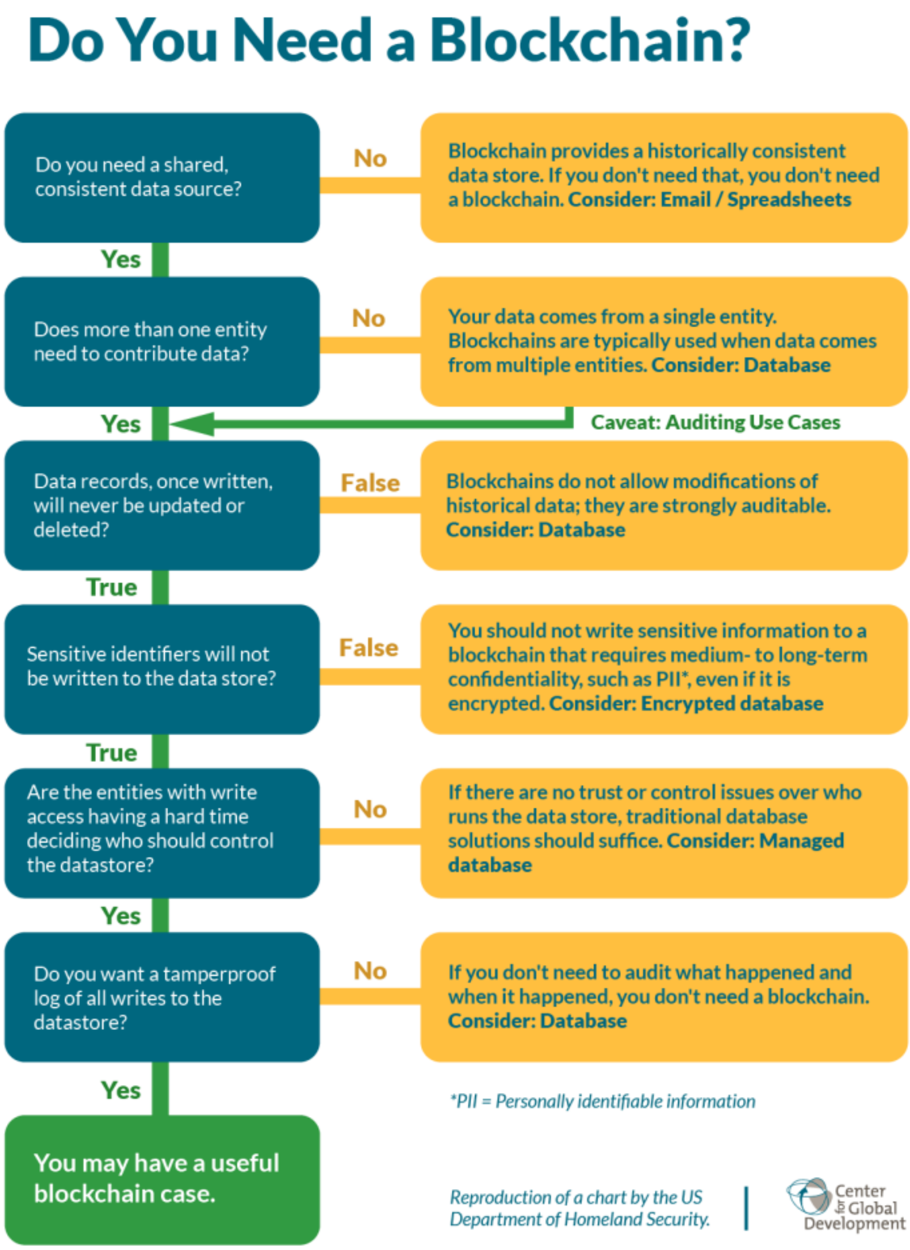



**Figure 13** – The Framework for Blockchain Suitability Evaluation

## 2.10 Cryptographic Tokens


**Figure 14** - Non-fungible (left) vs fungible (right) classes

As alluded to in the previous section, designing a successful cryptographic incentive must take into account certain aspects of monetary theory, financial economics, and game theory. Cryptographic tokens can also have strictly controlled supply mechanisms and complex real-world relationships, thus facilitating coordination among stakeholders when network effects are present, i.e. they can be coded in a way that links their value to the number of token holders and the number of total tokens in circulation. A token is fungible if its individual units are essentially interchangeable, and each of its parts is indistinguishable from another part. A non-fungible token represents a unique entity (or ownership of a unique physical world item), their main goal is to create verifiable digital scarcity.

### 2.10.1 Fungible token standards

Tokens that are fungible behave like currencies. Fungibility means that one instance of a particular class of token is virtually indistinguishable from another instance of the same class. This concept is illustrated in **Figure 14**.

#### 2.10.1.1 ERC-20

**ERC-20** is a class of identical tokens that can be issued and exchanged by smart contracts on the Ethereum blockchain. It is a standard for implementing money (currencies) and currency-like tokens. In fact, most tokens issued on the Ethereum blockchain are ERC-20 compliant and almost 200,000 different types are currently present on the main network.

All **ERC-20** tokens are purported to be fungible, so that one is never expected to add a history, provenance, or identity to any **ERC-20** token. This fungibility means that there are very limited ways of adding metadata to **ERC-20** tokens, a strictly controlled mechanism that differs from Bitcoin's coloured coins approach. The Ethereum **ERC-20** token standard became popular with crowdfunding companies working on initial coin offerings due to the simplicity of deployment. The most successful **ERC-20** tokens include the **EOS ERC-20** (now frozen and mostly swapped to mainnet EOS) and Tezos.

#### 2.10.1.2 ERC-223 (upgraded ERC-20)

**ERC-223** is a backwards-compatible upgrade to the **ERC-20** token standard. It eliminates the problem of lost tokens which happens during accidental transfers of ERC-20 tokens to contract addresses when people mistakenly use the instructions for sending tokens to a wallet. The **ERC-223** specification allows users to send their tokens to either wallets or contracts within predefined protocols, thereby eliminating the potential for confusion and lost tokens.

This upgraded standard also allows Ethereum developers to handle incoming token transactions in a way that rejects non-supported tokens. In this case, you won't lose the unsupported tokens as they will be automatically refunded back to you, minus the gas (the cost of processing your instructions), something that was not possible within the **ERC-20** framework.

Finally, this new protocol offers significant energy savings. The transfer of **ERC-223** tokens to a contract is a one-step process rather than 2 step process (as was the case for **ERC-20**), and this means two times less gas and no extra blockchain bloating (consider that in a blockchain every party gets the copy of every transaction incurring additional network congestion). This, as a result, also lowers the transaction fees one pays for the transfer of tokens.

#### 2.10.1.3 ERC-621 and other extensions

The **ERC-621** token standard was proposed in early 2017 and has since gained some momentum as an extension of the **ERC-20** token standard that caters for some particular niche applications. This specification adds two additional contractual functions that enable token governance to easily increase and decrease the total supply of tokens in circulation, as contrasted to **ERC-20** which only allows a single token issuance event which restricts the supply to a certain amount which can't be changed. Since **ERC-621** proposes that the total token supply can be modified, this standard has proven useful in implementing certain fiat pegged tokens (tokens representing the value of a real-world currency, for example the British Pound).

Another extension of the **ERC-20** standard is **ERC-827**. It allows for the transfer of tokens and allows tokens to be approved by the holder to be spent by a third party. Tokens on Ethereum can be reused by other applications, including wallets and exchanges. This could be very useful for spending a dynamic amount that is up to a third party based on some criteria both parties have agreed to.

### 2.10.2 Non-fungible Token Standards

Non-fungible tokens are used to create tokens that differ from each other. This can be used as a mechanism to create verifiable digital scarcity meaning a resource with limited supply of its different variants, with some variants potentially more desirable than others.

#### 2.10.2.1 ERC-721

Ethereum community created an open standard for issuing non-fungible tokens (NFT), called **ERC-721**. Introduced in 2017 (finalised in 2018), the standard exhibits some important properties for conforming tokens:

- Token cannot be divided or combined.
- Token can only belong to a to a physical address - to an account - either user's wallet or another smart contract; each token can thus have one (and only one) owner;
- Tokens, minted after contract creation, must follow a special protocol for any transfer of ownership, ensuring the safety of the transfer.

To implement that standard, we can use a so-called first-party solution (the **ERC-721** minting contract performs the sale) or a third-party solution (another contract performs the sale.) Non-fungibles have since become actively used for varying use cases, and evolved in many directions.

#### 2.10.2.2 Non-Securities Non-Fungibles

**ERC-994** was subsequently introduced as an extension of **ERC-721**, and quickly dubbed "Delegated NFTs". This upgrade was specifically designed with the use case of Ethereum-based registration of land and physical property in mind. Delegated **ERC-994** NFTs (DNFTs) are arranged in a federated, tree-like format, similar to a domain name system, where each NFT can "delegate", or sub-contract other NFTs within a certain space ("zone"). Unlike many digital assets, like currencies or collectibles, physical property needs to be valid within the context of the physical scheme that governs it. Thus, DNFT "zones" can be established by different land registry authorities around the world, through enabling the creation by each such authority of a root-level Delegated Non-Fungible Token encompassing a wide area. Children DNFTs can then be created as subdivisions of this root.

**ERC-998** was subsequently introduced as an extension of the **ERC-721** and became the "Composable NFT Standard". This has found its use in massive multiplayer online games, allowing players to not just purchase individual items (like villages) but to grow collections of them (like empires)—all through a single token of ownership. In one such game, in-game characters can be composed of all of its underlying NFTs: shield, sword, boots, special items, and even other **ERC-20** tokens. When you are ready to sell or trade the character, it takes just one blockchain transaction, after which all underlying assets belong to the new owner. The advantage of composable NFTs is thus reduced transaction costs.

**ERC-948** is a recently proposed standard protocol for Subscription Services on the blockchain. Under this scheme, the user can create a new subscription on a blockchain smart contract, permitting "x" tokens to be withdrawn from his/her wallet every "y" time period by "z" Service until this user cancels the subscription.

The massive growth of the non-fungible token industry, as illustrated by the multitude of the standards, of which the main ones were described above, has also led to the emergence of blockchain securities that have since received legal recognition. These are described in the next subsection

### 2.10.3 Securities Non-Fungibles

The promise of implementing securities on the blockchain has generated discussion about a lot of potential benefits, such as reduced costs, automated compliance, rapid settlement, increased transparency, better liquidity and more. A security token offering (STO) is an offering of traditional securities in a digital token format in order to raise funds. Competing standards have emerged, with various degrees of recognition worldwide. A Security Token shares many of the characteristics of both fungible and non-fungible tokens. In particular, security tokens are designed to represent complete or fractional interests in assets and/or entities (“having a stake").

An early standard, **ERC-884**, or “tokenised shares", takes advantage of a recent Delaware blockchain-friendly Senate Bill (n. 69). The **ERC-884** is a Ethereum token spec allowing any Delaware corporate entity to use a smart contract to create and maintain an official share register on the Ethereum blockchain. Essentially, this is a legally compliant standard for tokenized equity. Successful **ERC-884** are SEC approved and can be traded on traditional financial markets as securities. However, in order to comply with securities laws, issuers of **ERC-884** must also maintain an off-chain private database which makes it more of a hybrid approach.

The **ERC-1400** standard introduced the concept of a partially fungible token that provides transparency over the partitions of a token holder’s balance that may be treated differently by the security token for the purposes of transfer restrictions. The term “tranches" is used to describe these partitions. Since its introduction, the single ERC-1400 standard has developed into a suite of several standards with separate specialisations:

- The **ERC-1594,** which was designed to provide the core functionality needed for all security tokens;
- with the **ERC-1410**, a user’s balance can be divided to accurately reflect the different specifications that come along with token ownership;
- the **ERC-1644** - a method for controlling and managing security tokens;
- last but not least, **ERC-1643** - a method for document management.

The idea behind this suite of solutions is to propose a common framework so that investors, issuers KYC suppliers, exchanges and wallets can work under the same conditions, increasing the democratisation of securities in the digital world (just like the **ERC-20** has democratised utility tokens).

Out of non-Ethereum ecosystems, notable is **SRC-20**, a standard developed by Swarm Fund, an asset tokenisation platform that runs on a utility token (SWM). The trading of **SRC20** tokens also occurs on the private blockchain to ensure that Swarm can monitor trades and ensure compliance. To address the concern about the lack of common blockchain standards for regulatory purposes, Swarm have created an interoperability-focused security token protocol called Market Access Protocol (MAP) that acts as a tool to determine whether a wallet is compliant with securities regulations, and only allows transactions to occur if compliance is verified.

### 2.10.4 Refungibles and Token Bonding Curves

Token holders could also hold a quantity of fungible tokens that represent in some way the original non fungible token. This structure has been dubbed "Refungibles" and potentially interesting applications were described in curation markets for art, Intellectual Property (for example, big pharma and other types of innovation) and digitally paywalled content. Of particular interest, first developed by Simon de La Rouviere[77], is the innovative proposal for a bonding (or bonded) curve that represents the price per token as a function of to the number of tokens in circulation (x-axis) by a predefined formula.

The function (slope) can be linear, exponential, logarithmic or arbitrary, allowing the token governance to control how the token price increases with the number of tokens in circulation. The value derived from curved bonding is that

[77] https://medium.com/@simondlr/tokens-2-0-curved-token-bonding-in-curation-markets-1764a2e0bee5

rewards participants for buying tokens at an early stage and encourages them to participate in curation activities. Several projects have started to integrate bonding curves into their tokens, for example, Ocean Protocol[78], a decentralised data exchange protocol aimed at providing an ecosystem for the data economy and associated services aimed at curating and monetising data. Data and service providers publish their services in the platform, other actors can decide to become servers of data or executors of algorithms and services for a fee, and consumers can buy them in a decentralised environment. From a network perspective, the community is interested in maximising the number of relevant AI and data services. But how to decide on relevancy? This is where bonding curves become useful, as each dataset or service is assigned its own bonded token (called a drop). Drops can be acquired by users and servers, representing a stake on the value of datasets and services, and the expectation is that users will be incentivised to find and stake for the most useful services, that will eventually prevail. The bonding curve manages the relationship between the token supply, the underlying data/service, and the price of the token, making speculation difficult, and incentivising high-quality curation.

[78] https://blog.oceanprotocol.com/introducing-the-equilibrium-bonding-market-e7db528e0eff

## 2.11 The Blockchain Hype and the Fallacies of Blockchain

There has been much criticism of blockchain from those sceptical about its usefulness and suspicious about its popularity. Bloomberg Businessweek, in 2018[79], called it a "mind virus", seemingly quoting from Richard Dawkins' 1991-1992 *Viruses of the Mind* essay (Dawkins, 1993), where the author discredits certain types of faith-based thought as propagating in the fashion of a "memetic" (imitable) virus, analogous to how biological and computer viruses spread, in a seemingly capricious and arbitrary way. Dawkins further distinguished the pointless and self-serving mind-virus propagation process from how noteworthy scientific ideas spread - which requires testability, evidential support, and so on. In the same year, Hanna Halaburda claimed in a widely cited paper (Halaburda, 2018) that *"most of the suggested benefits of blockchain technologies do not come from elements unique to the blockchain,"* noting also how much of the blockchain hype seems to be driven by the over-estimation of the usefulness of eliminating the middleman, while underestimating the gateway problem (how data makes its way to the blockchain in the first place). Halaburda famously wrote that the blockchain revolution may not require blockchain, essentially attempting to discredit the value of this emerging technology.

The mind-virus narrative can be, however, refuted upon dissecting it – blockchain technology not only exists in many different shapes, but it also does have many applications from crypto-currencies to DAO's, as outlined by the evidence presented in the present section of this thesis. These applications are also testable, we know for example how much the Ethereum market is worth[80]. However, could a pre-conceived notion of "blockchains' uniqueness" be a memetic virus tainted by Halaburda's sentiment that the value added by the blockchain comes from elements not actually unique to it? First of all, Halaburda's essay relied on her qualitative argument whose power was weakened by the use of tentative language; and it further suffered from not having any of the described impacts quantified, let alone tested. Secondly, even if Halaburda's thesis was to be taken at face value, it can still be refuted, as there are papers confirming the value added by specific blockchains, seen as unique technologies, as opposed to their individual elements (such as timestamping or smart contracting) - examples include UkraineDAO "raising millions dollars in support of Ukrainian fighters" (Nabben & Rennie, 2022), and OpenSea, the world's

---

[79] https://www.magzter.com/stories/Business/Bloomberg-Businessweek/The-Blockchain-Is-A-Mind-Virus [Accessed August 2022]
[80] $196,614,686,887 *from* https://coinmarketcap.com/currencies/ethereum/ [Accessed August 2022]

“largest NFT marketplace” with over 160,000 weekly traders[81]. Furthermore, many reports of blockchain usefulness need to be considered that come from outside academia, for example Accenture’s analysis showing that blockchain adoption is likely to reduce investment banks’ infrastructure costs by 30 percent[82], or the McKinsey research that estimates blockchains applied to cross-border payments saving banks ca. $4 billion a year [83].

It needs to be mentioned here that the year 2018 saw such heavy critique of blockchain, as it was the lowest point of the so-called Gartner Hype Cycle for blockchain technologies[84], entering the phase of the so-called *trough of disillusionment* according to Gartner, the influential technology consultancy. The hype cycle is part of Gartner’s research methodology for adopting emerging technologies and is assumed by Gartner to be a natural phenomenon resulting from the failing expectations over time and is typically seen for any emerging technology according to the temporal pattern shown in **Figure 15:**

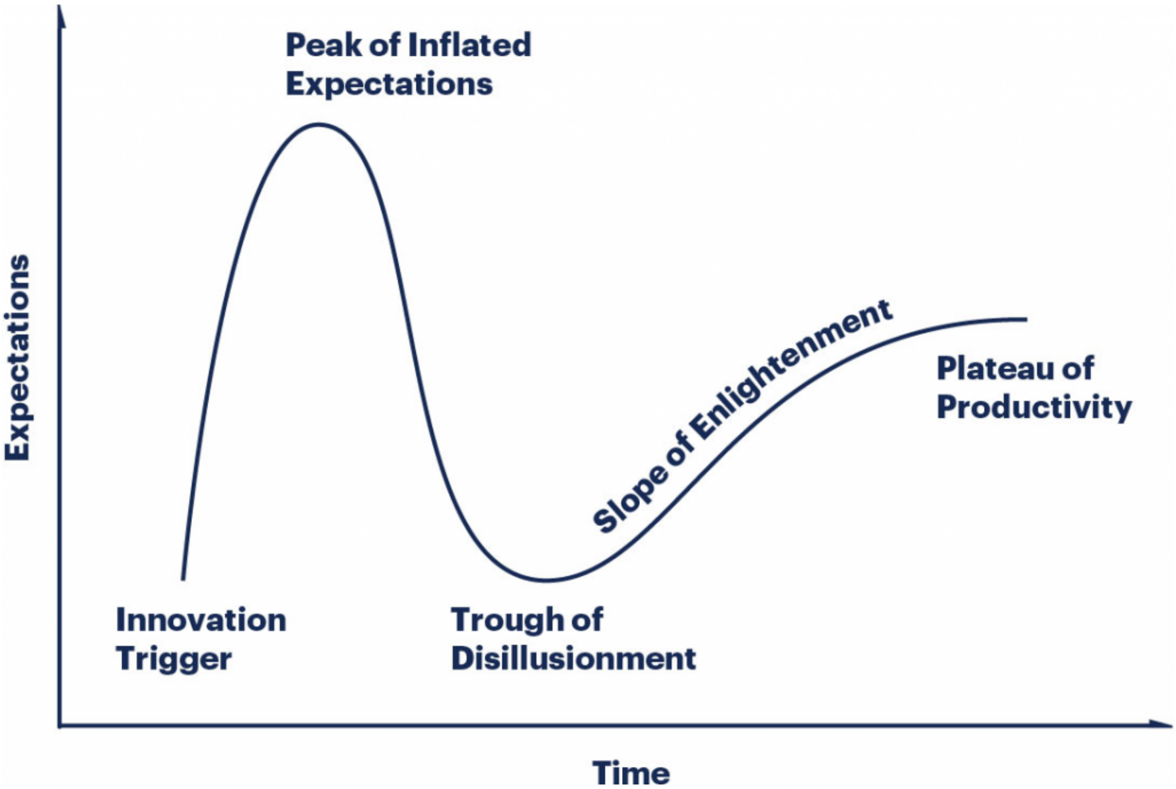



**Figure 15** - Gartner Hype Cycle for Emerging Technologies

[81] https://dappradar.com/multichain/marketplaces/opensea [Accessed August 2022]
[82] https://newsroom.accenture.com/news/blockchain-technology-could-reduce-investment-banks-infrastructure-costs-by-30-percent-according-to-accenture-report.htm [Accessed August 2022]
[83] https://www.mckinsey.com/industries/financial-services/our-insights/blockchain-and-retail-banking-making-the-connection [Accessed August 2022]
[84] https://www.ledgerinsights.com/gartner-blockchain-hype-cycle/ [Accessed August 2022]

Nevertheless, some more recent critiques of the blockchain and its applicability to different use cases have been performed more rigorously. Specifically, the authors Hawlitschek, Notheisen and Teubner (2020) wrote about the three concrete fallacies that may lead to misapplying blockchains:

- **trust-free fallacy** – *"blockchain technology does not inherently solve all trust issues when it comes to real-world interactions between human beings"*
- **disintermediation fallacy** – *"the role and importance of platforms as the broker and curator must not be neglected, but often is"*
- **consumer will fallacy** – *"unlike tech enthusiasts, main street consumers do not put much weight on the underlying technology".*

Defining these fallacies sheds light on what blockchains don't do very well despite the early promises of their proponents – they do not remove the need for human trust, they do not remove all intermediaries, and they do not make sense to the general public. But blockchains are still useful despite that - many corporations have been announcing recently that their blockchain-based products have been contributing to their bottom lines, e.g. the global financial services firm J.P. Morgan, who reported that it is *"using blockchain technology to improve funds transfers between banking institutions globally, including payments originating from Taiwan banks to beneficiary banks in other markets"*[85], or EY (consultancy), whose blockchain offerings include supply chain management for "*product traceability and inventory management,"* incorporating a privacy solution on a public blockchain[86].

---

[85] https://www.jpmorgan.com/news/jpmorgan-uses-blockchain-technology-to-help-improve-money-transfers [Accessed August 2022]

[86] https://www.ledgerinsights.com/ey-simplifies-public-blockchain-transactions-with-api-launch/

Knowing which fallacies to avoid allows us to refocus on the real strengths of blockchain applications. It is my opinion that the value added by blockchain platforms really resides not just in the technical features but also in the fact that the blockchain communities, similar to Linux communities, are strong believers in open source[87]. Blockchains don't remove the need for trust, but they can re-focus us to trust open-source smart contracts over closed-source corporate implementations. Blockchains don't remove intermediaries, but they can introduce coded intermediaries that can be audited, and whose behaviour may be more easily predicted than closed-source ones. Blockchains won't be understandable by the average user, but they will be valued by so-called power users, passionate about technology, often visible in open-source communities. Finally, the value apparently lost by the gateway problem according to Halaburda (loss of trust at the blockchain point of entry), can be addressees by new research into Nash equilibria for truthful reporting of data through crowd-sourced oracles (third-party data sources that attest that the data put on the blockchain is true), where reporters, or certifiers, have a monetary stake when they post the data onto the blockchain - see (Nelaturu, 2020) – which incentivises them to be trustworthy when putting data on-chain.

[Accessed August 2022]

[87] https://opensource.com/article/20/10/open-source-blockchain [Accessed August 2022]

## 2.12 Related Work on Blockchain use in Academia

There have been several initiatives put forward to utilise the promise of blockchains and distributed ledger technologies in modernising the operations of science. Whilst nearly all of such work focuses on the narrative of publishing currently being an "exploitative system" (Teixeira da Silva & Dobránszki, 2015), currently many researchers focus only on registering research artifacts online and improving the peer review system, highlighting the need to increase transparency, and stamp out fake peer review. The most comprehensive proposal for blockchain in peer review was drafted by the late, learned, Jon Tennant, a pioneer of Open Science, and an inspiration to many, who died tragically at the time my thesis was being written. His blog[88] presents a vision adapted from a paper published by a team of 30 researchers in F1000 Research (2017), "*A multi-disciplinary perspective on emergent and future innovations in peer review*". Among other things, Tennant proposes rewarding reviewers with Science Coins. Meanwhile, (Jan, et al., 2018) proposed a digital cryptographic token, ScienceMiles, to incentivise peer reviewers towards better quality reviews and to increase reviewer's recognition. As a pioneering solution, ScienceMiles introduces the much needed notarisation of peer review artefacts and providing reviewers with cryptographic tokens, but it lacks clear governance and privacy-enabling mechanisms, such as the ones that I develop in Chapter 5 of my Thesis.

An early challenger in the space of blockchain for academia was also the Pluto Network (Inc., 2018), but there is a lack of a working end-to-end solution as of the time of writing this thesis. Another 2018 addition to this landscape, DEIP.world[89] is a project whose goal is to turn research projects into digital assets, to make them more "self-sustaining" and "democratic". DEIP has reportedly launched a public beta on a proprietary blockchain[90] which is not open-source, as the project's GitHub repository appears mostly empty of content apart from the rudimentary API[91]. Rather than relying on traditionally well-understood mining protocols like Proof of Work, DEIP's blockchain introduces "Delegated Proof of

[88] http://fossilsandshit.com/26-a-blockchain-based-model-of-peer-review/ [Accessed August 2022]
[89] https://deip.world/deip-about-us [Accessed August 2022]
[90] https://medium.com/@deip/deip-first-blockchain-for-science-to-launch-public-testnet-9506ae19dd4a
[91] https://github.com/DEIPworld/ [Accessed August 2022]

Expertise Contribution" as its block production schedule. Google Scholar search yields no results on DEIP, and their website suggests that they have now shifted their focus to the wider "creator economy" rather than just academic publishing, to enable the "discovery, evaluation, licensing, and exchange of intangible assets"[92]. Other similarly themed projects such as Blockchain For Peer Review[93] are still in their infancy, lacking a working solution. Another publishing platform that was being developed that leverages blockchain was Aletheia, which uses the technology to "achieve a distributed and tamper proof database of information, storing document metadata, vote topics, vote results and information specific to users such as reputation and certifications" [94] and its development ceased in 2019 according to the code commits available online.

There are more specialised initiatives as well. For example, VitaDAO[95] dedicated to raising funds for drug development in the field of longevity research; GenomeDAO that focuses on auditable monetisation of genomic data; Opscientia whose flagship project is Holonym, a DID protocol for bridging scientists' credentials across blockchains and Web apps[96]. Nature, in their Correspondence section, called these, and similar initiatives DeSci – short for decentralised Science[97], with the promise of scientists being able to create their own incentive structures.
Further examples of DeSci prototypes include ARTiFACTS, aimed at timestamping research artifacts; and "decentralized.science", providing a public repository of Open Peer Reviews and a reviewers' reputation network [98]. One very exciting initiative is https://www.researchhub.com which is a research community powered by a ResearchCoin, an ERC20 token that is distributed to users for uploading new content, as well as summarizing and discussion research, as

[92] https://www.globenewswire.com/en/news-release/2022/07/26/2486154/0/en/DEIP-Partners-With-Roketo-And-Sets-Up-Over-1000-Financial-Streams-to-Power-The-Creator-Economy.html [Accessed August 2022]
[93] https://www.blockchainpeerreview.org [Accessed August 2022]
[94] https://github.com/aletheia-foundation/aletheia-whitepaper [Accessed August 2022]
[95] https://www.vitadao.com [Accessed August 2022]
[96] https://pulse.opsci.io/provable-and-computable-identity-for-future-proof-scientific-workflows-b020cdea11e3 [ Accessed August 2022]
[97] https://www.nature.com/articles/d41586-021-03642-9 [Accessed August 2022]
[98] https://decentralized.science/sample-page/papers/ [Accessed August 2022]

measured by up-votes, also known as “REP”. The main workflow of ResearchHub is shown in **Figure 16**.

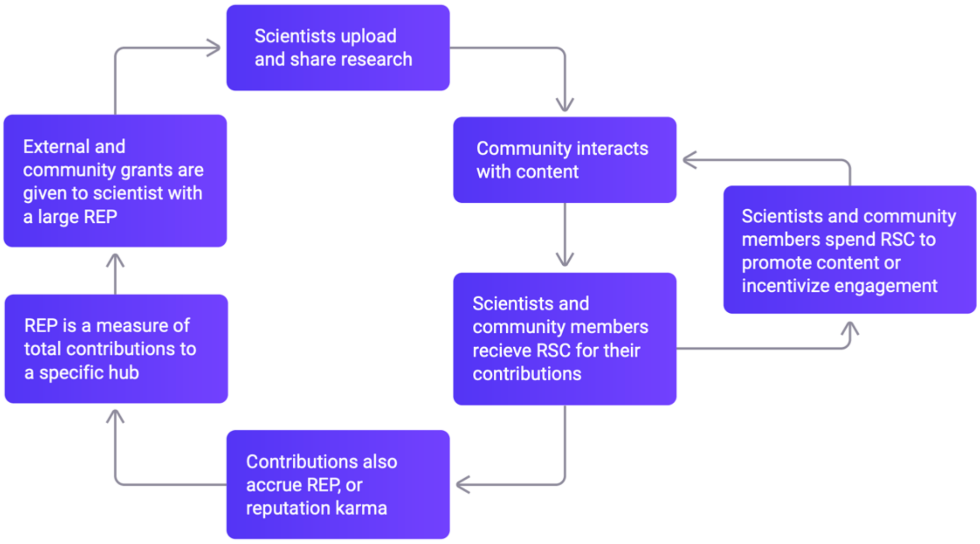



**Figure 16** ResearchHub workflow

Despite the interesting nature of the projects that have been reviewed, the common problems with most of the related work identified in this section include:

- Too focused on one particular aspect of blockchain, such as cryptographic tokens (coins)
- Very vague on the implementation, or implementation is lacking
- Software engineering approaches are not described
- Data modelling approaches are not described
- Little or no focus on privacy and no mention of privacy techniques
- Lack of governance structure such as DAO, or the DAO implementation is unclear
- Lack of evaluation of performance, cost and speed
- Lack of evaluation of the usefulness and social impact

The contribution of the current thesis is that it addresses all of the above concerns.

### 2.12.1 Comparison of Centralised and Decentralised Approaches to Scholarly Communications across its Various Facets

The facets of scholarly communications include research production, research publication, verification of research papers and artefacts, feedback provision, impact monitoring, reputation monitoring, data archival, search and discovery. Most of these facets have been described by Capadisli (2016, 2017, 2020). The following Table demonstrates the difference in what it could possibly mean for a particular facet to be approached in a "centralised" or "decentralised" manner, as it synthesises the knowledge from the previous subsections of this thesis (2.1 – 2.12).

| **SC Facet** | **Approach to Implementation** | |
|---|---|---|
| | *"Centralised"* | *"Decentralised"* |
| Production | using closed-source authoring software for manuscripts, following closed file format specifications for manuscripts and data, using centralised repositories for storing working versions of research, following the idea of research being a "paper" | using open-source software, following open file format specifications, using decentralised repositories, following ideas like liquid publications (Casati, Giunchiglia, & Marchese, 2007), dynamic publications (Heller, The, & Bartling, 2014), and nano-publications (Mons & Velterop, 2009). |
| Publication | following the idea of "papers" being published in a "journal", using closed journal workflows and systems | allowing research to be published outside of traditional journals, using open workflows for the lifecycle of a scholarly publication |
| Verification | downloading papers from a trusted source such as a website of a mega-publisher | retrieving papers using on-chain links (hashes) through a decentralised file system |
| Feedback | following a closed peer-review workflow through a mega-publisher website | using a DAO to participate in an open workflow to provide trustworthy feedback |
| Impact & Reputation | calculating traditional scientometrics in a corporate-owned database | using smart contracts to keep track of scientometrics and altmetrics, tracking genuine *interest* and *engagement* |
| Search & Discovery | a commercial (ad-funded) search engine displays results from megapublisher websites | a dApp aggregates on-chain data and makes it searchable |

**Table 3** – comparison of centralised and decentralised approaches to SC

As mentioned in the Table above, decentralised approaches to the production facet of Scholarly Communications often assume that it is possible to do away with the form of a "paper" and that it is possible to assume an approach that is collaborative, evolutionary, with a new way of crediting authors for their granular contributions. These approaches also abstract away journals and conferences and simply treat them as collections of research artifacts. Whilst the benefits of adopting such a new way of thinking about research production and publication appear to be ample (quicker turnaround, increased collaboration, easier crediting, more accessible composition of smaller elements into bigger pieces, etc.) most of these approaches have now been waiting for more than 15 years for a workable implementation, suggesting that changes in this area are very slow.

## 2.13 Social Machines and Sociotechnical Systems

In terms of nomenclature, a blockchain-based system designed to facilitate scientific collaboration via the Web can be either seen as an example of a Social Machine, or a Sociotechnical System (STS). Smart, Shadbolt and Simperl (2014) noted how the notion of social machines, describing Web applications and services that support user participation in the creation of online content (e.g., Wikipedia), serves to emphasize the socio-technical nature of the Web, and it provides a conceptual anchor for multidisciplinary collaboration within the scientific community. "Social machines" are thus goal-specific, functionally-closed engineered constructs embedding social processes and relations take place in routine and predictable ways (Afeltowicz & Pietrowicz, 2011).

Social machines are often discussed interchangeably with sociotechnical systems (STS), software-supported systems in which individuals and organisations interact to exchange information and services digitally. The authors (Chopra & Singh,

2016) report that existing approaches for social machines over-emphasize their technical aspects and inadequately support the meanings of social processes, leaving human interactions inadequately analysed. This finding inspires the heavy emphasis on social methods in my Methodology chapter.

## 2.14 Coordination Theory

*Coordination theory* was proposed as the interdisciplinary study of coordination, "the act of working together" (Malone & Crowston, 1991), drawing from computer science, operations research, economics and psychology. The usefulness of this framework is that it formalizes and illuminates the following processes of coordination, and appropriate methods for studying them:

- Goal decomposition, for example object oriented programming
- Resource allocation, for example analyses of markets
- Synchronization, for example workflow analyses
- Group decision-making, for example behavioural analyses
- Communication, for example establishing common vocabularies
- Perceptions of common objects, for example developing frameworks

-

Coordination theory flirts with the idea of applying new information technologies to reduce the costs of coordination but notes the following effects:

1st order: *replacing middle management with computer systems*

2nd order: *applying the freed-up managerial resources to more complex analysis tasks*

3rd order*: promoting coordination-intensive structures in reshaping organisations*

Coordination theory also proposes the following axioms of centralisation (Malone & Crowston, 1991):

- When IT-driven coordination decreases decision information costs, this leads to more centralisation. I use this finding to create recommendations for centrally co-ordinating change.
- When IT-driven coordination decreases agency costs, this leads to more decentralisation. This finding is used to propose the Smart Papers contribution of this thesis (Chapters 5 and 6), using DAOs, decentralised autonomous organisations.

Even though coordination theory was not applied explicitly as a method in the thesis, its way of framing goal decomposition, group decision-making and co-ordinating change has aided the researcher's understanding of the context in which scholarly communication happens, which inspired the application of Actor-Network Theory and Activity Theory, as described in the Methodology chapter.

## 2.15 Constructionism

The main dialectical tension of this thesis is that we all make assumptions about SC as a *socially constructed* reality agreed upon between researchers who experience it together, but at the same time, we also want to measure many aspects of SC *empirically*, in ways that are objective and quantifiable. This is a philosophical tension, as philosophy is the study of fundamental questions, such as those of knowledge and reality. The arguments hereby follow a constructionist worldview that is not anti-empiricist per se, but it could be called empiricism-within-constructionism, as I believe that academia performs an essential social function, providing to scholars and the general public alike meaning that is derived from its social construction. Many aspects of its social function, however, can be measured, which influenced my choice of research topic, research questions, and the methodology used. In other words, the social construction of scholarly communications did not prevent me from pragmatically identifying metrics and scales that were useful in evaluating proposed solutions.

One branch of philosophy, the philosophy of science, implicitly plays a key role in the writing of this thesis. Firstly, because it deals with the system of values that scholars and non-scholars hold around science, for instance, whether it should be product-driven or process-driven, exploratory vs explanatory, public or private, etc. (Bridgman, 1940). It also deals with the fundamental question of whether science can provide the truth about the world, or just the understanding of the world, and whether it is possible to have the latter without the former. This thesis assumes that we can know the truth about the understanding, being able to trace beyond reasonable doubt how that understanding was generated and how it evolved.

## 2.16 Interdisciplinary Research

The authors (Kaptelinin & Nardi, 2006) argue that Computer Sciences, Human-Computer Interaction (HCI) and Social Sciences researchers *"need to know more about the sociocultural contexts of other researchers' use of theory, in the same way that designers need to know users' context of use in order to design systems and products for them."* The authors point towards the importance of interdisciplinarity, such as combining engineering with philosophically grounded discussions of human-technology relations. As this thesis is informed by the discipline of Web Science, which studies the World Wide Web as an evolving first-class phenomenon using interdisciplinary approaches and methods (De Roure & Willcox, 2020), the methodology employed hereby is also interdisciplinary. This is because the Web faces complex problems that are increasingly beyond the ability of any single discipline to address adequately (Repko, 2008). One also must note that the purpose of any research process is cognitive advancement, and integration of disciplines is a means to that end, especially when the disciplines are unable to address increasingly complex issues (Repko, 2008, p. 30) Interdisciplinarity promotes non-linear thinking and creativity (Repko, 2008, p. 46). One key metaphor that combines the non-linear and creative attributes of interdisciplinary research is that of *bricolage* and the *bricoleur.* Lévi-

Strauss first introduced the idea of the *bricoleur* and juxtaposed it with the idea of the engineer, as two different ways of creating new things. The engineer has access to a vast selection of precise instruments and resources, while the bricoleur must manage with the tools and materials at hand. The engineer follows a structured process of planning and executing, while the bricoleur must make compromises, and modify his or her goals according to the constraints being faced (Johnson 2012, 361-364.) The *bricoleur's* competence "resides in his know-how and ability to 'cobble things together'" (Duymedjian & Ruling, 2010, p. 141). The bricoleur is also optimistic in that they can achieve something creatively, producing an arrangement that fits the problem they are faced with. Like any creative process, bricolage involves trial and error, and matches a widely accepted understanding that research is iterative (Spradley, 1979; Pratt, 2022). This resonates with the assumption of there being a feedback loop (Repko, 2008, p. 46) in interdisciplinary research, that requires the researcher to periodically revisit earlier activity.

The bricoleur is considerably more at the mercy of her or his environment than the engineer, and this framing is highly relevant to this thesis - consider for a moment how the beliefs, traditions and behaviours surrounding scholarly communications are ready-mades, as are the protocols and even file formats used to disseminate scholarly knowledge. Furthermore, scholarly communication practices, such as peer review, can be understood as a battlefield of multilateral forces and motivations, sometimes conflicting. To re-think the whole system in a centrally planned manner and according to a pre-defined specification, as the engineer would want to, may yield a system that is too far detached from the social reality in which scholars are comfortable operating.

Pratt (2022) defines *methodological bricolage* as the combining of analytic moves for the purpose of solving a problem or problems tailored to one's own research project. This combined approach works particularly in Web Science which is not only interdisciplinary but studies ever-evolving phenomena whose dynamics change at a fast pace. Therefore, bricolage and interdisciplinarity underpin the next chapter – Methodology.

# Chapter 3

# Methodology

## 3.1 Overall Approach

To answer the research questions, I follow a mixed-methods approach. The purpose of this Chapter is to discuss the methods that were selected and how they were applied. As reproducibility is a major problem in science, one must be mindful that the methods need to be laid out in a way that is detailed enough to allow other researchers to replicate them. This would be useful, for instance, if there was another attempt to develop a novel scholarly communications platform, albeit on a different type of blockchain, for example using Proof-of-Stake rather than Proof-of-Work. For a researcher trying to replicate my results, I first discuss the overall approach in this section and then focus on each individual method, grouping them by the Research Questions that they were used to answer.

The overall research approach begins with analysing the meanings of "decentralisation" by means of a systematic review into the meaning of this concept. What is required next is identifying the stakeholders involved in scholarly communications. Once they are categorised, one needs to examine the workflows that they use in their SC efforts and analyse which ones can be transferred to a new system in a decentralised manner, and how. As this requires the understanding of the social context of scholarly communications, Actor-Network Theory is used to capture the importance of the relevant socially constructed phenomena, and Activity Theory is used to formalise the use cases for the novel SC software platform. Subsequently, Action Research (learning by doing) is applied – a system is built, so that I can learn from the process of creation, from the artifact itself, and from how it is received by the relevant stakeholders. Methods are borrowed from software engineering, as the software development life cycle (SDLC) is applied to build object-oriented software. Agile project methodology is used to manage requirements and to deliver code in a timely manner. Smart contract design, execution and evaluation is used to evaluate the system technically. Questionnaire-based surveys and Thematic Analysis are then used to evaluate how the research artifacts are perceived by the relevant stakeholders, and what their attitudes are towards decentralised Open Science software.

In the following sub-sections, I map each method individually onto the three research questions that they were used to answer.

## 3.2 Methods Applied to Research Questions

### 3.2.1 Methods for Research Question 1

*(RQ1) Within the context of scholarly communications (SC), (a) what is the meaning of centralisation/decentralisation, and (2) what are those components of SC that exhibit centralisation that can be re-designed, re-imagined or re-formulated in a decentralised manner?*

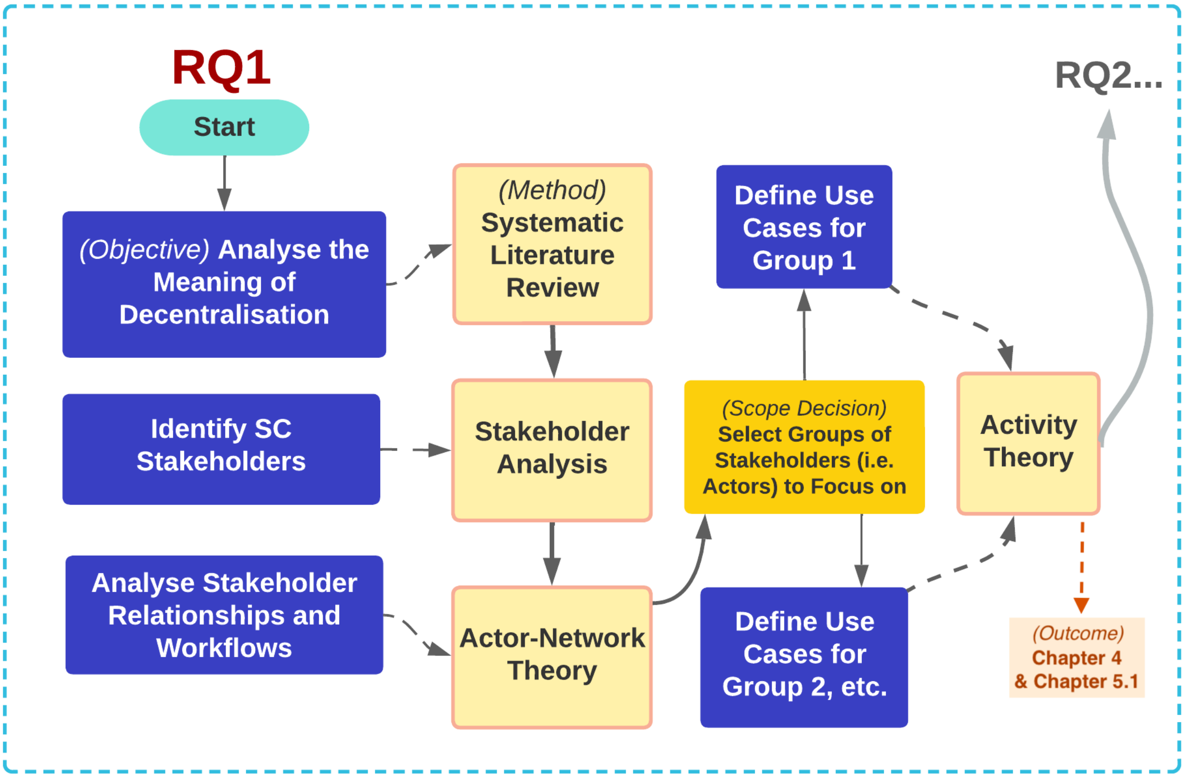



**Figure 17 -** Methods for answering RQ1

**Figure 17** depicts the overall approach for answering Research Question 1. The following sub-sections add detail to the individual methods presented in the Figure.

#### 3.2.1.1 Systematic Literature Review

The literature-based part of the thesis advances the understanding of blockchain decentralization discourses by providing the meanings of decentralization in general scholarly discourses and comparing them to the usages of the word within the context of blockchain-related papers. A literature review was performed based on the systematic approach similar to the Systematic Literature Review (SLR) as outlined by Kitchenham et al. (2010). The SLR research question hereby was

> *"what contexts, meanings and sentiments have been historically ascribed to decentralization by the authors of highly influential publications?"*

To best assist with answering this question, the data source of choice was the Web of Science set of databases hosted by Clarivate Analytics, which was queried using its Web portal[99] to carry out citation analysis and to complete the systematic analysis on the most influential articles. The alternative would have been Google Scholar, which several academics have been critical of on grounds of poor sourcing quality and incorrect citations (Jacso, 2006). Web of Science was ultimately elected due to its accuracy and the higher level of reproducibility that it facilitated.

The search results were manually inspected, and papers were read by the researchers to build a relevant understanding. The review technique relied on analyzing the sentiments, meanings and themes associated with the usage of decentralization across most influential scholarly articles. Only those articles became considered where "*decentralization*" (or "*decentralisation*") was mentioned explicitly in the title, abstract or keywords, or where there was an attempt to provide a definition for it in the body of the paper. If "decentralization" was contained in the body of the article but explicitly missing from the abstract, the abstract was screened for the inclusion of a related term, such as "centralization," "decentralizing," etc. (i.e.,

[99] http://webofknowledge.com

the declination of “decentralization” or “centralization”). Only articles with at least one citation were considered. More than 4000 articles were screened for this purpose. The findings were compiled into the relevant tables in this paper. Secondly, the same approach was repeated whilst additionally searching for blockchain as one of the required keywords. For the first task, the initial search query was:

“(*decentralization OR decentralisation*) *NOT blockchain*”

and the source was all Web of Science databases. For the second task, the initial search query was:

“(*decentralization OR decentralisation*) *AND blockchain*”

and the source was, like before, all databases included in Web of Science. The choice of the two Boolean expressions, yielding disjoint sets of papers, was motivated by the need to focus on decentralization primarily, and blockchain secondarily, and to speed up the filtering process because of the vast number of articles mentioning decentralization. The authors also made their choice to disregard articles just mentioning blockchains but not discussing decentralization, as they were deemed as not relevant to the systematic review question[7]. The review was followed by an analysis and inclusion of additional sources beyond the scholarly ones. Furthermore, if lower-citation articles were found to contain a repeated combination of context/meaning/sentiment that was identical to a previously found higher-citation article, then only the higher-citation article would have been included in our reported results, to maintain the conciseness of our findings. This does not contradict the author’s view that citation counts are far from a perfect measure of research impact and research quality, but they were still the most convenient measure available at the time. According to Leydesdorff and others (2016), citation counts are composed indicators making them convenient for evaluations of textual and socio-cognitive dimensions of document sets, sufficient to establish an inference about their “quality”, unless otherwise rebutted.

One key limitation of SLR is related to the use of citation counts as a *prima facie* proxy for document quality. This indicator is not normalized and contains information about non-quality aspects such as short-term impact, long-term impact, differences in citations across disciplines, prior popularity of the researcher, short-term trends in topic popularity, etc. (Leydesdorff, Bornmann, Comins, & Milojević, 2016). A more robust approach in future work of this type would involve normalizing citation counts and using them in tandem with other scientometrics and altmetrics. It is important to remember that research quality evaluation is a complicated subject because research quality cannot be captured by any single metric or evaluation method, and qualitative approaches may be key to consider, such as expert evaluation, societal impact evaluation, etc.

Another limitation associated with this methodology is that search reproducibility is limited by the subjectivity of discarding articles that mentioned decentralization too briefly without making a case for it, or without defining it, as these aspects were subjectively evaluated and decided upon manual inspection and screening by a human researcher. This human-based evaluation approach also poses a limitation in how the sentiments (positive, negative, or neutral) were subjectively decided based on the double-hermeneutic, i.e., the researcher's individual interpretation of the context and how it was evaluated by previous researchers.

Finally, using Web of Science may limit the researchers' visibility of the most recent search results given the length of the academic cycle required to publish and for the work to be indexed, in which case using Google Scholar would have presented some advantages due to Scholar's broader inclusivity of the more recent scholarly efforts as well as pre-prints and gray literature.

#### 3.2.1.2 Stakeholder Analysis

Stakeholder theory has been a useful framework which recently attained more theoretical status. Stakeholder identification and prioritisation is crucial for the early stages of undertaking any systematic work aimed at solving wicked problems of socio-economic and cultural change. The stakeholder theory focuses on one or more of three relationship attributes: power, legitimacy, and urgency (Mitchell, Agle, & Wood, 1997). By combining these attributes, we generate a typology of stakeholders and propositions concerning their salience. One way of identifying relevant stakeholders is through the application of the stakeholder matrix, also known as the influence-interest or power-interest grid (Ackermann & Eden, 2011), as illustrated in **Figure 18** - the power-interest matrix from Ackermann & Eden (2011)below:

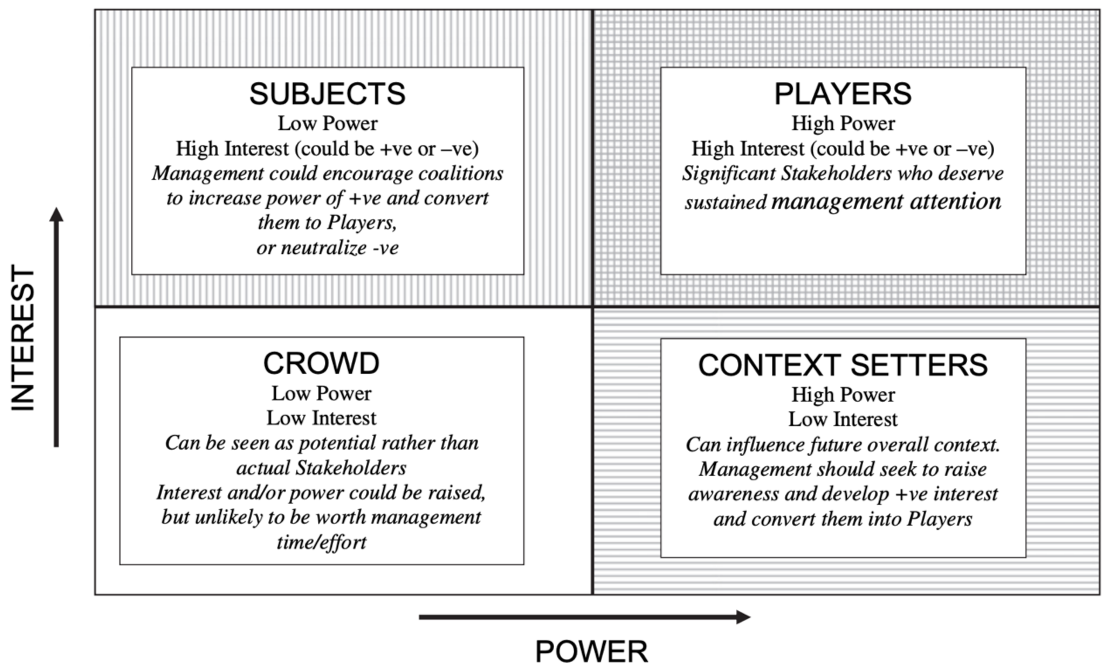



**Figure 18** - the power-interest matrix from Ackermann & Eden (2011)

Stakeholders in the upper two classes have the most stake in the issue, and the difference lies in how much power they have to affect (whether support or sabotage) the relevant strategies.

The input to the matrix was taken from the literature review carried out by me as part of creating the Background section (Chapter 2 of the Thesis). Reviewing the stakeholders that played a key role in the Background generated the initial set of stakeholders for the analysis. By applying the matrix framework (subjectively assessing the interest and power) the *subjects* were identified to be:

scholars (including authors and reviewers), students, university administrations, university associations[100], early-career researchers, researchers from emerging economies, and independent researchers;

and the *players*:

funders, academic publishers, policy-makers, tenured academic staff and senior scholars (including editors and chairs), the taxpayer

The largest projected group of subjects are scholars (academic researchers), who want to increase the transparency of scholarly communications, and to better monitor the lifecycle of their own research and how it is being used, strengthening control over their work and career. Reviewers also belong to that group, with an added competency of peer review participation. They want to increase the visibility of their peer review efforts. University and Research Administrators, as well as Library / Repository Managers form a group of subjects that carry governance privileges. They influence decisions related to core infrastructure setup.

One strategically important group of players is editors and chairs, who are appointed to oversee grouping of papers into topics and controlling the Conference and Journal workflows. These will be important to directly address when promoting the adoption of a new scholarly communications system. Another important group of players are the Funding Bodies. Funders will be interested in aggregated statistics of research papers mapping onto their projects. They will be interested in the visibility of their projects, organised by funding streams. Finally, a very important player is the Taxpayer who abstractly pays for the bulk of the research work that gets funding. This player is interested in a public view into the platform, its source code, the data it stores, the quality of the data, the quality of the platform, especially how much it costs and how it performs. She is best understood as a public observer of the system.

[100] For example, Universities UK - https://universitiesuk.ac.uk, an alliance of 140 British Universities, and Una Europa - https://www.una-europa.eu, an alliance of 9 European Universities.

As stakeholders are involved in business and organisational processes, communications and workflow analyses are used as approaches of modelling the relevant processes, so that questions of task dependence and information flow can be addressed through formal operations, leading to more effective design of these processes (Basu & Blanning, 2000). By exploring, uncovering and formalising academic publishing workflows, we gain a better understanding of which task sequences can be transferred from players to subjects, to increase the power of subjects and convert them to players. The relevant control flows (workflows) will be modelled and extracted from existing set-ups such as Elsevier's[101] and Clarivate's[102].

BPMN is a widely understood graphical notation for modelling business processes (workflows) via business Process Diagrams (BPD), flow-chart representations that depict activities, control, data, and helpful information about the process. The core elements of the visual language are discussed in (Rospocher, Ghidini, & Serafini, 2014) and presented in **Figure 19.** The peer review business process modelled using BPMN is presented in **Figure 20**.

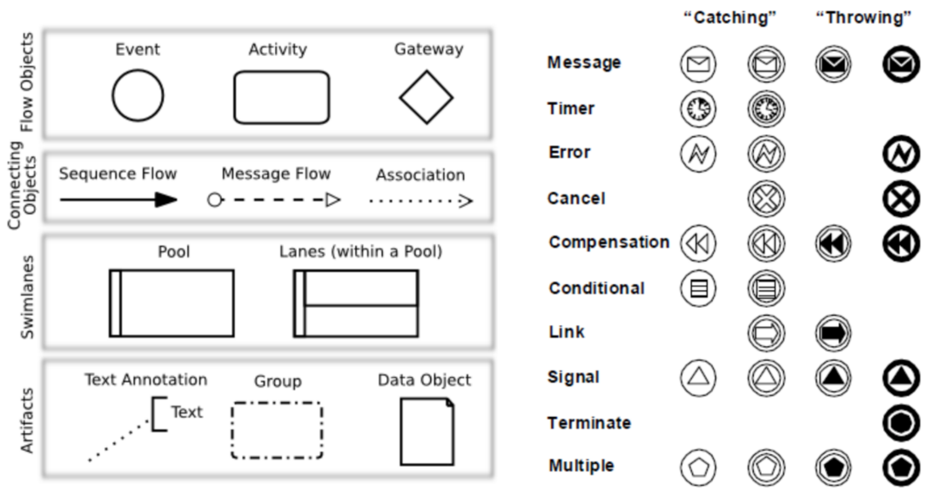



**Figure 19** - BPMN core and extended visual notation from Rospocher et al. (2014)

[101] https://www.elsevier.com/?a=91173 *{accessed April 2022}*
[102] https://clarivate.com/blog/tag/journal-workflow/ *{accessed April 2022}*

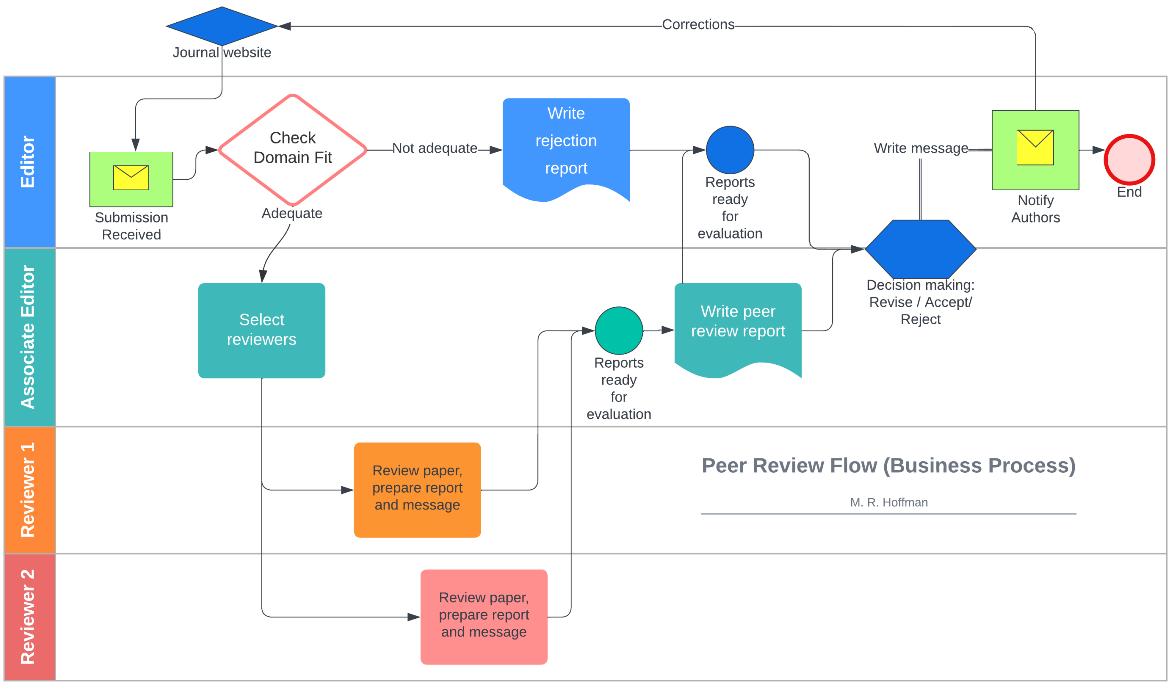



**Figure 20** - BPMN modelling of the peer review process

#### 3.2.1.3 Actor-Network Theory

The framework for mapping the human and non-human constituents of systems and how they relate to each other diagrammatically will be the Actor Network Theory (ANT) which provides detailed accounts of how human and non-human actors interact in the networked world. In this framework, human stakeholders are identified through an iterative, interpretive, dynamic and context-contingent process, and are extended to include the identification of nonhuman actants (Pouloudi, Gandecha, Atkinson, & Papazafeiropoulou, 2004), which means that Actor-Network Theory (ANT) can be used synergistically with Stakeholder Theory. The usefulness of ANT is that it provides a lens through which we observe how technology shapes human interactions and relationships.

Actor-Network Theory (ANT) posits that nothing exists outside of the relationships that are formed within the networks between objects, ideas, processes and social situations which are all seen as belonging to the same level of phenomena. The purpose of ANT is thus to describe the reality through describing the networks. Latour (2005) further posits that for a network to exist, it needs to be "translated" into a single entity, through problematisation (the definition of the nature of the problem by one of the actors from within the network); it also needs to be defended, so the 'main' actor must be able to explain why the network is 'worth of existing'. (Latour, Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford: Oxford UP., 2005). Thus, the burden of the "translation" of the decentralised scholarly communication actor-network belongs to the author of this thesis.

One criticism of ANT is it does not have predictive power, only descriptive. Furthermore, testing of the network cannot be done quantitatively, but is done by telling a story about the network, thus becoming enrolled into the network (Leydesdorff,, 2021). This enmeshment is something that one must be aware of, and any assumptions resulting from it must be documented as part of the thesis.

ANT has been previously used to analyse and conceptualise the landscape of Scholarly Communications, most recently by (Praun, Cunningham, & Pieperhoff, 2022) who mapped current disruptions and disruptors in this area and identified Open Access and Open Science as relevant actors. An earlier study by Rieger (2008), which was focusing on the social construction on innovation in scholarly communications, identified publishers, research institutions, end users as well as the Open Source community as the primary relevant actor groups. A 2007 publication by Leydersdorff further identified Scientometrics as a key actant in the ANT of scientific communication. A more recent paper by Lukka and Vinnari (2017) suggests that ANT is a valid methodological choice for "interventionist" research, because it provides theoretical grounding for studies on dynamic processes, such as the emergence and implementation of Scholarly Communications.

As ANT allows us to better understand the social context for emergent technologies, including scientific software and scholarly communication utilities, and since it illuminates the social construction of "requirements" (or determinants) for those technologies, it is useful to integrate it with the software development life cycle. This will allow us to map the relevant actors, actants, and their relationships, onto software requirements, in an Agile manner – to allow appropriate contingencies for the dynamic nature of those requirements.

As there is an apparent lack of a well-established formal methodology for this sort of mapping, a simple technique is proposed hereby, with the following assumptions:

- ANT actors may become Scrum/Agile personas (ANT actor X in the "As an X" part of a user story)
- The relationships between ANT actors and actants may be used in the rest of the user story (ANT relationship Y in "I want Y" or "So that I can achieve Y").

#### 3.2.1.4 Activity Theory

Activity Theory (AT), as originally proposed by Vygotsky is recommended as a theoretical framework for grounding the understanding of human intentionality, modelling activities by following a human-centric top-down approach. This framework accounts for environment, motivations, and complexity of activities, and emphasises the role of learning (Roth & Lee, 2007). As activities are enacted by human constituents in ever-changing, complex and self-organising systems. On the one hand, Actor-Network Theory is often suggested as a tool to snapshot those systems visually, even though ANT does not take into account human intentionality, treating human actors on the same level as non-human ones. On the other hand, AT is concerned with documenting the motives of individuals within their communities, also documenting rules and agreements that community members agree to while engaging in activities. Activity theory places focuses on identifying activities as first-class citizens, and organizes other components (such as tools and people) around the activities that they are meaningfully involved in. When more detail is required, additional activities may be identified and analysed. Finally, there seems to be overlap of Activity Theory with Object-oriented analysis and design which means that it has been successfully applied to analyse collective software development efforts (de Souza, Cleidson, & Redmiles, 2003).

Engström presented 5 principles of activity theory that have become widely accepted: a) an activity system is the primary unit of analysis, b) activity systems always contain many "voices" – i.e. perspectives, traditions, and viewpoints, c) an activity system takes shape over time and its history affects the activity system, d) contradictions are the driving forces of change, and e) expansive transformations occur when motives change to accommodate new possibilities. Furthermore, he proposed a hierarchy of mediating artifacts (instruments, signs, language, machines, programs)


**Figure 21 -** Engström's Hierarchy of Mediating Artifacts

that can be categorised according to their relevance to supporting a particular activity. This categorisation is achieved by asking questions outlined in **Figure 21** (Georg, 2011)(Georg, 2011). I use these questions to develop a blockchain decentralization framework in the next Chapter.

When analysing the environments and communities in which blockchain operates, there are a number of websites and online platforms that have to be observed, considering not only technical aspects, but also discussions and debates on policy, regulation, ethical concerns, and social implications.:

- GitHub: GitHub is a popular platform for developers where many blockchain projects are open-sourced. You can monitor the development and changes to various projects, as well as community discussion through issues and pull requests.
- Reddit: Reddit has several communities devoted to blockchain technology and cryptocurrencies. Some of the largest include r/Blockchain, r/ethereum, r/Bitcoin, r/CryptoCurrency, and many more. These communities often

discuss recent developments, and include both technical and non-technical discussions.

- Blockchain and Cryptocurrency News Websites: Websites like CoinDesk, Cointelegraph, and The Block are frequently updated with news about blockchain projects, market trends, regulatory changes, etc.
- Medium and Other Blogs: Many blockchain developers and enthusiasts write on Medium or maintain their own blogs. Platforms like Hackernoon also frequently feature articles about blockchain technology.
- StackExchange: Websites like StackExchange have sections for blockchain and cryptocurrencies where developers ask technical questions and discuss solutions.
- Twitter: Many blockchain developers, enthusiasts, and thought leaders are active on Twitter. Tracking their discussions can provide insight into what's currently important or controversial in the community.
- Academic databases: Websites like Google Scholar, ResearchGate, and the ACM Digital Library publish papers from researchers exploring blockchain technology.
- Forums and Online Groups: BitcoinTalk, Stack Overflow, and other more specific forums like Ethereum Stack Exchange, or even LinkedIn groups and Facebook pages related to blockchain could provide valuable information.
- Crypto Marketplaces and Exchanges: Websites like Binance, Coinbase, and Kraken often have blogs and community sections where users discuss the latest trends in blockchain technology and cryptocurrencies.
- Official Websites and Blogs of Blockchain Projects: Websites of blockchain projects such as Ethereum Foundation, Bitcoin.org, Ripple, Cardano Foundation, Polkadot Network, and others often have blogs and news sections, where they post about their latest developments, partnerships, and events.

In the case of creating software for scholarly communications, the principles of Activity Theory can be applied in the following manner:

1. Identifying subjects and their objectives: The "subjects" are the users, in this case, scholars or academic researchers. Their objectives might be to find relevant literature, to publish their work, to collaborate with others, to seek peer reviews, etc. Understanding these objectives will help define the main use cases.
2. Understanding Tools: Tools are the artifacts that subjects use to achieve their objectives. In the context of scholarly communication, tools can be existing software platforms, databases, digital libraries, reference managers, etc. Analyzing how scholars use these tools will inform the design of the new software.
3. Examining Rules and Community: Rules are the formal or informal regulations that govern how subjects achieve their objectives, while the community refers to all the people involved in the activity. In scholarly communication, the community can include researchers, peer reviewers, editors, etc., and rules may encompass publication guidelines, citation styles, etc. Understanding these dynamics can help identify additional use cases.
4. Division of Labor: This refers to how tasks are distributed among members of the community. For instance, in scholarly communication, authors write articles, reviewers critique them, and editors make final decisions. By analyzing the division of labor, you can develop use cases that streamline these workflows.
5. Outcome: The expected outcome of the activity. For scholarly communications, it will be the documentation of the specific use cases that will be then used to create the novel scholarly communications software.

Thus, to generate our initial use cases using activity theory, the following concrete steps will be followed:

1. Identify the various activities that the user (the scholar) undertakes.

2. Analyse the context of these activities, including the tools they use, the rules they follow, the community they engage with, and how the work is divided among this community.

3. Identify blockchain tools and blockchain communities, and how they can enrich the original context of scholarly communications (as introducing blockchain introduces a change in the context in which scholarly communication operates).

3. Formulate use cases based on relevant activities that can be decentralised using blockchains.

Activity theory thus provides a holistic perspective that considers the multiple dimensions of user interaction, making it a potent tool for preparing the ground for software development in the realm of scholarly communications.

### 3.2.2 Methods for Research Question 2

*(RQ2) What are the requirements for a decentralised academic scholarly communications platform - including (a) what features are relevant? (b) how can they be implemented? (c) how can they be evaluated?*

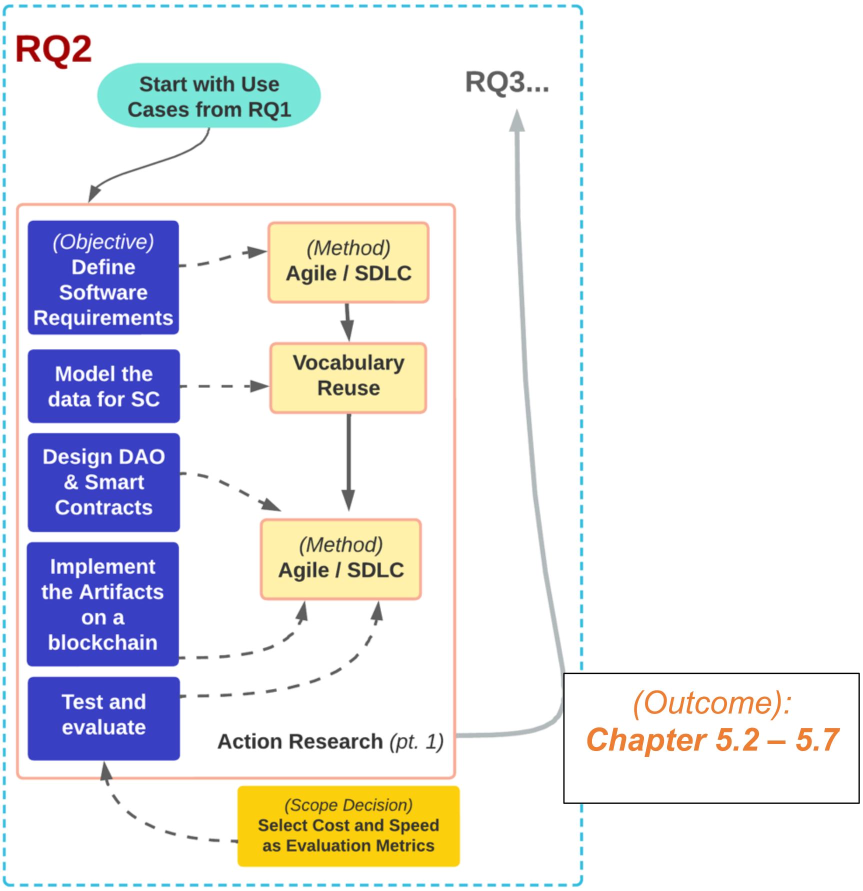



**Figure 22** - Methods for answering RQ2

**Figure 22** depicts the overall approach for answering Research Question 2. Cost and speed were selected as an evaluation metrics based on long-standing standards in quantitative evaluation of software quality (Boehm, Brown, & Lipow, 1976). The following sub-sections add detail to the individual methods presented in the Figure.

#### 3.2.2.1 Action Research

The purpose for **action research** is socially motivated and it is to collect evidence or data to expose unjust practices or environmental dangers and recommend actions for change (Bogdan and Biklen, 1982, p. 223). According to (Santos & Travassos, 2009), there is an increasing tendency of action research use in combination with software engineering. According to a recent review, computer scientists use action research to construct artifacts and to improve their communication capabilities with users (Maria Orquiza, Sánchez García, & Gabriel Costelini, 2022). Action research is *"a form of self-reflective enquiry … in social situations … to improve the rationality and justice of their own practices, their understanding of these practices and the situations in which the practices are carried out*" (Carr and Kemmis 1986: 162). This term was first used by Kurt Lewin (Lewin, 1946). His approach is an iterative one, involving the following steps:

**A. Identify the problem / initial idea / situation**
**B. Establish facts**
**C. Plan a course of action**
**D. Take initial action**
**E. Evaluate the outcomes of the initial action**
**F. Amend the plan**
**G. Take next action, etc.**

Since action research is a form of reflective practice (Schon, 1983), it may be seen as "thinking on your feet" (Lowe, 2007 "Beginning Research", p. 105) whereby the practitioner executes experiments in order to test their "theories" to gain a new understanding of the situation and the effect of constructive change. Therefore, in the pipeline above, there is a possibility of feedback loops, so the initial idea or situation could change, and so would the action plan. Kemmis (1995) sees all reflective research, and thus also action research, as an active process in which experiences are reviewed, analysed and evaluated, with a goal to provide an action plan for the future. Reflection starts out when the researcher encounters some problematic aspect of practice (this could be, for example, educational or research practice), and attempts to make sense of it.

Kolb (1984) sees this type of research as a learning process, where new knowledge is generated by critically reflecting upon actions and experiences, thus forming new abstract conceptualisations that can be, subsequently, actively experimented upon. All in all, when reflecting on experimental observations, action research encourages the practitioner to reflect on what participants have been doing and saying and to focus on any issues or problems therein. Other key characteristics of action research are the empowerment of participants, and the central importance of communication (Lowe, 2007, p. 106).

In implementing Action Research, the step "D" (*"Take initial action"*) would hereby take the shape of building a novel SC software platform on a blockchain. The authors Pfeifer & Bongard (2007) note that *"understanding by building"* is widely applied in cognitive and computer sciences, and is based on the following premise:

> *"If we are interested in how … humans walk or recognize a face in a crowd, we build a system … that mimics certain aspects of the behavior we wish to study… you have to build something that actually works in the real world, there is no way of glossing over details, which is possible when you formulate a theory abstractly [Pfeifer & Bongard, 2007, p. 78]*

By applying this premise in this thesis, one constructs software artefacts behaving in ways that are relevant to answering the Research Question, by constructing representations of relevant concepts and processes on a blockchchain.

In implementing Action Research step "E" (*"Evaluate initial outcomes"*), a scoping decision is arbitrarily made to focus on evaluating the speed and the cost of running the software artifacts on the blockchain.

#### 3.2.2.2 Ontology and Vocabulary Reuse

Ontology (Allemang & Hendler, 2011), in the context of computer science and information science, is a framework for representing knowledge about a particular domain in a formal, machine-interpretable way. When applying ontology to model linked data for a scholarly communications platform on a blockchain, we first have to identify the types of entities (or "classes") that we will be dealing with, such as researchers, scholarly articles, citations, peer reviews, etc. This is informed by the previous Stakeholder Analysis, Actor-Network and Activity Theory research We also have to specify the properties that these entities can have, and the relationships that can exist between them. The ontology then forms the basis for a shared vocabulary that can be used to annotate the linked data, which in turn facilitates semantic querying and interoperability across different systems.

To leverage blockchain technology in this context, we would create a decentralised platform where each piece of scholarly communication (e.g., an article or a peer review) is a transaction on the blockchain, and the linked data about the transaction is annotated using the ontology. This way, each transaction is permanently recorded and publicly verifiable, ensuring transparency and accountability in the scholarly communications process. The ontology also allows the platform users to efficiently retrieve, link, and analyse the data. For example, it may now support advanced queries like "find all articles by a given author that have been cited by more than 100 other articles", by combining the ontological knowledge about authors, articles, and citations with the linked data recorded in the blockchain transactions.

Vocabulary reuse is intimately linked to the process of ontology creation and management. An ontology is a way of defining a set of concepts and the relationships between them for a specific domain. In essence, it's a model for understanding a certain field or topic. Vocabulary reuse is a powerful method for identifying appropriate ontologies in any domain, including scholarly communications software. It involves the reuse of existing vocabularies, concepts, and ontologies to better define and understand a particular domain.

In the context of scholarly communication software, ontology can provide a shared and common understanding of the domain that can be communicated between people and application systems. It can assist in identifying the semantics of information in a machine-readable format.

The following steps are needed to apply this method:

Identify Existing Vocabularies: Begin by identifying the existing vocabularies used in scholarly communication software. This may involve reviewing the terms used in current literature databases, repositories, data sets, software documentation, user interfaces, and application program interfaces (APIs).

Analyze the Domain: Analyze the domain and the relationships between its entities. This process will help the researcher understand the context in which the vocabulary is used.

Align Vocabularies with Existing Ontologies: After identifying the existing vocabularies, align them with existing ontologies. This process involves mapping each vocabulary term to its equivalent term in an existing ontology.

Choose Appropriate Ontologies: Not all ontologies are created equal. When choosing which ontologies to reuse, I should consider factors such as their expressivity, whether they are actively maintained, how widely they are used, and whether they are endorsed by relevant authorities.

Adaptation: It's unlikely that an existing ontology will cover all the nuances of the scholarly communication domain perfectly. As such, adaptation is usually necessary. This might involve creating new classes or properties, or even developing a new ontology that builds on existing ones.

Implementation and Testing: Once the researcher has selected and adapted your ontology, the next step is to implement it within scholarly communication software as part of the software development life-cycle. Once implemented, it must be tested to ensure it functions as intended.

Feedback and Iteration: building an ontology is not a one-time task. One should continuously update and improve the ontology based on feedback from users and changes in the domain itself.

By reusing vocabularies and aligning them with appropriate ontologies, we can ensure consistency and interoperability across different systems, improve data sharing and discovery, and facilitate semantic integration.

#### 3.2.2.3 Software Engineering and Design

Software engineering methodology (Wieringa, 2014) is based around the "**design cycle**" idea that contains certain crucial stages:

- Specification of requirements
- Stakeholder and goal analysis
- Implementation and Evaluation

The design cycle idea is known in the software engineering industry as SDLC – Software Development Life Cycle – and is shown in **Figure 23**:

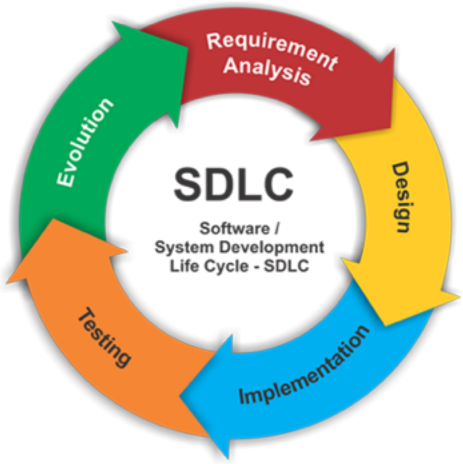



**Figure 23** - Software Development Life Cycle

In scientific and research software development, apart from the design-cycle, Wieringa (2014) also specifies the existence of the "empirical cycle" that focuses on the following:

- Formulating the research problem;
- Creating the research setup;
- Execution and data collection and analysis.

This thesis makes use of both the design cycle, as well as the empirical cycle, as the main method to answer the research questions will be through engineering a software system to collect new research data (a blockchain-based suite of smart contracts designed to facilitate decentralised academic publishing). The outcomes of the interactions transacted on this new platform will be tested and documented. One immediate outcome of this approach is that the performance and cost of candidate solutions will be known.

Furthermore, modern software engineering usually follows object-oriented structure that emphasizes the importance of modular planning and design, reusability, genericity, design-by-contract and abstract data types (Meyer, 1997). Important is also the idea of *separation of concerns*, which allows us to separate parts of the software that exhibit different rates of change or different types of change – this is applied by applying abstraction to different classes of objects, their business rules, algorithmic logic and data models (Hürsch & Lopes, 1995).

#### 3.2.2.4 Agile Software Development

Agile is an umbrella term for iterative development methods and processes. These iterations (shown in Figure 24) also known as sprints, often last from 1 to 3 weeks, and after each sprint, requirements may be introduced, modified or removed. Agile differs from traditional "waterfall-style" software project management methods as it refrains from specifying all the requirements up front, leans towards small but regular software releases, with cooperative and straightforward methods and adaptive scopes of features. Agile systems are defined by shared metaphors between the stakeholders and the programmers, known as "stories", which are then group into "epics", and are prioritised regularly (Abrahamsson, 2017). The use of stories and epics is how software requirements are formalised and prioritised in Agile, thus it serves a purpose of a requirement analysis framework, whilst also serving project management purposes. The use of specialist tooling such as JIRA[103] is common in Agile, and these can be used to create tickets with tasks assigned to project team members which helps organise work. The principles of Agile have been compiled into the so-called Agile Manifesto and are available online at https://agilemanifesto.org/principles.html. They

[103] A dedicated JIRA instance was created for this thesis at https://smartjournals.atlassian.net

include welcoming changes in requirements even when they appear late in the process, reflexivity, valuing face-to-face conversations, sustainable development, simplicity, good design and attention to technical excellence.

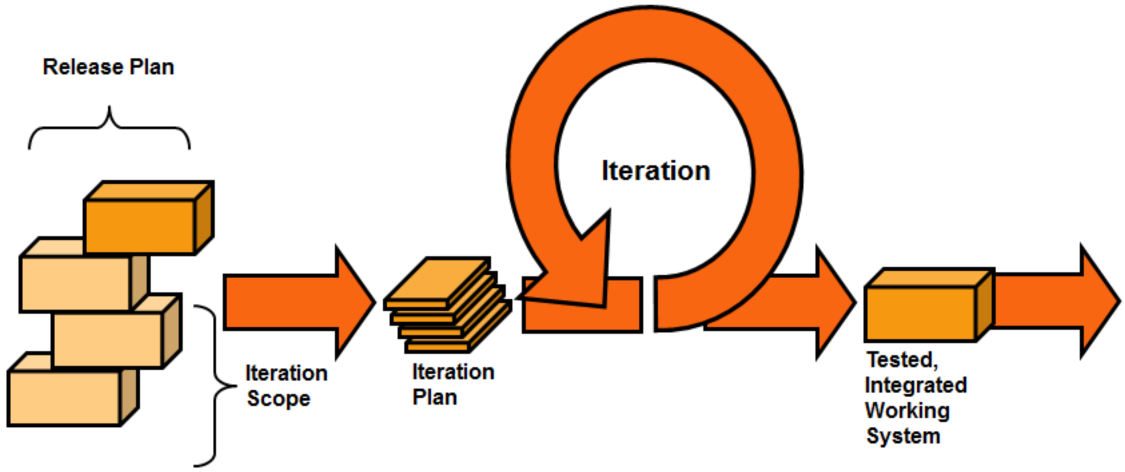



**Figure 24** - Agile Software Development Methodology

#### 3.2.2.5 Smart Contracts Design and Blockchain Implementation

Smart contracts are written in a high-level language, Solidity. When compiled, they are stored on the Ethereum blockchain using the special contract creation transaction by being sent to the creation address 0x0. Each contract then receives its own Ethereum address (that can be considered alike to a URI in Ethereum's namespace), and contracts can be executed if they are called by a transaction using this address, so that a client can call functions defined in a smart contract, in a similar fashion to a web service; contracts can also call each other. Ethereum blockchain was designed to be deterministic. This means, that everyone should always end up with the same, correct state, if they try to replay the history of Ethereum transactions.

In Ethereum, the code execution layer is provided by the Ethereum Virtual Machine (EVM), a Turing complete 256-bit VM that allows anyone to execute code that references and stores blockchain data in a trustless environment. Furthermore, every contract on the Ethereum blockchain has its own storage which only it can write to; this is known as the contracts state and it can be seen as a flexible database albeit at a high cost. That is why, when implementing our models, I chose to store all the artefacts using IPFS (Benet, 2014), the InterPlanetary File System. IPFS[104] is used for efficiently distributing and referencing hash-linked data in a way that is not centralised and does not necessarily involve blockchain transactions, thus avoiding the economic penalties associated with on-chain storage (Eberhardt & Tai, 2017). In many ways, IPFS is similar to the World Wide Web, but it could be also seen as a single BitTorrent swarm for exchanging objects. Furthermore, the IPFS specification contains a special *commit* object which represents a particular snapshot in the version history of a file. This allows us to reference resources in an immutable way, akin to Trusty URI functionality (Kuhn & Dumontier, Trusty URIs, 2014). Using IPFS we can, therefore, limit the role of Ethereum, so that it only deals with the application logic; the data layer is provided by the IPFS stack, and the two layers are integrated via hash references. Another design choice is to employ SafeMath[105], which is a Solidity math library, belonging to the OpenZeppelin framework, specially designed to support safe mathematical operations, which is used to decrement and increment citation counts on existing papers being referenced - changes that are triggered by publishing new papers referencing them.

#### 3.2.2.6 Testing the Smart Contract Implementation

To test the implementation of the smart contracts, the reader will need a POSIX-compliant platform with NodeJS (at least version 8.0.0) and NPM packager installed. The NPM package **@aragon/cli** installs Aragon CLI, a tool for creating, testing and publishing Aragon applications. Aragon[106] is a Web-based DAO

[104] https://ipfs.io

[105] https://docs.openzeppelin.com/contracts/2.x/api/math

[106] https://aragon.org
[Accessed August 2022]

deployment platform that helps Internet users collaborate remotely and freely organise without intermediaries, to transfer data and value, and to make joint decisions.

When starting a new project on Aragon, a DAO (decentralised autonomous organisation, (McGregor-Lowndes & others, 2019)) is created. The DAO will be a set of smart contracts that define a governance structure and rules for user interaction via a web interface. In an abstract sense, a DAO represents a group of people working together towards a shared goal.

One further advantage of using Aragon to manage our smart contracts is that it manages the upgrading of major, minor and patch versions of contracts in a way that is seamless (using a proxy pattern) and was unachievable in our original implementation – upgrading old-style smart contracts would have required a completely new contract address – which now can be avoided with Aragon.

A test DAO identity called "soton" is already registered on the Rinkeby Ethereum testnet[107]. A *soton* project can be initialised locally by typing in "*aragon init soton*" which will prepare the ground for local testing. The code for the Smart Papers app can be checked out from GitHub[108]. The app can then be compiled and launched with "*npm run start:app*", whereas a local Aragon environment can be launched through "*npm run start:aragon:http*" and subsequently navigated through the web browser on localhost port 3000. This will start a development chain you can interact with ganache-core[109], so it is a full testrpc instance[110]

Essentially, any Smart Papers transaction can now be signed by the user in their Web browser with MetaMask[111] extension, and all the interactions requiring multiple signatures can be signed individually by different users as part of a single Aragon vote. Available as either a Web browser extension or a mobile app, MetaMask is a tool for

---

[107] https://rinkeby.aragon.org/#/soton.aragonid.eth/
[108] https://github.com/mikehoff/SmartPapersApp
[109] https://github.com/trufflesuite/ganache-core
[110] https://github.com/trufflesuite/ganache-cli
[111] http://metamask.io

navigating Web3, incorporating a key store, secure login, cryptographic token wallet, and token exchange functionalities, which makes the DApp ecosystem accessible to Web users.

Furthermore, unit testing is used, which is a software testing method by which the smallest possible chunks (units) of code are tested, typically in an automated manner, to ascertain that they perform as expected, and whether they are fit for purpose. The Truffle suite[112] is used to automate the unit testing for our contracts. Solidity test contracts live alongside other code as *.sol* files. They contain assertions that are checked for being true at runtime. For more details on how these tests are written, consult https://trufflesuite.com/docs/truffle/testing/writing-tests-in-solidity/

As blockchain technology focuses on creating trust in an untrusting ecosystem, it belongs to the field of cybersecurity technologies (Taylor, Dargahi, & Dehghantanha, 2020). In secure applications, it is standard practice to perform security audits of the system to identify vulnerabilities, bugs and to determine the semantic soundness of the implementation. Smart contract audits allow dApp developers to persuade users that the dApp is trustworthy and win their confidence that their assets are safe. To standardise the evaluation of smart contracts during an audit, the following measures are based on the open-source OWASP Risk Rating Methodology[113] which is in line with the Risk Management Guide for Information Technology Systems of the National Institute of Standards and Technology (Stoneburner, Goguen, & Feringa, 2002):

- Likelihood, which represents how likely a particular vulnerability is to be exploited by malicious actors
- Impact, which measures the potential losses and damages caused by a successful attack

[112] https://trufflesuite.com/docs/truffle/testing/testing-your-contracts/

[113] https://owasp.org/www-community/OWASP_Risk_Rating_Methodology [Accessed August 2020]

- Severity, which demonstrates the overall priority of the risk

Likelihood and impact are classified into three levels: High, Medium and Low. Severity is the product of likelihood and impact and can be additionally classified into Critical (High x High). To evaluate risks, a checklist[114] is applied – for each item, if the analysis does not identify any issue, the smart contract is considered safe with respect to that item. For any discovered bugs or vulnerabilities, these must be documented in an audit report, and can be tested further.

The auditing report contains the name and position of the person performing the audit, the details of where the source code was accessed from, the names of each smart contract .SOL (Solidity) file, report version, and report date. For each contract file, the auditing section follows a specific structure:

- Coding Bugs: By statically analysing the smart contracts, bugs can be identified quickly and then manually verified. The useful tools here are Slither[115] (static analysis framework for Solidity), Mythril[116] (bug hunting framework for Ethereum) and Solgraph[117] (for visualising the control flow of the smart contract).

- Semantic Consistency Checks: the auditor reads the source code and notes whether the structure and logic of implemented smart contracts (including the data stored by the contract) are relevant to the purpose of the contract and consistent with the description in the white paper.

- Additional Recommendations: additional suggestions regarding programming style, best practices, and gas usage.

A security audit for the main contribution of this thesis is included in the Appendixes.

---

[114] https://github.com/runtimeverification/verified-smart-contracts/wiki/List-of-Security-Vulnerabilities [Accessed August 2022]
[115] https://github.com/crytic/slither [Accessed August 2022]
[116] https://github.com/ConsenSys/mythril [Accessed August 2022]
[117] https://github.com/raineorshine/solgraph [Accessed August 2022]

### 3.2.3 Methods for Research Question 3

*(RQ3) What are the attitudes of scholars towards decentralised Open Science software, and, towards decentralising science with blockchains?*

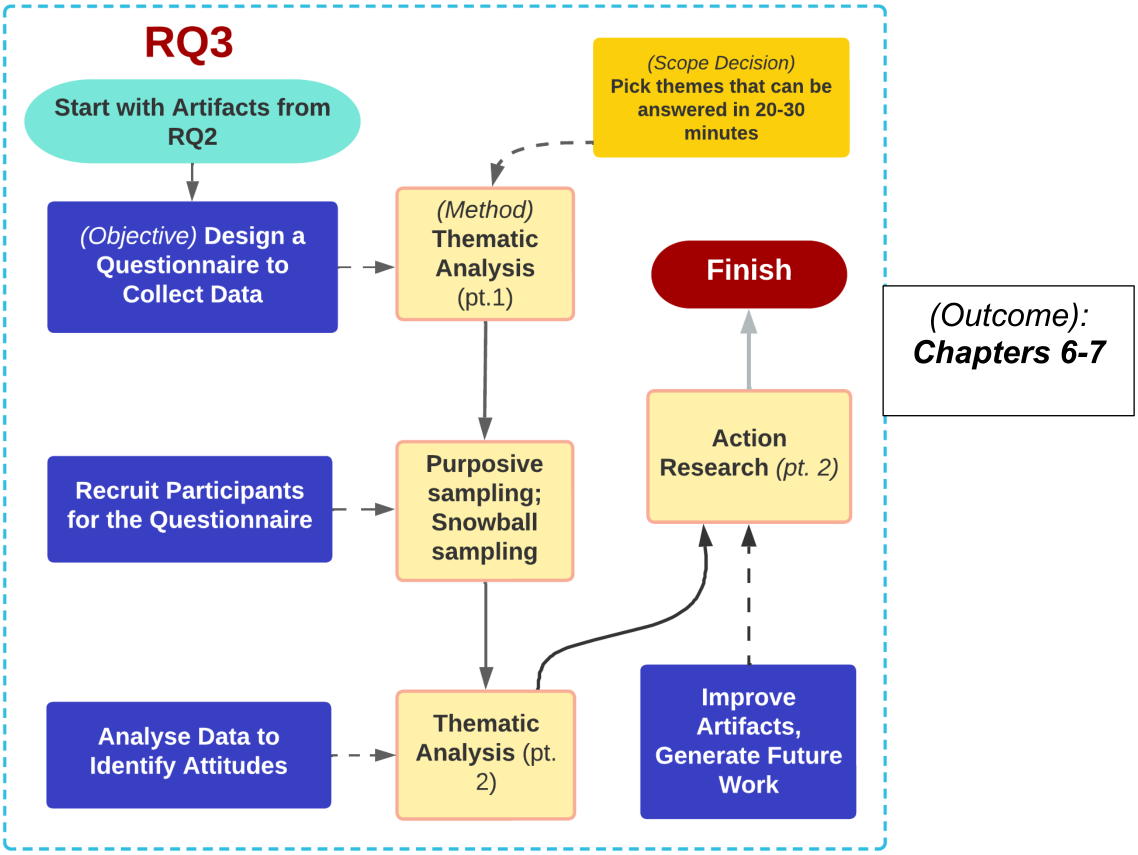



**Figure 25** - Methods for answering RQ3

#### 3.2.3.1 Questionnaire-based Survey and Thematic Analysis

In the second iteration of Action Research (*Step G – "Take the next action"),* a questionnaire is designed to ask researchers to reflect on decentralised approaches to scholarly communications. It is expected that the survey will encourage its participants to critically analyse their own practices as researchers and scholarly communicators and thus provide scope for answering RQ3. A crucial outcome is that there will be a better understanding of the wider, social implications of a novel decentralised approach to SC using a blockchain, by evaluating the artifacts from the earlier stages of the research in a social setting, and asking the stakeholders about their perceptions related to decentralised SC.

As a huge chunk of insightful data from the survey will be qualitative in nature, it is therefore important to select an appropriate qualitative research method to extract relevant insights from this data. Thematic Analysis (TA) is one widely used foundational method in the field of qualitative research (Braun & Clarke, 2006) and it is particularly helpful, as it can be used for both generating the initial questionnaire, as well as for analysing the data collected from the survey (after the questionnaire has been deployed). It is a method for "identifying, analysing, organizing, describing and reporting themes found within a data set" (Braun & Clarke, 2006). From their examination of different approaches to qualitative analysis, Nowell et al. (2017), have outlined a practical procedure for carrying out TA that meets the trustworthiness criteria by Lincoln and Guba (1985), which include credibility, transferability (ability to generalise inquiry), dependability (quality of documentation) and confirmability of the research, all of which require an audit trail specifying the choices made by the researcher. Tobin and Begley (2004) specify that any such audit trail should be reflexive, and include the internal and external dialogue of the researcher. Nowell et al. (2017) thus specify the following phases of TA presented in **Table 4** below:

1. **Familiarising yourself with the data**
2. **Generating initial codes**
3. **Searching for themes**
4. **Reviewing themes**
5. **Defining and naming themes**
6. **Producing the report**

**Table 4** - The steps of Thematic Analysis prescribed by Nowell et al. (2017)

Thus, TA involves identifying patterns, themes and meanings emerging from qualitative information. In my methodology, I concretely implemented the TA model in the following specific way:

1. Review the artifacts from the earlier stages of the research (research documents, literature reviews, observations, designs, models, test results, use cases). Reinforce the understanding of the social setting in which the decentralised scholarly communications approach is being introduced and the key themes and concepts that have emerged from the earlier stages of the research.

2. Identify preliminary themes or concepts related to decentralised scholarly communications and their wider social implications from the artifacts being reviewed. Look for recurring ideas, concerns, or perspectives expressed by stakeholders/actors involved in the research or relevant to the research context.

3. Generate Questions: Based on the initial themes, generate open-ended questions that encourage stakeholders/actors to express their perceptions, opinions, and experiences related to the decentralised scholarly communications approach and its social implications. Use language that is clear, concise, and easily understandable for the target audience of the questionnaire. Ensure that the questions are specific and focus on capturing stakeholders' views on key aspects of decentralisation and related themes, such as openness (Open Science, Open Access), fairness, transparency, reputation, impact, and any other relevant themes from the Actor-Network. Make a scoping decision to limit answering the questions to 20-30 minutes to maximise completed surveys.

4. Review and Refine Questions: Review the generated questions and assess their relevance and clarity in addressing the research objectives. I did

this by discussing the questions regularly with my PhD Supervisor. Ensure that the questions align with the identified themes and capture the nuances and complexities of the wider social implications of the decentralised approach. Seek feedback from colleagues, experts, or pilot participants to refine and improve the clarity and appropriateness of the questions.

5. Organise Questions into Sections: Group the questions into logical sections that reflect the different dimensions or themes identified from the artifacts and earlier stages of the research. Consider the flow and coherence of the questionnaire, ensuring that the questions progress in a logical and meaningful sequence. The resulting survey design is given in Chapter 6 section 3.

6. Pilot Testing. Conduct a pilot test of the questionnaire with a small sample of stakeholders/actors to assess its clarity, comprehensibility, and effectiveness in eliciting the desired information. Seek feedback from pilot participants to further refine and improve the questionnaire, if necessary. By applying thematic analysis to the artifacts and earlier stages of the research, you can identify key themes and generate relevant questions that capture stakeholders' perceptions regarding the wider social implications of the decentralised scholarly communications approach using blockchain. This approach helps ensure that the questionnaire aligns with the research objectives and provides valuable insights into the social dynamics and implications of the novel approach.

#### 3.2.3.2 Survey as the Link between Actor Network Theory and Action Research

Action Research and Actor-Network Theory (ANT) are two distinct research approaches that can complement each other when designing a questionnaire-based survey into the attitudes towards novel scholarly communications software. One way in which action research can link to ANT is by adopting a network perspective when analysing social phenomena. This means that researchers should pay attention to the relationships between human and non-human actors, and how they contribute to the emergence of social structures and processes. By doing so, researchers can gain a deeper understanding of the complexity and dynamics of the social world and identify leverage points for change.

Regarding the design of a survey-based questionnaire following both action research and ANT principles, it would be essential to involve stakeholders in the design process. Thus, relevant actors, such as community members, should be engaged. Additionally, attention should be paid to the relationships between human and non-human actors when designing survey questions. This means that we should consider how technologies, policies, and non-human actors shape social phenomena and how they can be used to facilitate change. This means that the survey should consider the role of human and non-human actors in shaping attitudes towards the novel, decentralised scholarly communications platform. Throughout the survey design and implementation process, it is important to engage stakeholders in a collaborative and participatory manner. This can help ensure that the survey captures the key issues and concerns of stakeholders and that the findings are relevant and actionable.

## 3.3 Summarising the Methodology

The methodology used to reproduce the results of this thesis is summarised below for conciseness. As the methodology was iterative in nature, there was a feedback loop (Repko, 2008, p. 46) making me periodically revisit earlier parts of the research.

1. Perform Systematic Literature Review, Stakeholder Analysis and **Actor-Network Theory** research on SC
    a. Add the findings to the requirements for the newly proposed software to decentralise SC by applying Activity Theory to generate relevant use cases for the two groups of actors that have been chosen (Authors and Reviewers);
    b. Revise the ANT as appropriate when new information becomes available to the researcher.
2. Plan **Action Research** informed by (1)
    a. Build the software to decentralise SC by applying software development lifecycle and Agile methods to build and test smart contracts in a DAO set-up on a blockchain;
    b. Evaluate the software focusing on the two chosen metrics – speed and cost.
    c. Revise the research artifacts as appropriate when new information becomes available to the researcher.
3. Continue Action Research with a **Survey** (based on a Questionnaire) to evaluate the attitudes of the relevant actors (personas) to the findings from step (1) and the software from step (2)
    a. Use Thematic Analysis twice, the first time to generate the initial set of questions for the survey based on the previous

research artifacts, and the second time to generate a list of findings after participants attempted the survey;

b. Analyse the validity of the concepts, scenarios and the most significant evaluation metrics for the software;
c. Revise the ANT as appropriate when new information becomes available to the researcher;
d. Revise the software requirements and the software build as appropriate when new information becomes available to the researcher.

In the next Chapter (4) of this thesis, I will illustrate how throughout many scholarly interactions, intermediaries can be removed, and a new mechanism can be created requiring that trust is derived from the collaborating parties instead of arbitrary middlemen. This requires arriving at a novel method of **decentralising with blockchains**, which I design and document. The framework presented in the next Chapter (4) will bring us closer to describing and understanding the main implementation of the Smart Papers system (Chapter 5) that is the main contribution of this Thesis, and also to its evaluation, as presented in further Chapters (5, 6, 7).

# Chapter 4

# Designing a Method to "Decentralise" a System

*It will be said, of course, that people will go to the best shop... I deny that the big shop is the best shop; and I especially deny that people go there because it is the best shop.*

-**G.K. Chesterton** in *The Outline of Sanity (1926)*

As outlined in Chapters 1 and 2, there have been calls for improving the current ways in which scientific research is disseminated. Modern research has been criticised for its reliance on big publishers whose interests are not always necessarily in line with academia's interests (Smith R. , 2006). Whilst centralised operations provide economies of scale, this benefit must be balanced against the danger of a centralised organisation becoming too dysfunctional, through diseconomies of scale. Many cautionary tales immediately come to mind, stories of banks that are too big to fail, big brother governments, record labels telling artists what they can or cannot record, and publishing houses able to restrict public access to scientific literature based purely on profit motives. To find the ways of addressing the dangers of centralised scholarly publishing and scholarly communications, Chapter 3 has outlined the methodology used to arrive at a detailed framework that can be used to decentralise scholarly communications. This chapter models the status quo, discusses centralisation and decentralisation, and offers a decentralisation step-by-step approach, explaining how it applies to subsequent contributions of this thesis.

## 4.1 Modelling a Scholarly Sharing Economy with Blockchains

There is a potential for re-negotiating the role of scholars and universities in today's global market economy. (Bullini Orlandi, Ricciardi, Rossignoli, & De Marco, 2018) make their case for the self-organising capabilities of the global scholarly system according to its unique set of institutions, norms and values that have evolved throughout centuries. Meanwhile, in (Beverungen, Böhm, & Land, 2012), another strong argument is delivered for the further development of open access repositories, a fair trade model of publishing regulation, a renaissance of the university presses, and, finally, a move away from private, for-profit publishing companies toward autonomous journal publishing by editorial boards and academic associations. These narratives point towards the function of the University as an "anchor of trust" that can foster scholarly interactions that do not rely on a centralised publisher. Autonomous journal publishing could be achieved by empowering scholars within the realm of academia, by using sharing economy principles without relying on publishing houses. Blockchain, as a technology used to allow economic interactions in a so-called "trustless" environment (assuming that participants may be competing but some sort of trust will happen as a by-product of everyone's incentives being aligned, and it will be guaranteed through the use of cryptographic techniques), should be evaluated as a means towards achieving this goal.

In their survey, (Hawlitschek, Notheisen, & Teubner, 2018) discuss the potential of blockchain technology in today's sharing economy landscape by following an interdisciplinary approach. In their dual literature review of blockchain and sharing economy papers, they find that blockchain technology can be used to provide a trusted platform for scholarly economy interactions. I argue that the mix of permissioned and permissionless blockchains running specially designed smart contracts, can be used to model an autonomous blockchain-mediated digital journal environment that can operate in a decentralised manner. This can be curated by the scientific community by reusing the existing structures of trust within Universities themselves, without relying on a publishing house. Before such an environment can be designed, the existing environment must be mapped and a method for change must be defined, arriving at a new model.

## 4.2 The Actor Network Modelling the Status Quo

The actor network in **Figure 26** has been mapped for the existing scholarly communication network. It is aimed at capturing the most salient actors and actants, with attention drawn to incentives and relevant artifacts:

- The most significant human (pragmatic) actors are scholarly authors, scholarly editors, scholarly leadership and grant funding bodies. Note that these have been identified by the Stakeholder Analysis part of the Methodology.

- There are also complex-concept (non-pragmatic) actors that shape scholarly communication - the most important ones being policy and reputation, but also impact, prestige and discipline-based knowledge organisation. These have been identified by the ANT process of *translation*, i.e. identifying the concepts needed to align the human actors' motivations; identifying how the pragmatic actors' interests are expressed in the world.

- Examples have been provided for most commonly used open-source manuscript management systems and the proprietary ones, highlighting the variety of offerings in this space. This expresses the ANT idea of *black-boxing*, e.g. identifying the relevant concepts that are in themselves "sealed actor-networks".

- The purpose of the connections (arrows) in the diagram is to show the influences exerted by the actors onto the world and on each other. The disconnect between the open-source-inspired and closed-source (for-profit) scholarly worlds is reflected through the central symmetry in this network.

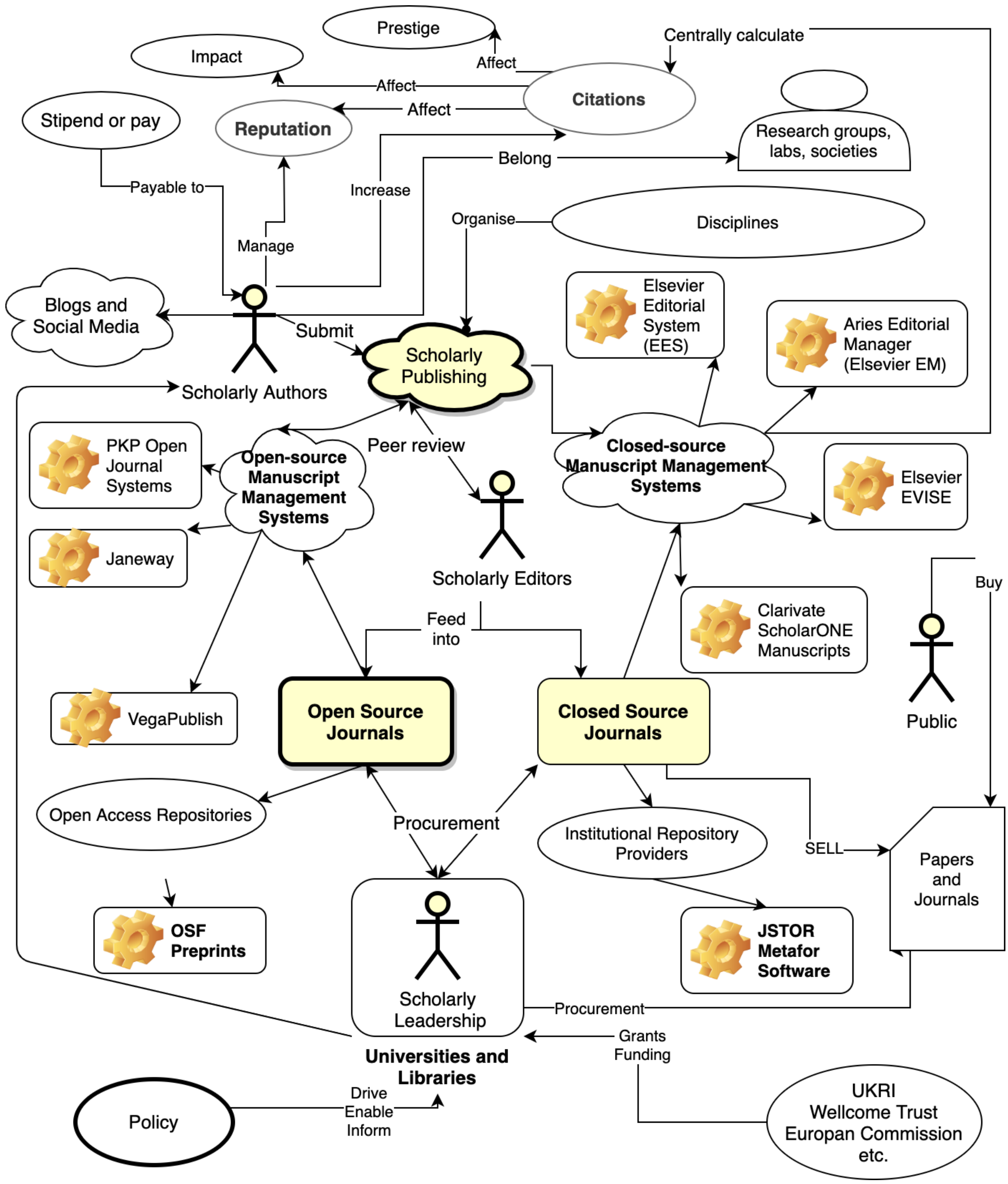



**Figure 26** - Actor Network for Scholarly Communication

## 4.3 Degrees of Centralisation

The actor network developed in the previous subsection can be described as centralised via the use of closed-source manuscript management systems that translate the interests of scholarly editors, without translating the interests of other pragmatic actors. But to understand the idea of *centralisation* in itself, a multi-faceted analysis is required. One common understanding of centralisation is technological, and is rooted in computer network analysis. Baran (1964) proposed that all computing networks can be classified as *distributed*, *decentralised* or *centralised* (see **Figure 27**). Whilst this classification was originally used in a survivability discussion (pointing to a superiority of a distributed design of a network without a single point of failure), in recent years, it has been increasingly used as a template to discuss network power dynamics (again, pointing to a superiority of a distributed design of a network in which value is exchanged in a peer-to-peer fashion). It must be pointed out that modern Web communications are layered, so that a service provider could provide a centralised service (for example, a pay-walled website of a publishing house backed by a central database) even though it is running on top of the underlying Internet protocols, like TCP/IP, which are fairly de-centralised. Therefore, in the remaining discussion, I will only focus on the top layer (the service layer).

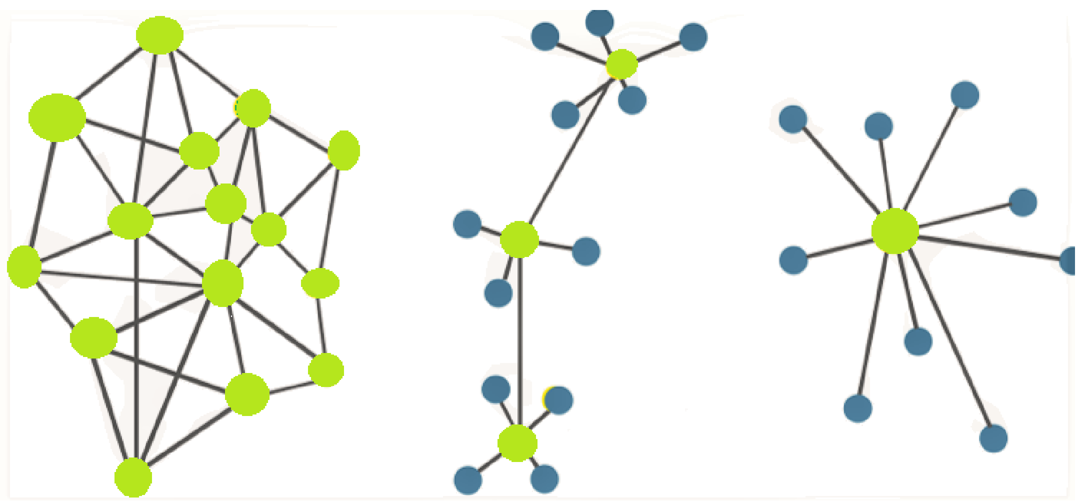


**Figure 27** - Distributed, decentralised and centralised networks.

A crucial advantage of distributed-ledger systems over their traditional alternatives that rely on centralised databases has been decentralisation - and it has certainly become a buzzword in the crypto-economic community, giving rise to the closer scrutiny of this term by researchers. In `Deconstructing Decentralization', Walch (2019) argues that the word decentralisation is used in blockchain discourse to describe distributed systems that *are resilient and lack concentrated power centers.* She identifies the distinct sources of meaning of decentralization (decentralisation) as coming ffromlaw, political theory, computer science and networks theory, potentially yielding various interpretations of the word.

Of utmost importance is the observation that decentralisation is not just elusive (a moving target) but also lies on a spectrum rather than being binary. From an ethical angle, the author hypothesises a moral risk that she calls a **'veil of decentralisation'** that dilutes human accountability in a decentralised system. Despite the importance of the centralisation-decentralisation continuum, most discourse on this topic lacks agreed-upon metrics and is therefore devoid of proper quantification. We can, nevertheless, look at some ways of measuring it in the context of distributed-ledger ecosystems.

## 4.4 Measuring and Interpreting Decentralisation

Despite the importance of the centralization-decentralization continuum (defined as the optimal allocation of authority in the "political compromise" outlined by Atzori, 2015), most examples of the discourse on this topic lack agreed-upon metrics and are therefore devoid of proper quantification. To measure the degree of the decentralization of power it is necessary that one first quantifies this power and analyse its distribution, both of which are complex tasks and not easy to achieve as researchers usually focus only on specific facets of decentralization whilst ignoring its other aspects (Pina Sanchez, 2014, p. 12). Since decentralization is not only defined but also described and measured in a plethora of ways, this methodological hurdle contributes to inconsistent research outcomes (Sharma, 2006, pp. 53, 55). Outside of the blockchain context and to build a working definition of technological centralization and decentralization, King (1983) introduces three decentralized dimensions that have to be measured – the locus of decision making (is it concentrated in one person, a small group, or rather dispersed across various levels), the placement of facilities (are the facilities in one place or spread around), and the locus of function (operations), observing that there is no universal "best" solution and that each organization has to find its unique mix of parameters. In the context of blockchain, there are a couple of issues with transplanting King's framework, the key issue being that the approach is not a technologically oriented one.

Another problem with King's framework's applicability to blockchain is that the author assumed a level of tangibility of "facilities," the meaning of which appears elusive in the current era of cloud computing, whereby IT facilities are now both global and on-demand. One other, similar, criticism would be the immateriality of the locus of operations in today's era of globalization, given that the largest corporations are not taxed locally, and their operations cannot be easily traced to particular geographical locations. It can be, therefore, argued that a more viable blockchain decentralization framework should be simpler, mostly focusing on the locus of decision making, to measure the bargaining power of the individual in their personal self-determination. It is not particularly surprising to note that no such framework had existed as of yet, but a contribution of this thesis was developing one.

Nevertheless, some very specific metrics in the blockchain space have been introduced to measure blockchain decentralization. In their online post, the authors Srinivasan and Lee (2017), propose the minimum Nakamoto coefficient as a simple, quantitative measure of a system's decentralization inspired by the Gini coefficient. The metric focuses on the size of each of the network's subsystems, and in the case of Bitcoin, these are the clients, developers, miners, exchanges, nodes and accounts. If any of these six subsystems becomes unreasonably centralized, the total metric will reflect that centralization. As the minimum Nakamoto coefficient increases, the minimum number of actors required to compromise the whole system thus also increases. Looking at decentralization metrics from an analytics angle, the founder of Dogecoin, Jackson Palmer, created the AreWeDecentralizedYet.com (AWDY) website. The AWDY website lists in real time, and for each major blockchain, the values for the following metrics:

***(1)*** *The number of client codebases that account for 90% of nodes;*
***(2)*** *The number of entities in control of more than half of total mining power;*
***(3)*** *The number of public nodes;*
***(4)*** *The amount of money supply held by top 100 accounts.*

If the Web data published by AWDY is to be trusted, major blockchains appear to be highly centralized, often in more than just one respect, failing to live up to their decentralization aspirations and their social media claims promoting the "veil" of decentralization. For example, just four Bitcoin mining pools comprise more than 50% mining power of the blockchain. Meanwhile in Ethereum, approximately one-third of the total money supply is in the hands of top 100 stakeholders. The above examples illustrate that it is an important responsibility for blockchain developers and researchers to regularly monitor, report and reflect upon these diverse metrics reflecting the reality of the decentralized/centralized spectrum, and to gauge whether any trends are present or emerge in the long term. More discussion, involving new emerging scholarly discourses, should be based on the nascent blockchain decentralization metrics to counter the decentralization hype which is often spread on social media without any facts to back up common blockchain "decentralization" claims.

Due to the multi-layered nature of the Internet, network decentralisation of communications and even individual transactions does not necessarily lead to disintermediation of services associated with peer-to-peer value exchange. Therefore, the re-emergence of `decentralised intermediaries' acting as gatekeepers of decentralised networks is a threat, in that it may introduce unwarranted complexity without providing a clear benefit to end users of those networks. We have recently seen closed-source crypto exchanges appear, which is an example of non-disintermediated (siloed) decentralisation that replicates the power dynamics of centralised services, in which trust is brokered by special third parties that quickly gain an unfair advantage over other players.

Even though we have just shown that decentralization can be measured in numerous ways, understanding decentralization is not merely a quantitative problem. As a social construct, it is qualitative and subtle, meaning that one must be mindful not just of the different metrics, but also of the different interpretations thereof. Zuboff and Maxmin (2004) point out that the "false" dichotomy of centralization and decentralization should not be confused with the fundamental problem of opposing concentrated authority inherent to the hierarchical managerial structure of the capitalist firm which has its origins in concentration. They posit that any quantitative measuring of "decentralization" could be a red herring, simply resolving the level of delegation of concentrated authority, without any consideration given to the breaking up of that concentration on a more essential level. The authors propose that instead of simple decentralization, stepping beyond the constraints and limitations of managerial capitalism would need to involve a more fundamental shift in enterprise logic through the creation of a new paradigm. With that in mind, they propose a framework of "distributed capitalism" where individuals are recognized as the source of all value and all cash flow. Because value is distributed and lodged in individuals, what this necessitates is distributed production, distributed ownership and distributed control. Even though Zuboff and Maxmin wrote their pivotal work a couple of years before first blockchains emerged, the so-called "federated support networks" proposed by them are organized in a way that can link enterprises under "distributed capitalism" in a manner "providing unique aggregations of products and services" whilst benefiting from the fact that "infrastructure convergence dramatically lowers operating costs and working capital," all of which is essentially describing an archetypical structure similar to that of a blockchain-based ecosystem.

## 4.5 Decentralization Framework With Blockchains

Anfara and Mertz (2014) define a theoretical framework as any empirical theory of social processes that can be applied to gain an understanding of a phenomenon. The extensive systematic review reported earlier in this work suggested that a framework for blockchain decentralization does not currently exist. The present paper has, thus far, provided a theoretical reflection on the meanings and understandings of the concepts of decentralization across scholarly and blockchain-related contexts, revealing decentralization as a multifaceted concept with several dimensions.

Based on the discussion above, one may envision a tentative framework which will help the future researcher direct their focus in their future work by breaking down decentralization into its dimensions and examining which ones can be tackled by means of blockchains and smart contracts. The main mechanism for gaining that focus is through clarifying the relative importance and meaning of technological, social, political and administrative imperatives and incentives, which can be achieved by answering the following questions: what is being decentralized and why (the case for decentralization), and how much of it are we going to decentralize and how will blockchains help (the scope and the action plan for decentralization with blockchains). The resultant frameworks is represented visually as two workflows: in **Figure 28,** the individual steps are discussed for establishing the context and formulating the decentralisation problem, so that it's well-formulated, justified, quantifiable and scope-constrained. Additionally, **Figure 29** provides much detail on how the framework should be used to develop an action plan for decentralising a system.

### 4.5.1 Problem Definition and Establishing the Case for Blockchain Decentralization

The "what" and "why" questions are the most essential ones to answer, so as to avoid an incorrectly designed proposal that does not address any issue in particular, which can be described as a "*solution looking for a problem*"[11]. There are, in fact, many good reasons for decentralizing the status quo with blockchains, for example, to improve the bargaining power of the end-users of traditionally provided financial services, whole marketplaces, and cloud-based IT applications. This is because most centralized socio-economic institutions and systems such as governments, banks and corporations, may share one or more negative issues with the way in which they handle, store, process, share and give access to individuals' data records. These issues can involve a perceived loss of control by the individual (low bargaining power, feeling of exploitation), loss of trust, unnecessary complexity (bureaucracy, overhead), vulnerability (security breaches), inefficiency (slowness and unresponsiveness), perceived injustice (unwarranted profits, lack of redistributive efforts), opaqueness (lack of transparency and accountability), as well as a propensity to corruption (decay and degradation of service quality), which are the themes identified in the SLR.

As each institution is different in terms of size, purpose and structure, and the issues outlined above may manifest themselves differently under different scenarios, it is essential that the researcher first defines which issue it is exactly that is being tackled, which should point them at the most suitable course of action. For example, when tackling a market complexity issue, we may want to "decentralize" the market to improve the coordination between the different stakeholders to better enable them to match their preferences to the options that are available to them. Opaqueness, in turn, calls for an increase in the transparency and availability of the data flows, such as a provision of user-friendly search and indexing functionalities for reviewing live transactional data and historical audit trails. However, institutional opaqueness may be caused by one of two factors, either an institution deliberately trying to conceal information, or inadvertently, when poor informational design prevents users from successfully consuming the relevant search and retrieval functionalities (for instance, lack of open data indexing and/or a lack of accessibility due to poor UX design). As many blockchains are notoriously difficult to use for non-technical audiences, a blockchain may only ever partially help opaqueness, until there is more effort in the blockchain sphere that focuses on providing an adequate level of user experience and accessibility, particularly so for the digitally excluded populations. Therefore, blockchain limitations must be kept in mind whilst designing an appropriate solution.

When it comes to vulnerabilities and inefficiencies of traditional siloed institutions, there is no "one size fits all" blockchain solution to addressing those maladies. Nevertheless, appreciating that many vulnerabilities, such as customer data leaks, and inefficiencies, such as long turnaround times for processing requests in bureaucracies, are often caused by a mixture of causes, from the "single point of failure" factor to inadequate system design, can offer some hope that a better system can be designed, possibly with blockchains. Whilst small-scale complexities and vulnerabilities lend themselves to "quick fixes," outright institutional corruption calls for outright institutional reform, meaning that any decentralization approach would need to be complex and concerted, and not necessarily focused on blockchains, although they would certainly play a role for encouraging *trust* between the transacting parties. Multiple scenarios revolving around trust inspired the creation of blockchains in the first place (Gaggioli, Eskendari, S, Cipresso, P, & Lozza, 2019) such as enabling transactions in environments where people distrust each other, minimizing trust on central parties, and enhancing trust through transparency and computational verifiability. According to Golbeck's (2008) survey, online trust is required to "*foster successful interactions and to filter the abundance of information*," it also has three main targets – content, services and people. Therefore, when establishing the case for blockchain-mediated decentralization, one must be clear about the sort of trust that is required and the relevant reasoning. As depicted in

**Figure 28**, only by first establishing the reason for decentralization (the problem, the system and the case), can the right approach be sought, and the complexity of the efforts estimated.

### 4.5.2 Defining the Scope of Blockchain Decentralization

Defining the motivation for decentralization and clarifying the "business case" should naturally help the researcher who attempts to design a decentralized system using blockchains, equipping her with a better idea of what the scope of decentralization should be. When analyzing the reason for decentralizing, the researchers' attention will be ultimately brought to the ultimate human-centric constituents of socio-technical systems, i.e., the stakeholders, main actors and beneficiaries, all of whom will now be well-defined and documented, their relationships delineated, and incentives mapped out. Next, it will be necessary to define which elements of the system require decomposition and/or which workflows and processes may benefit from disintermediation. Decentralizing significant chunks of infrastructure may be achieved by the introduction of peer-to-peer infrastructures and distributed systems in place of server-based solutions but there needs to be a level of clarity as to which parts of the complete solution may have to remain centralized, and, as no blockchain is an island, which parts of this solution will need to be made interoperable, and to what extent. In terms of disintermediation, one must plan for second system effects such as the scope in which previously non-existent intermediaries may be incentivized to form, and the negative "bloating" impact that this emergence would have on decentralization outcomes.

One must also keep in mind, however, that decentralization is as much technological as it is sociopolitical in nature. Thus, one will never just decentralize the infrastructure. What may need to be decentralized additionally are the modes of creation (for example through peer production), the rights and responsibilities of the stakeholder (such as fiscal responsibility), managerial control (through a decentralized autonomous organization), and governmental power (e.g., through devolution, citizen education and activation, or via co-management approaches). Regardless of the above, it could be the size of the organization that becomes an issue if the system becomes "too big to fail," which would lead to an antitrust sort of decentralization, so as to prevent predatory practices such as monopolies and/or oligopolies, and to encourage healthy competition that is fair and stimulating.

### 4.5.3 How Are We Decentralizing with Blockchains?

The final consideration of the decentralization action plan should be to establish the manner and mode of the proposed change in power/authority structure. Broadly speaking, decentralization can be implemented as either a top-down or a bottom-up exercise. As both approaches are vertical in nature, and decentralization emphasizes horizontal relationships, one must be mindful of the "exit strategy" for the temporary verticalisation of efforts. Any top-down method will require an existing critical mass that may already be centralized, and to which it can be applied. In that approach, one has a choice of administrative reform, antitrust proceedings, deconstruction, deconcentration, decomposition, decoupling, co-management and power transfers. In contrast, bottom-up methods can be used to design decentralized systems from scratch and are more applicable to the sphere of blockchain. In those approaches, the available tools and approaches include mechanism design (which is about creating markets or other mechanisms matching individuals through the reverse application of game theory, i.e., through designing individual incentives and modeling the related payoffs), distributed systems design (which is about creating the necessary infrastructure to relay and process the communications), encouraging active participation (which may involve education and campaigning activities), designing a decentralized autonomous organization (requiring software and Web skills to implement) or creating simple voting-oriented "democracy" dApps (distributed applications) by means of coding smart contracts and deploying them on blockchains. When designing these completely new mechanisms, infrastructures and applications from scratch, it is necessary to carry out enough experimentation and observation to rule out any negative externalities and second-system effects, such as new centralized loci of novel middlemen appearing in place of old ones that were removed.

On a technical level, smart contracts may only be viable if their intentions and actions are understandable to the wider population. This accountability can be achieved through coupling them with legal prose. This is achievable through establishing so-called Ricardian contracts (Grigg, 2004) which can link the legal intent to smart contracts in the form of tuples: *{legal prose, transaction parameters, smart contract code}* (Clack, Bakshi, & Braine, 2016). Additionally, the design of a decentralized system comprised of Ricardian-style smart contracts should involve the definition of stakeholders, their rights and responsibilities, economic incentives and preferences, and what top-down or bottom-up changes will be required. As there is currently no standardized format for this, a Linked Data ontology should be created to manage decentralized sociotechnical initiatives. An example of that can be found in a vision paper by Hoffman (2018) which proposes the use of smart contracts used in tandem with Linked Data documents to improve the transparency of global taxation.

Finally, all blockchain-mediated decentralization requires a flexible yet structured approach to governance design and implementation. The governance of a blockchain platform is not necessarily the same as that blockchain's rule-based operational logic. Instead, by blockchain governance we understand conflict resolution, dispute resolution, crisis management and all sort of sociotechnical mechanisms that facilitate higher-order decision-making regarding those cases that day-to-day operational logic does not address, including resolving uncertainties under incomplete contracts. Well-designed blockchain governance covers a plethora of aspects (Barrera, 2019) from proposal-making mechanisms to voting rights. It is also essential to keep in mind my findings that decentralised governance has different levels, and if a decentralized blockchain solution can successfully govern itself (higher-order governance) then it can be successfully used to implement other types of governance such as fiscal or environmental governance (first-order governance). As a final note here, it is important to consider the conclusions of (Arruñada & Garicano, 2018) that in order to protect individuals and help coordinate their efforts with blockchains, new forms of decentralized governance are required.

### 4.5.4 Blockchain-Mediated Decentralization

I will build on top of the definition by (Arruñada & Garicano, 2018) of a "platform" as the "*combination of software and hardware resources enabling the functioning of an exchange network*," where "network" means "*the community of individuals exchanging goods or services through a platform.*" With those definitions in mind, drawing from my systematic review findings, and taking into consideration the framework proposed in the previous chapter of this work as well as the distributed capitalism framework from (Zuboff & Maxmin, 2004) discussed earlier, I define "**blockchain-mediated decentralization of a system**" as (**Def. 1**):

> *the technique for designing a new federated support network using a combination of blockchain and P2P platforms, as a means for a particular community to address one or more shortcomings in terms of inefficiency, opaqueness or vulnerability of the existing system used by that community for a particular purpose, by introducing tamper-proof records, incentives, rules and workflows aimed at breaking up the inadequate concentrations of power in the existing system, in a way that allows any subsequent improvement to be reported with agreed-upon metrics.*

As an example, consider the creation of Bitcoin, whereby a novel network powered by a blockchain was designed, in the midst of the financial crisis and bank bailouts, for the global community to pay for products and services without relying on controversial central banks, utilizing the novel incentives of cryptocurrency mining, yielding an independent global payment system. Alternatively, consider MakerDAO running on top of Ethereum, where users can take out loans backed by crypto-backed collateral, in a way that is dynamically federated by multiple organizations, thus yielding a system that creates value by enabling independent access to credit, addressing the barriers to entry associated with traditional lenders, without the need for middle-man arbitrators. Also consider Steem (or its competing fork Hive, accessed through Peakd), a blockchain-based content rating dApp, where users vote on the content and stake their cryptocurrency, increasing the pay-out available to the creators of the most popular content, without biases associated with traditional media and publishing outlets. Using the above-introduced definition, it can be said that Bitcoin attempts the blockchain-mediated decentralization of currency, MakerDAO attempts the blockchain-mediated decentralization of credit, and Hive attempts the blockchain-mediated decentralization of web content curation and publishing. The success metrics for Bitcoin may include one of the AWDY metrics introduced earlier, such as the amount of money supply held by the top 100 accounts. Success metrics for MakerDAO and Hive may focus on the level of perceived improvement in accessibility and efficiency of the respective solutions, as compared to their centralized counterparts.

In light of the understanding of the meaning of *decentralisation* as offered by this chapter, the next chapter will focus on the innovative distributed-ledger technologies of blockchain, smart contracts and distributed autonomous organisations, to outline how I successfully use them to coordinate transactions (information and value exchange) and governance decisions in scholarly communications.

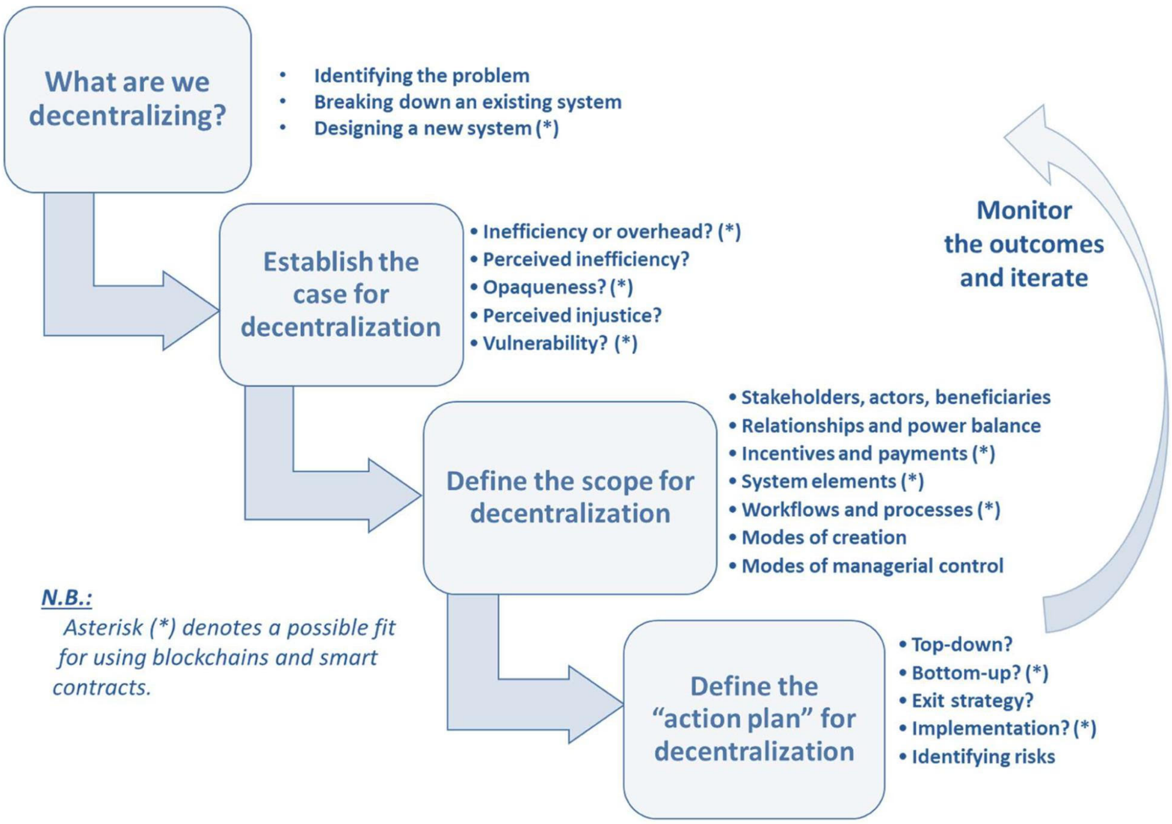



**Figure 28** - Framework for decentralizing a socio-technical system, with the option of blockchain application.

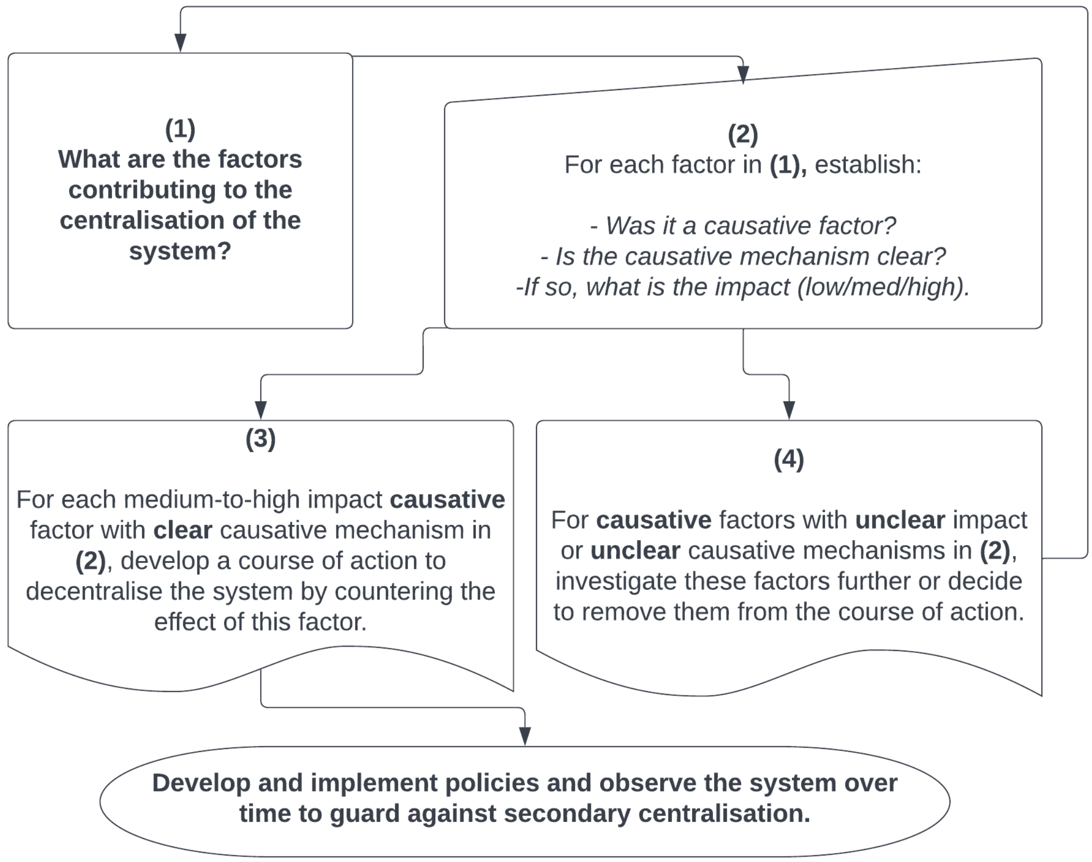



**Figure 29** - Framework details on how to develop the "Plan of Action" for decentralising a system.

## 4.6 Towards Decentralised Open Science With Blockchains

By following my framework described in the previous subsection and removing the need for traditional centralised trusted third-parties, blockchain technology can be used to decentralise systems used to perform activities (including collaborative activities) in a way that eliminates the middle-man. Whilst blockchains have been criticised as overhyped, with hundreds of blockchain-based projects popping up recently, offering solutions to everything from efficient road repairs to improved electronic voting, I frame blockchain technology as a type of foundational technology that provides trust management for the academia without a big publishing house mediating that trust. Hence, the use case is very clear, justified, scope-constrained, and much of the hype associated with blockchains is avoided. Within this well-defined setting, blockchains can provide a single version of truth to everyone involved, making all sorts of agreements much easier to manage and more auditable. Blockchains can also be used to model state machines, which makes them ideal for modelling control flows (workflows) such as the publishing flow, which consists of multiple stages that need to be completed in a particular order for a paper to be published.

### 4.6.1 New Actor Network Proposed for Change

The actor network in **Figure 30** has been mapped for the proposed scholarly communication network utilising Smart Papers and DAO4PR (a peer-review Decentralised Autonomous Organisation, such as a journal or a conference editorial group), the two major contribution of this thesis. The "plan of action" framework depicted in **Figure 26** was instrumental in creating this network. The "centralising factors" of the framework were identified as closed-source manuscript management systems and these were then broken down, or "decentralised". This was achieved by applying the workflow depicted in **Figures 28-29**, with the following steps outlined below for completeness:

- Identifying the problem
    - Re-design of an existing complex system, focusing on the following key actants from the original ANT:
        - Manuscript management systems
        - Papers and journals
        - Repositories
        - Reputation
- Establishing the case for decentralisation
    - Perceived inefficiency and opaqueness
        - Closed-source publishing workflows
        - Suboptimal formats for authoring and publishing
        - Centralised storage and gatekeeping
- Define the scope for decentralisation
    - Stakeholders
        - Scholars, Editors, Funders, Libraries, Universities, Public
    - Workflows and processes
        - Collaboration & authoring, peer-review, publishing
    - Modes of managerial control
        - Distributed Autonomous Organisations (DAO Research Groups, DAO editorial groups, DAO "journals")
    - Incentives
        - Tokens, TCRs

The new network is thus aimed at capturing the most salient actors and actants, with attention drawn to public participation, realigned incentives and relevant artifacts. The most significant human actors are scholars, who can act in different roles (as authors, editors etc., based on their reputation), scholarly governance mechanisms through the DAO, and grant funding bodies interacting through blockchain.

DAOs, explored in the previous sections in this thesis, are used to store scholarly data and metadata, help scholars govern their team efforts and introduce a tokenised way of exchanging value - tokens are separated from monies. Tokens have been introduced for multiple purposes, the main one being tracking the 'value' (not necessarily monetary) accumulated by a Smart Paper, and the expected value staked against it.

Tokens can be further aggregated into TCR's - token curated registries, which are lists of Smart Papers (and also of scholars) that can be sorted according to different criteria. Specially designed non-transferrable tokens are also used for tracing a scholar's reputation. Tokens may also be useful for editors to help them express their interest in particular Papers. The role of Open Science in shaping policy and behaviour change has been highlighted. Reputation remains a very important non-pragmatic actor in the new network.

The actor network also assumes a possibility of a set-up in which journals are not the main media for aggregating papers - as these can be aggregated by keywords (discipline or topic based) or manually by members of a DAO, such as a conference DAO or a special interest editorial DAO. The list of key elements for this Actor-Network was, thus, derived based on a comprehensive understanding of the scholarly communications landscape, as well as a consideration of the different actors and their roles in the system. The reasoning behind the selection of these elements was based on their importance and impact on scholarly communications in the context of decentralising the overall system. It places key emphasis on some common actors that have already been identified by other ANT researchers of SC, for example Latour (1987), Wagner (2010), Boyd Rayward (2005), and others, but it could be improved by introducing more complex actors, for example communities such as research communities and open-source communities, and more discrete actants such as regulations, patents and knowledge claims. The following Table (5) allows the reader to review the relevant ANTs in the literature to make it easier to perform comparisons with the ANT presented in this thesis.

| **Paper details** | **SC Actors identified** | **SC Actants identified** |
|---|---|---|
| (Latour, Science in Action: How to Follow Scientists and Engineers Through Society, 1987) | Scientists, engineers | Instruments, texts, laboratories, scientific journals |
| (Callon, Law, & Rip, 1986) | Scientists, engineers, policymakers, funding agencies | Instruments, texts, laboratories, scientific journals, funding policies, scientific committees |
| (Bowker & Star, 1999) | Experts, librarians, scientists, publishers | Documents, classification systems, technologies, standards, databases |
| (Hessels & van Lente, 2008) | Scientists, knowledge brokers, policymakers, users | Scientific publications, patents, knowledge networks, funding agencies, regulations |
| (Gieryn, 1983) | Scientists, lay public, policymakers | Scientific facts, knowledge claims, scientific methods, social norms, values |
| (Svensson, 2010) | Digital humanities scholars, computer scientists, information professionals | Digital tools, databases, digitized texts, metadata standards, scholarly communication platforms |

**Table 5** – Literature comparison on ANT use in SC research

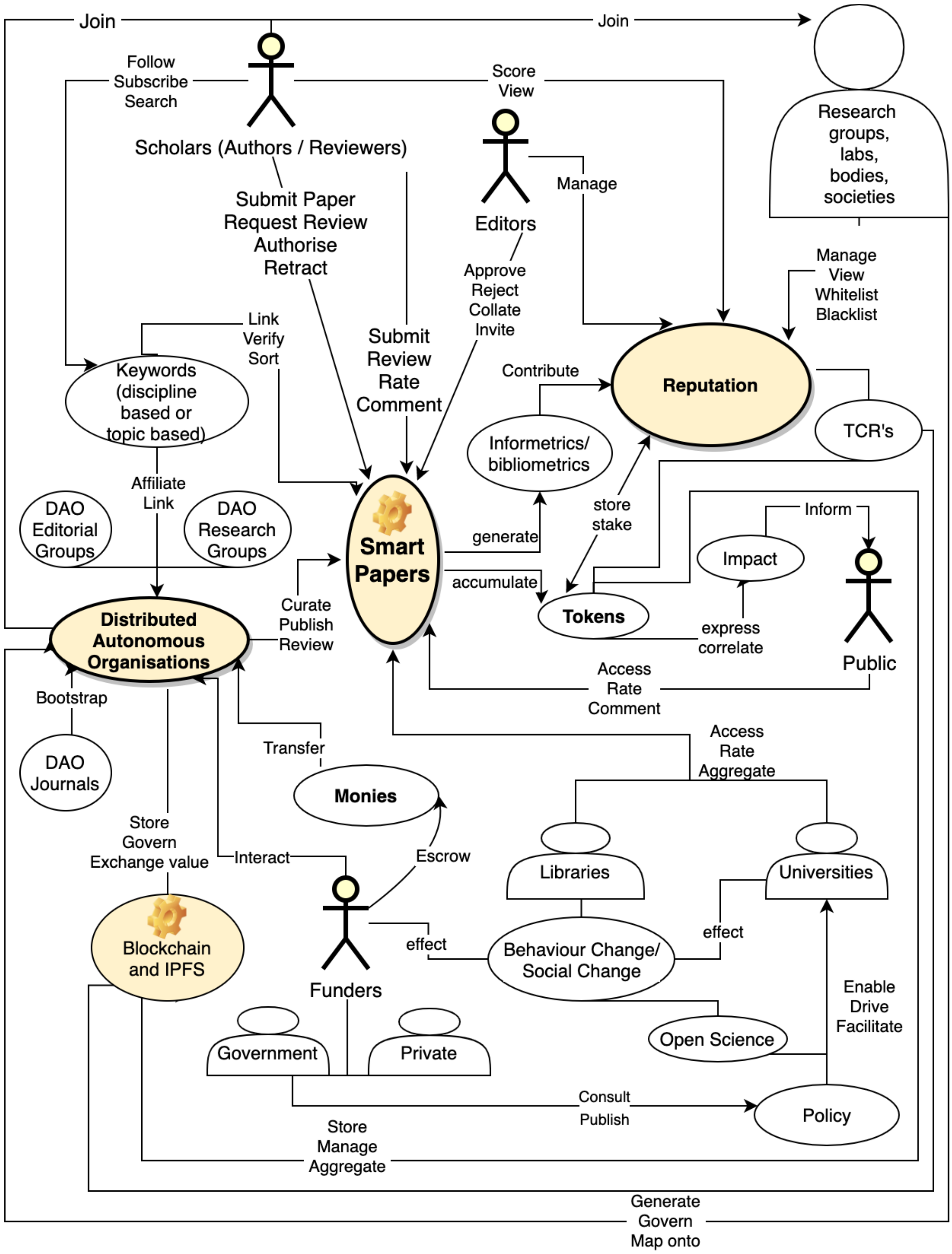



**Figure 30** - Proposed Actor Network for Scholarly Communication with Smart Papers

In the literature, the main actors and actants vary depending on the context of the study. However, some common actors and actants include scientists, engineers, policymakers, funding agencies, documents, technologies, scientific publications, knowledge networks, and institutions. ANTs highlight the importance of understanding how these actors and actants are interconnected and how they shape the production and dissemination of knowledge in scholarly communication. The new actor network for Scholarly Communications with Smart Papers has informed the design of the Smart Papers suite of smart contracts, whose technical modelling is documented in the Appendices, whilst the next Chapter discusses in a step-by-step fashion how Smart Papers were implemented from the modelling phase to the empirical testing stage. The non-pragmatic actors in the network in **Figure 29** were translated into software concepts by using the Agile and Object-Oriented Methodologies (see Chapter 3). The next chapter also documents how the human actors' motivations have been mapped out to allow the translation of their interests into software artifacts.

# Chapter 5

# Smart Papers – Decentralised Publishing on Ethereum

In (Hoffman M. R., Ibanez, Fryer, & Simperl, 2018) and subsequently (Hoffman, Ibanez, & Simperl, 2019), I have contributed and evaluated smart contracts developed on the Ethereum platform to manage the attributions and annotations of scholar publications, filling the gap of existing open decentralised publishing models. In our approach, that we dubbed **Smart Papers,** a suite of smart contracts was deployed and reusability was achieved by an unbounded number of research objects calling those contracts, and storing publication metadata in a distributed ledger.

In the extended version of this paper, I described how a decentralised autonomous organisation can be created that not only lets users vote on papers, but also enables transparent scientometrics that are tamper-proof on the blockchain. We describe a use case demonstrating how the nature of scientific publishing would benefit from storing agreements and artefacts in a verifiable distributed database that does not rely on a centralised party to provide proofs. We found that blockchains, by their design, are appropriate for this use case.

The second research question of this thesis, RQ2, was already defined at the beginning of the document as *what features are needed in a decentralised academic scholarly communications platform, and how to implement and evaluate them.* The following two research sub-questions were added to RQ2:

**RQ2.1.** *How to manage releases of scholarly papers and their attribution agreements in a trusted way?* and

**RQ2.2.** *How to avoid malicious/accidental modifications in remote data stores affecting the computation of bibliometrics for scholarly work?*

## 5.1 Use Case Description

We use examples of Alice and Bob - we imagine that they both represent "typical" scholars from two separate institutions, who agree to work together on a new publication. The purpose of our motivating examples is to give the reader an idea of what sort of problems those scholars might currently encounter within the constraints of traditional ways of organising research and disseminating findings.

Bob and Alice begin working by employing their collaborative authoring tool of choice to start a draft version of their paper. After a few weeks of writing, they decide to release a public version in order to start receiving open comments and reviews. Charlie is a scholar from a third institution that finds Bob and Alice's release through an aggregator website or a search engine. He reads the article and leaves some comments on it that are stored in his personal data store and linked to the release, for instance, using the Web Annotation ontology[118].

Alice and Bob then integrate Charlie's comments in their working version. They continue their work and eventually publish a second release. This time, they submit it to the Call for Contributions of a conference that uses open reviewing. Diane is one of the assigned reviewers. Her review is linked to the release which she read, as stored in the conference's data store. When it is finally time for Bob, Diane and Charlie's appraisal meeting, their employers ask them for the dynamic publications that they have been involved in. Bob shows the full sequence of releases of the publication, while Charlie shows the comment he made on Bob and Alice's paper, and Diane shows the review she made for the conference. Employers apply their preferred credit models to assign weights to each type of attribution described in the attribution metadata and quantify their values. However, when reputation, credit, and ultimately, jobs are involved, social interactions can lead to problematic outcomes, with people trying to game the system in their favour, or to disfavour others. Below, we outline some examples of when things can go awry:

[118] https://www.w3.org/TR/annotation-vocab/

### 5.1.1 Example 1 - Single Version of Truth

Alice trusts Bob for creating the releases and their attribution metadata, as Bob controls the data store. However, Bob can publish a release with the metadata giving more attribution to himself. In a decentralised authoring tool like Dokie.li, each author would hold a copy of the working version, and they could independently generate the release, but if the attribution metadata differs between them, who solves this disagreement? How does an external agent know which copy to trust?

### 5.1.2 Example 2 - Trusted Protocols

Bob and Alice could collude to show different versions of the attribution metadata (see **Figure 31**). For example, consider that employers use two different services to query dynamic publications linked to their faculty members. It is not hard to imagine a semantic store that returns a different version of the attribution metadata, depending on which agent is asking.

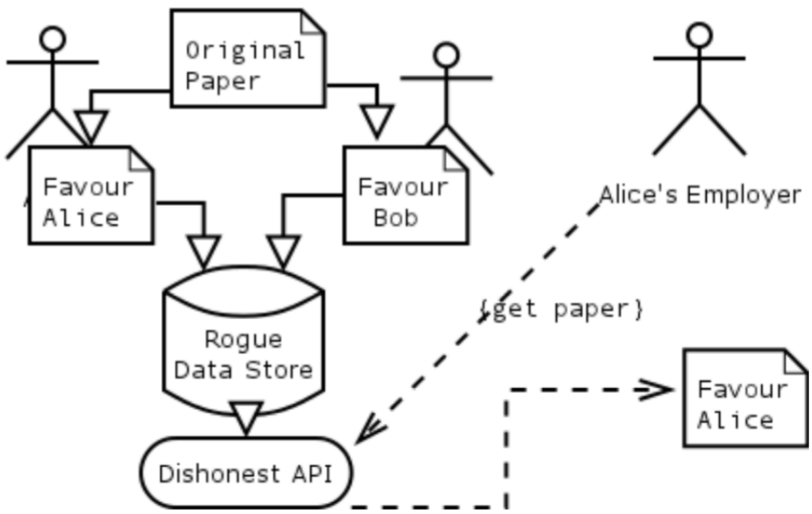



**Figure 31** - Alice and Bob collude through a dishonest protocol

### 5.1.3 Example 3 - Tamper-proof Records

Bob and Alice could collude to ignore Charlie's comment, in an attempt to not share part of the credit with him. In a decentralised model, a link to the comment and Charlie's identity should be stored in Bob and Alice's data store; however, if Bob and Alice control the data store, nothing prevents them from deleting the link. Charlie would have the copy of the comment and the link to the release, but he might have a hard time convincing a third party (his employer for example), that the comment was not forged.

### 5.1.4 Example 4 - Non-repudiation

If Diane's review is considered unfair, the persons in control of the data store might be tempted to make it disappear. A third-party agent querying the conference's data store would see nothing. An agent following links from Bob and Alice's data store would get a de-referencing failure (404). Even if Bob and Alice decide not to delete the unfavourable review from their data store, how can they prove that they are not forging fake reviews to damage Diane's reputation? **Figure 32** shows how such breaches of trust may occur when a single party controls data storage.

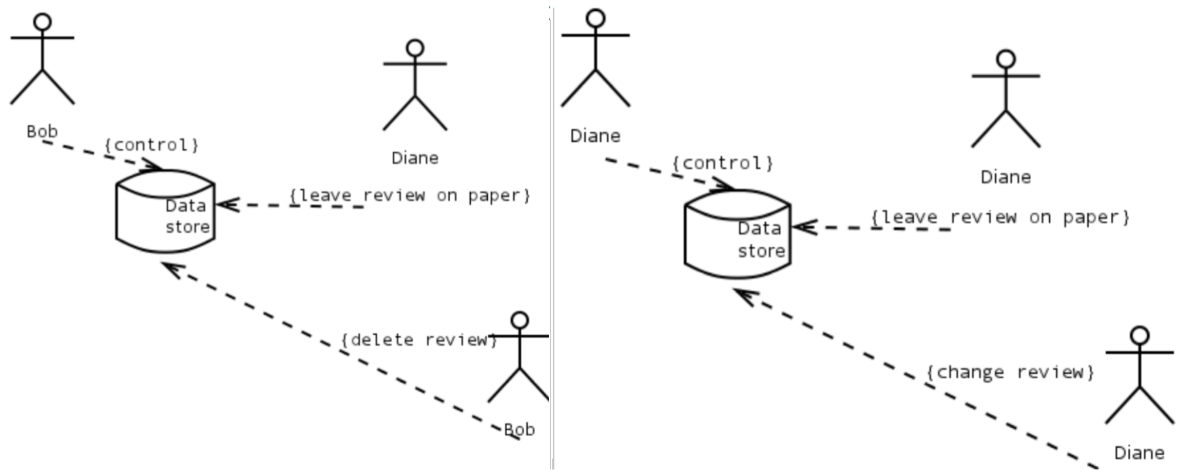



**Figure 32** - Breaches of trust in individually controlled data stores

### 5.1.5 Example 5 - Trusted Analytics

Bob and Alice's paper becomes so successful that hundreds of scholars start citing it in their work. However, different informetrics providers like Scopus and Google Scholar seem not to be able to agree on the number of citations for Bob and Alice's paper. One provider is missing a few high-impact conference proceedings papers quoting Bob and Alice's paper, whereas the other provider seems to have included lots of grey literature, resulting in two very different citation counts. How can third parties (most importantly, Bob and Alice's employers and funders) be sure of the true citation count of Bob and Alice's original work?

### 5.1.6 Example 6 - Governance

Alice and Bob subsequently work on a paper with 4 more co-authors from other institutions (Edna, Fred, Gina and Hal) who also happen to be based in different time zones. The six scholars now need to add a complex figure to their paper. There are a few alternatives for plotting the figure, and each approach has its own unique pros and cons. How can we carry out a vote to enable the whole team to reach a single decision?

### 5.1.7 Example 7 - Dispute Resolution

Alice, Bob, Edna, Fred, Gina and Hal now have to decide whether to publish the current version of the paper. Edna believes it is not camera-ready and submits corrections, whereas Hal pushes for immediate submission without corrections. How can the resolution of this conflict be facilitated in a transparent manner?

## 5.2 Proposed Solution

The common problem of these scenarios is that for all actors (Alice, Bob, Charlie, Diane and their employers, funders and contributors), the data that is crucial to show or measure their performance is not entirely under their control, making it vulnerable to manipulation. This approach addresses this problem by empowering all collaborators with the following:

- The notarisation of releases providing evidence that all the authors agreed to the release of a particular version of their paper.
- The notarisation of the attribution metadata linked to a release, ensuring that all authors have agreed on it, and guaranteeing to third parties that none of them can tamper with it.
- A mechanism that ensures that annotations made on releases by agents other than authors cannot be repudiated by annotators or their recipients, guaranteeing to both authors and third parties querying this data, that it was not tampered with.
- An index of links and data concerning a particular dynamic publication. This potentially facilitates the task of Web agents that compute bibliometrics, as there is no need to either trust the data store of the authors, or to crawl the Web in search of the comments and reviews to the publication.
- A mechanism for linking papers that reference each other.
- Transparent and trustworthy citation counts produced in a decentralised environment.
- A voting mechanism for facilitating the resolution of governance conflicts using pre-configurable quora.
- The notarisation of voting outcomes on the blockchain.

It is worth noting that through implementing the above requirements, the solution automates the fulfilment the Vancouver (ICMJE) guidelines on Authorship[119], also known as the Vancouver Convention. Its purpose is to emphasise how "authorship confers credit and has important academic, social, and financial implications", recommending that substantial contributions must be credited, final version must be approved before publishing, and that authors should have "confidence in the integrity of the contributions of their co-authors".

[119] http://www.icmje.org/icmje-recommendations.pdf

## 5.3 The Smart Paper Model

The motivating example above illustrates the importance of trust throughout the collaborative process. In particular, there is a strong need for making agreements and setting their outcomes in stone so that they cannot be later repudiated. Furthermore, all the essential artefacts associated with those agreements must be timestamped and securely stored in a truly permanent way. Currently available collaborative tools solve some trust issues, for example Dokie.li removes centralisation so that the authoring parties do not have to rely on an intermediary to publish and annotate their documents. This is a very welcome step towards removing the overhead associated with middleman activities (publishing house), albeit it merely shifts the trust towards the authoring parties (author, reviewer). It is easy to imagine a situation in which the authors destroy their data, the reviewers could do the same, and any track of their writing will be lost forever.

The purpose of our model is to provide trust where it has not existed before. Smart Papers provide a collaborative platform that preserves a single version of the truth throughout the collaborative process. This is somehow similar to employing a trusted third party (e.g. a notary public) to keep track of contracts signed by multiple parties, alongside with all the certified photocopies of all the evidence attached to the contracts as relevant appendices. An example of such notarised contract would be Alice and Bob signing an agreement specifying the ordering of their names on a paper (e.g. "Bob, Alice") and then attaching a certified photocopy of their paper in its current version as an appendix. We use smart contracts for maintaining all such signed agreements in order to implement Smart Papers. **Table 6** summarises how smart contracts can provide the functionality analogous to that of a traditional trusted third party.

| Notary public function | Blockchain function |
|---|---|
| *Authenticate parties using their legal identification* | *Identify parties cryptographically* |
| *Take statutory declarations, store them and certify photocopies* | *Store data permanently and securely and provide real time access* |
| *Prepare and certify contractual instruments* | *Store and execute smart contracts* |
| *Provide a trusted record for the above* | *Provide a trusted record for the above* |

**Table 6** - Blockchain smart contracts as compared to a traditional trusted third party

To design the Smart Papers model, we shall assume that all authors successfully identify through their ORCID (Open Researcher and Contributor ID) which is a non-proprietary alphanumeric code to uniquely identify scientific and other academic authors and contributors. ORCIDs are mapped to authors' signing and encryption keys using a smart contract. The main functionality in this model is then designed using the separation of concerns (SoC) (Hursch & Lopes, 1995) design principle, such that each contract file addresses a different concern, i.e. a different set of information that jointly affects the global state for the Smart Papers use case. We group these concerns into the following four categories: **Paper**, **Version**, **Annotation** and **Contributor**. We use UML to model the main classes corresponding to our smart contracts. It is important to note that smart contracts and OOP classes (as modelled by the UML) are not quite the same. The semantics are very similar in many cases, but some fundamental differences arise from the fact that smart contracts can store and send value and have a public address once deployed. In Appendix A, the Smart Papers model is shown, which implements an approach inspired by Dynamic Publications through the design of smart contracts that control the workflow for a Smart Paper as it evolves over time.

One of the core requirements of the SmartPaper model is the ability to provide a tool for all collaborators to agree with the result of a certain interaction. The number of collaborators can be unbounded, but certain decisions need to be reached jointly. An example is calling the *PublishVersion()* function on the *Paper* contract.

### 5.3.1 Bibliometrics

To solve our Motivating Example 5 - Trusted Analytics (in which citation inconsistencies were present across different centralised providers), we can entice scholars to store references as links between Smart Papers using our Smart Papers web interface. To store a reference, a living Smart Paper (A) can, at any time, register a citation of another Smart Paper (B) through calling the *registerCitation*(B-paperId) function on A's Paper contract. Whilst the Smart Paper A is in its working version, and unpublished, this only updates the references internally. Nevertheless, publishing the Paper A will trigger an event *PaperPublished* which, in turn, triggers the *CitationUpdate* functionality. PaperB's *incrementCitationCount* will be called at that point. This chain of events will result in Paper B's citation count being updated with the new reference. However, if at any point A is retracted, the *PaperRetracted* event deals with calling previously cited papers to decrement their counts.

### 5.3.2 Votes

For those Smart Papers use cases that require a quorum to reach a decision, we replaced the low-level multi-signature approach of our original implementation, with a new Vote-based approach in the extended version. Most importantly, once more than 1 author has been added to a Smart Paper instance, a Vote now needs to be carried out before the paper can be published.

Votes are also carried out for other common authoring scenarios that require a decision-making protocol, such as:

- adding a new contributing author to an existing Smart Paper instance,
- approving a new significant revision to the existing Smart Paper,
- choosing the name of a paper - an example of how Votes can be used for brainstorming as well as signing approval,
- voting on the listing order of authors' names.

The quorum for most of our use cases is simple democracy (more than 50% voters agree with a particular outcome). However, the quorum requirement should be set to 100% for publishing a final version of the paper. Submitting papers without the knowledge and permission of co-authors appears to be a worrying issue that does sometimes happen[120]. Taking governance to a meta-level, the attractiveness of using a framework like Aragon for managing governance is that we can even schedule a Vote to decide about what the thresholds (quora) should be for all future Votes.

## 5.4 The Smart Paper Workflow in Ethereum

To begin with, an article and its metadata is submitted by a writer (we shall refer to her as Alice, from my motivating examples earlier), and stored in a distributed file store, all of which is recorded on the blockchain. Alice will have been set up in the system through the use of the *Contributor* smart contract. In my implementation, the *Contributor* contract requires Alice to have a valid ORCID as well as an IPFS node identity belonging to her. The default type for Alice is "author". Bob is also set up as an "author", but Diane uses a different argument for the *Contributor* contract, and so she becomes registered as a "reviewer".

Smart contracts often act as state machines, meaning that they have certain stages making them act differently, and in which different functions can be invoked. A function invocation often transitions the contract into the next stage which can be used to model workflows. It is also possible for stages to be automatically reached after a certain period elapses. An example for this is a crowd-funding contract which starts in the stage of "not accepting donations" then progresses to "accepting donations", then transitions to "releasing funds". Function modifiers can be used in such a use case to model the states and safeguard the user against incorrect usage of the contract.

[120] https://www.elsevier.com/connect/co-authors-gone-bad-how-to-avoid-publishing-conflicts

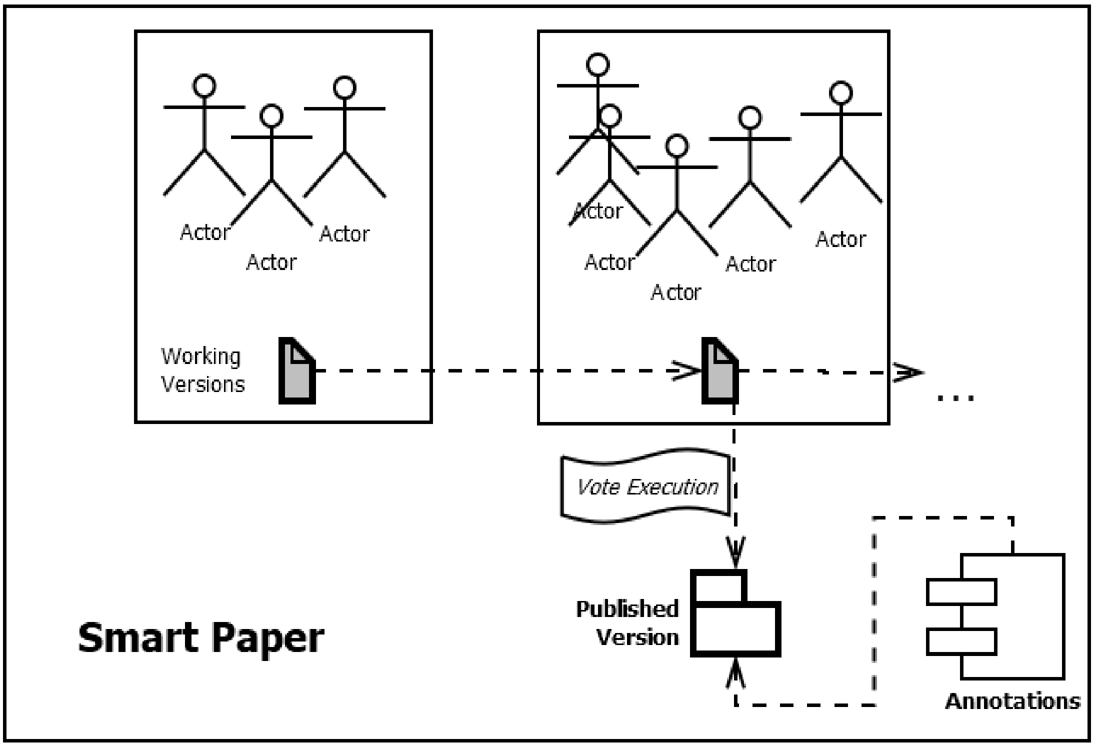



**Figure 33** - The Smart Paper workflow involves multiple working versions with dynamic collaborators. Versions can become published and made available for annotating.

The Smart Paper workflow allows the participants to release new versions of their paper and to publish versions when enough authors agree to do so. Papers can also be retracted.

Figure 33, once instantiated, a Smart Paper becomes a dynamic list of versions, each of which can exist in a working state or become published. The number of contributors and their formal ordering is allowed to change on a per-version basis. Annotations can be left by reviewers on published versions.

To create a new Smart Paper, either Alice or Bob call *createNewPaper* in the *Paper* Contract which will return a valid *PaperId* that uniquely identifies their new publication (see **Figure 34**). This also instantiates the workflow with an initial, blank, working version of this paper manufactured by the *Version* contract. Bob and Alice work on their preferred authoring tool to produce a first draft (e.g., to show to a trusted colleague), to register it in the Smart Paper, Bob calls *addNewVersion* in the Version contract, including the artefact, its metadata and his signature. Before committing the transaction, the Smart Paper will wait for Alice (marked as contributor of the paper) to also perform a call to *addNewVersion* using the same artefact, metadata and her signature.

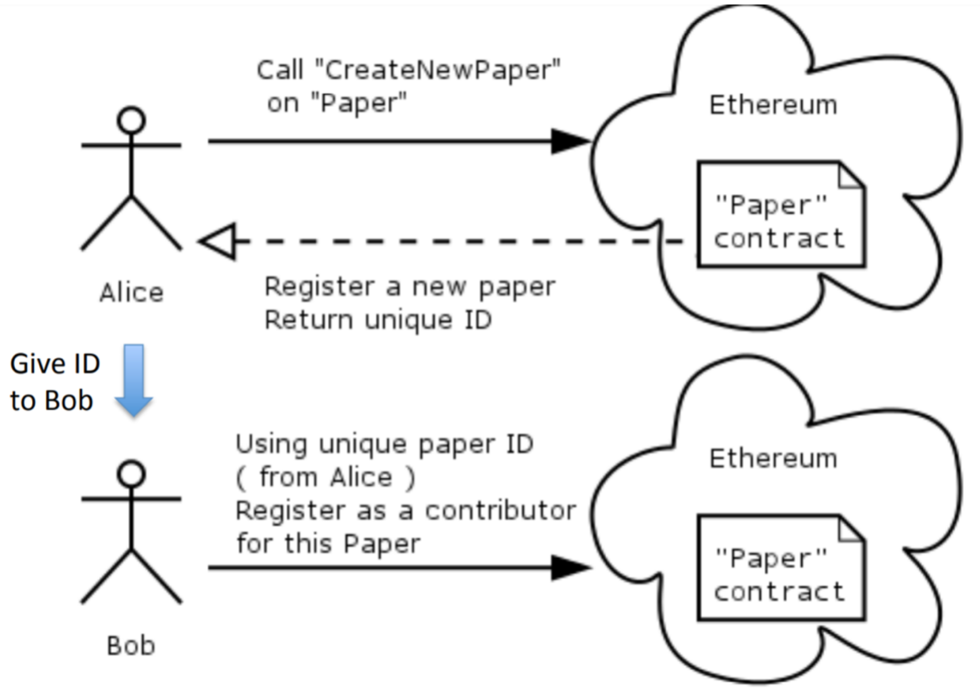



**Figure 34** - Unique paper identifiers on Ethereum blockchain

The procedure is repeated each time Bob and Alice want to register a new version. For marking a version as public, Bob calls *publishVersion* in the *Paper* contract, providing the *versionID* and his signature. Like *addNewVersion*, Alice needs to issue her signature through a function call to *publishVersion* before the Smart Paper commits the transaction. The *getCurrentPublishedVersion* and *getCurrentWorkingVersion* return a versionID that can then become the input to the *getIPFSLink*. The Smart Paper only commits a version (including metadata) if all authors sign their agreement to it. An external agent that gets a version from a Smart Paper instance has the assurance that it was approved by all authors, and that the Smart Contract consistently returns the correct version and metadata.

Interactions with external actors like reviewers or annotators, are abstracted as Annotations. When Cynthia or Dean want to leave their comment or review, they call *addAnnotation* using the *versionID* of the version they want to comment on, and their signatures. Contrary to the *Version* functions, no approval from authors is needed. The annotation is registered in Ethereum's Blockchain and can be retrieved by calling *getAnnotation*.

## 5.5 Implementation

Although in theory, the Smart Papers model could be implemented on any smart contract-enabled platform, the choice of the implementation framework dramatically impacts development time and costs. Whilst there are multiple distributed ledger technologies, such as Corda[121] or HyperLedger[122] that could be utilised to develop trusted smart contract code that runs on top of the blockchain, for this thesis, I elected to develop on top of the Ethereum platform (Wood, 2014) which is the most commonly used technology of its kind (Alharby & van Moorsel, 2017).

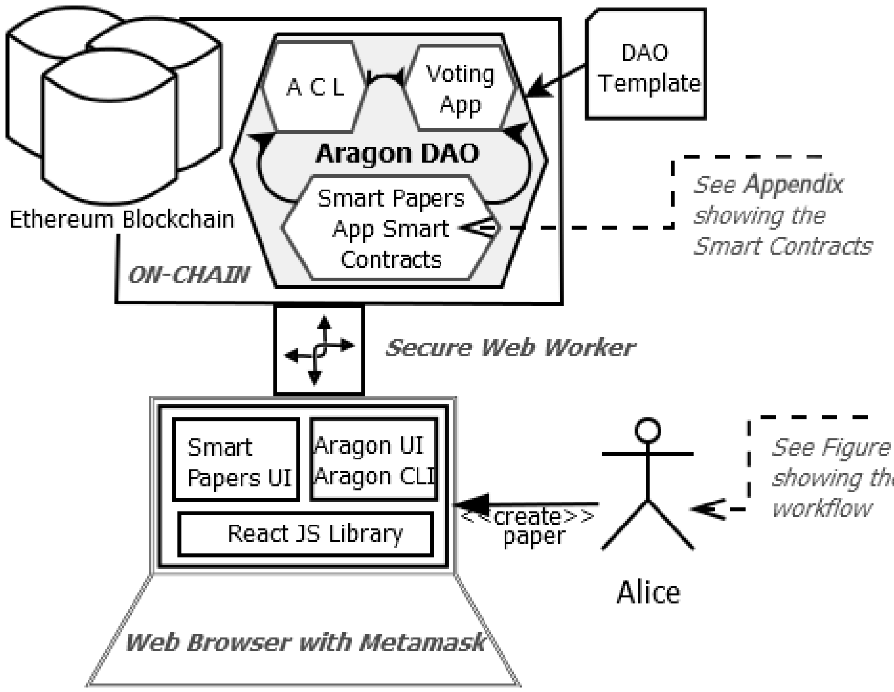



**Figure 35** - The architecture of the Smart Papers implementation involves frameworks and features that work on-chain, as well as those that run locally in the browser. The secure web worker mediates communications.

121 https://github.com/corda/corda
122 https://www.hyperledger.org/

### 5.5.1 Smart Papers as a Web App

One of the core requirements of the SmartPaper model is the ability to provide a tool for all collaborators to agree with the result of a certain interaction. Decision making can be implemented in different ways. In our implementation, the number of collaborators can be unbounded, but to make a decision, an agreement needs to be reached by a quorum of authors that is configurable on a per-paper basis. This means that you could have a Smart Paper, in which only one person's vote is enough to decide, but you could also have a Smart Paper where everyone is required to agree on a particular course of action. The default quorum for Smart Papers is 50%, which we call 'simple democracy'. As a Smart Paper evolves, its collaborators can change the quorum at a later time through a vote. Interacting with Smart Papers is done through a Web browser, which is why we call it a Web app.

### 5.5.2 How the Aragon Framework helps with Governance

Each DAO also has a creator account who initially has all the voting power. This voting power is expressed through percentage DAO token ownership. This is because DAOs have their own tokens, in our case it is the University of Southampton (SOT) token. New members joining the DAO are provided tokens enabling their participation in the DAO through a process that requires existing members to vote on approving the newcomer. Introducing a custom cryptographic token can help with administrative processes, including governance and HR processes - here are a few ways in which the SOT token could be helpful:

1. Governance: With the help of a DAO, the SOT token holders could have a say in the decision-making process for the scholarly publishing system. They could vote on important issues such as editorial policies, peer-review processes, and the allocation of funds. This could lead to a more transparent and democratic governance structure.

2. Incentivisation: By using a custom cryptographic token, the university could incentivise scholars and researchers to contribute to the scholarly publishing system. For example, they could reward researchers with tokens for publishing high-quality research papers or serving as peer reviewers. This could help to attract more high-quality research to the platform.

3. HR Processes: The custom cryptographic token could also be used for HR processes such as performance evaluation and compensation. For instance, the university could reward researchers and editors with tokens for their contributions to the platform, and use those tokens to determine bonuses or promotions.

4. Funding: The custom cryptographic token could be used to raise funds for the scholarly publishing system. The university could conduct an initial token offering (ITO) to raise funds for the platform's development, or sell tokens to fund ongoing operations.

5. Security: By using a custom cryptographic token, the university could ensure the security of the platform. For instance, the token could be used to provide access to certain features of the platform or to authenticate users. Additionally, the blockchain technology underlying the token could help to prevent fraud or hacking attempts.

The SOT token is also required for scholars to participate in voting on the future of the platform. Aragon DAOs support apps that are custom bundles of smart contracts closely integrated to the Aragon platform via their bespoke secure Web interfaces. The Voting app is an example of an app that acts as an entity that will execute a set of actions on other entities (e.g., other apps, smart contracts, user's balances) if the token holders of a particular governance token (i.e., members of the DAO, like SOT holders for the Soton DAO) decide to do so. The Smart Papers app is the name of the suite of smart contracts that has been integrated into Aragon to support its front- and backend. Essentially, any Smart Papers transaction can now be signed by the user in their Web browser with a Metamask extension, and all the interactions requiring multiple signatures can be signed individually by different users as part of a single Aragon vote. The overall architecture of this set-up can be seen in **Figure 35**, which also shows how a new DAO can be deployed from a DAO template, which will install the custom Smart Papers app interacting with both the ACL and Voting components of Aragon.

Aragon DAOs are always deployed with ACLs (Access Control Lists). This is a computer security mechanism (Sandhu i Samarati, 1994) whereby permissions are attached to users or processes, or in the context of blockchain, to user accounts.

Within the realm of Aragon, a permission is defined as the ability for users to perform actions (that are grouped by roles) in a certain app, for example, the Voting app[123] or our custom Smart Papers app. ACLs are grouped for a particular DAO. New members can be granted permissions in an ACL from the creator and then grant permissions to others (if permitted).

The Access Control List remembers who can perform what actions where. Sometimes users are not allowed to perform actions directly, but only through a decision-making process, such as a vote. The ACL handles this use case through a concept called forwarders.  A forwarder is an app that can execute actions on someone's behalf if the ACL permits it. An example of a forwarder is the Voting app, and we can specify an ACL entry allowing a user to carry out a certain operation if a Vote is carried out. The Voting app will only execute the user's desired function call if the vote passes. One advantage of using Aragon to manage our smart contracts is that it manages the upgrading of major, minor and patch versions of contracts in a way that is seamless (using a proxy pattern) and was unachievable in our original implementation - upgrading old-style Smart Papers would have required a completely new contract address - which now can be avoided with Aragon.

[123] https://wiki.aragon.org/dev/apps/voting/

## 5.6 Evaluation

Two important factors affecting the performance of a distributed application like Smart Papers are firstly, how much it costs to run this app's operations like creating new papers and registering references, and secondly, how long these operations take to complete in the decentralised environment. We set out to evaluate both the costs and the speeds for four main use cases for Smart Papers:

1. creating a new Smart Paper on the Ethereum blockchain from the Smart Papers app,
2. adding a citation (a link between two Smart Papers) through the Smart Papers app,
3. creating a new open vote using the Voting app, and
4. casting a ballot / participating in an open vote in the Voting app.

### 5.6.1 Notes on Cost

Execution of a smart contract begins with a transaction that is sent to the blockchain. This transaction specifies the address for the contract, the arguments, and an amount of Ethereum's currency to pay for the execution. It is commonly observed in small-to-medium size contracts that most of the cost is taken up by a fixed base fee. This base fee of 21,000 is expressed in *gas* which is an abstract unit. While gas is fixed per each transaction, it is additionally fixed for every operation called from within the smart contract. Each low-level operation available in the EVM is called an OPCODE. These include operations such as *ADD* - adding two integers together, *BALANCE* - getting the balance of an account, and *CREATE* - creating a new contract with supplied code. Each of these OPCODEs has a fixed amount of gas that it costs to execute. The fixed amount of gas has been chosen by the designers of Ethereum for each OPCODE in a way that reflects the relative complexity of that OPCODE.

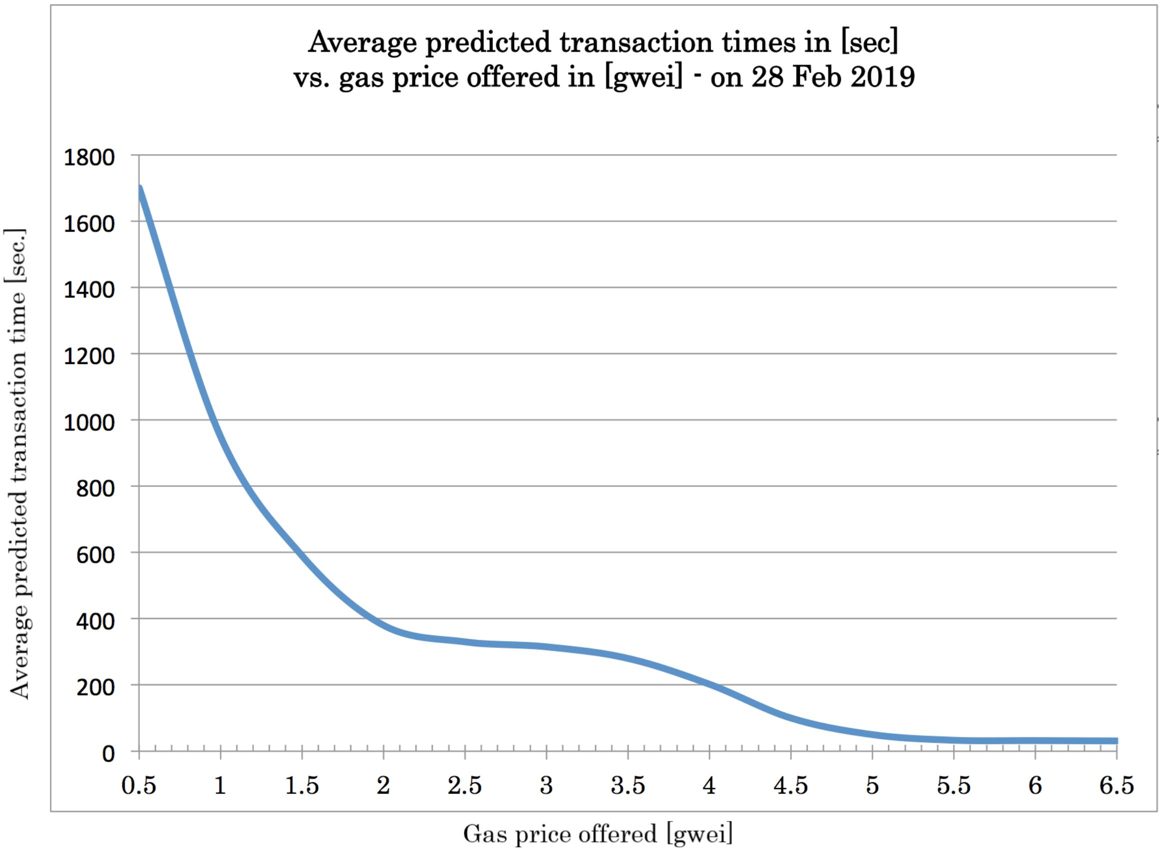



**Figure 36** - The influence of gas price on execution times

Whereas gas is fixed and predictable, the amount a user pays per gas, the *gas price*, is dynamic and dictated by market conditions. The price is usually given in subunits of *ether*, such as *gwei* (1 Gwei = 0.000000001 Ether). Miners receive ether fees based on the amount of gas multiplied by the gas price, which incentivises them to prioritise those transactions that attract higher fees. It also follows that the higher gas price you are willing to pay, the faster your transaction will be processed, and the sooner your contract will be allowed to execute. While offering a high gas price can speed things up, there is a limit to the acceleration.

### 5.6.2 Data Collection and Replication

The gas prices are estimated with (ETH Gas Station, accessed January 2018), the *de facto* reference for current gas market conditions and miners' current policies. The *Recommended User Gas Prices* section of ETH Gas Station shows the range of gas prices one might pay for an expected transaction commitment time. We compile and graph the gas price predictions from the time of writing this chapter in **Figure 36**. This data has been collected using the ETH Gas Station Application Programming Interface[124]. Data collection can also be performed manually using the *Metamask* extension and the relevant steps are documented online[125].

### 5.6.3 Smart Papers Evaluation Results and Discussion

Evaluation results are presented in **Table 5** and sorted by average transaction cost, with adding citations being the cheapest (costing as low as $0.04) and creating a new open vote being the most expensive (costing up to $0.32). The more expensive operations are those that use Aragon's Voting app, due to the complexity of setting up and executing Votes. It is worth noting that the absolute costs are as much as 57-fold smaller than in our original implementation. This can be attributed to Ether (the Ethereum currency) now being much cheaper than at the time of printing the previous version. There have also been bytecode optimisations to the Solidity compiler yielding faster and cheaper code.

[124] https://docs.ethgasstation.info/

[125] https://github.com/MetaMask/metamask-extension/issues/3037

| Use Case / Gas Cost | Fee Grade | Transaction cost | Speed | Transaction time |
|---|---|---|---|---|
| Adding a Citation / Cost: 104,438 [gas] | Low (3 gwei) | $0.04 | Slow | 5 min 16 sec |
| | Average (4 gwei) | $0.06 | Average | 3 min 13 sec |
| | High (5 gwei) | $0.07 | Fast | 31 sec |
| Participating in an Open Vote / Cost: 146,151 [gas] | Low (3 gwei) | $0.06 | Slow | 5 min 15 sec |
| | Average (4 gwei) | $0.08 | Average | 3 min 13 sec |
| | High (5 gwei) | $0.10 | Fast | 31 sec |
| Creating a New Paper / Cost: 204,966 [gas] | Low (3 gwei) | $0.08 | Slow | 5 min 15 sec |
| | Average (4 gwei) | $0.11 | Average | 3 min 13 sec |
| | High (5 gwei) | $0.14 | Fast | 31 sec |
| Creating a New Vote / Cost: 315,818 [gas] | Low (3 gwei) | $0.13 | Slow | 5 min 30 sec |
| | Average (4 gwei) | $0.17 | Average | 3 min 22 sec |
| | High (5 gwei) | $0.22 | Fast | 32 sec |
| Adding a New Version to Paper / Cost: 475,281 [gas] | Low (3 gwei) | $0.19 | Slow | 5 min 30 sec |
| | Average (4 gwei) | $0.24 | Average | 3 min 22 sec |
| | High (5 gwei) | $0.32 | Fast | 32 sec |

**Table 7**– Cost and Speed Results for the Main Use Cases

Based on Web of Science data, Marx and Bornmann estimated that, in 2010, the average number of references in a scholarly paper was 32.21, with a low average of 27 for Engineering papers, and a high average of 42 for Social Sciences papers (Bornmann & Marx, 2015). Assuming Small's view that references represent intellectual debt (Small, 2004), this debt can be symbolically paid off by the co-authors of a new Smart Paper through covering the fees associated with registering the citations resulting from their newly published work. Therefore, for the most costly case of a new Smart Paper created in the Social Sciences domain, paying for all the citation counts that were changed by references from that Smart Paper, would cost 42 (average reference count) times $0.04 (transaction cost for registering a citation), which would cost an average of $1.68 per paper in Ethereum fees. NB. in addition to Ethereum transaction fees, there may be other fees incurred by Smart Papers, including:

1. Storage fees: a scholarly publishing system that stores a large amount of data on-chain could incur significant storage fees.
2. Gas fees: these can increase as the complexity of the smart contract code increases.

3. Development fees: maintaining a novel scholarly publishing system that uses blockchain technology requires costly skills and expertise.
4. Integration fees: integrating a blockchain solution with existing academic publishing solutions requires additional resources and consultancy fees.
5. Network fees: finally, the cost of operating will require fees for network access, infrastructure costs, and maintenance costs, including support. These fees may be passed on to users as part of the cost structure.

Apart from the transparency, publishing speed and cost-effectiveness angles, a fair viability comparison that is recommended as part of future work, should moreover include user adoption measurement, citation impact, accessibility, and framework integration.

In terms of our Motivating Examples, we have seen that the notarisation of scholarly papers and their associated metadata on a blockchain, using Smart Papers, can solve the problems presented in Examples 1-4. We have also shown how the governance benefits of Aragon open votes help us solve the problems described in Examples 6-7. Finally, the citation-counting functionality of Smart Papers resolves the dilemmas presented in Example 5.

Regarding cost, our result suggest that a new Smart Paper can be created for as little as $0.08. A vote with three participants can be carried out for as little as $0.31 ($0.13 to create a vote and 3 * $0.06 for each ballot cast). We already estimated the cost of storing references as $1.68 (the costly case). Adding all these costs together yields a paper that, throughout its total life cycle, costs around **two dollars** in total to govern, update and disseminate. These costs are in stark contrast to publishing houses' costs that have been previously estimated to be in the range of **$3,500-$4,000** per article (Van Noorden, 2013). When comparing these dollar figures, the reader must be mindful of what exactly is being compared. For example, services like typesetting and printing would not be provided by Smart Papers, even though they may be still included by traditional publishing house in their cost estimates. The discussion of whether traditional print services are needed to successfully disseminate a paper in today's digital era is beyond the current scope, and I assume a paper to be successfully published after it's made its way to the blockchain. However, the additional costs of IPFS file storage are a completely valid concern. IPFS has no concept of transaction or storage costs, but also lacks an incentive layer. So currently, a researcher may elect to run a local IPFS node, the cost of which could vary wildly depending on the machine specifications and the local networking fees. The real costs may be inferred from those quoted by cloud IPFS providers (around $0.14/GB/mo[126]). All in all, in the era of decentralised Web, Smart Papers offers a cost-effective alternative to the services being offered by big publishers.

It is worth noting that for most purposes, the *Low Cost* options should be used for all use cases, as the slowest observed transaction time in that instance is 5 minutes 30 seconds which we believe is still acceptable for most scholarly communications use cases, especially considering that it would typically take weeks, if not months, to get a scholarly paper successfully disseminated through a publishing house.

In the Smart Papers approach, a valid question arises about who should pay for the transaction fees that we estimated. Should it be the scholars, their universities, funders, etc? Should the payment mechanism depend on the use case - for example, should adding references be funded differently than creating votes? Currently, we allow for as much flexibility as possible. Aragon offers a multi-token Finance App[127] that can be used to manage budgets, to track income and expenses in an organisation, as well as to perform payments transparently according to customisable rules that can be different for each DAO.

---

[126] pinata.cloud, eternum.io – these costs are fairly stable and haven't changed in 5 years

[127] https://wiki.aragon.org/dev/apps/finance/

## 5.7 Decentralising Peer Review with Smart Papers

The Smart Paper suite of smart contracts described in the previous sections can be used to realise a vision for decentralised science as set out in the opening chapters of this thesis. However, as all journal and conference workflows have peer review at their heart, peer review is a crucial element to be implemented as part of the proposed decentralised platform. Peer review is a family of "quality control" mechanisms that promote the establishment and dissemination of trustworthy scientific knowledge (Chubin, Hackett, & Hackett, 1990). In the Background chapter, I mentioned that it has been described as a "flawed process", hard to define and relatively unstudied, with the systems of every journal and every grant giving body differing in detail, as the system keeps "operating on trust" (Smith R. , 2006). Regardless, academics believe that that peer review "helps scientific communication" thus "improving the quality of published papers" (Ware, 2008). Most prevalent modern models for peer-review operation include single-blind (reviewer is anonymous, often seen as the norm), double-blind (both reviewer and author are unknown to each other), open and post-publication (Ibid.) Nature conducted a trial of open peer review in 2006, and the feedback suggested 'that there is a marked reluctance among researchers to offer open comments'[128].

As science keeps progressing in the direction of a more data-centric model, the amount of data linked to a typical scientific paper keeps increasing. The problem arises as to whether the data itself should be subject to review. This may add additional economic strain on reviewers within the existing *pro bono* models. Importantly, some elements of the peer review processes have been proposed to be automated. In particular, checking for spelling mistakes, redundancies or even plagiarism does not in fact require a human peer to perform the task and can be achieved through software. One can thus envision a peer-review smart contract that performs these tasks automatically. Finally, Post-publication peer review (PPPR) must be mentioned in more depth. The process takes place after a paper has been

[128] Overview: Nature's peer review trial. Nature (2006) | doi :10.1038/nature05535 ; accessible via https://web.archive.org/web/20150325164021/http://www.nature.com/nature/peerreview/debate/nature05535.html

published. It involves peers who assess the quality and content of a paper based on their experience. PPPR potentially is a significant form of review because it sets out to correct the biases caused by veiled and vested interests (Teixeira da Silva & Dobránszki, 2015).

As Smart Papers was already shown to be a viable mechanism for scholarly self-governance through enabling the scholars to vote on decisions relating to their publishing work, letting them timestamp and tamper-proof their research outcomes, and enabling trustworthy scholarly communications through Ethereum blockchain, thus removing the reliance on centralised publishing houses, it can be directly extended with modular peer review functionality, and in the next subsections, I present a model in which a Smart Paper can be directly sent for decentralised peer review to a Reviewing DAO, which I also call DAO4PR (DAO for Peer Review). The DAO4PR blueprint can be used by journals, conferences, etc.

### 5.7.1 The Workflow Extension

Peer Review Smart Contracts expose a workflow with the following elements:

1. Submit Smart Paper for Review (from Research DAO to Reviewing DAO)
2. Select Peer Reviewers (inside Reviewing DAO)
3. Revisions Required (from Reviewing DAO to Research DAO)
4. Resubmit for Review (from Research DAO to Reviewing DAO)
5. Decision (Accept/Decline) (inside Reviewing DAO)

### 5.7.2 Proposed 'Smart Review' Extensions

Although blockchain transactions in Ethereum are public, as they are in Bitcoin, users remain "pseudonymous" by keeping their public keys anonymous. Thus, a privacy mechanism is needed, such as ring signature. The main research sub-question (belonging to RQ2 of the thesis) was therefore:

**RQ2.3**

*when decentralising peer review, what is the viability of using ring signatures for ensuring double-blindness of the process?*

As Zero-knowledge proofs and Ring signatures are two most commonly used mechanisms for preserving privacy on blockchains (Mahmood i Vacius, 2020), I compare and contrast their usefulness for the double-blind peer-review use case.

### 5.7.3 Privacy and Ring Signatures vs ZKPs

Cryptographic signatures, also known as digital signatures, enable the recipient of the data to verify the authenticity of this data, giving this recipient high confidence who the data originated from. The concept of signing and verifying data is shown in **Figure 37:**

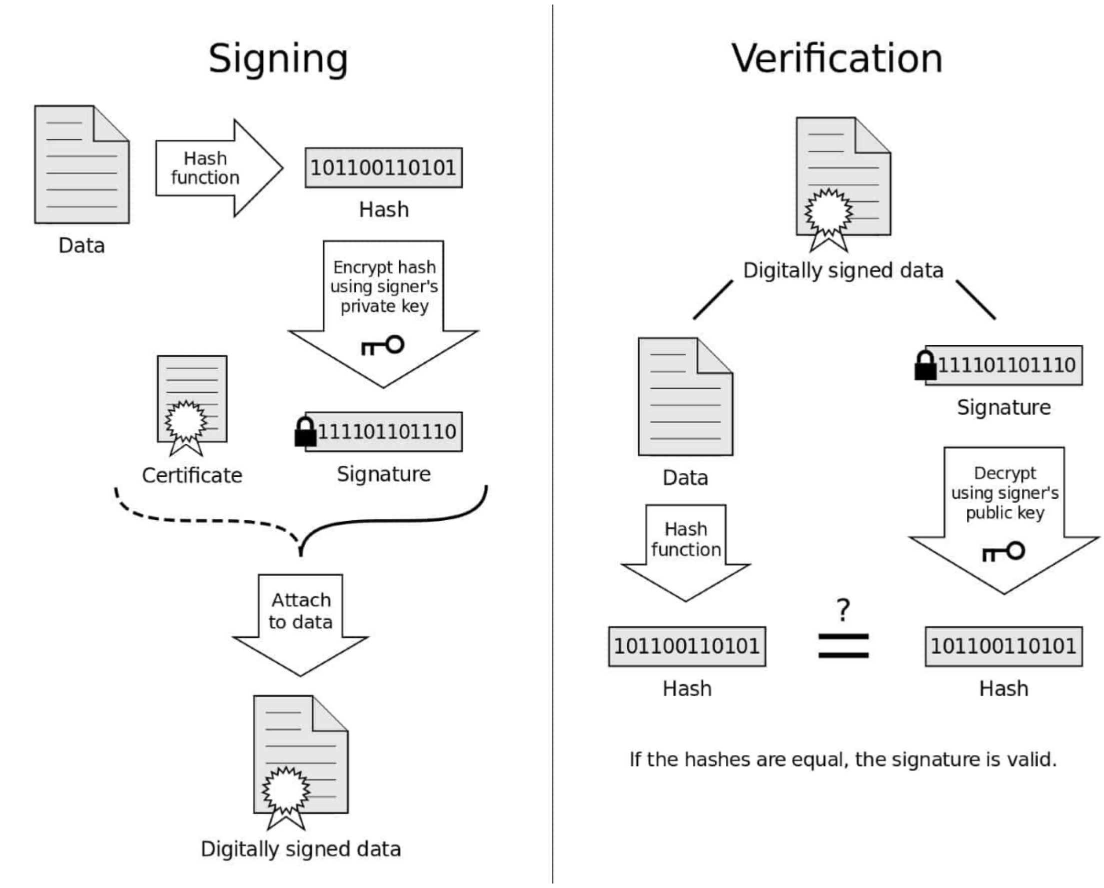



**Figure 37** - Signing and Verifying Data with Cryptographic Signatures

Extending the basic concept of cryptographic signatures, ***ring signatures*** enable the signer to create a signature that doesn't fully reveal their identity. More precisely the signer bundles their public key alongside the public keys of entities that are not involved at all and uses them for creating the signature. The signature then proves to the verifier that the corresponding message was signed by one of the owners of the public keys. Nevertheless, the verifier is not able to tell who exactly the signer is. Currently, no ring signature implementation can scale to anything beyond 10-15 participants per ring, per block, but this is sufficient for many use cases. In particular, ring signatures are less computationally and memory- intensive than an alternative privacy solution - Zero Knowledge Proofs (ZKPs). In **Table 8**, I contrast Ring Signatures with ZKPs. Another disadvantage of ZKPs like zk-SNARKs is that the initial setup is more complicated, and any future provers/verifiers, have to trust that the party who performed the initial setup did so properly and in good faith. On the other hand, ZKPs provide much stronger privacy guarantees.

Nonetheless, ring signatures can be shown to provide a "good enough" level of privacy for simple identity masking use cases. Mathematically, a ring signature algorithm involves a function which can be fully computed with just a public key, but, where knowing the private key allows you to add a ``seed" to the input that allows you to specify an arbitrary value for the output. The signature then becomes a list of values, where each value is computed as a function applied to the previous value, plus the seed. Producing a complete valid signature requires someone's private key to "close the ring" (**Figure 38**), yielding the last computed value to be equal to the first one. During validation, a ring can be verified, but there is no way to tell at which point in the ring a private key was used.

Since the execution of smart contracts on a public blockchain is not free and relies on fees that are paid to block miners, it is therefore essential to discuss who pays for peer review. Two main schemes could be proposed:

- The DAO that created the Smart Paper and submitted it for review.
- The DAO that received the Smart Paper and performed the Smart Peer Review.

In both cases, the fees would be the same, because the complexity of the underlying code and its execution does not change, it is simply a matter of policy to select whose account covers the fees.

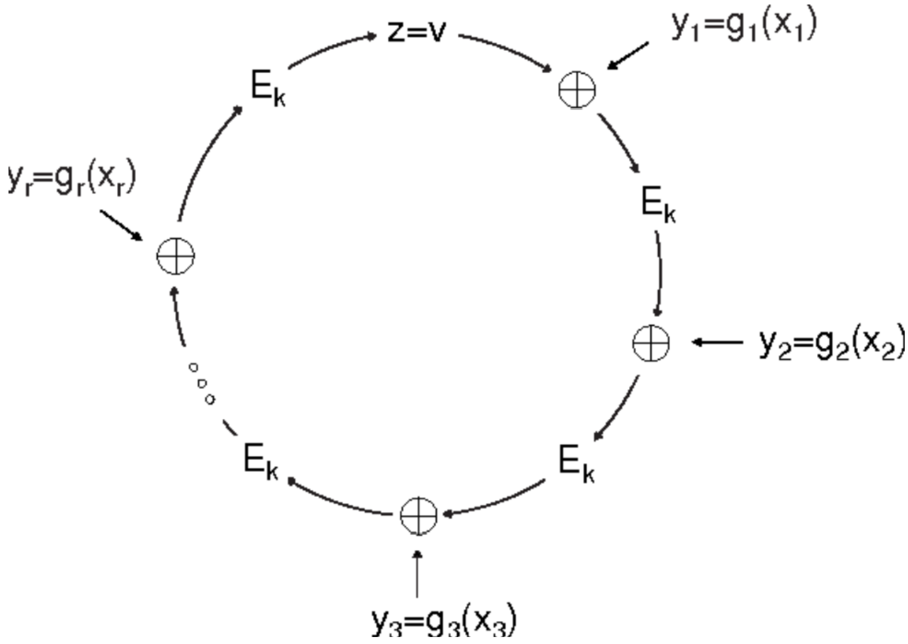



**Figure 38** - A ring signature requires a private key to close the ring

| **Feature** | **Ring Signature** | **ZKP (zksnark, zkspark)** |
|---|---|---|
| RAM utilisation | Low | High |
| Computational expense | Low | High |
| Maturity of solution | High | Medium |
| Privacy level | Medium | High |
| Trust required for setup | Low | High |

**Table 8 -** The main differences between ring signatures and zero-knowledge proofs (ZKPs)

### 5.7.4 Linkable and Unique Ring Signatures

Linkable ring signatures enable users to sign on behalf of a group, without revealing the individual signer's identity, but with the additional property that any signatures from the same signer (whether signing 1 or more messages) have a tag linking them together. Through referencing this tag, any third party is able to verify that those signatures were produced by the same signer without disclosing his or her actual identity. The tag (or link) is constructed using the signer's private key, message, and a description of the ring (this would be usually a list of public keys).

Unique ring signatures were introduced by (Franklin & Zhang, 2013) offering unforgeability, secure linkability, and restricted anonymity, meaning that if you sign twice with the same private key, that fact can be detected - but no other information is exposed. The unique ring signature (URS) scheme produces an efficient linkable ring signature implementation under random oracle model. Mobius[129] is one Smart Contract that runs on Ethereum that offers trustless autonomous tumbling (transaction mixing) using linkable ring signatures. The Mobius Mixer contract places user's token into an unfilled ring that is specific to that token type and value denomination, and provides a unique identifier for the ring, while the ring size is fixed at 4 paricipants. A caveat of the current version of the solution is that whoever ends up paying for the gas needed by the contract has their actual identity exposed. Currently the Gas usage is 150k for deposits and 725k for Withdrawals. There are opportunities for reducing these costs further. Other generic-use implementations of linkable (AOS type) ring signatures include the HarryR's SolCrypto library[130] and MattDF's RingCrypto contract[131].An example of

---

[129]https://github.com/clearmatics/mobius

[130] https://github.com/HarryR/solcrypto

[131] https://github.com/mattdf/RingCrypto/blob/master/secp256k1/ringsig-with-eclib.sol

a practical application of linkable ring signatures is Heiswap DApp[132], an Ethereum mixer for sending fixed-amount amounts of Ether with zero-knowledge privacy and stealth addresses. The ring signatures are generated in the user's browser[133] and verified on the blockchain[134].

The ease of design of linkable ring signatures, their cheap cost, and acceptable privacy levels, all contributed to my choice of this mechanism for implementing double-blind peer reviews on top of Smart Papers.

### 5.7.5 Identity Management

Whereas most blockchain applications generally guarantee user anonymity, the peer review use case calls for verifying collaborators' identities. Whilst different digital identity schemes exist, the most popular form seems to be digital certificates used to prove ownership of a public key associated with someone's private key. Even though public-private cryptography can exist in a decentralised environment, digital certificates are always issued by authorised entities. There exist multiple such authorities which makes it difficult to implement a universal solution. Due to the complexity of this issue, the logic for liaising with various types of digital certificates to verify parties' identities is normally moved to the client's user interface, as it would be too costly to include in smart contracts. New approaches such as Self-Sovereign Identity (Tobin & Reed, 2016) appear useful, with relevant open-source frameworks such as Sovrin[135] available for reuse.

In blockchain applications such as the present one, the notion of identity is a hard problem as it requires establishing a certain level of trust in a trustless

---

[132] https://heiswap.exchange

[133] https://github.com/kendricktan/heiswap-dapp/tree/master/src

[134] https://github.com/kendricktan/heiswap-dapp/tree/master/contracts

[135] https://sovrin.org/

environment. In Ethereum, and Aragon in particular, all identity is *opt-in*, being implemented as an abstraction layer situated on top of pseudonymous addresses belonging to transacting entities. In my contribution, I employ **Keybase**[136] as an opt-in identity provider. This enables peer reviewers to securely (and in a provable way) link their *soft* blockchain identities to one or more of their *hard* social identities (for example, their Twitter accounts, their Github code repositories, and their ORCID identities. Keybase, despite not being fully decentralised but rather composed of a mixture of centralised and decentralised components, has become a well-established and reliable implementation of a blockchain-enabled public-key infrastructure. Since 2014, it has serviced almost 3 million keys belonging to half a million users. The service offers a user-friendly website, mobile apps and an API that we used for automation purposes.

I deploy a smart contract that maps Keybase identities onto Ethereum accounts used in Aragon DAO. From Keybase, we use its Team functionality to pool researchers' public keys together. Those public keys are used to create rings that are used for the anonymous signing of reviews. Additionally, ENS[137] (Ethereum Name Service) is used to eliminate the need to type long Ethereum addresses when identifying blockchain entities.

---

[136] https://keybase.io

[137] https://ens.domains

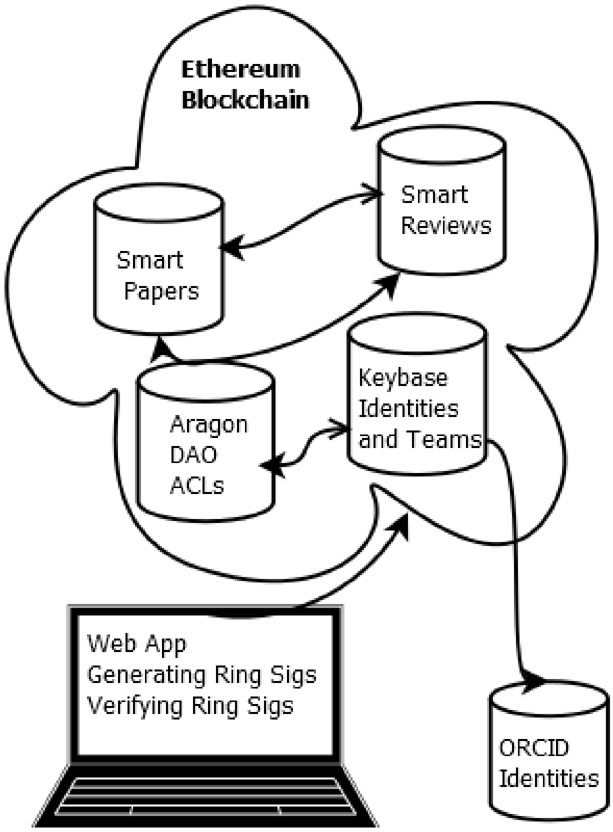



**Figure 39** - Integrating the various storage and computational components of Smart Papers and Smart Reviews

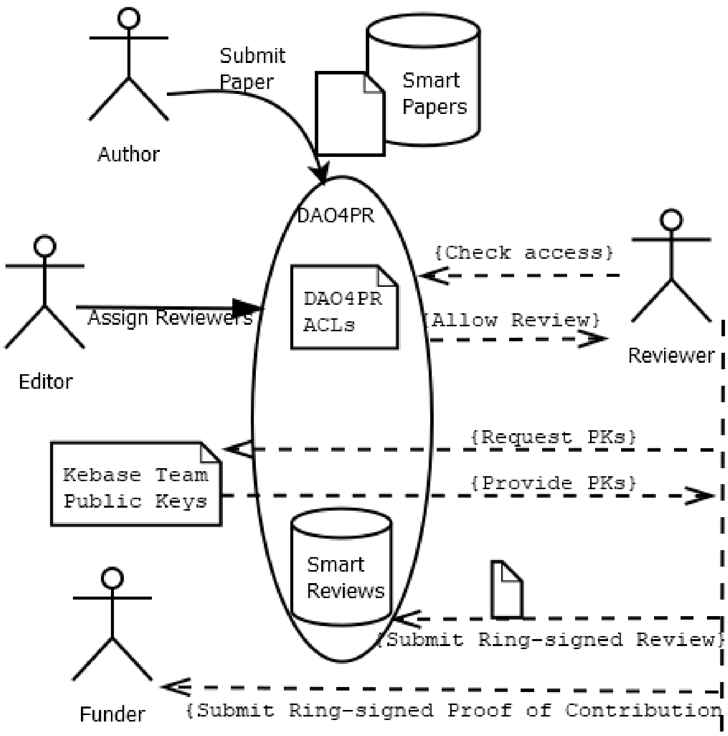



**Figure 40** - DAO4PR workflow for the reviewer

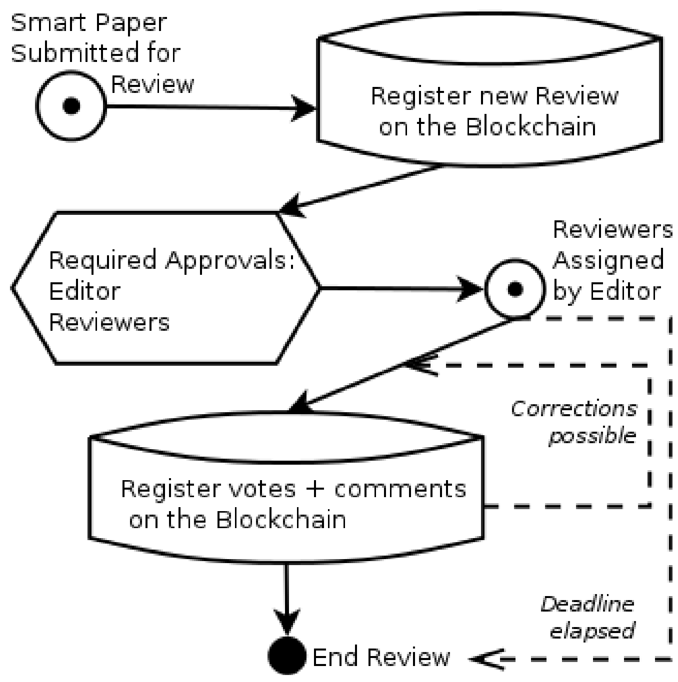



**Figure 41** - DAO4PR workflow for a new paper

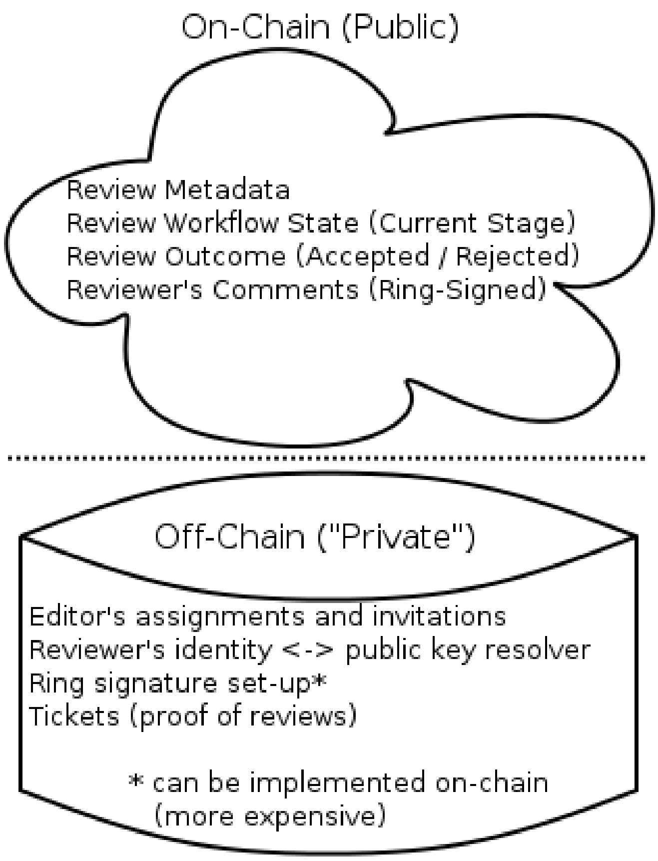



**Figure 42** - On-chain vs Off-chain data in DAO4PR

### 5.7.6 The `DAO4PR' Smart Review Model

The model in **Figure 39** shows how Smart Reviews can be implemented by a mixture of smart contracts and client-side logic and integrated with Smart Papers, such that the workflow for peer review is organised as an app in a DAO that we call DAO4PR, as shown in **Figure 40**. This app can be freely employed by journals, conferences, pre-print repositories etc. so that they integrate it with their APIs and workflows.

**Figure 42** shows how the various on-chain and off-chain data elements integrate to enable the linking of Smart Papers to Reviews, signing those reviews, and resolving participant identities. In the current implementation, the generation and verification of ring signatures and their linkages are performed online by the client code. ORCID identities are also stored off-chain due to the design of ORCID infrastructure, which is external to our solution.

**Figure 41** focuses on what happens with a new paper that needs to be reviewed. A new review is automatically created in Ethereum once an author submits their manuscript in the form of Smart Paper (which is a special kind of digital asset that implements the "Reviewable" interface) that can be identified by the *InstanceID* of the Smart Papers DAO that released it, and the *PaperID* from the said DAO. Upon being submitted for review, the Smart Paper asset is only responsible for storing its associated *Review Process ID* alongside the rest of its contents (by calling *addNewReviewProcessID* on the paper). All the other responsibilities of reviewing including review workflow and signed reviews storage are now delegated to the DAO4PR dealing with the review. The DAO4PR in receipt of the new manuscript will now progress to the approval-required state. The editor will be able to set the deadline for the review. The editor will also use a traditional (off-chain) process to send invitations, which can be done via emails and is beyond the scope of this paper. However, once invitations are accepted, the editor will assign the Reviewer permission in the DAO's ACL to the designated reviewers.

#### 5.7.6.1 Signing Reviews

To generate a ring signature, a reviewer accesses a list of public keys of other academics that are part of his or her research DAO. As such a list cannot be easily "downloaded" from Aragon, we use the *Teams* feature in Keybase which we can then map onto DAO membership lists. This is automated through the use of Keybase API. Having requested the keys from the team, the reviewer is now able to ring-sign her review. This means that she calls a *sign* JavaScript function that takes her plain-text review, her private key, and the ring of all public keys, including her own. Artificial private keys are generated for all the other reviewers, and the only genuine private key is the one from the signing reviewer. The genuine private key is required to close the ring to generate the signature. There would be no way for the outside world to use the resulting ring signature to find out the signing reviewer's true identity unless they were to explicitly reveal themselves through linkability.

The essence of this approach is that DAO team members have been vetted as scholars, so the review will carry a weight of coming from an academic, without revealing that academic's personal details. The resulting files can be stored in different ways on the blockchain. We use a *Forwarder* functionality of the DAO to pseudonymously post written reviews to the blockchain by using the DAO account address rather than reviewers' individual addresses. Moreover, the whole review process associated with a particular paper has a *ReviewProcessId* identifier that can be used to retrieve all the relevant written reviews and the accept/reject outcome for the paper. Additionally, for each review, a DAO4PR smart contract stores links to the public keys associated with the ring that was used for generating signatures. As the membership of the DAO could fluctuate, it is vital to have a snapshot of the public keys on a per-review basis.

#### 5.7.6.2 Verifying Reviews and Proving Contributions

In order to access the review, the verifier downloads three items from the review record that is stored on the blockchain: the contents of the review, the linkable ring signature and the list of public keys associated with the signing ring. During review validation, a ring-signed message can be verified against the list of public keys, and there is no way to tell whose private key had been used to sign the written review.

The reviewer has first generated her ring signature and uploaded the ring-signed review to the DAO, which published it on her behalf. By using the linkable variation of ring signatures from (Tsang & Wei, 2005), the reviewer can prove their contributions by ring-signing a "proof of contribution" letter and sending it to her funder or employer together with all her previous ring-signed contributions. As long as her private key remains the same, the interested party will be able to prove the link by repeatedly calling the *link-verify* function in their Web front-end. The first input to *link-verify* will always be the identity-containing proof, while the second input will be fed individual ring-signed contributions. If the reviewer tries to pass someone else's contribution, *link-verify* will return *false* as the private key would be different. In other cases, it will return *true*, enabling the verifier's (employers, funders) to honour the reviewer's verified contributions accordingly. The "proof of contribution" must reveal the reviewer's identity by linking to their Keybase, which is ORCID-validated.

#### 5.7.6.3 Bringing off-chain data and processes on-chain

Bringing off-chain parts of the scholarly publishing system on-chain could provide several benefits, including greater transparency, security, and efficiency. Here are some potential ways to accomplish this using no-code platforms:

1. Smart contract integration: Using a no-code platform that enables smart contract integration, the university could potentially create a custom smart contract that would handle the invitation process and link ORCID identities to on-chain data. This would allow the university to automate and streamline the invitation process while ensuring that ORCID identities are securely stored on-chain.
2. API integration: Many no-code platforms allow for easy integration with external APIs, including ORCID's API. By integrating the ORCID API with the no-code platform, the university could potentially bring ORCID identities on-chain by automatically creating and updating user profiles on the blockchain. This would allow for greater transparency and security in the verification of user identities.
3. Decentralized storage: The university could use a no-code platform that enables decentralized storage to store and manage off-chain data on the blockchain. This would provide greater security and transparency in the storage of ORCID identities, as well as other off-chain data related to the scholarly publishing system.

Oracles play an important role in the process of bringing off-chain data, such as ORCID identities, onto the blockchain. Oracles can be thought of as intermediaries between the blockchain and external data sources, providing a way to bring off-chain data onto the blockchain in a secure and reliable way. They typically use cryptographic techniques to ensure the integrity and authenticity of the data they provide and can be designed to trigger on-chain events automatically or based on certain conditions or triggers. An oracle could be used to provide a secure and reliable way to verify off-chain data and trigger on-chain events. Specifically, an oracle would be responsible for communicating with the external data source, in this case ORCID, and feeding relevant data into the blockchain. For example, an oracle could be used to verify the identity of a researcher or editor by accessing their ORCID profile and comparing it to the data stored on the blockchain. If the data matches, the oracle could then trigger an on-chain event, such as the issuance of tokens or the allocation of editorial responsibilities.

Overall, using no-code platforms to bring off-chain parts of the scholarly publishing system on-chain could provide several benefits, including greater transparency, security, and efficiency. By leveraging smart contracts, APIs, and decentralized storage, the university could create a more seamless and integrated system for managing scholarly publications and related administrative, governance and HR processes. However, it would be important to carefully evaluate the no-code platform and associated tools to ensure that they are secure and meet the needs of the university's unique use case. By providing a way to verify off-chain data and trigger on-chain events securely and reliably, oracles can help to bridge the gap between traditional data sources and the decentralized world of blockchain.

### 5.7.7 Proposed Standards

While PDF has been the dominant format for publishing scholarly papers for many years, HTML is becoming increasingly popular as a format for scholarly communication. Here are some reasons why:

1. Interactivity: HTML allows for greater interactivity and multimedia elements, such as videos, animations, and interactive figures. This can enhance the reader's experience and help to convey complex information in a more intuitive and engaging way.

2. Accessibility: HTML is more accessible than PDF, especially for readers with visual impairments or other disabilities. HTML can be easily formatted to be screen reader friendly and can be adapted for different devices and screen sizes.
3. Searchability: HTML is more searchable than PDF, allowing readers to easily search for and find specific information within a document.
4. Flexibility: HTML allows for greater flexibility in terms of layout and formatting and can be easily adapted for different types of content and devices.

In addition to HTML and PDF, there are several other formats that should be considered for scholarly communications, depending on the specific needs of the publication and its audience. These include:

1. XML: XML is a markup language that is commonly used in scholarly publishing for tagging and structuring content. It is particularly useful for machine readability and interoperability between different systems and databases.
2. JSON: JSON is a lightweight data interchange format that is becoming increasingly popular for publishing and sharing data in scholarly research.
3. Markdown: Markdown is a lightweight markup language that is easy to write and read, making it a popular choice for authors and publishers who want to create content quickly and easily.
4. EPUB: EPUB is an open standard for e-books that allows for greater interactivity and multimedia elements, similar to HTML. It is particularly useful for publishing scholarly books and other long-form content.

Overall, while PDF has been the dominant format for publishing scholarly papers for many years, HTML and other formats are becoming increasingly popular for their interactivity, accessibility, and searchability. Depending on the specific needs of the publication and its audience, publishers should consider using a variety of formats, including XML, JSON, Markdown, and EPUB, to create a more flexible and adaptable publishing system.

Furthermore, when discussing standards for publishing, all relevant data that was produced (or used) in the research paper must conform to FAIR guidelines and be compliant with the PLOS data availability guidelines prior to submission. Data must be made openly accessible and freely reusable via established institutions and standards, unless privacy concerns forbid its publishing. In any case, metadata must be made public and transparent.

Following from that, it is recommended that a Smart Reviews (DAO4PR) submission should be in one of the following formats, inspired by arXiv compatibility:

- (La)TeX, AMS(La)TeX, PDFLaTeX
- PDF
- PostScript
- HTML with JPEG/PNG/GIF images

The goal is to store articles in formats which are highly portable and stable over time. Currently, a great choice is TeX/LaTeX because this open format does not hide much information from the consumer. Note that for this reason, it is not acceptable to upload a PDF created from TeX/LaTeX source. Users of word processors such as Microsoft Word should save their documents as PDF and submit that. Note also that scanned documents should not be accepted, regardless of format. For submitting a document for review in a scholarly publishing system, there are several formats that can be considered, besides Latex, PDF, PostScript, and HTML. Here are some potential formats and their benefits:

1. Microsoft Word: While it is not a standard format for publishing scholarly papers, Microsoft Word is a widely used word processing program that many authors are familiar with. Accepting Word documents can make it easier for authors to submit their work without having to learn new tools or formats.
2. Markdown: Markdown is a lightweight markup language that is easy to write and read, making it a popular choice for authors who want to create content

quickly and easily. Accepting Markdown can help to attract authors who prefer this format and can help to streamline the submission process.

3. XML and JSON: Both XML and JSON are widely used for data exchange and can be useful for structuring and organizing content in a scholarly publishing system. Accepting XML and JSON can help to promote interoperability and data sharing between different systems and databases.

Regarding Latex, it should be acceptable because it is a popular tool among researchers for writing scientific and technical documents. Latex offers a high level of control over document formatting, allowing authors to create complex equations, tables, and figures. Accepting Latex can attract authors who prefer this tool and can help to ensure the quality and consistency of the documents being submitted.

Whether PDF should always be required to be submitted with the original document (like a Word document) from which it was generated is still an open debate. In some cases, it may be necessary to have access to the original document for editing or formatting purposes. In other cases, a PDF may be sufficient for review and publication.

In addition to the formats listed above, other formats that can boost interoperability and accessibility include:

1. RDF: RDF is a format for representing metadata and can be used to describe the relationships between different entities in a scholarly publishing system, such as authors, institutions, and publications.
2. TEI: The Text Encoding Initiative (TEI) is a standard for encoding and exchanging text-based documents, particularly those in the humanities.

Overall, accepting a variety of formats can help to attract a wider range of authors and promote interoperability and accessibility within a scholarly publishing system. It is important to carefully consider the specific needs and requirements of the system and its users when deciding which formats to accept.

#### 5.7.8 Evaluation

##### 5.7.8.1 Test set-up

I use cryptographic subkeys of size 2048-bit, which is a modern security standard. Ring signatures are generated and evaluated using Node v10.6.0 running in a Ubuntu Mate 18.10 virtual machine allocated 4GB of RAM and two i5 1.8GHz cores. When signing, we take the average length of anonymous review size from (Wang P. e., 2016) to be 477.16 words, or 2243 characters, and generate random strings of that length. Testing script can be accessed in the project's GitHub repository[138]. Blockchain smart contract transactions are executed in Ethereum Rinkeby test-net with an average block time of 15*s*, and signed with Metamask version 7.2.3. We offer the average gas price of 1 Gwei and note ETH price as $183[139] as of 13 Oct. 2019; block gas limit is 7,000,000 for Rinkeby and 8,000,000 in mainnet[140].

##### 5.7.8.2 Results

| **Linkable Ring Signatures for Anonymous Reviews** | | | | | **Storing full signatures and full reviews on-chain** | | **Storing hash-links to signed reviews** | |
|---|---|---|---|---|---|---|---|---|
| Ring size [participants] | Generate time [ms.] | Verify time [ms.] | Link-verify [ms.] | Signature size [bytes] | Ethereum [gas] | Cost [USD] | Ethereum [gas] | Cost [USD] |
| **5** | 436 | 379 | 7 | 1,795 | 1,225,135 | $ 0.22 | 73,896 | $ 0.02 |
| **10** | 712 | 683 | 11 | 3,077 | 2,086,897 | $ 0.38 | | |
| **20** | 1,261 | 1,221 | 18 | 5,642 | 3,844,424 | $ 0.70 | | |
| **30** | 1,752 | 1,767 | 25 | 8,206 | 5,633,080 | $1.03 | | |
| **42** | 2,489 | 2,434 | 36 | 11,283 | **Exceeds block gas limit** | | | |

**Table 9** - Results for linkable ring signatures of anonymous reviews stored on the Ethereum blockchain

138 https://github.com/mikehoff/DAO4PR/blob/master/test/testing-script.js
139 https://ethereumprice.org/
140 https://etherscan.io

I considered two options for storing the data on-chain, either storing full signatures and messages, or just their hashes linking to distributed file systems. I generated ring signatures for ring sizes ranging from 3 to 41; the cut-off size of 42 was found to cause block gas limit to be exceeded for direct blockchain storage. For the multihash storage case, the storage costs were fixed, as expected. The results for the *generate*, *verify*, *link-verify* times, ring signature sizes and the storage costs for both cases are given in **Table 9.**

### 5.7.9 Discussion

I have observed that signature generation, verification and link-verification times are linear with respect to the ring size, which was expected. Signature size follows the same expected trend. The cost is fixed and very inexpensive at **$0.02** when storing reviews and signatures in IPFS and storing the multihash-links on the blockchain. On the one hand, the multihash approach has the downside of introducing a dependence on IPFS availability for reviews and signatures, which increases the complexity of generating proofs of contributions. On the other hand, it seems surveycheaper to do so, even if the true storage costs are farmed out to IPFS nodes volunteering storage.

When electing to store full ring-signed reviews on the blockchain, the benefits are self-containedness of the workflow, transparency, security, and high availability, all derived from the underlying blockchain. However, this is only viable for rings up to 41 participants. Our results show that ring sizes between 10 and 20 are relatively inexpensive (between **$0.38** and **$0.70** respectively) and will offer good enough anonymity for practical purposes, considering that a pioneer ring-signature-based blockchain, Monero, fixed its ring size in 2018 at 11 participants[141] (bumped up from an earlier 5). We also note that for ring sizes up to 15, the ring signature client code should never take longer than 1 second to execute, making it suitable for Web deployment.

In terms of proving contributions, all identity-containing proofs need to be stored offline by the reviewers and must be securely communicated only to trusted verifiers who must not leak them, as to do so would jeopardise the anonymity property of the ring signatures. Although this may be seen as the weakest link of the entire set-up, such a risk that can be mitigated by signing proof-sharing agreements and/or designing a secure proof exchange protocol.

---

[141] https://github.com/monero-project/monero/issues/4229

#### 5.7.9.1 Transitioning from legacy systems to DAO-based ones

It can be difficult and time-consuming to switch from a conventional scholarly publishing system to an Ethereum DAO-based one, and it calls for careful preparation and collaboration. The following are some general actions that could be taken:

- Create a clear vision: The first step in switching to an Ethereum-based publishing system is to create a clear picture of the new system's appearance, functionality, and advantages. This should contain a review of the main components and capabilities of the current publishing system as well as a discussion of the advantages and disadvantages of adopting Ethereum.

- To find out whether Ethereum is a suitable platform for realising the intended results, a feasibility study should be carried out after the vision for the new publishing system has been formed. This study should evaluate the system's technological requirements, including the Ethereum network's scalability, security, and interoperability.

- Create the system architecture: After the feasibility study is finished, the system architecture must be formalised and finalised. Creating the system's technical requirements is required for it to store data and metadata on the Ethereum blockchain using smart contracts.

- Test the system in a limited setting with a small group of users before making the entire switch to the new publishing system. This can assist in locating any potential problems or difficulties that might need to be solved before the system is made available to a larger public.

- The system can be made available to a larger audience once it has been properly created and tested. Users may need to be trained on how to use the new system, and current data and metadata may need to be transferred to the Ethereum blockchain.

Finally, it is crucial to keep an eye on and assess the new publishing system to make sure it is achieving the anticipated results and resolving any problems that may develop. This may entail gathering user feedback, evaluating the system's performance, and making any necessary modifications to enhance the system's functionality and usability.

# Chapter 6

# Evaluating the Decentralisation of SC – a Survey

In the previous Chapter, I discussed how Smart Papers and DAO4PR, the two solutions proposed for decentralising scholarly communications, can be evaluated in technical terms of speed and cost. As explained in the Methodology chapter, Action Research methodology employed in this thesis places heavy emphasis on the empowerment of technology users and evaluating how they communicate their use of technology. Towards this end, a survey has been prepared to evaluate the decentralisation of SC.

## 6.1 Survey Purpose

The purpose of this survey was to evaluate the human factors and attitudes related to decentralising scholarly communications using computer software - with a particular focus on blockchain software. Smart Papers, the new open-source system for scholarly publishing on the blockchain, was already proposed in the previous chapter of this thesis. As this new system addresses several real-life use cases, including peer review and citation count generation, all of these needed to be evaluated in terms of their impact on academic work and culture. As the aim of Smart Papers has been to decentralise scholarly communications (to make them less reliant on big publishing houses), it is important to know the public sentiments and perceptions associated with this mode of decentralisation. The scenarios and use cases that were used to motivate and design Smart Papers were, therefore, explained in the survey and participants were asked to evaluate the key concepts from their own perspective as scholars. This survey can be seen as a tool of Action Research, which has been explained in the Methodology section of the Thesis. It is also informed by the theory of Social Machines and Sociotechnical Systems introduced in the Background section.

## 6.2 Survey Research Questions

**What is the awareness of blockchain among the scientific community?**

This question focuses on the perceptions of blockchain among scholars. It is important to measure their familiarity with this technology to estimate the effort involved in rolling out and popularising a blockchain-based system. It is also valuable to comprehend key perceptions, such as any negative pre-conceptions related to blockchain, as this will facilitate the planning of a course of action with respect to any proposed roll-out of new decentralised systems. Since the Smart Papers system involves blockchain technology, it would be essential to ask the researchers' opinion on it as it would give insight into their level of familiarity and acceptance of this emerging technology. By asking the researchers' opinion on blockchain technology, we gain insights into their level of familiarity with the technology and their potential reservations or concerns about its use in scholarly publishing. Their responses help us determine if blockchain technology is a viable solution for the proposed publishing system and identify any potential barriers to adoption. Furthermore, understanding the researchers' opinion on blockchain technology enables us to tailor communication and education efforts around the new publishing system, addressing any misunderstandings or misconceptions that may exist about the technology. Overall, asking for researchers' opinion on blockchain technology could provide valuable insights into the feasibility and potential success of the innovative scholarly publishing system.

Hypotheses: *There would be a high familiarity with the term "blockchain" but low technical awareness, requiring more work in the fields of blockchain education and usability. There may be some inaccuracies associated with how blockchain is perceived that may need to be corrected. There may also be blockers or negative factors perceived to be decreasing blockchain's suitability for decentralisation efforts – these could be social, cultural or technological.*

### What attitudes do academics present towards Smart Papers use-cases?

This question focuses on the evaluation of key use cases associated with Smart Papers usage (the "initial action" of Action Research). These were the "early" use cases designed to motivate the design of the system. Presenting them to the public, therefore, acts as a sanity check (the "evaluate the outcomes of the initial action" step of Lewin's framework). In particular, it would be useful to identify the most desirable features of Smart Papers, to be able to prioritise them for future development.

Hypotheses: *Smart Papers initial use cases would be seen as generally useful but they may be experienced differently by scholars of different disciplines or seniority levels. Some use cases may be quite niche, in which case other use cases may crystallise.*

### What attitudes do academics present towards Open Source Software in academia?

As decentralisation can be achieved in many ways, not just by means of blockchain, I would like to capture the significance of Open Source in decentralising scholarly communications. A focus on experiential accounts of using Open Source in Academia may reveal additional details about its perceived importance among scholars. As blockchain and open source are not exclusive but overlap, it would be useful to analyse what common denominators these two abstract concepts possess as viewed by the lens of scholars.

Hypotheses: *Open Source would be regarded with positive sentiment but there may be differences according to different disciplines and seniority levels. There may be some context-specific details provided by participants to help me understand the importance of blockchain and open source overlapping. Open-source smart contracts could play an important role.*

**What are the key human factors influencing the decentralisation of scholarly communications using blockchain?**

As Web Science operates within a feedback loop of technology and society, I would like to capture and identify the social elements driving the decentralisation of scholarly communications using blockchain.

Hypothesis: *Open Access and Open Science are important social themes that will be mentioned and guiding details would be provided by participants in open-ended text.*

## 6.3 Survey Design

This Survey is hosted online, via the iSurvey[142] online platform, made available by the University of Southampton. The questionnaire contains five sections related to the research questions and a sixth section capturing basic respondent information (anonymously). The named sections are:

1. **Technological Awareness and Software Needs in Academia**
2. **Attitudes and Values of Scholars Related to Decentralised Publishing**
3. **Scenario Evaluation**
4. **Factors in Adopting Open Source and Open Access**
5. **Desirable Features of Smart Papers**
6. **Anonymous Respondent Information**

The questions are a combination of closed and open-ended questions, using a variety of answering options, appropriate for each section, including Likert scales. It would take ca. 30 minutes for a participant to answer all the questions. The ethics approval was granted for the survey by the University of Southampton.

[142] https://www.isurvey.soton.ac.uk [Accessed August 2022]

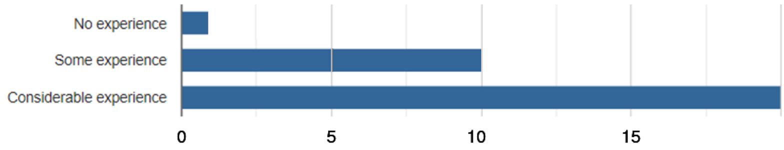
## 6.4 Participants

A cohort of 173 academics was invited to participate in this survey, from the UK, USA, Canada, Belgium, Germany, Poland, Ireland and the Netherlands. The participants were mostly recruited using the snowball approach. The inclusion criterion was defined as the participant publicly demonstrating some basic interest in scholarly communications (via Open Science, open-source software, blockchain and/or Open Access publishing). This was confirmed via their public profile (Twitter, ResearchGate, LinkedIn). Participants were approached individually by private messages, and no mailing lists were used. An exclusion criterion applied where participant would not be included if they were not currently pursuing career in academia. Participants belonged to the fields of Computer Science, Engineering, Life Sciences, Law, Physics, Psychology and other disciplines. 71 participants have responded and attempted the survey. 31 participants then provided full answers to the complete survey.

The mean of participants' years of experience of conducting research (including PhD experience) was 19 years. The median was 16 years, standard deviation was 13 years. Therefore, participants in this sample had plenty of research experience. Participants had the most experience as peer reviewers, associate editors, and as lead authors of papers. There was a general lack of experience as journal Editors-in-chief. Experience breakdown by role can be revealed in the following bar charts:

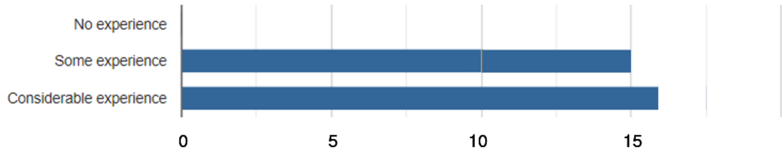
Question 5.1

Lead author of a published paper



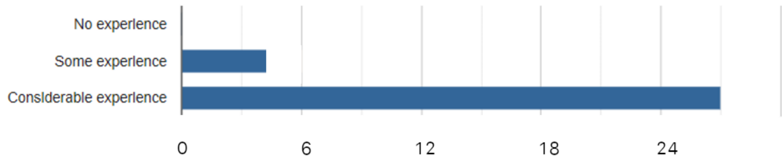

Question 5.2

Co-author of a published paper



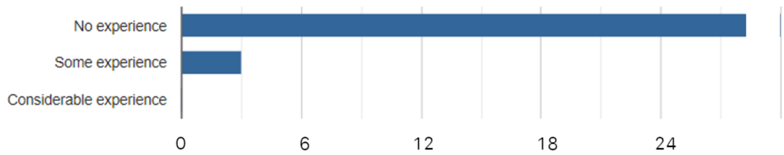

Question 5.3

Peer Reviewer



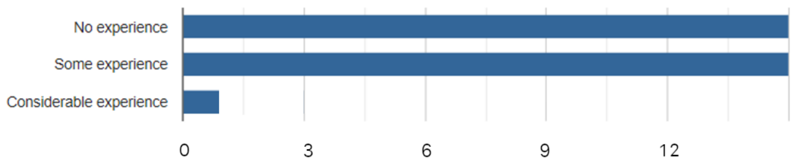

Question 5.4

Journal Editor-in-chief



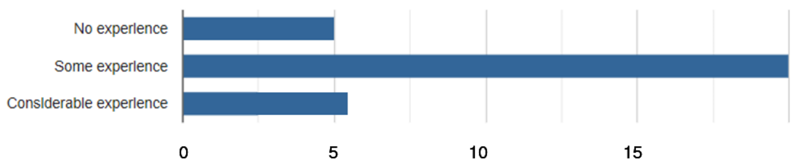

Question 5.5

Associate Editor, or other journal editor




Question 5.6

Conference chair, co-chair or assistant chair

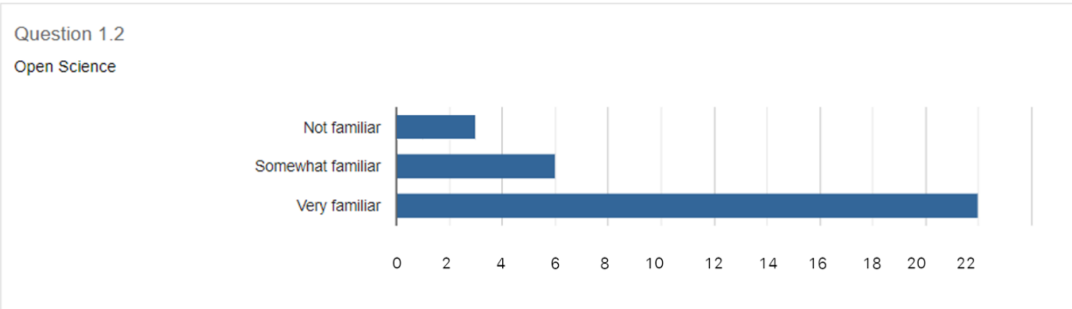


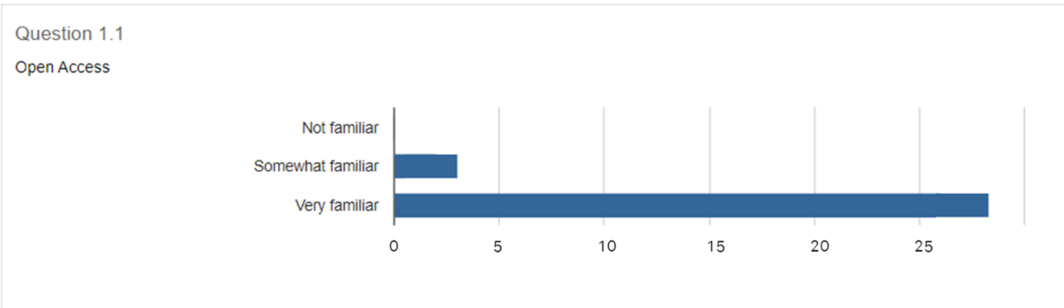
## 6.5 Findings

Key findings are summarised in the following subsections corresponding to relevant parts of the survey.

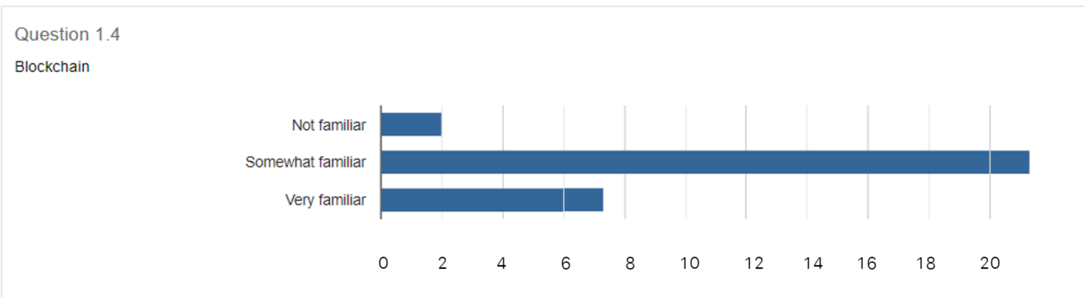
### 6.5.1 Technological Awareness and Software Needs in Academia

Participants were predominately familiar with Open Science, which was expected from the inclusion criteria of the survey. However, Open Access is even more widely recognised than Open Science.

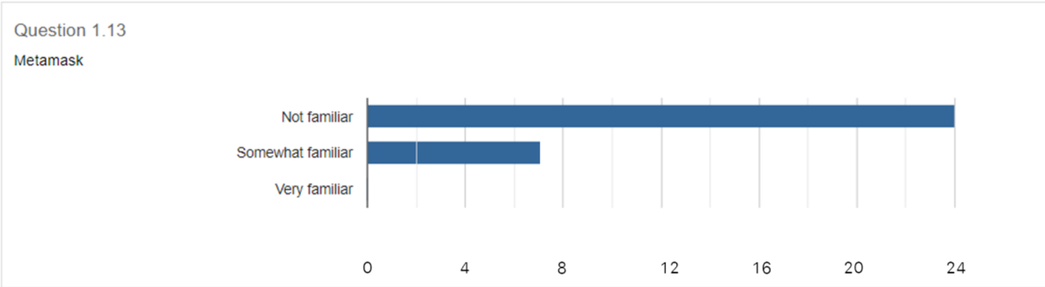


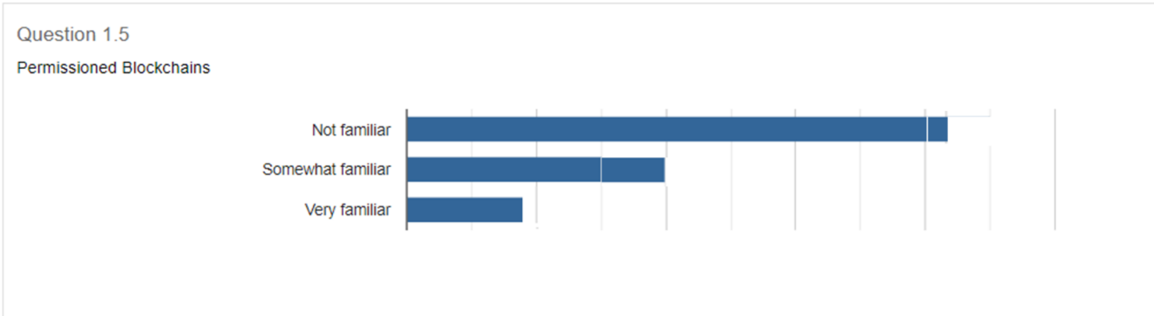

Technological awareness of blockchain technology is not as high, with only 7 people reporting being "very familiar" with blockchain technology (compared to as many as 28 people reporting being "very familiar" with Open Access), but it can be described as moderate, with 22 participants "somewhat familiar" with blockchain, as the following bar graph demonstrates:

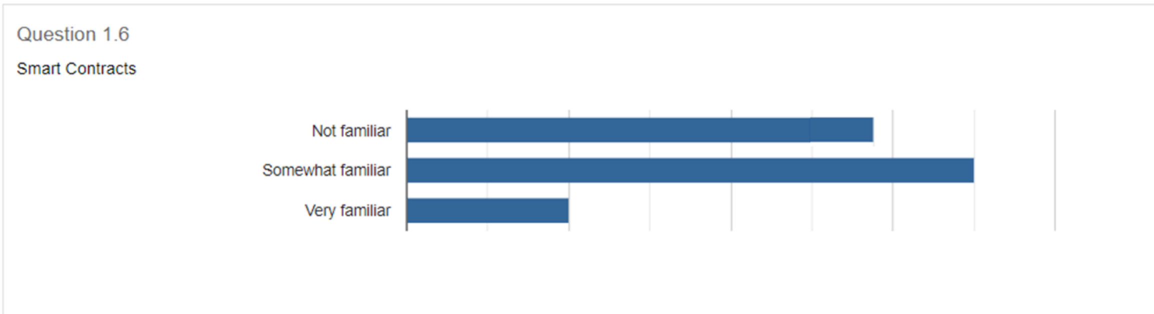


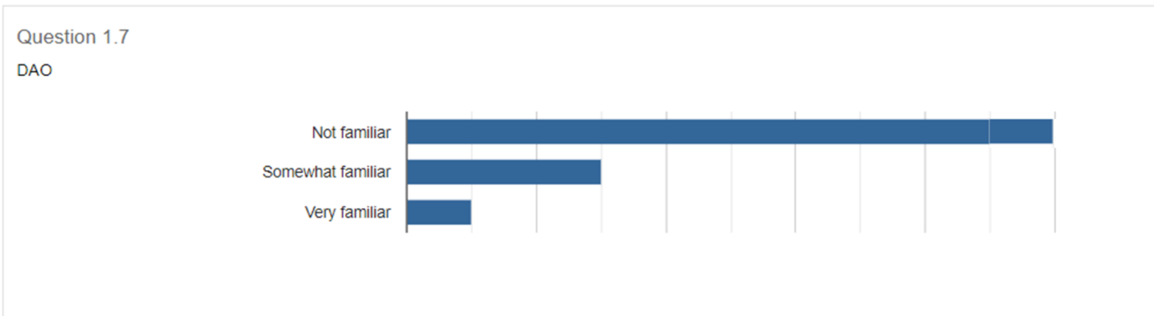
However, the self-reported blockchain knowledge decreases with the amount of specificity of this technology, so *Metamask* (a key software component for signing Ethereum blockchain transactions) is not familiar to most participants, with only 7 of them saying that they are "somewhat familiar" with it.



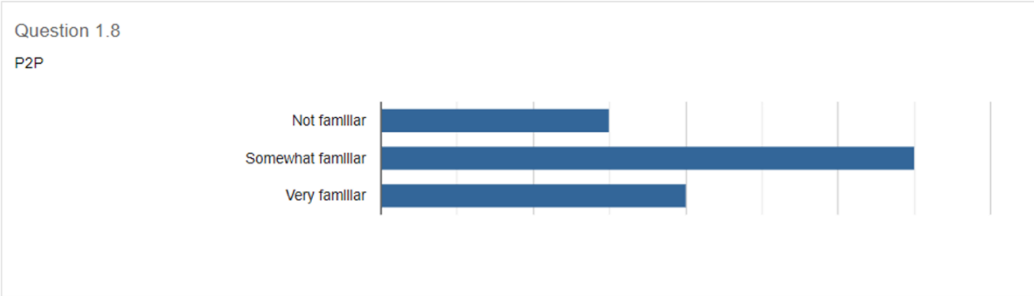

There is some low-to-moderate familiarity with the high-level concept of smart contracts, but much lower familiarity with the technological niche of DAO's and permissioned blockchains.

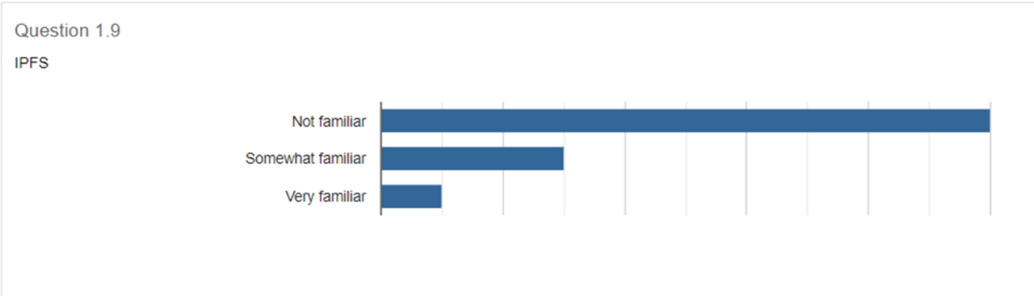


There is also some familiarity with the idea of peer-to-peer (P2P) content sharing but this familiarity is again lower when faced with a more specific Interplanetary File System (IPFS).




### 6.5.2 Required New Software

As this section asked whether the participants could see themselves personally benefitting from new software tools or services, the answers can be used to prioritise use cases from the most useful ones to the least. Note that this does not focus on blockchain software, but rather any viable software.

In terms of the absolute number of participants that directly answered "*I could benefit from a new software tool or service*", whilst ignoring all other types of responses, we have got the following sorted list of top 3 items:

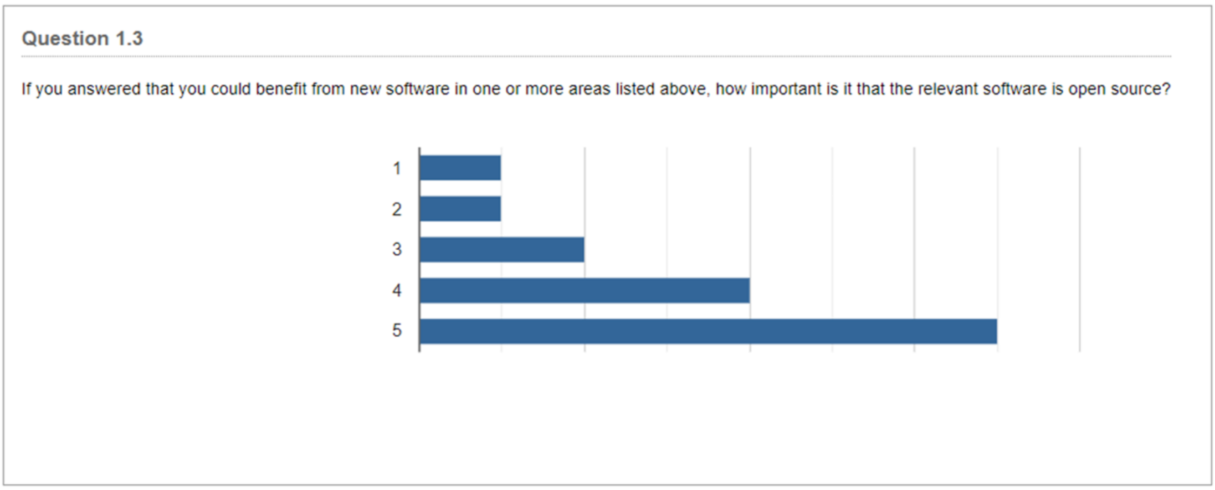
| | | |
|---|---|---|
| 1.2.13 | 19 people | verifying/tracking researchers' **reputation** |
| 1.2.11 | 17 people | identifying **experts** in particular disciplines |
| 1.2.10 | 15 people | tracking and **recognising** reviewer time and effort |

Another way of sorting (prioritising) items would be in terms of the ratio of the number of participants who responded: "*I could benefit from a new software tool or service*" divided by the number of participants who responded: "*My existing software is sufficient for this use case*", whilst ignoring all other types of responses. This seems like a rather intuitive measure of "relative importance" that yields the same sorted list of top 3 items.

| | | |
|---|---|---|
| 1.2.13 | 19/1 | verifying/tracking researchers' **reputation** |
| 1.2.11 | 17/1 | identifying **experts** in particular disciplines |
| 1.2.10 | 15/1 | tracking and **recognising** reviewer time and effort |

Note that the top three items (verifying/tracking researchers' reputation and identifying experts in particular disciplines, as well as tracking and recognising reviewer time and effort) do not change regardless which metric is used. These three top items reflect the need for new software that can facilitate the objective reporting on researcher's reputation, expertise and contribution across scientific domains, to boost verifiability of identity, tracking of productivity and recognition of effort and merit.

Furthermore, the following bar chart demonstrates that it is important that this new software should be *open source* in way of licensing:



### 6.5.3 Attitudes and Values of Scholars Related to Decentralised Publishing

This section of the survey was based on pre-determined statement shown to participants, as they indicated their attitude on a scale (see Appendix).

The top statements with which most participants strongly agreed were:

- *Making research **Open Access** should be encouraged*
    - **25** participants strongly agreed.

- *It is important to use **Open Source** software in academia even if commercial alternatives exist*
    - **17** participants strongly agreed.

- *Promoting **Open Science** can help in decentralising scholarly publishing to make it less reliant on publishing giants*
    - **15** participants strongly agreed.

The top statements with which most participants usually agreed:

- *There should be a new independent mechanism to publicly track negative reviews of a paper and how they were addressed / solved*
    - **16** participants usually agreed.
- *I am able to verify the authenticity and provenance (lineage) of the papers that I read*
    - **16** participants usually agreed.

The top statements with which most participants usually disagreed:

- *I am able to access and **verify the data used in the papers** that I read*
    - **20** participants usually disagreed.
- *Academics **receive a fair share of benefits from the papers** they publish*
    - **18** participants usually disagreed.
- *I trust how **bibliometrics (scientometrics, such as citation counts)** are calculated*
    - **16** participants usually disagreed

The top statements with which most participants strongly disagreed:

- *Academics receive a fair share of benefits from the papers they publish*
    - **16** participants strongly disagreed

- *I would have more trust in Peer Review if it relied more on algorithmic logic to calculate whether to accept a paper*
    - **13** participants strongly disagreed

- *I trust leading academic publishers in how they manage scholarly communications*
    - **13** participants strongly disagreed

- *Academic publishing works efficiently overall*
    - **10** participants strongly disagreed

The statement with the most *"don't know"* responses:

- *Using **blockchain** can help in decentralising scholarly publishing to make it less reliant on publishing giants*
    - **11** participants responded *"do not know".*

### 6.5.4 Risk Scenario Evaluation and Mitigation

The proposed risk scenarios were seen as unlikely to happen and the value of blockchain solution was unclear to the participants. To further understand why this was the case, thematic analysis was used to analyse the themes from the free-text commentaries provided by participants.

#### 6.5.4.1 Thematic Analysis of the Free-Text Responses

The following themes were identified via thematic analysis:

- **Transparency whatever the mechanism**

    - More transparency needed on contributions to prevent underplaying one researchers' contributions

- More transparency needed on editorial decision-making logic (acceptance/rejection)
- More transparent peer-review is a good idea, whatever the mechanism
- Publicly viewable reviews should be encouraged
- Editorial decisions should be trackable (to view how they were reached)
- Different transparency options should be offered, as one size does not fit all (linked to the idea of partial privacy in reviews).
- More transparency into reviewers' efforts needed, to make sure their time spent reviewing is recognised
- More transparency would be a good idea for decisions related to awarding grants to institutions

- **Integrity with blockchain**

    - Although participants view ethical breaches of trust as unlikely in scholarly communications, blockchain is seen as useful in those unlikely situations where fraudulent practices may happen, in order to improve the integrity of the publishing process
    - Blockchain, without doubt, satisfies the proof that a particular researcher had a particular idea at a specific time, helping with integrity of research provenance
    - Detecting plagiarism may be a use case for blockchain, when coupled with Artificial Intelligence / Machine Learning
    - Blockchain may be a good way of tracking true impact of articles, helping with integrity of scientometrics (as opposed to "gamed" impact).

- **Blockchain weaknesses**

    - Smart contracts are too rigid in that they lack common sense, and the current dispute resolution mechanisms are nascent or non-existent

    - It is still unknown whether blockchains will not become more and more centralised in the future

    - Trust is a complex phenomenon and can be achieved in multiple ways, so blockchain value proposition must rely on more than just "trustlessness"

### 6.5.5 Factors in Adopting Open-Source Software in Research

**Thematic Analysis of the Free-Text Responses**

The following themes were identified via thematic analysis:

- Open Source software is mostly used for typesetting and referencing
    - LaTeX, Paperpile given as examples
- Proprietary software is mostly used for collaboration and data search
    - Microsoft Office 365 given as an example
    - Google used as an example
- Closed-source is more likely to be used
    - When it's easier to use (lower learning curve)
    - Where the license was already provided (by the university, the publisher, or the organiser of a conference)
    - When the support offered is better / more responsive
    - For reasons of convenience, reliability, familiarity
- Open-source is more likely to be used
    - When you have time and resources to adopt it
    - When you rely on the community for support
    - For certain highly specialist research tasks
    - For individuals who are values-driven (idealistic)

### 6.5.6 Desirable Features of Smart Papers

The following blockchain-based features that were seen as the most desirable:

- *Time-stamped* ***pre-registration*** *of research*
    - **19** participants answered "desirable"
- *Calculating* ***bibliometrics*** (e.g., citation counts) transparently by means of smart contracts
    - **17** participants answered "desirable"
- ***Grouping papers by any combination of keywords*** *rather than organising them into journals*
    - **12** participants answered "desirable"
- *A standardised* ***linked-data format*** *for publishable papers*
    - **11** participants answered "desirable"
- ***Creating papers as records on a blockchain***
    - **10** participants answered "desirable"
- ***Managing peer reviews by privacy aware smart contracts***
    - **9** participants answered "desirable"
- ***Being able to trace a paper's evolution by viewing its different versions/iterations***
    - **9** participants answered "desirable"
- ***Creating lists of researchers relevant to a particular discipline or keyword***
    - **9** participants answered "desirable"

The blockchain-based feature with the most "somewhat desirable" responses:

- ***Using tokens to reward scholars for their work and efforts***
    - **11** participants answered "somewhat desirable"

The blockchain-based feature that was seen as the least desirable:

- *Facilitating online votes for decision-making (within research groups and labs)*
    - **13** participants answered "not desirable"

The blockchain-based feature with the most "don't know" responses:

- *Self-sovereign researcher identities (not using a corporate server to verify)*
    - **10** participants responded "do not know".

### 6.5.7 Useful Free-Text Responses

Participants described in their free-text responses, what sort of software-based approaches they would like to see more of:

Participant A: *"A standards-based approach that moves publishing in the digital age but also doesn't lock us into a technological solution.*

*As research is an human endeavour largely paid for and in service of the society (and not for corporate profit)."*

Participant B: *"Experimentation. Supporting new ideas (even if don't see their benefits) to help the good ones get off the ground."*

Participant C: *"It is very important that we focus on features that boost usability. Nobody wants to be wasting time on navigating command-line interfaces and obscure scripting languages."*

Participant D: *"The smart contracts should be able to run on any blockchain, to avoid locking into any particular blockchain platform."*

### 6.5.8 Problems or Blockers Identified

Participant E: *"Energy consumption inflexibility;*

*The creation of new de facto authorities that won't necessarily have better governance than we have currently but which may be even more centralised in terms of their practical control;*

*Requirements for new standards and workflows that are not supported by current tools;*

*Large costs to re-tool processes that will make switching unattractive for e.g. learned societies or journal/conference organisers;*

*Great difficulty in creating a single standard for published papers that recognises the very different conventions and requirements across disciplines - and the (likely) possibility that an eventual standard would not meet the needs of some disciplines and would make future innovation in the format and structure of academic communications more difficult."*

Participant F: *"What I fear in block chain mostly is the technological complexities which are build into the algorithms used (with proof of work as an extreme example) and the fragmentation of the many blockchain solutions that are available. It is very costly to add data on a blockchain not only in energy but also cost wise I don't see how large institutions can afford to put data on the chain.*

*Offchain solutions like IPFS seem to be more interesting for decentralization of data. In IPFS there is an issue that data can't be deleted which is quite problematic when data enters a network for a wrong reason (e.g. mistakes in data processing putting harmfull data online etc)."*

Participant G: *"The existing system of incentives. If promotion and tenure committees won't give due credit for works published this way then the idea is dead in the water. Building powerful new platforms does not solve cultural problems. Building them naively without regard to cultural problems is a waste of energy and money."*

Participant H: *"It requires a critical mass of researchers using it which I don't think will happen soon given the conservative nature of academia".*

# Chapter 7 - Conclusions

## 7.1 Contributions and Limitations

In this thesis, I argued why a decentralised solution is needed for scholarly communications, then proposed a solution built on top of a blockchain and evaluated it. I have shown how the Ethereum blockchain running Solidity smart contracts can foster trust and decentralise collaborative processes. As demonstrated through the Smart Papers case study, there is a lot of potential in DAOs (decentralised autonomous organisations) as a model for implementing co-operative workflows, such as research dissemination and evaluation in academia. In a broader sense, science can now be reimagined as not just relying on journals, but in a true spirit of Open Science, relying on peer-to-peer scholar interactions that are less arbitrary and more open to algorithmic logic and coded incentives. I have contributed an open-source solution for organising peer review on a public blockchain that is decoupled from existing infrastructures and overcomes problems with tracking anonymous reviews, proving their ownership and securing their availability and integrity.

The contribution enables academics to identify with ORCID up to the point of signing their reviews anonymously with linkable ring signatures. Linkability is used to prove their track records to funders and employers. I showed the viability of this solution at the cost of $0.40 per anonymous review and I believe it to be a workable alternative to current centrally operated peer-review modules running on traditional websites. Given the limitations of my workflow compared to elaborate commercial workflows that reviewers may be used to, I suggest that future work should focus on integrating this functionality into existing open-source tools such as OJS[143]. Double-blind reviews should also be implemented. As regards future qualitative work, this should focus on assessing academics' experiences of DAOs and the platform's usability.

More attention should be given to users who are the transacting parties we want to empower, with a particularly strong need for trusted Web interfaces that are user-friendly and bug-free. It is crucial to note that market conditions for smart contract execution may impact user behaviour through the variable nature of smart contract pricing and transaction completion times. One also must remember that markets do not have the upper hand and should not dictate the operations of science, as I argued in the early chapter of this thesis. Towards that goal, we must offer new mechanisms to counter the inefficient price mechanism for rewarding scientific research, focusing on altmetrics and newer forms of scientometrics. More work is needed to understand what protocols need to be developed for governance and resolving disputes. By negotiating and building on the current ways in which science is organised, rather than outright rejection of them, we will be able towhetherady journey towards the future of scholarly communications that will benefit Open Science and, ultimately, *open science.*

Survey results tell us that experienced scholars are highly familiar with Open Access publishing, and they believe that making research Open Access should be encouraged. They are also highly familiar with the umbrella term of Open Science and strongly believe that promoting Open Science can help in decentralising scholarly publishing to make it less reliant on publishing giants. For achieving openness through

143 https://docs.pkp.sfu.ca/learning-ojs/en/editorial-workflow

technology, experienced academics tend to believe that it is important to use Open Source software in academia even if commercial alternatives exist. On the other hand, they also value a pragmatic approach saying they will use whatever software serves their research goals best. They have also expressed a desire for any new scientific software developed for their purposes to be made open source by design. This gives us more confidence in that focusing on blockchain is not misplaced, because the most powerful blockchains like Ethereum, are indeed open source.

Regarding blockchain, the results suggest that there is a moderately high familiarity with the term “blockchain” but low technical awareness of the blockchain specifics (such as platforms and tools), which suggests that there is work required in the fields of blockchain education and usability among scholars. There is some low-to-moderate familiarity with the high-level concept of smart contracts, but much lower familiarity with the technological niche of DAO’s and permissioned blockchains, and virtually no familiarity with tools like Metamask (used to sign transactions on the Ethereum blockchain). In light of this low familiarity with blockchain, it is perhaps not surprising that most academics express they do not currently know whether blockchain can help them in decentralising scholarly publishing to make it less reliant on publishing giants, suggesting that a proof of concept may be needed to demonstrate the value of a blockchain-based solution. Based on the analysis of the software needs in academia (not necessarily on the blockchain) experienced scholars require new software to verify other researchers' reputation, track their reputation and identify experts in particular domains. They would also benefit from new software to facilitate data sharing and reuse and to track and recognise reviewer time and effort.

Some suggest that a standards-driven approach (designing standards such as schemas, APIs, protocols etc.) would be useful to prevent academia from being locked into any single technological solution. A culture-first approach was suggested, where cultural issues are researched and addressed first, such as the need to provide scholars with incentives to experiment with new scholarly communications systems.

Time-stamped pre-registration of research was the most desirable feature of Smart Papers. What was also seen as highly desirable included a standardised linked-data format for research, computing bibliometrics on a blockchain and making papers discoverable by any combination of keywords rather than grouping them into journals. Participants also usually agreed that there should be a new independent mechanism to publicly track negative reviews of a paper and how they were addressed / solved, and similarly that there should be a better way of tracking the provenance of data used in research. Receiving (or being able to receive) a fair share of benefits from the papers that authors publish was also highlighted as an area urgently requiring future work.

Blockchain was seen as a good candidate for solving problems related to insufficient transparency and integrity of existing scholarly communication practices, but the issues highlighted included the rigidity of smart contracts and the perceived costliness of engineering and implementing a blockchain-based solution, and the high energy consumption of Proof-of-Work blockchains (suggesting a Proof-of-Stake solution instead). Participants agreed that many elements of scholarly communications should be made more public including reviewers' efforts (for example, time spent reviewing papers), most (but not all) reviews, the exact nature of individual contributions to group research and the logic of how editors make their decisions on whether to accept or reject a paper, so that it may appear less arbitrary.

It must be noted that Smart Papers, Smart Reviews and the Peer Review workflow on the Ethereum blockchain, have proven to be extremely inexpensive, where each individual operation would cost much less than one pound sterling, even considering the fluctuating and somewhat volatile nature of cryptocurrencies. This has been compared to much costlier centralised operations of publishing houses. The code, designs, and security audits for these blockchain solutions have been made available as part of this thesis' contribution.

I have also introduced the definition for "blockchain-mediated decentralization of a system" as (Def. 1):

> *the technique for designing a new federated support network using a combination of blockchain and P2P platforms, as a means for a particular community to address one or more shortcomings in terms of inefficiency, opaqueness or vulnerability of the existing system used by that community for a particular purpose, by introducing tamper-proof records, incentives, rules and workflows aimed at breaking up the inadequate concentrations of power in the existing system, in a way that allows any subsequent improvement to be reported with agreed-upon metrics.*

and developed a framework for decentralising social machines and socio-technical systems by using blockchains, and a framework guiding the researcher in how to develop an "action plan" for their decentralization efforts. A limitation of this definition is that it is difficult to measure quantitatively how decentralized any new such created system becomes. Despite the importance of the centralization-decentralization continuum, I had noted in my literature review the lack of agreed-upon metrics to measure the degree of the decentralization of power in a system, so it is necessary that one first quantifies this power and analyse its distribution, both of which are complex tasks and riddled with methodological hurdles (Sharma, 2006, pp. 53, 55). However, using King's (1983) three decentralized describable dimensions – the locus of decision making (is it concentrated in one person, a small group, or rather dispersed across various levels), the placement of facilities (are the facilities in one place or spread around), and the locus of function (operations), my contribution – the Smart Paper model – can be described as having a decentralised locus of decision making (de-concentrated in potentially thousands of small groups – research DAOs), a decentralised locus of facilities (de-concentrated in thousands of Ethereum nodes) and a decentralised locus of function (de-concentrated in multiple instances of DAO smart contracts that can be deployed individually by different research DAO's whilst still connecting to the main Smart Papers source contract).

One shared limitation of this thesis is related to the use of bricolage (mixed methods), as well as Action Network Theory. Because both approaches rely on the researcher drawing some decisions from their experience, they both acknowledge that the researcher is enmeshed in the domain that they are studying and therefore need to be as self-reflective as possible, and there is no guarantee that another researcher would construct the same bricolage or translate their experience into the same actor-network. However, the best defense of the bricoleur was formulated by Derrida, who said that the Engineer is a myth produced by the Bricoleur, i.e., there cannot be any pure engineering approaches, especially when applied to social problems. A subject who would be the absolute origin of their own discourse is not only impossible, but she is also seen by Derrida as absurd, since “something cannot be created out of nothing”. However, this puts responsibility on the bricoleur in that they must be as transparent as possible in their methodological choices. That is why, if done again, I would keep a reflective journal of all the choices made in this thesis.

Another limitation of this thesis is related to how due to time constraints, the Action Research part only had 2 iterations – the initial Smart Papers system, the Survey, and some additions to the Peer Review functionality. Action Research works best as an iterative research method, where the researcher creates a feedback loop of multiple iterations, each evaluating the previous steps and informing the changes to be implementing in the next step. This could be done in the future.

There are also limitations of the survey itself, mostly related to how it was created. As there is no one-size-fit-all method for creating surveys evaluating sociotechnical systems and social machines, the survey was created by the researcher auditing his question choices with his supervisor via multiple Teams meetings, which was a learning process and an exercise in justifying which elements should be in and out of scope (as surveys should not take too long to fill in for the fear of losing participants or receiving incomplete answers). One major limitation is the lack of detail in the answers provided and low response rate. If done again, the survey would be followed up by interviews for more detail, and the survey itself would be shorter to increase the response rate.

Finally, the limitation of the software development cycle of this project is that despite having researched different token types on Ethereum that can be used to design incentives (mechanism design), a working prototype has not been developed due to time constraints. However, the relevant incentives have been modelled as tokens in the relevant actor network, and the general concept to be implemented (the TCR) is provided in the next section as future work remaining to be done.

Despite its limitations, the thesis successfully demonstrated how a blockchain-based sociotechnical system for scholarly communications can reduce the number of intermediaries required in academic publishing and peer review. This is achieved by providing new efficient modes of coordination through the use of blockchain smart contracts, whose execution is validated globally by the blockchain (Ethereum) network, thus providing a single version of the truth to all scholars - with fairness, transparency, and efficiency in mind.

The viability of these smart contracts, containing the logic for disintermediated AP (Smart Papers), decoupled PR (DAO4PR) and transparent scientometrics computations, was successfully evaluated computationally and sociologically. For this, I did data modelling, workflow modelling, systems modelling, sociological modelling, building the system, executed an empirical experiment and collected data from it, and carried out a questionnaire on potential system users.

The results told us that scholarly communications need to be modernised to stay fit for purpose and that blockchain solutions are useful and viable towards that end - in particular, the hereby proposed solution of open-source smart contracts running atop a Proof-of-Stake[144] blockchain, Ethereum.

[144] Ethereum is moving from Proof-of-Work to Proof-of-Stake at the time of finalising this thesis (August 2022), see "The Merge" (*"The Merge is the most significant upgrade in the history of Ethereum. Extensive testing and bug bounties were undertaken to ensure a safe transition to proof-of-stake."):*
https://ethereum.org/en/upgrades/merge/ [Accessed August 2022]

## 7.2 Summary of Answers to Research Questions

*RQ1.*

*Within the context of scholarly communications (SC), (a) what is the meaning of centralisation/decentralisation, (2) what are those components of SC that exhibit centralised tendencies and can be re-designed, re-imagined or re-formulated in a decentralised manner?*

The meaning of "decentralising SC" was found to be the re-design of an existing complex system of scholarly communications, focusing on the following key actants from the original actor-network: manuscript management systems, papers and journals, repositories, and reputation management. The reason for this was the perceived inefficiency and opaqueness of closed-source publishing workflows, the perceived inefficiency and opaqueness of scientometrics, the suboptimal formats for authoring and publishing, and the centralised storage and gatekeeping of scholarly artefacts. Please refer to Chapters 2-4 for more information.

*RQ2.*

*What are the requirements for a decentralised academic scholarly communications platform - including (a) what features are relevant? (b) how can they be implemented? (c) how can they be evaluated?*

The features of SC workflows focusing on collaboration & authoring, peer-review, publishing, scientometrics (implemented through smart contracts), modes of managerial control (through DAO Research Groups, DAO editorial groups, DAO "journals") and incentives (tokens). Evaluated based on their cost and speed, and the public perceptions / attitudes of the relevant communities. Please refer to Chapter 5 for more information.

*RQ3.*
*What are the attitudes of scholars towards decentralised Open Science software, and, towards decentralising science with blockchains?*

Survey results showed scholarly communications need to be modernised to stay fit for purpose and that blockchain solutions are useful and viable towards that end - in particular, the proposed solution of open-source smart contracts, Smart Papers, designed in line with Open Science values, in a way that promotes Open Access principles. Blockchain was a good candidate for solving problems related to insufficient transparency and integrity of existing scholarly communication practices. Please refer to Chapter 6 for more information.

## 7.3 Future Work Directions

There is scope for integrating Smart Papers and DAO4PR into existing Editorial Workflows, such as the workflow of the Open Journal Systems (OJS) software created by the Public Knowledge Project [145]. The OJS workflow is shown below:

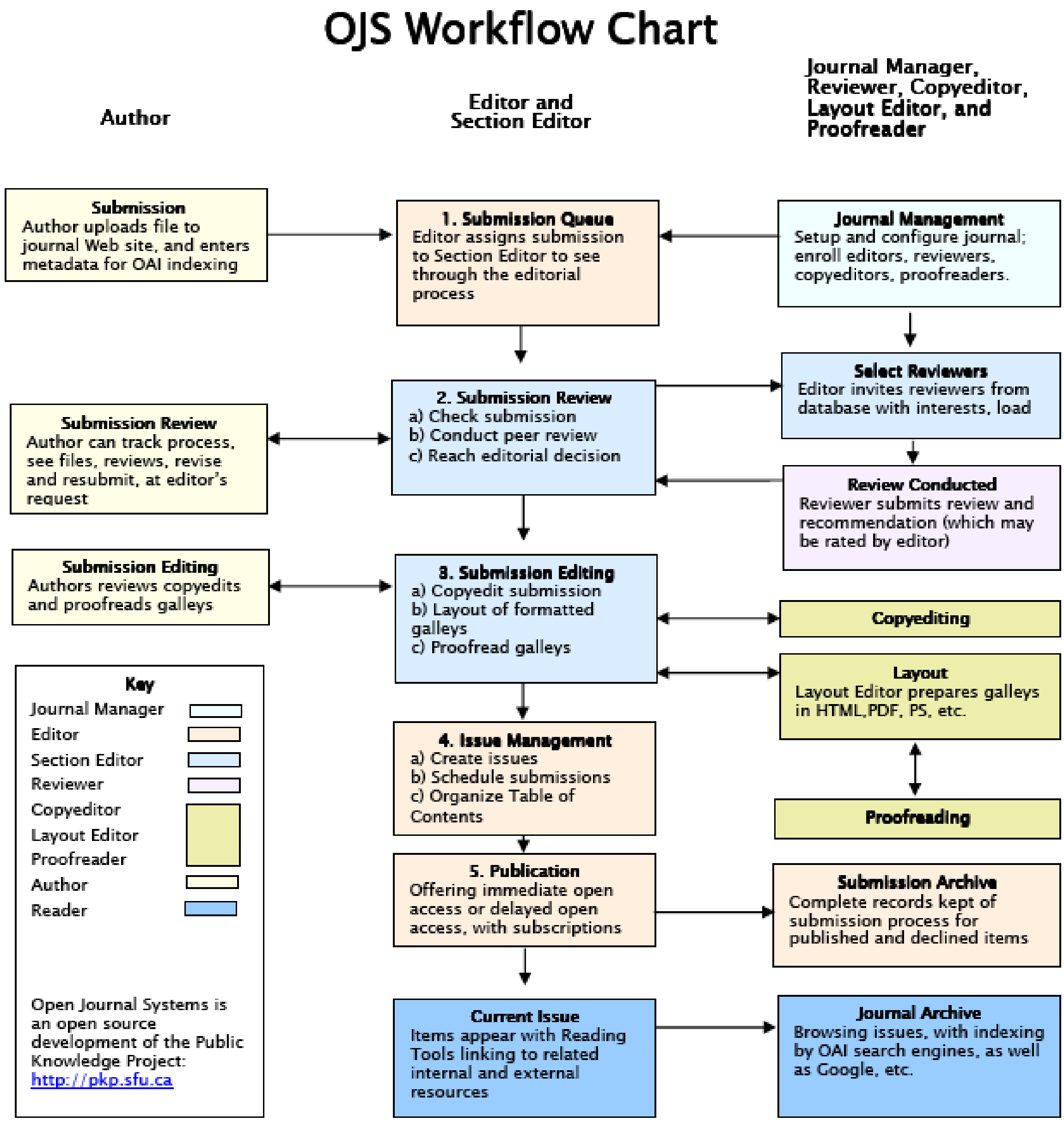



**Figure 43** - Open Journal Systems Editorial Flow

OJS is free software. distributed based on the GNU General Public License, and is written in the PHP programming language. As of 2018, OJS was being used by at

[145] https://docs.pkp.sfu.ca/learning-ojs/en/editorial-workflow

least 7,000 journals worldwide[146] A survey in 2010 found that about half of these journals originated in developing countries[147].

A successful distributed journal design should involve exploring the incentives for the storage of papers, especially the less popular ones which may be prone to disappearing. Public file storage incentive systems such as FileCoin[148] can be explored and compared with other theoretical and practical approaches.[149]

Furthermore, a successful product should rely on its client interfaces to be usable and secure. A focus needs to be placed on removing usability barriers to the greatest extent possible, to foster transparency and accountability whilst promoting human-centric governance and enforcement mechanisms, thus empowering scholars in academic settings. Another important avenue of work in this project, from a socio-technical point of view, manifests as a need to develop an understanding of how wallet system should be integrated into existing University governance and accounting structures.

Work remains to be done on the mechanisms that track reputation and incentivise fair play. Tokens have been introduced for multiple purposes, the main one being tracking the 'value' (not necessarily monetary) accumulated by a Smart Paper, and the expected value staked against it. Tokens can be further aggregated into TCR's - token curated registries, which are lists of Smart Papers (and also of scholars) that can be sorted according to different criteria. Specially designed non-transferrable tokens are also used for tracing a scholar's reputation. Tokens may also be useful for editors to help them express their interest in particular Papers. Token-Curated Registries should be managing the incentives for the purpose of listing papers, authors and reviewers according to their reputation. A Token-Curated Registry, also known

[146] https://pkp.sfu.ca/ojs/ojs-usage/ojs-stats/
[147] https://src-online.ca/index.php/src/article/view/24/41
[148] https://filecoin.io/filecoin.pdf
[149] (e.g. KopperCoin and Swarm)

as a TCR, is an incentivised game of voting that helps create trusted lists of reputable items. TCR users collectively vote, using their tokens, to decide which proposals are valid and should be added to the list. There is no TCR owner and each list has its own native token. A newcomer needs a deposit to propose a new addition to the TCR. Existing token holders can then issue a challenge against a newcomer that requires another deposit to initiate. If the application is accepted, the newcomer keeps the deposit, otherwise it's forfeited and split between the challenger and the existing token holders. Abstractly, TCR are one way of injecting trust into a blockchain. TCRs introduce the concept of reputation, and the research questions when deploying TCRs include - should there be a barrier to entry (a trusted community), how are the tokens distributed at the outset, should they be refreshed periodically and how, should they be gamified and how, etc. These are good ideas for future research papers based on Smart Papers.

Related to TCRs, keyword-based functionality should be enhanced. Currently Smart Papers can be tagged with any text-based tags, but these should become more robust and searchable by something like the Solr [150]open-source enterprise search platform, that will mean that papers on a particular subject can be discovered much faster, and that analytics on different topics can be provided much more readily.

All the contracts that we design as part of the suite of contracts that model Smart Papers and DAO4PR need to be made governable by higher-order "governance" smart-contracts, and these should reflect existing relationships, hierarchies and synergies within academia and be flexible and upgradeable so that changes to the configuration and implementation can be implemented when necessary. As blockchain-based systems evolve, we can also anticipate the need to

---

[150] https://solr.apache.org [Accessed August 2022]

update the configurations of the underlying blockchains, and not just the data stored on it. Permissioned blockchains like the ones belonging to the Hyperledger family of blockchains are more flexible in this regard, as opposed to a public blockchain like Ethereum which is more general-purpose and thus rigid. Permissioned blockchains normally enable users to define data models and transaction languages tailored to their particular use cases. For instance, Hyperledger Sawtooth, which we already introduced earlier as a permissioned consortium blockchain, provides a customisable on-chain governance mechanism. This mechanism stores configuration settings (such as which consensus mechanism is used) on the blockchain whilst enabling the users to use smart contracts (in a programming language of their choice) to vote on the changes to those configuration settings (using a voting scheme of their choice). The consensus mechanisms that already exist in Ethereum should be sufficient for building the basic smart contracts, but governance may be boosted by having a parallel governance infrastructure on a Hyperledger and integrating the two blockchains.

Furthermore, conflicts of interest need to be considered for peer-reviewing, where an off-chain mechanism needs to be provided to mediate these. Conflicts of interest (COIs) are a potential problem because they can compromise the impartiality and overall fairness of the review. COIs can arise whenever reviewers have a personal/financial stake in the outcome, such as an affiliation with a company who benefits from, or offers, related solutions. A blockchain-based system may require reviewers to disclose their COIs upfront and then record this information on-chain. This would create a transparent, auditable record that may be used to ensure that the review was conducted as objectively as possible.

Within the context of scientometrics/informetrics, crucial work remains to take the requirements further so that they encompass calculations of more substantial metrics such as citation counts per author, h-Index values (despite their attached controversy), altmetrics and to provide trustworthy decentralised computations of many useful author-, journal- and article- based bibliometrics. With the growth of distributed data analytics on the horizon, solutions are slowly appearing that enable big data analytics on blockchains, which shows real promise for developing my scientometrics/informetrics use case.

What is also important to keep in mind are the advances in privacy. As Polygon network have only just released their groundbreaking zkEVM[151] at the time of finalising this thesis (August 2022), this came in too late to change the privacy solution proposed hereby, which I based on ring signatures due to their cheaper cost. Polygon's zkEVM inherits all the security guarantees of Ethereum, whilst making it much cheaper to create highly robust and private transactions using zero-knowledge proofs, that I previously described as "expensive". The authors describe zkEVM as able to reduce the relevant transaction fees by 90%, by using Polygon Zero technology, "the fastest ZK proof in the world"[152], as the zero-knowledge proofs are calculated off-chain and then rolled-up (compressed) for storing on Ethereum.

Furthermore, crowdfunding, which is a new Internet-based form of financing a project, may also be used with DAO4PR and modelled as a worklet that is implemented using smart contracts belonging to the funding DAO. The power of science crowdfunding lies in its flexibility and independence from traditional funding structures, as it connects scientists and citizens directly in new and exciting ways. Scientists who spend time nurturing their relationships with the public will likely find this process rewarding, not just in monetary ways, and should increase the trust and understanding that the public has in scholarly processes. In an era of

---

[151] https://polygon.technology (Accessed August 2022)

[152] https://polygon.technology/solutions/polygon-zkevm/(Accessed August 2022)

budget austerity, it is also a good back-up option to explore for projects that failed on traditional grant applications. The use of blockchain Smart Papers in crowdfunding will mean that the process is fully transparent to all the parties involved.

Last but not least, evaluating the adoption of a custom cryptographic token linked to Smart Papers (the SOT token described in Chapter 5) would require carefully planned future research. The viability of this solution would depend on several factors. Here are a few key considerations:

1. Token value: The value of the custom token would need to be established and agreed upon by all stakeholders. This would require a clear understanding of the token's utility and how it could be used within the scholarly publishing system.
2. Performance metrics: The university would need to establish clear performance metrics for researchers and editors and determine how these metrics would translate into token rewards. This would require a fair and transparent evaluation process, and clear communication of expectations and goals.
3. Token distribution: The university would need to determine how the tokens would be distributed to researchers and editors, and how they could be used for compensation. This would require careful consideration of tax and legal implications, as well as the potential impact on traditional compensation structures.
4. Token governance: The governance of the token and its use within the scholarly publishing system would need to be established, potentially through a DAO or similar decentralized structure. This would require careful consideration of the rights and responsibilities of token holders, as well as mechanisms for decision-making and dispute resolution.

In terms of evaluating the usefulness of this approach, the university would need to consider several factors:

1. Cost-benefit analysis: The university would need to weigh the potential benefits of using a custom token for HR processes against the costs and resources required to implement and maintain the system.
2. Employee engagement: The university would need to evaluate the level of employee engagement and motivation resulting from the use of the token. This could be measured through employee surveys or other feedback mechanisms.
3. Impact on performance: The university would need to assess whether the use of the token has a positive impact on employee performance and the quality of research produced. This could be evaluated through metrics such as publication output, citation rates, and peer-review quality.
4. Regulatory compliance: The university would need to ensure that the use of the token complies with relevant tax, labour, and securities regulations.
5. Mechanism design: The field of economic mechanism design is of particular importance for analysing, modelling and managing incentives (Varian, 1995), as it uses the tools of economics and game theory to design rules of interaction for technology mediated economic transactions that will yield some desired social outcome. The theory of mechanism design can be described as the engineering side of economic theory, or the "reverse" of game theory, where researchers begin by defining their desired outcome or social goal. We then ask whether or not a mechanism could be designed to attain it. If the answer is positive, then we want to know what form that mechanism might take (Maskin, 2008). Usually, in economics, outcomes are predicted in markets where the rules are already known or inferred. In mechanism design, we reverse that order, by first defining the outcomes that we desire. In economics, if everyone's preferences were public knowledge, there would be no need for policing socially desirable outcomes, as they could be easily calculated using optimisation

techniques. The problem at hand is then caused by most preference information being in fact private. The 2007 Nobel Prize in Economic Sciences was awarded to Hurwicz, Maskin, and Myerson "for having laid the foundations of mechanism design theory" that deals with maximising likelihoods of socially desirable outcomes by tailoring incentives to individuals in situations where preference information is kept private by those individuals. Where real-world economic games use salaries (direct monetary incentives), fines (direct monetary disincentives), reputational boosts and hits, rights and permissions to police certain human behaviours, traditional software can only track these mechanisms. However, blockchain is a type of software that can directly enact these mechanisms using tokens and other smart contract logic. Smart contract designers, therefore, have the responsibility to design systems that incentivise ethical behaviour among participants. University would have to ensure that enough staff, students or contractors with relevant expertise are available to ensure the success of the mechanism design approach.

Overall, the use of a custom cryptographic token for HR processes in the context of a scholarly publishing system could be a useful approach, but would require careful planning, execution, and evaluation to ensure its success. Scaling the use of a custom cryptographic token for HR processes in a scholarly publishing system would also require careful planning and execution to ensure that the system can handle increasing numbers of users and transactions. Here are some key considerations for scaling:

Token supply: The university would need to carefully manage the token supply to ensure that there are enough tokens to incentivize and reward a growing number of researchers and editors. This might require periodic adjustments to the token issuance schedule and the criteria for earning tokens.

Scaling token distribution: As the number of researchers and editors on the platform grows, the university would need to determine how to fairly distribute tokens to new users while maintaining a balance of incentives and rewards. This might require the development of more sophisticated algorithms for token distribution, based on factors such as publication quality, peer-review ratings, and editorial responsibilities.

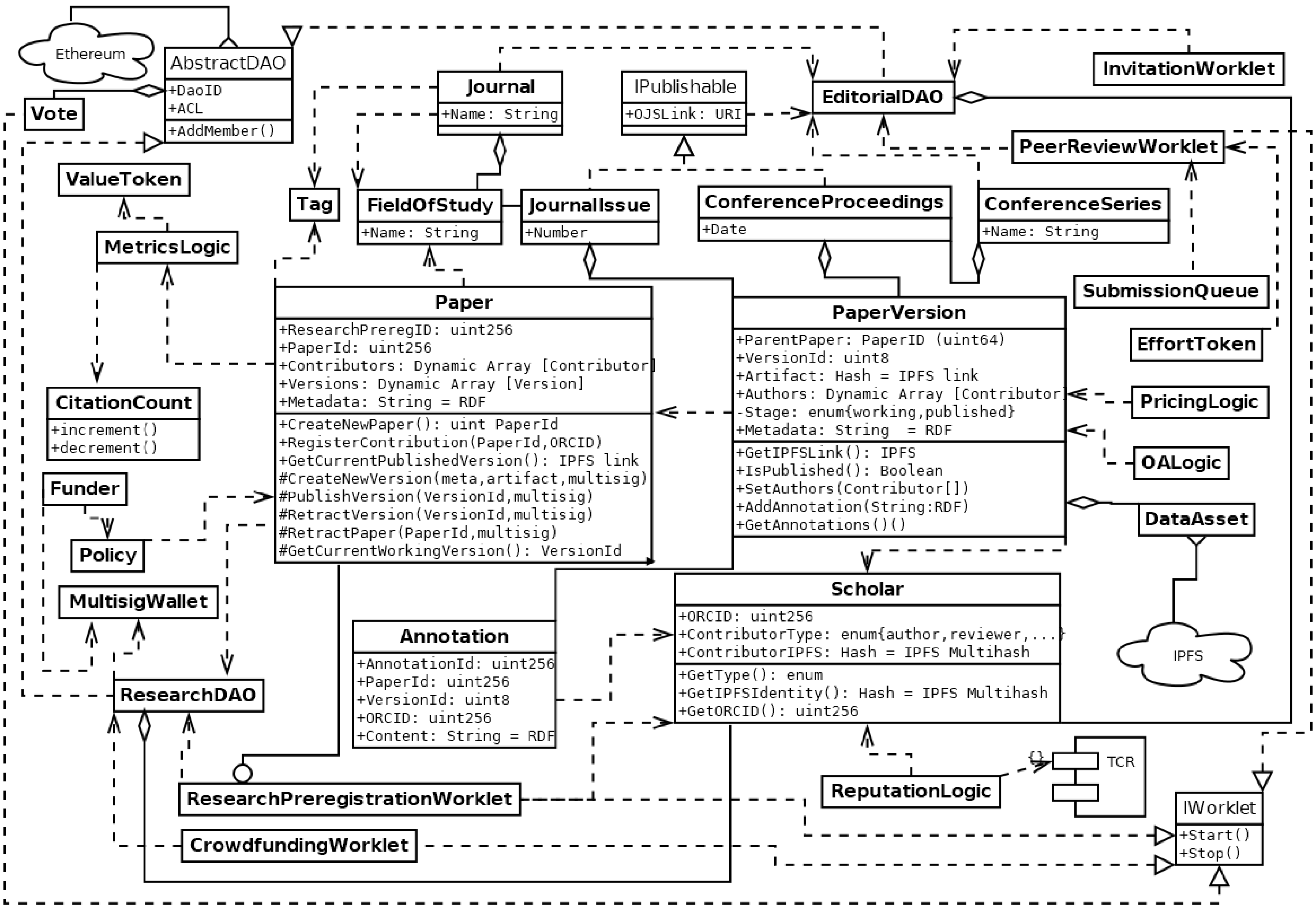
Scaling the technical infrastructure: As the number of users and transactions on the platform grows, the technical infrastructure of the system would need to be able to handle the increased load. This might require scaling up servers, databases, and other components of the system, as well as implementing more efficient algorithms for transaction processing.

Scaling governance: As the number of token holders and stakeholders on the platform grows, the governance structure of the system would need to be able to handle the increased complexity of decision-making and dispute resolution. This might require the development of more sophisticated voting mechanisms and dispute resolution procedures, as well as the creation of more specialized roles within the DAO or other decentralized governance structure.

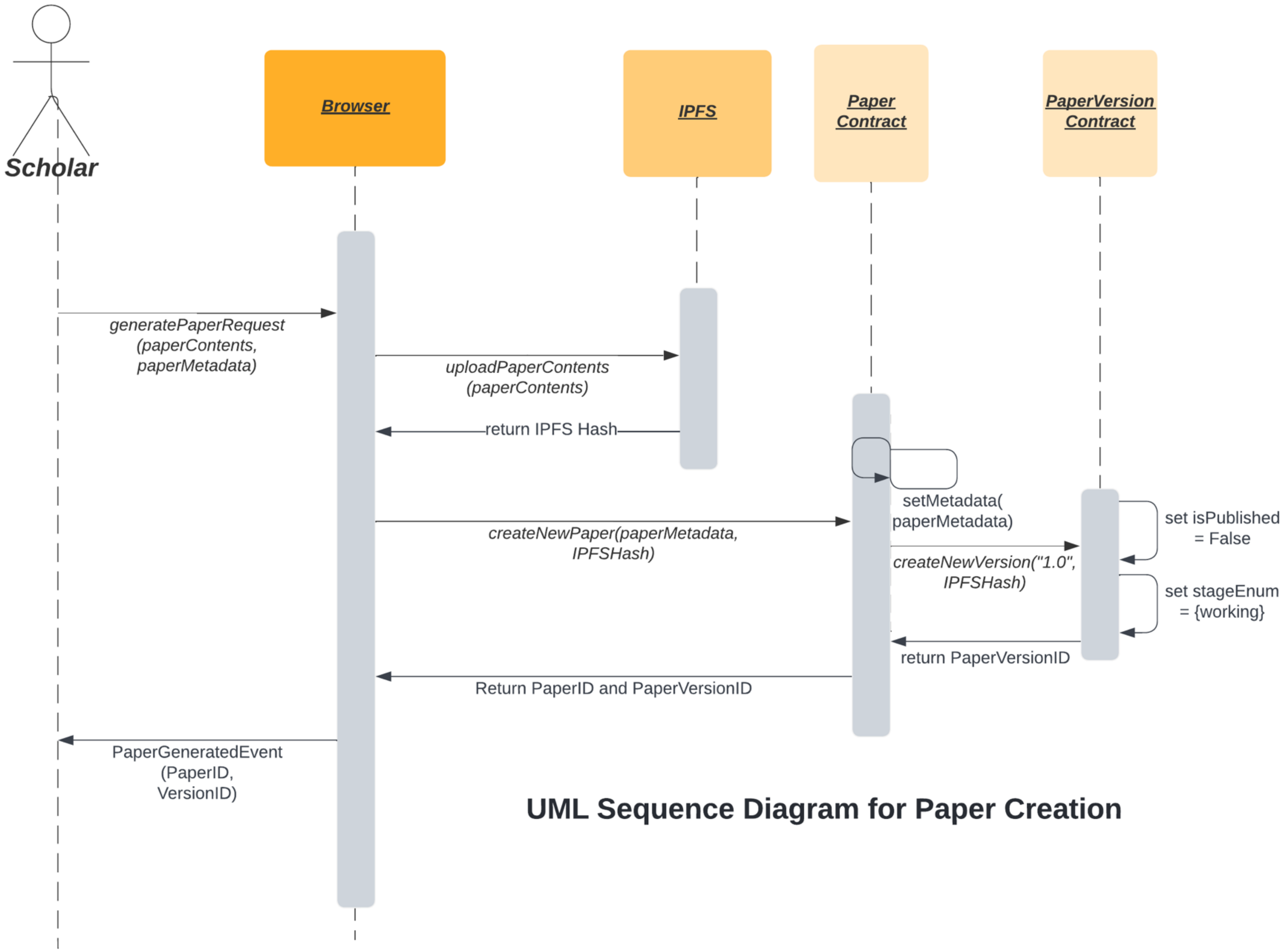
Adoption and user engagement: Ultimately, the success of the token-based HR system would depend on the adoption and engagement of users on the platform. To scale effectively, the university would need to continue to attract high-quality researchers and editors to the platform, while also ensuring that the token-based incentives and rewards remain compelling and meaningful.

Overall, scaling the use of a custom cryptographic token for administrative and governance processes in a scholarly publishing system would require the University to carefully consider technical, governance, and adoption-related factors. By addressing these challenges effectively, the university could create a token-based system that is scalable, transparent, and effective at incentivizing and rewarding high-quality research and editorial contributions. It is possible to implement such an approach on top of the system proposed by this thesis.

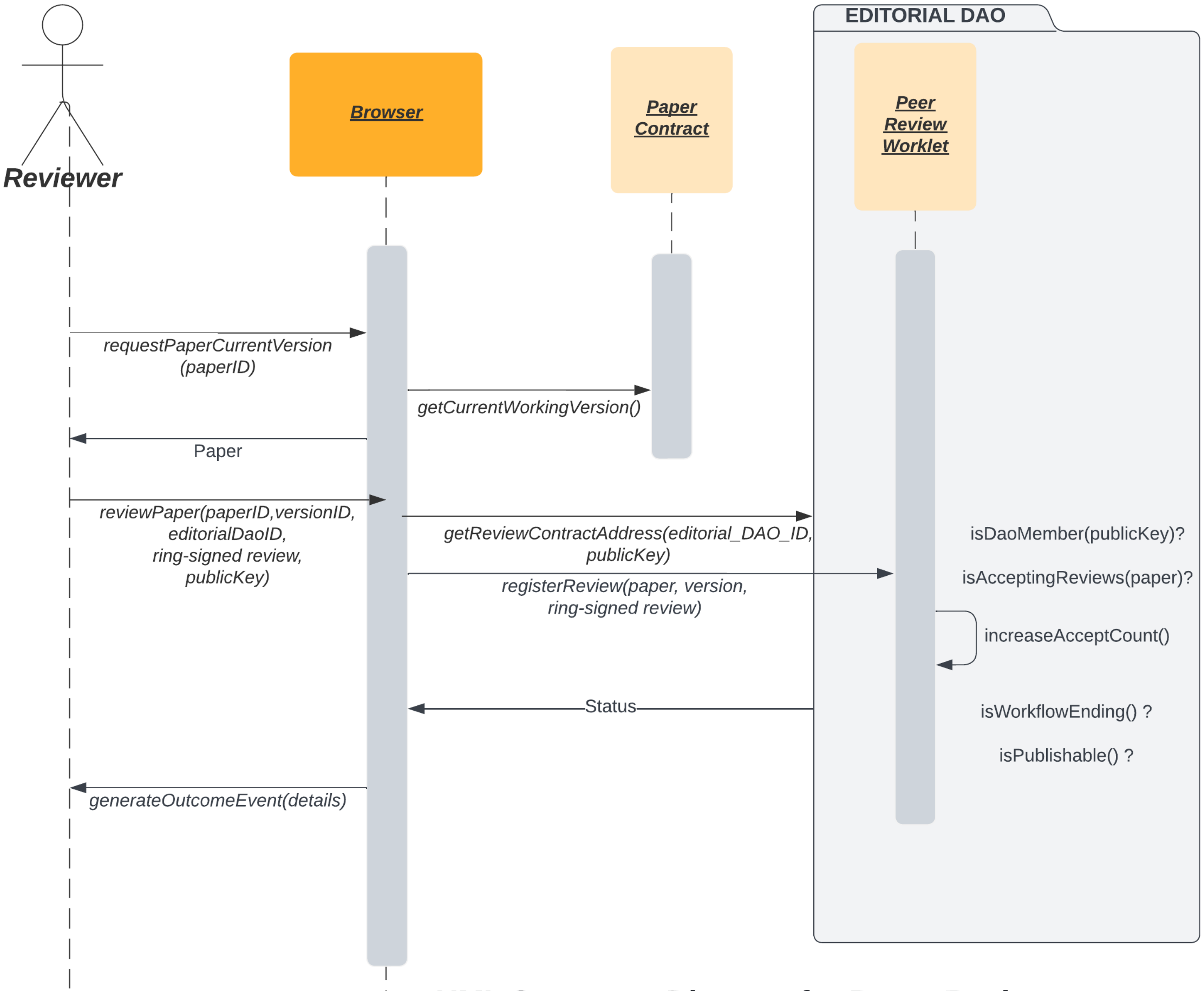
# Appendix A UML Diagram for Smart Contracts System Design



# Appendix B UML Sequence Diagram for Paper Creation



# Appendix C UML Sequence Diagram for Paper Review




UML Sequence Diagram for Paper Review

# Appendix D Security Audit for Smart Papers

**Security Audit**
**for *Smart Papers* dApp / Aragon App**
**(Ethereum Smart Contracts written in Solidity)**

**REPORT DETAILS:**

| | |
|---|---|
| **Report Version:** | 1.0 |
| **Report Date:** | Monday, 29th August 2022 |

| | |
|---|---|
| **Code Location:** | https://github.com/mikehoff/SmartPapersApp<br>https://github.com/mikehoff/SmartPapers<br>https://github.com/mikehoff/DAO4PR |
| **Description:** | Ethereum dApp for Academic Publishing |
| **Programming language:** | Solidity (pragma 0.4.24) |
| **Aragon version:** | aragon-cli 5.2.2 (aragon 0.6 / aragonOS 4) |

**Files Audited:**

**SmartPapersApp.sol**

---

**Semantic consistency checks:**

---

The purpose of this contract is to deploy an Aragon App on a testnet, whose aim is to support the creation and management of Smart Papers. This contract stores data about its owner and other related smart contracts. It also stores a structure of the Smart Paper, consisting of creator, collaborators, creation time, citation counts and other relevant metadata.
The contract exposes important functions **newPaper(), addVersion(), paperExists(), versionExists(), incrementCitations(), decrementCitations(), publishPaperExecuteVote()** which are all semantically related to paper management and creation. It also correctly extends **AragonApp**.

There are no other relevant data or functions in this contract. Based on the above analysis, the contract is **semantically consistent** with its defined purpose.

---

**Potential bugs and vulnerabilities:**

---

**Severity: LOW** **Likelihood: MEDIUM** **Risk: LOW**
**Description:** Saving potentially unencrypted confidential data on the blockchain. As Ethereum data is publicly visible, the users of this contract should be made aware that even an unpublished version of their paper may be visible to any third party. This can be easily mitigated using non-public storage for non-published versions of papers.

---

**Severity: MEDIUM** **Likelihood: MEDIUM** **Risk: MEDIUM**
**Description:** Reentrancy
As the main functions aren't modified in any way to be called only by the owner, and there are no locks, the functions could be called by multiple parties at the same time, potentially while the owner is calling other functions, which could corrupt the contract state. This is especially risky with decrementing and incrementing citations at the same time. This can be resolved by using Aragon's ReentrancyGuard.sol.

**PeerReviewWorklet.sol**

The contract is semantically consistent.

No vulnerabilities found.

General recommendation:

*Explicitly define visibility for all state variables*

**ArtifactWorklet.sol**

The contract is semantically consistent.

No vulnerabilities found.

**IWorklet.sol**

The contract is semantically consistent.

No vulnerabilities found.

**ITaggable.sol**

The contract is semantically consistent.

No vulnerabilities found.

**Artifact.sol**

The contract is semantically consistent.

No vulnerabilities found.

**IPublishable.sol**

The contract is semantically consistent.

No vulnerabilities found.

**Contributor.sol**

The contract is semantically consistent.

No vulnerabilities found.

**Dao4PRApp.sol**

---

**Potential bugs and vulnerabilities:**

---

**Severity: MEDIUM** **Likelihood: MEDIUM** **Risk: MEDIUM**

**Description:** Authorisation checks are disabled if AragonApp not configured with a kernel. This could allow unauthorised DAO members to perform actions that they are not supposed to (e.g. approving a paper).
This can be mitigated by upgrading to a newer version of Aragon that requires apps to always be attached to a kernel.

# Appendix E Ethics Approval

## Approved by Faculty Ethics Committee - ERGO II 67394

E ERGOII
To: Mike Hoffman

Sat 30/10/2021 16:59

Approved by Faculty Ethics Committee - ERGO II 67394

UNIVERSITY OF Southampton

ERGO II – Ethics and Research Governance Online https://www.ergo2.soton.ac.uk

Submission ID: 67394
Submission Title: Evaluating the Decentralisation of Scholarly Communications with Blockchain
Submitter Name: Mike Hoffman

Your submission has now been approved by the Faculty Ethics Committee. You can begin your research unless you are still awaiting any other reviews or conditions of your approval.

Comments:

-

Click here to view the submission

*TId: 23011_Email_to_submitter___Approval_from_Faculty_Ethics_committee__cat_B___C_ Id: 422938*

*M.R.Hoffman@soton.ac.uk coordinator*

# Appendix F Information Sheet and Questionnaire for the Survey

Evaluating the Decentralisation of Scholarly Communications with Blockchain

**Important Information About This Survey**

You are welcome to participate in this **anonymous** survey that aims to evaluate the human factors and attitudes related to decentralising scholarly communications using computer software, with a particular focus on **blockchain** software. The survey should take **around 30 minutes** to complete and your participation is very much appreciated, as it helps with the progress of a research project to improve scholarly communications (scholarly publishing, peer review, etc.) to make them work better for everyone.This study was **approved** by the Faculty Research Ethics Committee (FREC) at the University of Southampton

**Why have I been asked to participate and what happens next?**

You have been asked to take part because of your public profile (Twitter, ResearchGate, etc.) showing that you are interested in one of the following: Open Science, Open Access, blockchain or open-source software. Between 80 and 100 participants will be invited to participate. As this survey is anonymous, the researcher will not be able to know whether you have participated, or what answers you provided.

**What information will be collected?**

- You will be asked a number of questions about your perceptions and opinions in relation to "Smart Papers", a new open-source system for scholarly publishing. This new system addresses a number of real-life use cases including peer review and citation count generation; and its aim is to decentralise scholarly communications, to make it less reliant on big publishing houses.
- You will be asked about your awareness of blockchain and Open Science, and your attitudes towards Smart Papers use cases and Open Science in general. You will be also asked to elaborate on any factors that may be key in understanding your attitudes towards open source software and Open Access publishing.
- Anonymous information related to your research experience will also be collected.

**What will happen to the information collected?**

All information collected for this study will be stored securely on a password protected computer and backed up on a secure server. In addition, all data will be pooled and only compiled into data summaries or summary reports. Only the researcher and their supervisor will have access to this information. The information collected will be analysed and written up as part of a PhD thesis.The University of Southampton conducts research to the highest standards of ethics and research integrity. In accordance with our Data Management Policy, data will be securely destroyed after conferment of the researcher's degree in Web Science in 2022/23.

**What happens if there is a problem?**

If you are unhappy about any aspect of this study and would like to make a formal complaint, you can contact the Head of Research Integrity and Governance, University of Southampton, on the following contact details: Email: rgoinfo@soton.ac.uk, phone: + 44 2380 595058. Please quote the Ethics/ERGO number **67394**. Please note that by making a complaint you might be no longer anonymous. More information on your rights as a study participant is available via this link: https://www.southampton.ac.uk/about/governance/participant-information.page

**Thank you for reading this information sheet and considering taking part in this research. We hope you enjoy this survey!**

**Section 1. Technological Awareness and Software Needs in Academia**

Question 1.1
How familiar are you with the following:

| | Not familiar | Somewhat familiar | Very familiar |
|---|---|---|---|
| Open Access | ◯ | ◯ | ◯ |
| Open Science | ◯ | ◯ | ◯ |
| Open Source | ◯ | ◯ | ◯ |
| Blockchain | ◯ | ◯ | ◯ |
| Permissioned Blockchains | ◯ | ◯ | ◯ |
| Smart Contracts | ◯ | ◯ | ◯ |
| DAO | ◯ | ◯ | ◯ |
| P2P | ◯ | ◯ | ◯ |
| IPFS | ◯ | ◯ | ◯ |
| Decentralisation | ◯ | ◯ | ◯ |
| Scholarly Communications | ◯ | ◯ | ◯ |
| Academic Publishers | ◯ | ◯ | ◯ |
| Metamask | ◯ | ◯ | ◯ |
| Cryptographic wallets | ◯ | ◯ | ◯ |

Question 1.2.
Can you see yourself personally benefitting from a new software tool or a service designed to assist with any of the following use cases?

| | I could benefit from a new software tool or service | My existing software is sufficient for this | Use case does not concern me | I do not use software for this |
|---|---|---|---|---|
| research pre-registration | ○ | ○ | ○ | ○ |
| research project management | ○ | ○ | ○ | ○ |
| research data collection & storage | ○ | ○ | ○ | ○ |
| research version control | ○ | ○ | ○ | ○ |
| paper publishing/dissemination | ○ | ○ | ○ | ○ |
| organising and conducting peer review | ○ | ○ | ○ | ○ |
| tracking and recognising reviewer time and effort | ○ | ○ | ○ | ○ |
| transparency of research funding and lab spending | ○ | ○ | ○ | ○ |
| calculating citation counts/ bibliometrics/scientometrics | ○ | ○ | ○ | ○ |
| facilitating data sharing and reuse | ○ | ○ | ○ | ○ |
| identifying experts in particular disciplines | ○ | ○ | ○ | ○ |
| searching and sorting scholarly papers | ○ | ○ | ○ | ○ |
| verifying/tracking researchers' reputation | ○ | ○ | ○ | ○ |
| vote-based decision making for research groups | ○ | ○ | ○ | ○ |

Question 1.3.
If you answered that you could benefit from new software in one or more areas listed above, how important is it that the relevant software is open source?

Not important                    Very important

1 2 3 4 5

**Section 2. Attitudes and Values of Scholars Related to Decentralised Publishing**

Question 2.1.
Please indicate the extent to which you agree with the following statements.

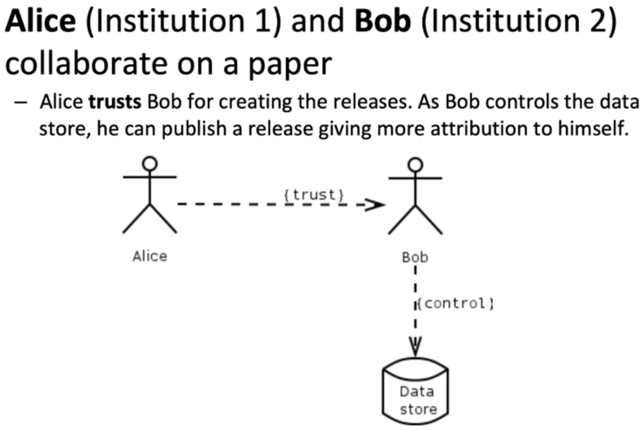
| | Strongly (always) disagree | Usually disagree | Neither disagree nor agree | Usually agree | Strongly (always) agree | Dont know |
|---|---|---|---|---|---|---|
| Open Science movement is important to me. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Peer review works efficiently overall. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Making research Open Access should be encouraged. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Blockchain looks like a useful building block for academic software. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Scholarly communication needs to be decentralised. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| I trust leading academic publishers in how they manage scholarly communications. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Peer-to-peer file sharing may be useful for research dissemination. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| It is important to use Open Source software in academia even if commercial alternatives exist. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Peer Review should be as transparent as possible. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Promoting Open Science can help in decentralising scholarly | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |

| | | | | | | |
|---|---|---|---|---|---|---|
| publishing to make it less reliant on publishing giants. | | | | | | |
| I rely on commercial tools or interfaces (including websites ran by for-profit corporations) to verify other researchers' reputation. | ○ | ○ | ○ | ○ | ○ | ○ |
| Academic publishing works efficiently overall. | ○ | ○ | ○ | ○ | ○ | ○ |
| Academics receive a fair share of benefits from the papers they publish. | ○ | ○ | ○ | ○ | ○ | ○ |
| I am able to verify the authenticity and provenance (lineage) of the papers that I read. | ○ | ○ | ○ | ○ | ○ | ○ |
| Academics receive a fair share of benefits from their peer review efforts as reviewers. | ○ | ○ | ○ | ○ | ○ | ○ |
| I would have more trust in Peer Review if it relied more on algorithmic logic to calculate whether to accept a paper. | ○ | ○ | ○ | ○ | ○ | ○ |
| Using blockchain can help in decentralising scholarly publishing to make it less reliant on publishing giants. | ○ | ○ | ○ | ○ | ○ | ○ |

| | | | | | | |
|---|---|---|---|---|---|---|
| I would have more trust in Peer Review if it relied more on cryptography to verify the reviewer's standing even if they remain anonymous. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| It is straightforward for me to identify relevant experts (or suitable peer reviewers) in my discipline. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| Mandating Open Access can help in decentralising scholarly publishing to make it less reliant on publishing giants. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| I am able to access and verify the data used in the papers that I read. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| There should be a new independent mechanism to publicly track negative reviews of a paper and how they were addressed / solved. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| I trust how bibliometrics (scientometrics, such as citation counts) are calculated. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| I would like to develop my knowledge of blockchain-related technologies. | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ |
| There should be a new independent mechanism to publicly track/verify researchers' reputation. | | | | | | |

Question 3.1.

Consider the following scenario.



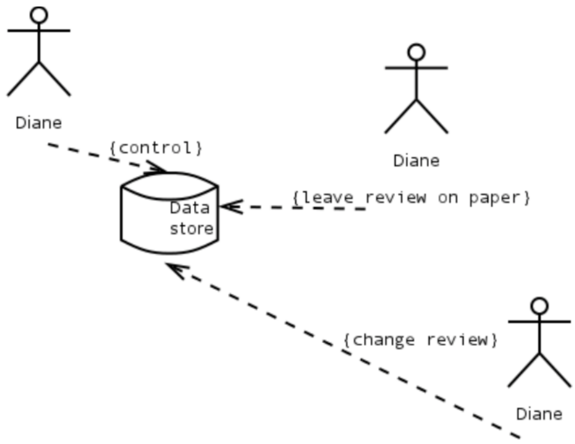

Has a similar situation ever happened to you (or someone else that you're familiar with, even if some of the details were different)? For example, you could select 'Yes' if you or your colleague were collaborating on a paper (or ideas, data) with someone who created a version of your paper (or ideas, data) that was different (or slightly different) to the version you all agreed on and then used it to their own advantage.

Question 3.2.

Consider the following scenario.

Alice and Bob submitted their article to a conference. Diane reviewed their paper.

If reviewers or editors have access rights that enable them to modify or delete reviews or acceptance decisions, those reviews or decisions could change.

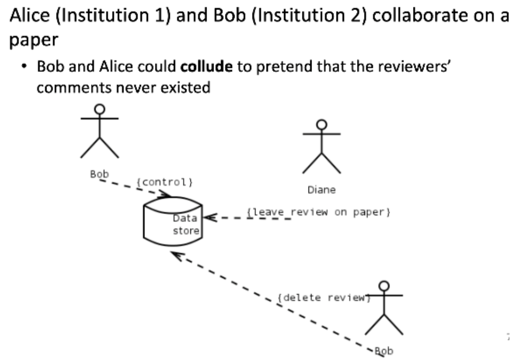



Has a similar situation ever happened to you (even if some of the details were different)?

For example, you could select 'Yes' if you, or someone you know, had a paper reviewed in a journal or at a conference, where the reviews later changed or unexpected reviews appeared, or the final decision made on the paper was unexpected on balance of the reviews that were given earlier.

Question 3.3.

Consider the following scenario.




Has a similar situation ever happened to you (or someone else that you're familiar with, even if some of the details were different)?

Question 3.4.
A blockchain timestamps events and is tamper-proof, therefore it guarantees a single version of the truth of the data stored on it. Additionally, blockchains allow smart contracts - algorithmic logic, automatically executed without relying on a central server to run on.
Based on the above, or your own knowledge, can you think of any other scenarios based on your publishing/reviewing experience, where blockchain would be beneficial?

---

## Section 4. Factors in Adopting Open Source and Open Access

Question 4.1.
Answer this question if you have authoring experience.
Think back to the last couple of papers that you worked on (if applicable) as an author of a scholarly paper.
If you used any proprietary, closed-source or commercial software to assist with your research activities, please enter the reasons why you chose that software, rather than open-source software.

Question 4.2.
Answer this question if you have editorial experience in a scholarly journal or at a conference.
Think back to the last couple of papers that you worked on (if applicable) as an editor or chair.
If you used any proprietary, closed-source or commercial software to assist with your activities, please enter the reasons why you chose that software, rather than open-source software.

Question 4.3.
Answer this question if you have peer review experience.
Think back to the last couple of papers that you worked on (if applicable) as a peer reviewer.
If you used any proprietary, closed-source or commercial software to assist with your peer review activities, please enter the reasons why you chose that software, rather than open-source software.

Question 4.4.
Answer this question if you recently used any open source software to help with your research work.
What were the benefits of using open source, rather than commercial software?

Question 4.5.
Answer this question if you have experience with open access publishing.
Compared to publishing with a traditional publishing house, is publishing fully open-access papers more difficult and/or time-consuming from a procedural (processes or workflows) perspective, and if so, why?

Question 4.6.
Answer this question if you have experience with open access publishing.
Personally, have you experienced any benefits of publishing your papers in a fully OA manner?

Question 4.7.
Answer this question if you have experience with open access publishing.
Can you think of any negative implications, or blockers, when publishing your papers as fully Open Access?

Question 4.8.
Answer this question if you have experience with open access publishing.
Can you think of any possible changes or improvements to alleviate any of the problems you have experienced with OA, if any?

**Section 5. Desirable Features of Smart Papers**

Question 5.1.
A blockchain-based Open Science Tool is being created called "Smart Papers". It leverages beneficial properties of blockchains (non-repudation, single version of truth, timestamping, decentralisation) to enable authors to share their research, and enable reviewers to share their reviews in a novel way that does not rely on existing intermediaries or trusted third parties. Please indicate which features of this tool you would find desirable.

| | Not desirable | Somewhat desirable | Desirable | Do not know |
|---|---|---|---|---|
| Creating papers as records on a blockchain | ◯ | ◯ | ◯ | ◯ |
| Calculating bibliometrics (e.g. citation counts) transparently by means of smart contracts | ◯ | ◯ | ◯ | ◯ |
| Managing peer reviews by privacy-aware smart contracts | ◯ | ◯ | ◯ | ◯ |
| Timestamped pre-registration of research | ◯ | ◯ | ◯ | ◯ |
| Peer-to-peer dissemination of research | ◯ | ◯ | ◯ | ◯ |
| A standardised linked-data format for publishable papers | ◯ | ◯ | ◯ | ◯ |
| Self-sovereign researcher identities (not using a corporate server to verify) | ◯ | ◯ | ◯ | ◯ |
| Grouping papers by any combination of keywords rather than organising them into journals | ◯ | | ◯ | ◯ |
| Curating lists of researchers relevant to a particular discipline or keyword | ◯ | ◯ | ◯ | ◯ |
| Being able to trace a paper's evolution by viewing its different versions/iterations | ◯ | ◯ | ◯ | ◯ |

| | | | | |
|---|---|---|---|---|
| Using tokens to reward scholars for their work and efforts | ○ | ○ | ○ | ○ |
| Managing the budget of a research project using shared wallets | ○ | ○ | ○ | ○ |
| Facilitating online votes for decision-making (within research groups and labs) | ○ | ○ | ○ | ○ |
| Creating journals as decentralised autonomous organisations on a blockchain | ○ | ○ | ○ | ○ |

Question 5.2.
Is there any other feature (not mentioned above) that you would find desirable?

Question 3.
Apart from those features that you believe are desirable, is there anything else that would motivate you to adopt and move onto a novel scholarly communications system?

Question 4.
What problems or blockers, if any, do you foresee when moving authoring/publishing activity to a blockchain-based scholarly communications system?

Question 5.
What problems or blockers, if any, do you foresee when moving peer reviewing and editorial activities to a blockchain-based scholarly communications system?

---

**6. Anonymous Respondent Information**
This section is used to capture basic demographic and other relevant basic data related to the respondent.

Question 6.1.
Please tell us your level of academic seniority.

Question 6.2.
Please tell us in which country you are currently basing your career.

Question 6.3.
How many years of experience of conducting research do you have (including PhD experience)?

Question 6.4.
What discipline do you specialise in

Question 6.5.
In which of these scholarly roles do you have experience?

| | No experience | Some experience | Considerable experience |
|---|---|---|---|
| Lead author of a published paper | ○ | ○ | ○ |
| Co-author of a published paper | ○ | ○ | ○ |
| Peer Reviewer | ○ | ○ | ○ |
| Journal Editor-in-chief | ○ | ○ | ○ |
| Associate Editor, or other journal editor | ○ | ○ | ○ |
| Conference chair, co-chair or assistant chair | ○ | ○ | ○ |

# Appendix G (SLR Results)

Top 16 relevant mentions

of decentra-lisation in block-chain contexts

| Context title | Citaions | Sentiment | Meaning/perspective | Themes/findings | References |
|---|---|---|---|---|---|
| Blockchain technology and decentralized governance: Is the state still necessary? | 278 | +/− | Decentralization of government services through *permissioned* blockchains is possible and desirable; decentralization is advocated by individuals and groups that fail to see the value-adding contribution of centralized institutions and the State in particular, seeing governments as "too slow, to corrupt, too lacking in innovation and benefiting too few; "blockchain-based governance" is the final stage of decentralization | Blockchain enables decentralized domain names, decentralized voting for tamper-proof ballots, decentralized autonomous organizations/corporations/societies (DAOs, DACs, DASs), and the disintermediation of all transactions on a global basis; potentially allows humans to redesign their interactions in politics; decentralization enhances "deliberative democracy"; decentralization plays a role in "multi-stakeholder" and "collaborative" governance; current state of blockchain decentralization is *pre-political* because it does not solve conflicts; decentralized platforms still have a tendency to elitism and centralization | Atzori, 2015 |
| Blockchain disruption and smart contracts | 215 | + | Decentralization means the dispersed keeping and verification of records in a public information distribution setting | Decentralization relates to consensus quality; the features of blockchain remold the landscape of competition; blockchains sustain new market equilibria with a wider range of economic outcomes | Cong and He, 2019 |
| Pervasive decentralization of digital infrastructures: a framework for blockchain enabled system and use case analysis | 201 | +/− | Technological decentralization is a driving force in the ongoing evolution and increasing openness of digital infrastructures and services | Decentralized blockchain systems could replace platform providers underlying all market models | Glaser, 2017 |
| An overview of blockchain technology: architecture, consensus, and future trends | 188 | + | Decentralization means the removal of the central trusted agency such as the central banking authority | Consensus algorithms are used to manage data consistency required for decentralization, without a trusted third party | Zheng et al., 2018 |
| decentralized applications: harnessing Bitcoin's blockchain technology | 165 | + | Decentralization of transactions requires decentralization of applications | Decentralized applications (dApps) require open-source code, cryptocurrency support, decentralized consensus and transparency with no single point of failure | Raval, 2016 |
| The invisible politics of Bitcoin: governance crisis of a decentralized infrastructure | 109 | − | Decentralization requires peer-to-peer architectures | The conception of Bitcoin as a decentralized platform was compromised by social and cultural factors; even a decentralized technology designed to promote disintermediation is unable to protect itself from capitalist tendencies to concentrate power and wealth; Bitcoin has centralized governance and oligopolistic market structure; technocratic power structure should be replaced with an institutional framework | de Filippi and Loveluck, 2014 |
| Governance in blockchain technologies and social contract theories | 93 | +/− | Blockchain decentralization has anarchist and libertarian roots; centralized powers like states and banks are easily corrupted; blockchain lets individuals create self-governing communities with enforceable rules without any centralized (hierarchical) power | Blockchain technology is not politically neutral, but transformative; political implications of blockchain are significant; decentralization requires governance; blockchain contracting offers "veil of ignorance" but lacks the idea of "common good" and distributive justice; modeling governance on the blockchain and how to govern the blockchain itself requires further research in political philosophy | Reijers et al., 2016 |
| Blockchain solutions for big data challenges a literature review | 34 | + | Decentralization requires decentralized trust which is the consensus of nodes | Decentralized trust means the opposite of client-server architecture | Karafiloski and Mishev, 2017 |
| Redecentralizing the web with distributed ledgers | 20 | + | Two axes of decentralization: architectural and application decentralization | Distributed ledgers will continue to support decentralized communities with different needs of privacy, verifiability and trust | Ibáñez et al., 2017 |
| A critical review of blockchain and its current applications | 10 | +/− | Decentralized means distributed | Decentralization means the distribution of markets, money and payments | Adhi Tama et al., 2017 |
| Blockchain: the birth of decentralized governance | 4 | +/− | Decentralization can protect individuals but also hinders coordination | New forms of "soft" decentralized governance (anarchic, aristocratic, democratic and autocratic) are required to avoid bad equilibria | Arruñada and Garicano, 2018 |
| Collusion by blockchain and smart contracts | 2 | +/− | Decentralization requires decentralized regulatory mechanisms | Blockchain decentralization complicates the work of antitrust and competition agencies | Schrepel, 2019 |
| Scholarly publishing on the blockchain–from smart papers to smart informetrics | 2 | + | A decentralized environment is an alternative to centralized publishing houses and large technology providers | Decentralization allows authors to retain the ownership of, and sovereignty over their data, and for others to calculate trustworthy computations of analytics that do not rely on any centralized data aggregator | Hoffman et al., 2019 |
| Deconstructing "Decentralization": Exploring the Core Claim of Crypto Systems | 1 | − | "Decentralization" functions as a liability shield for those operating the blockchains | The "veil of decentralization" leads to the misunderstanding of the power dynamics within blockchain systems, and faulty risk assessments | Walch, 2019 |
| Web 3.0: the decentralized web, blockchain networks and protocol innovation | 1 | +/− | Decentralization is about connecting people | The development of the decentralized web is focused on developing protocols that may not be noticed by end users | Alabdulwahhab, 2018 |
| Decentralization: an incomplete ambition | 0 | +/− | The rhetoric of decentralization diverts focus from where "concentrations of power" are operating | For decentralization to be useful in formulating future social orders, it needs to become a much more specific concept | Schneider, 2019 |

# Appendix H (SLR Results part 2)

| Context category | Citations | Sentiment | Meaning/perspective | Themes/findings | References |
|---|---|---|---|---|---|
| Governance | 1009 | + | Optimal allocation of authority | Multi-level governance, accommodating diversity, second-order coordination, "how" vs. "for whom" | Hooghe and Marks, 2003 |
| Health administration | 877 | + | Fiscal (expenditure/revenue) and administrative reform | Political redesign, transferring responsibility through policy and frameworks, allowing more stakeholders to take part in decision-making processes | Paim et al., 2011 |
| Health administration | 607 | +/− | A trade-off in ensuring healthcare quality | A multilevel approach to change; groups, teams and microsystems | Ferlie and Shortell, 2001 |
| Governance | 251 | + | Allocation of control rights under incomplete contracts | Determining levels of fiscal transfers between localities; preventing governments from appropriating resources | Seabright, 1996 |
| Environmental governance | 642 | + | Shift away from centralized form of governance | Loss of faith in the state as a custodian of nature; community-based management, hybrid forms of governance | Lemos and Agrawal, 2006 |
| Business management | 628 | +/− | Decentralization merely means that concentrated authority is delegated | Delegated authority can also be recentralized; organizations often go through these pendulum swings but both ends of the spectrum are simply different manifestations of concentrated authority | Zuboff and Maxmin, 2004 |
| Governance | 536 | − | Decentralization refers to both a state and a process | Decentralization can increase disparities, jeopardize stability, undermine efficiency | Prud'Homme, 1995 |
| Governance | 522 | + | Decentralization = major governance reform/major institutional framework | Making governance more responsive to the "felt needs" of the population, introducing checks-and-balances, technology makes it easier to provide public services | Bardhan, 2002 |
| Societal structure | 477 | +/− | Decentralized systems are flexible, networked forms of power | Networked power structures are superior to top-down chains, the state tries to regain legitimacy by decentralizing responsibilities and resources | Castells, 2000 |
| Fiscal governance | 427 | + | Decentralization means fiscal decentralization (shifting fiscal responsibility downwards) | Estimates suggest that fiscal decentralization in government expenditure is strongly and significantly associated with lower corruption | Fisman and Gatti, 2002 |
| Fiscal governance | 400 | +/− | Decentralization is not about fiscal decentralization itself, but what form does it take; paradoxically, decentralization may require a strong central government to be effective | It is difficult for a central authority to determine the particular preferences of the residents in the myriad of decentralized jurisdictions; there exists an asymmetry of information: local governments know the preferences of their residents, but the central government does not | Oates, 2005 |
| Social aspects of computing | 379 | + | Decentralization is P2P; it is community driven and depends on distributed data and distributed indexing | Decentralization enhances fault tolerance and security, but it may make regulation of the content almost impossible | Parameswaran et al., 2001 |
| Social aspects of computing | 304 | +/− | Decentralization is never complete as there is continuous struggle between the forces of centralization and decentralization | This struggle can help blurring the boundary between man and machine, and embrace social computing in which humans are part of the computation and decision-making loop, resulting in a human-centered system design | Garcia Lopez et al., 2015 |
| Environmental governance | 284 | − | Decentralization proponents are too infatuated with the local sphere | International central organizations have a critical role to play in natural resource governance | Andersson and Ostrom, 2008 |
| Complex systems | 271 | + | Decentralization overcomes specific difficulties arising in large-scale complex systems | Decentralization as decomposition improves robustness by minimizing delays and structural constraints under uncertainty | Bakule, 2008 |
| Fiscal governance | 261 | +/− | Decentralization is fiscal decentralization | Fiscal decentralization hampers growth in developing countries, but doesn't have that effect in developed countries | Davoodi and Zou, 1998 |
| Social aspects of computing | 255 | +/− | Models of decentralization follow from the limited capacities of individuals for information processing and decision making | Hierarchical structures, which are often thought of central structures, are actually effective in decentralizing the activities of information processing | Radner, 1993 |
| Conflict and tribalism | 227 | +/− | Decentralization can reduce conflict in some countries | Decentralization may increase conflict indirectly by encouraging growth of regional parties | Brancati, 2006 |
| Marketing | 189 | + | Decentralization is a Nash equilibrium strategy | Strategic interaction is a prerequisite to decentralization being profitable | Moorthy, 1988 |
| Logistics management | 180 | + | Decentralization means decentralized production | Decentralized decision-making / self-regulation; human beings, machines and resources communicate with each other as naturally as in a social network | Hofmann and Rüsch, 2017 |
| Governance | 165 | +/− | Decentralization means devolution or the transfer of power to lower levels | Authority requires legitimacy; decentralization of resources is different from decentralization of authority | Rodríguez-Pose and Gill, 2002 |
| Urban planning and transport | 143 | − | Urban decentralization as a settlement pattern | Decentralization influences planning controls for public transport | Schwanen et al., 2004 |
| Environmental governance | 123 | + | Transferring power through decentralization requires coordination, civic education campaigns | Chicken-and-egg problem; decentralization requires a critical mass to decentralize; broad resistance of central governments to local democratization and decentralization of power; what's required is applying multiple accountability measures, in addition to elections, to support democratic local institutions; decentralization requires critical analysis and informed public debate | Ribot, 2003 |
| Social aspects of computing | 88 | +/− | Economic factors dictate whether to centralize or decentralize computing | Politics of organization and resources shape the debate, centering on the issue of control; a universally appropriate arrangement has never been found | King, 1983 |
| Meta-analysis | 87 | + | Decentralization means elected local governments | Decentralization addresses the disarticulations created by globalization; economic discourse of decentralization has emerged as a central justification for the decentralization of power, | Rodríguez-Pose and Sandall, 2008 |
| Meta-analysis | 54 | +/− | Technocratic decentralization appears to create disorder, consolidates authoritarian politics and predatory economic relationships | Decentralization did not change the basic frameworks of power which remain intact; market reforms have been resisted and hijacked to consolidate predatory state and private oligarchies; the factors of transparency and accountability, and other aspects of "good governance," are no more inherent within decentralized government than centralized government; power reforms need to be enforced through political struggle | Hadiz, 2010 |

# List of Tables

# List of Figures

# List of References


Abrahamsson, P. (2017). Agile software development methods: Review and analysis. . *arXiv preprint*.

Abramo, G. (2018). Revisiting the scientometric conceptualization of impact and its measurement. *Journal of Informetrics, 12*, 590–597.

Abu-Jbara, A., Ezra, J., & Radev, D. (2013). Purpose and polarity of citation: Towards nlp-based bibliometrics. *Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: Human language technologies*, (pp. 596–606).

Ackermann, F., & Eden, C. (2011). Strategic management of stakeholders: Theory and practice. *Long range planning 44, no. 3*, 179-196.

ACRL. (2003). Principles and strategies for the reform of scholarly communication: Issues related to the formal system of scholarly communication. *College & Research Libraries News, 64*, 526–547.

Adam, D. (2002). Citation analysis: The counting house. *Citation analysis: The counting house*. Nature Publishing Group.

Afeltowicz, Ł., & Pietrowicz, K. (2011). Social Machines and Patterns of Natural Sciences: On Some Implications of Science and Technology Studies. *Polish Sociological Review*, 469-491.

Aggarwal, D., Brennen, G. K., Lee, T., Santha, M., & Tomamichel, M. (2017). Quantum attacks on Bitcoin, and how to protect against them. *arXiv preprint arXiv:1710.10377*.

Alharby, M., & van Moorsel, A. (2017). Blockchain-based Smart Contracts: A Systematic Mapping Study. *CoRR, abs/1710.06372*. Retrieved from http://arxiv.org/abs/1710.06372

Allemang, D., & Hendler, J. (2011). *Semantic web for the working ontologist: effective modeling in RDFS and OWL.* Elsevier.

Allen, C., Brock, A., Buterin, V., Callas, J., Dorje, D., Lundkvist, C., . . . others. (2015). Decentralized public key infrastructure. A White Paper from Rebooting the Web of Trust. *Decentralized public key infrastructure. A White Paper from Rebooting the Web of Trust.*

Arnold, M. C., & Gillenkirch, R. M. (2011). Centralization versus delegation in an experimental capital budgeting setting. *Business Research, 4*, 10–30.

Arruñada, B., & Garicano, L. (2018). *Blockchain: the birth of decentralized governance, in Pompeu Fabra University, Economics and Business Working Paper Series 1608.* Barcelona: Pompeu Fabra University.

Artz, D., & Gil, Y. (2007, June). A survey of trust in computer science and the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web, 5*, 58–71. doi:10.1016/j.websem.2007.03.002

Asgaonkar, A., & Krishnamachari, B. (2018). Token Curated Registries-A Game Theoretic Approach. *arXiv preprint arXiv:1809.01756.*

Asor, O. (2015). About Tau-Chain. *arXiv preprint arXiv:1502.04120.*

Assange, J., Applebaum, J., Muller-Maughn, A., & Zimmermann, J. (2012). *Cypherpunks: Freedom and the Future of the Internet.* OR Books.

B8d5ad9d974a44e7e2882f986467f4d3. (2016, August). *Towards Open Science: The Case for a Decentralized Autonomous Academic Endorsement System.* Tech. rep., Zenodo.

Baez, M. (2012, December). *Knowledge Dissemination in the Web Era.* phd, University of Trento. Retrieved November 29, 2017, from http://eprints-phd.biblio.unitn.it/902/

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature 533*, pp. 452-454.

Baran, P. (1964). On distributed communications networks. *IEEE transactions on Communications Systems, 12*, 1–9.

Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century - A review. *Journal of informetrics, 2*, 1–52.

Barrera, C. (2019). *A Framework for Blockchain Governance Design.* Hong Kong: Prysm Group.

Bartling, S. (n.d.). *Blockchain for Open Science and Knowledge Creation.* Tech. rep., Blockchain for Science. Retrieved December 20, 2017, from http://www.blockchainforscience.com/2017/02/23/blockchain-for-open-science-the-living-document/

Basu, A., & Blanning, R. W. (2000). A formal approach to workflow analysis. *Information Systems Research, 11*, 17–36.

Baym, N. K. (2010). Rethinking the music industry. *Popular Communication, 8*, 177–180.

Beck, R., Avital, M., Rossi, M., & Thatcher, J. B. (2017). Blockchain Technology in Business and Information Systems Research. *Blockchain Technology in Business and Information Systems Research*. Springer.

Beltramin, E. (2020). Against technocratic authoritarianism. A short intellectual history of the cypherpunk movement. *Internet Histories (Park Drive: Taylor and Francis).*

Benet, J. (2014). Ipfs-content addressed, versioned, p2p file system. *arXiv preprint arXiv:1407.3561.*

Berberich, M., & Steiner, M. (2016). Blockchain Technology and the GDPR-How to Reconcile Privacy and Distributed Ledgers. *Eur. Data Prot. L. Rev., 2*, 422.

Berners-Lee, T., & Fischetti, M. (1999). *Weaving the Web; The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor (2 Cassettes).* Harper Audio.

Beverungen, A., Böhm, S., & Land, C. (2012). The poverty of journal publishing. *Organization, 19*, 929-938. doi:10.1177/1350508412448858

Bjork, B.-C., Roos, A., & Lauri, M. (2009). Scientific journal publishing: yearly volume and open access availability. *Information Research: An International Electronic Journal, 14*.

Boehm, B. W., Brown, J. R., & Lipow, M. (1976). Quantitative evaluation of software quality. *Proceedings of the 2nd international conference on Software engineering*, (pp. 592-605).

Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual review of information science and technology, 36*, 2–72.

Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of informetrics, 8(4)*, 895-903.

Bornmann, L., & Marx, W. (2015). Methods for the generation of normalized citation impact scores in bibliometrics: Which method best reflects the judgements of experts? *Journal of Informetrics, 9*, 408–418.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences.* MIT press.

Braley, M. (2005, January). *Reed Elsevier: Moving The Supertanker.* Company Focus: Global Equity Research Report #36: Deutsche Bank.

Brams, S. J., & Fishburn, P. C. (1978). Approval voting. *American Political Science Review, 72*, 831–847.

Bridgman, P. W. (1940). Science: Public or private? *Philosophy of Science, 7*, 36–48.

Budish, E. (2018). *The economic limits of bitcoin and the blockchain.* Tech. rep., National Bureau of Economic Research.

Bullini Orlandi, L., Ricciardi, F., Rossignoli, C., & De Marco, M. (2018, February). Scholarly work in the Internet age: Co-evolving technologies, institutions and workflows.

Buterin, V. (2015). On Public and Private Blockchains. *On Public and Private Blockchains*. Retrieved from "https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains/"

Callon, M., Law, J., & Rip, A. (1986). *Mapping the dynamics of science and technology: Sociology of science in the real world.* Macmillan Education UK.

Capadisli, S., Guy, A., Lange, C., Auer, S., Sambra, A., & Berners-Lee, T. (2017). Linked Data Notifications: A Resource-Centric Communication Protocol. In E. Blomqvist, D. Maynard, A. Gangemi, R. Hoekstra, P. Hitzler, & O. Hartig (Eds.), *The Semantic Web* (Vol. 10249, pp. 537–553). Cham: Springer International Publishing.

Capadisli, S., Guy, A., Verborgh, R., Lange, C., Auer, S., & Berners-Lee, T. (2017, June). Decentralised Authoring, Annotations and Notifications for a Read-Write Web with dokieli. *Web Engineering* (pp. 469–481). Springer, Cham. Retrieved December 7, 2017

Casati, F., Giunchiglia, F., & Marchese, M. (2007). *Liquid Publications: Scientific Publications meet the Web.* Technical Report, University of Trento.

Castro, J. C. (2016). Social networks as dispositives of neoliberal governmentality. *Journal of media critiques, 2(7)*, pp. 85-102.

Chan, L., & Loizides, F. (2017). *Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices: Proceedings of the 21st International Conference on Electronic Publishing.* IOS Press.

Chawla, D. S. (2019, January). *Open-access row prompts editorial board of Elsevier journal to resign.* doi:10.1038/d41586-019-00135-8

Chopra, A. K., & Singh, M. P. (2016). 2016. From Social Machines to Social Protocols: Software Engineering Foundations for Sociotechnical Systems. . *Proceedings of the 25th International Conference on World Wide Web (WWW 16).* Geneva.

Chubin, D. E., Hackett, E. J., & Hackett, E. J. (1990). *Peerless science: Peer review and US science policy.* Suny Press.

Churchman, C. W. (1967). Wicked problems. *Management Science, 14(4)*, pp. 141-142.

Clack, C. D., Bakshi, V. A., & Braine, L. (2016). Smart contract templates: foundations, design landscape and research directions. *arXiv preprint arXiv:1608.00771.*

Clancey, W. J. (2006). Observation of work practices in natural settings. *The Cambridge handbook of expertise and expert performance*, 127–145.

Coase, R. H. (1937). The nature of the firm. *economica, 4*, 386–405.

Conallen, J. (2003). *Building Web applications with UML.* Addison-Wesley Professional.

Cong, L. W., & He, Z. (2019). Blockchain disruption and smart contracts. *The Review of Financial Studies, 32*, 1754–1797.

Cong, L. W., Li, Y., & Wang, N. (2018). Tokenomics: Dynamic adoption and valuation.

Contractor, F. J. (2016). Tax avoidance by multinational companies: Methods, policies, and ethics. *AIB Insights, 16*, 10.

da Silva, J. A., & Al-Khatib, A. (2017). The ClarivateTM Analytics acquisition of Publons – an evolution or commodification of peer review? *Research Ethics, 0*, 1747016117739941. doi:10.1177/1747016117739941

Dai, J. (2017). *Three essays on audit technology: audit 4.0, blockchain, and audit app.* Ph.D. dissertation, Rutgers University-Graduate School-Newark.

Davenport, E., & Cronin, B. (2000). Knowledge management: semantic drift or conceptual shift? *Journal of Education for library and information Science*, 294–306.

Davis, P. M. (2011). Open access, readership, citations: a randomized controlled trial of scientific journal publishing. *The FASEB Journal 25 (7)*, 2129–34.

Dawkins, R. (1993). In e. B. Dalhbom, *Dennett and his critics: Demystifying mind.* Blackwell.

de Pedro, A. S., Levi, D., & Cuende, L. I. (2017). Witnet: A decentralized oracle network protocol. *arXiv preprint arXiv:1711.09756.*

De Roure, D., & Willcox, P. (2020). Scholarly Social Machines: A Web Science Perspective on our Knowledge Infrastructure. *12th ACM Conference on Web Science*, pp. 250-256.

de Souza, C., Cleidson, R., & Redmiles, D. (2003). Using activity theory to understand contradictions in collaborative software development. *Automated Software Engineering.*

Deci, E. L., & Ryan, R. M. (2013). Intrinsic motivation and self-determination in human behavior. *Springer Science & Business Media.*

Dubois, , H., & Fattore, G. (2009). Definitions and typologies in public administration research: the case of decentralization. . *Int. J. Public Admin. (32)*, 704–727.

Eberhardt, J., & Tai, S. (2017). On or off the blockchain? Insights on off-chaining computation and data. *European Conference on Service-Oriented and Cloud Computing*, (pp. 3–15).

Editorial, N. (2019, April). Three-year trial shows support for recognizing peer reviewers. *Nature, 568*, 275–275. doi:10.1038/d41586-019-01162-1

ElSabry , E. (2017). Who needs access to research? Exploring the societal impact of open access. *Revue française des sciences de l'information et de la communication (11).*

Ensor, P. (1988). The functional silo syndrome. *AmE Target (16).*

(accessed January 2018). *ETH Gas Station.* Retrieved from https://ethgasstation.info/FAQpage.php

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB journal, 22(2)*, pp. 338-342.

Fanelli, D. (2018). Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences 115(11)*, (pp. 2628-2631).

Fecher, B., & Friesike, S. (2014). Open science: one term, five schools of thought. *Opening science*, 17–47.

Fecher, B., Friesike, S., Hebing, M., & Linek, S. (2017). A reputation economy: how individual reward considerations trump systemic arguments for open access to data. *Palgrave Communications, 3*, 17051.

Fehr, E., & Fischbacher, U. (2002). Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *The economic journal, 112*, C1–C33.

Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of economic perspectives, 14*, 159–181.

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016). The museum of errors/horrors in Scopus. *Journal of Informetrics, 10*, 174–182.

Frankenhuis, W., & Nettle, D. (2018). Open Science Is Liberating and Can Foster Creativity. *Perspectives on Psychological Science.*

Franklin, M., & Zhang, H. (2013). Unique ring signatures: A practical construction. *International Conference on Financial Cryptography and Data Security*, (pp. 162–170).

Fresco-Santalla, A., & Hernández-Pérez, T. (2014, November). Current and Evolving Models of Peer Review. *The Serials Librarian, 67*, 373–398. doi:10.1080/0361526X.2014.985415

Fuchs, C. (2007). *Internet and society: Social theory in the information age.* Routledge.

Gaggioli, A., Eskendari, S, Cipresso, P, & Lozza, E. (2019). The middleman is dead, long live the middleman: the "trust factor" and the psycho-social implications of blockchain. *Front. Blockchain 2:20.*

Gangemi, A., Peroni, S., Shotton, D., & Vitali, F. (2017). The Publishing Workflow Ontology (PWO). *Semantic Web, 8*, 703–718.

Garcia-Castro, A., Labarga, A., Garcia, L., Giraldo, O., Montaña, C., & Bateman, J. A. (2010, July). Semantic Web and Social Web heading towards Living Documents in the Life Sciences. *Web Semantics: Science, Services and Agents on the World Wide Web, 8*, 155–162. doi:10.1016/j.websem.2010.03.006

Garijo, D., van Hage, W. R., Kauppinen, T., Kuhn, T., & Zhao, J. (Eds.). (2017). Proceedings of the First Workshop on Enabling Open Semantic Science (SemSci). *First Workshop on Enabling Open Semantic Science (SemSci).* Aachen. Retrieved from http://ceur-ws.org/Vol-1931/

Georg, G. (2011). "Activity theory and its applications in software engineering and technology. *Colorado State University Technical Report.*

Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American sociological review, 48(6)*, 781-795.

Glaser, F., & Bezzenberger, L. (2015). Beyond cryptocurrencies-a taxonomy of decentralized consensus systems. *Proceedings of the 23rd European conference on information systems (ECIS), Münster, Germany.* eds J. Becker, J. vom Brocke, and M. de Marco.

Golbeck, J. (2008). Trust on the world wide web: a survey. *Foundations and Trends in Web Science, Vol. 1*, 131–197.

Goldbeck-Wood, S. (1999). Evidence on peer review—scientific quality control or smokescreen? *Bmj, 318*, 44–45.

Goodman, L. M. (2014). Tezos: A self-amending crypto-ledger position paper. *Tezos: A self-amending crypto-ledger position paper.*

Grigg, I. (2004). The ricardian contract. *In Proceedings. First IEEE International Workshop on Electronic Contracting,. (pp. 25-31). IEEE.*

Hürsch, W. L., & Lopes, C. V. (1995). *Separation of concerns.*

Haak, L. L., Fenner, M., Paglione, L., Pentz, E., & Ratner, H. (2012). ORCID: a system to uniquely identify researchers. *Learned Publishing, 25*, 259–264.

Halaburda, H. (2018). Blockchain revolution without the blockchain? *Communications of the ACM, 61(7), 27-29.*

Harzing, A.-W. K., & Van der Wal, R. (2008). Google Scholar as a new source for citation analysis. *Ethics in science and environmental politics, 8*, 61–73.

Hasselbring, W., Carr, L., Hettrick, S., Packer, H., & Tiropanis, T. (2020). Open source research software. *Computer (53)*, pp. 84-88.

Hawlitschek, F., Notheisen, B., & Teubner, T. (2020). A 2020 perspective on "The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy". *Electronic Commerce Research and Applications, 40*.

Hawlitschek, F., Notheisen, B., & Teubner, T. (2018). The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy. *Electronic Commerce Research and Applications*.

Hayek, F. (1944). *The Road to Serfdom.* Chicago: University of Chicago Press.

Hayek, F. A. (1960). v.(1960): The constitution of liberty. *Chicago: University of Chicago.*

Haythornthwaite, C. (2009). Crowds and communities: Light and heavyweight models of peer production. *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*, (pp. 1–10).

Heller, L., The, R., & Bartling, S. (2014). Dynamic Publication Formats and Collaborative Authoring. In S. Bartling, & S. Friesike (Eds.), *Opening Science* (pp. 191–211). Cham: Springer International Publishing. Retrieved December 20, 2017

Hess, C., & Ostrom, E. (2004). Studying Scholarly Communication: Can Commons Research and the IAD Framework Help Illuminate Complex Dilemmas?

Hess, C., & Ostrom, E. (2007 a). Introduction: An overview of the knowledge commons.

Hess, C., & Ostrom, E. (2007 b). A framework for analyzing the knowledge commons.

Hessels, L. K., & van Lente, H. (2008). Re-thinking new knowledge production: A literature review and a research agenda. *Research policy, 37(4)*, 740 - 760.

Higginson, A. D., & Munafo, M. R. (2016). Current Incentives for Scientists Lead to Underpowered Studies with Erroneous Conclusions. *PLoS Biol 14 (11).*

Hill, S. A. (2016). Making the future of scholarly communications. *Learned Publishing, 29*, 366-370. doi:10.1002/leap.1052

Hindriks, J., & Lockwood, B. (2004). Centralization and political accountability.

Hoffman, M. R. (2018). Can blockchains and linked data advance taxation. *Companion Proceedings of the The Web Conference*, (pp. 1179-1182).

Hoffman, M. R., Ibanez, L. D., & Simperl, E. (2019). Scholarly publishing on the blockchain - from smart papers to smart informetrics. *Data Science, 2*, 291-310.

Hoffman, M. R., Ibanez, L. D., Fryer, H., & Simperl, E. (2018). Smart papers: Dynamic publications on the blockchain. *European Semantic Web Conference* (pp. 304-318). Springer, Cham.

Hoffman, M., Ibanez, L. D., & Simperl, E. (2020). Toward a Formal Scholarly Understanding of Blockchain-Mediated Decentralization: A Systematic Review and a Framework. *Frontiers in Blockchain, 3.*

Holbrook, J. B. (2019). Open science, open access, and the democratization of knowledge. *Issues in Science and Technology, 35*, 26–28.

Hull, R., Batra, V. S., Chen, Y.-M., Deutsch, A., Heath, F. T., & Vianu, V. (2016). Towards a Shared Ledger Business Collaboration Language Based on Data-Aware Processes. *ICSOC.*

Hursch, W. L., & Lopes, C. V. (1995). *Separation of Concerns.* Tech. rep., NorthEastern University.

Ibanez, L. D., Hoffman, M. R., & Choudhry, T. (2019, June). *Blockchains and Digital Assets.* Retrieved from EU Blockchain Forum: https://www.eublockchainforum.eu/sites/default/files/research-paper/blockchains_and_digital_assets_june_version.pdf

Inc., P. (2018, January). Pluto Network. Retrieved from https://assets.pluto.network/Pluto_white_paper_v04_180719_1355_BSH.pdf

Innis, H. (2007). *Empire and Communications.* Toronto: Dundurn Press.

Jan, Z., Third, A., Ibanez, L.-D., Bachler, M., Simperl, E., & Domingue, J. (2018). Sciencemiles: Digital currency for researchers. *Companion of the The Web Conference 2018 on The Web Conference 2018-WWW'18*, (pp. 1183–1187).

Janara, L. A. (1998). *After the Mother: Authority, Autonomy and Passion in Tocqueville's Democracy in America.* Minneapolis, MN: University of Minnesota.

Jang, M., Patted, T., Gil, Y., Garijo, D., Ratnakar, V., Ji, J., . . . Jahanshad, N. (2017). Automatic Generation of Portions of Scientific Papers for Large Multi-Institutional Collaborations Based on Semantic Metadata. *Proceedings of the First Workshop on Enabling Open Semantic Science (SemSci)*, (pp. 63–70).

Jefferson, T., Alderson, P., Wager, E., & Davidoff, F. (2002). Effects of editorial peer review: a systematic review. *Jama, 287*, 2784–2786.

Jennings, C. (2006). Quality and value: The true purpose of peer review. *Nature.* doi:10.1038/nature05032

Jha, R., Jbara, A.-A., Qazvinian, V., & Radev, D. R. (2017). NLP-driven citation analysis for scientometrics. *Natural Language Engineering, 23*, 93–130.

Kaptelinin, V., & Nardi, B. A. (2006). *Acting with technology: Activity theory and interaction design.* MIT press.

Kiayias, A., Koutsoupias, E., Kyropoulou, M., & Tselekounis, Y. (2016). Blockchain Mining Games. *Proceedings of the 2016 ACM Conference on Economics and Computation* (pp. 365–382). New York, NY, USA: ACM. doi:10.1145/2940716.2940773

Kiktenko, E. O., Pozhar, N. O., Anufriev, M. N., Trushechkin, A. S., Yunusov, R. R., Kurochkin, Y. V., . . . Fedorov, A. K. (2018). Quantum-secured blockchain. *Quantum Science and Technology, 3*, 035004.

Kim, A. (2008). *Decentralization and the Provision of Public Services: Framework and Implementation.* Washington, DC: The World Bank.

Kitchenham, B. (2010). Systematic literature reviews in software engineering–a tertiary study. *Inform. Softw. Technol. 52*, 792–805.

Kling, R., & McKim, G. (1999, March). Scholarly Communication and the Continuum of Electronic Publishing. *arXiv:cs/9903015.* Retrieved June 28, 2017, from http://arxiv.org/abs/cs/9903015

Kopp, H., Bösch, C., & Kargl, F. (2016). KopperCoin – A Distributed File Storage with Financial Incentives. In F. Bao, L. Chen, R. H. Deng, & G. Wang (Ed.), *Information Security Practice and Experience* (pp. 79–93). Cham: Springer International Publishing.

Krapivin, M., Marchese, M., & Casati, F. (2010). Exploring and understanding citation-based scientific metrics. *Advances in Complex Systems 13 (01)*, pp. 59-81.

Kuhn, T., & Dumontier, M. (2014). Trusty URIs: Verifiable, Immutable, and Permanent Digital Artifacts for Linked Data. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern, . . . A. Tordai (Eds.), *The Semantic Web: Trends and Challenges* (Vol. 8465, pp. 395–410). Cham:

Springer International Publishing. Retrieved May 22, 2017, from http://link.springer.com/10.1007/978-3-319-07443-6_27

Kuhn, T., & Dumontier, M. (2015, September). Making Digital Artifacts on the Web Verifiable and Reliable. *IEEE Transactions on Knowledge and Data Engineering, 27*, 2390–2400. doi:10.1109/TKDE.2015.2419657

Kuhn, T., Chichester, C., Krauthammer, M., & Dumontier, M. (2015). Publishing Without Publishers: A Decentralized Approach to Dissemination, Retrieval, and Archiving of Data. In M. Arenas, O. Corcho, E. Simperl, M. Strohmaier, M. d'Aquin, K. Srinivas, . . . S. Staab (Eds.), *The Semantic Web - ISWC 2015* (Vol. 9366, pp. 656–672). Cham: Springer International Publishing. Retrieved May 22, 2017, from http://link.springer.com/10.1007/978-3-319-25007-6_38

Laakso, M., Welling, P., Bukvova, H., Nyman, L., Björk, B.-C., & Hedlund, T. (2011). The development of open access journal publishing from 1993 to 2009. *PloS one, 6*, e20961.

Lagoze, C., Edwards, P., Sandvig, C., & Plantin, J.-C. (2015). Should I stay or should I go? Alternative infrastructures in scholarly publishing. *International Journal of Communication, 9*, 20.

Larivière, V., Haustein, S., & Mongeon, P. (2015 a). The oligopoly of academic publishers in the digital era. *PloS one, 10*, e0127502.

Larivière, V., Haustein, S., & Mongeon, P. (2015 b). "Big publishers, bigger profits: how the scholarly community lost the control of its journals. *MediaTrope 5, no. 2, Libraries in crisis*, pp. 102-110.

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society.* Harvard University Press.

Latour, B. (2005). Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford: Oxford UP.

Lawrence, P. A. (2007). The mismeasurement of science. *Current Biology, 17*, R583–R585.

Lemieux, V. L. (2016, July). Trusting records: is Blockchain technology the answer? *Records Management Journal, 26*, 110–139. doi:10.1108/RMJ-12-2015-0042

Leydesdorff, L., Bornmann, L., Comins, J. A., & Milojević, S. (2016). Citations: Indicators of quality? The impact fallacy. *Frontiers in Research metrics and Analytics, 1.*

Leydesdorff,, L. (2021). The Evolutionary Dynamics of Discursive Knowledge, Qualitative and Quantitative Analysis of Scientific and Scholarly Communication.

Lightfoot, G., & Wisniewski, T. P. (2014). Information asymmetry and power in a surveillance society. *Information and Organization 24(4)*, pp. 214-235.

López-Cózar, E. D., Robinson-Garcia, N., & Torres-Salinas, D. (2012). Manipulating Google Scholar citations and Google Scholar metrics: Simple, easy and tempting. *arXiv preprint arXiv:1212.0638.*

Luu, L., Chu, D.-H., Olickel, H., Saxena, P., & Hobor, A. (2016). Making Smart Contracts Smarter. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 254–269). New York, NY, USA: ACM. doi:10.1145/2976749.2978309

Luu, L., Chu, D.-H., Olickel, H., Saxena, P., & Hobor, A. (2016). Making Smart Contracts Smarter. (pp. 254–269). ACM Press. doi:10.1145/2976749.2978309

Mühlberger, R. (2020). Foundational Oracle Patterns: Connecting Blockchain to the Off-Chain World. . *Business Process Management: Blockchain and Robotic Process Automation Forum. BPM 2020. Lecture Notes in Business Information Processing.*

Mahmood, Z., & Vacius, J. (2020). Privacy-Preserving Block-chain Framework Based on Ring Signatures (RSs) and Zero-Knowledge Proofs (ZKPs). *2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies*, (pp. 1-6).

Malone, T. W., & Crowston, K. (1991). *Toward an interdisciplinary theory of coordination.*

Maria Orquiza, L., Sánchez García, L., & Gabriel Costelini, B. (2022). How is Action Research Being Used in Computer Science? A Review. International Journal of Action Research, 18(1).

Maskin, E. S. (2008). Mechanism design: How to implement social goals. *American Economic Review, 98*, 567–76.

Maxwell, E. E. (2012). *The Future of Taxpayer Funded Research: Who Will Control Access to the Results?* Committee for Economic Development.

May, T. (1992). The crypto anarchist manifesto. *in High Noon on the Electronic Frontier: Conceptual Issues in Cyberspace, ed. P. Ludlow.*

McCorry, P., Shahandashti, S. F., & Hao, F. (2017). A smart contract for boardroom voting with maximum voter privacy. *International Conference on Financial Cryptography and Data Security*, (pp. 357–375).

McGregor-Lowndes, I., & others. (2019). The rise of the DAO disrupting 400 years of corporate structure. *Proctor, The, 39*, 32.

Meyer, B. (1997). *Object-oriented software construction.* Prentice Hall.

Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European journal of operational research, 246*, 1–19.

Mirowski, P. (2011). *Science-mart: privatizing American Science.* Harvard University Press.

Mitchell, R. K., Agle, B. R., & Wood, D. J. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of management review, 22*, 853–886.

Moedas, C., Vernos, I., Kuster, S., Nowotny, H., Saltelli, A., Mungiu-Pippidi, A., . . . Cunningham, P. (2019, May). Views from a continent in flux. *Nature, 569*, 481–484. doi:10.1038/d41586-019-01569-w

Mookherjee, D. (2006). Decentralization, hierarchies, and incentives: A mechanism design perspective. *Journal of Economic Literature, 44*, 367–390.

Nabben, K., & Rennie, E. (2022). Ad hoc network. *Internet Policy Review,[online] 11(2).*

Nakamoto, S. (2019). *Bitcoin: A peer-to-peer electronic cash system (2008).* Tech. rep., Manubot.

Nelaturu, K. (2020). On Public Crowdsource-Based Mechanisms for a Decentralized Blockchain Oracle. *IEEE Transactions on Engineering Management, vol. 67, no. 4*, 1444-1458.

Odlyzko, A. (2002). The rapid evolution of scholarly communication. *Learned Publishing, 15*, 7–19.

Pielke, R. (2010). In Retrospect: Science — The Endless Frontier. *Nature 466*, pp. 922–923.

Ponte, D., Mierzejewska, B. I., & Klein, S. (2017, May 01). The transformation of the academic publishing market: multiple perspectives on innovation. *Electronic Markets, 27*, 97–100. doi:10.1007/s12525-017-0250-9

Pouloudi, A., Gandecha, R., Atkinson, C., & Papazafeiropoulou, A. (2004). How stakeholder analysis can be mobilized with actor-network theory to identify actors. In *Information Systems Research* (pp. 705–711). Springer.

Praun, A., Cunningham, A., & Pieperhoff, M. (2022). Disruption, Disruptors and Disruptiveness of Scholarly Communication: An Actor-Network Theory Approach. *European Conference on Knowledge Management (Vol. 23, No. 2)*, (pp. 1452-1456).

Priem, J., & Hemminger, B. M. (2012, April). Decoupling the scholarly journal. *Frontiers in Computational Neuroscience, 6*. doi:10.3389/fncom.2012.00019

Priem, J., & Hemminger, B. M. (2012). Decoupling the scholarly journal. *Frontiers in computational neuroscience, 6, 19.*

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833.*

Proctor, R. N., & Schiebinger, L. (2008). *Agnotology: The making and unmaking of ignorance.* California: Stanford University Press.

Rabesandratana, T. (2019). Radical open-access plan is delayed a year. *Radical open-access plan is delayed a year.* American Association for the Advancement of Science.

Raskin, M., & Yermack, D. (2016, May). *Digital Currencies, Decentralized Ledgers, and the Future of Central Banking.* Working Paper, National Bureau of Economic Research. doi:10.3386/w22238

Raval, S. (2016). *Decentralized Applications: Harnessing Bitcoin's Blockchain Technology.* " O'Reilly Media, Inc.".

Repko, A. F. (2008). *Interdisciplinary Research.* Sage.

Rescher, N. (2013). *Dialectics.* De Gruyter.

Rospocher, M., Ghidini, C., & Serafini, L. (2014). An Ontology for the Business Process Modelling Notation. *FOIS*, 133-146.

Roth, W. M., & Lee, Y. J. (2007). “Vygotsky’s neglected legacy”: Cultural-historical activity theory. *Review of educational research, 77(2)*, pp. 186-232.

Roure, D. D., Goble, C., Bhagat, J., Cruickshank, D., Goderis, A., Michaelides, D., & Newman, D. (2008, December). myExperiment: Defining the Social Virtual Research Environment. (pp. 182–189). IEEE. doi:10.1109/eScience.2008.86

Rozas, D., Tenorio-Fornés, A., Díaz-Molina, S., & Hassan, S. (2021). When Ostrom Meets Blockchain: Exploring the Potentials of Blockchain for Commons Governance. *SAGE Open*.

S. McGuigan, G., & S. Russell, R. (2008, January). The Business of Academic Publishing: A Strategic Analysis of the Academic Journal Publishing Industry and its Impact on the Future of Scholarly Publishing. *9*.

Sandhu, R. S., & Samarati, P. (1994). Access control: principle and practice. *IEEE communications magazine, 32*, 40–48.

Santos , P., & Travassos, G. H. (2009). Action research use in software engineering: An initial survey,". *3rd International Symposium on Empirical Software Engineering and Measurement*, 414-417.

Sarabipour, S., Debat, H. J., Emmott, E., Burgess, S. J., Schwessinger, B., & Hensel, Z. (2019). On the value of preprints: An early career researcher perspective. *PLoS biology, 17*, e3000151.

Schinckus, C. (2021). Proof-of-work based blockchain technology and Anthropocene: An undermined situation? *Renewable and Sustainable Energy Reviews*.

Schminke, M., Ambrose, M. L., & Cropanzano, R. S. (2000). The effect of organizational structure on perceptions of procedural fairness. *Journal of Applied Psychology, 85*, 294.

Schneider, N. (2019). Decentralization: an incomplete ambition. *J. Cult. Econ. 12*, 265–285.

Schor, J. (2016). DEBATING THE SHARING ECONOMY. *Journal of Self-Governance & Management Economics, 4.*

Schumacher, E. (1973). F.(1973): Small is Beautiful: Economics as if People Mattered. *Blond & Briggs, London.*

Schwartz, C. A. (1994). Scholarly communication as a loosely coupled system: Reassessing prospects for structural reform. *College & research libraries, 55*, 101–117.

Seabright, P. (1996). Accountability and decentralisation in government: an incomplete contracts model. *Eur. Econ. Rev. (40)*, 61–89.

Shema, H., Bar-Ilan, J., & Thelwall, M. (2014). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. *Journal of the Association for Information Science and Technology 65(5)*, 1018–1027.

Simperl, E., Cuel, R., & Stein, M. (2013). *Incentive-Centric Semantic Web Application Engineering.* Morgan & Claypool.

Skolnikoff, E. B. (1994). *The Elusive Transformation: Science, Technology, and the Evolution of International Politics.* Princeton, NJ: Princeton University Press.

Small, H. (2004). On the shoulders of Robert Merton: Towards a normative theory of citation. *Scientometrics, 60*, 71–79.

Smart, P., Simperl, E., & Shadbolt, N. (2014). A taxonomic framework for social machines. In *Social Collective Intelligence* (pp. 51–85). Springer.

Smith, J. W. (1999). The deconstructed journal—a new model for academic publishing. *Learned Publishing, 12(2)*, 79-91.

Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the royal society of medicine, 99*, 178–182.

Smith, R. (2006). The trouble with medical journals. *Journal of the Royal Society of Medicine, 99*, 115–119.

Smith, W. D. (2000). Range voting. *The paper can be downloaded from the author's homepage at http://www. math. temple. edu/˜ wds/homepage/works. html.*

Sompel, H. v., Payette, S., Erickson, J., Lagoze, C., & Warner, S. (2004). Rethinking scholarly communication: building the system that scholars deserve. *D-Lib Magazine; 2004 [10] 9.*

Squazzoni, F., Grimaldo, F., & Marušić, A. (2017). Publishing: Journals could share peer-review data. *Nature, 546*, 352.

Srnicek, N. (2017). *Platform capitalism.* John Wiley & Sons.

Stigler, G. J. (1964). A theory of oligopoly. *Journal of political Economy 72, no. 1* , pp. 44-61.

Stoneburner, G., Goguen, A., & Feringa, A. (2002). Risk management guide for information technology systems. *Nist special publication 800(30).*

Strauss, A., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Sage Publications, Inc.

Suber, P. (1990). Appendix 3: Nomic: A Game of Self-Amendment.

Suber, P. (2007). Open access overview.

Svensson, L. (2010). The infrastructure of digital humanities. *Digital humanities quarterly, 4(1).*

Swanson, T. (2015). Consensus-as-a-service: a brief report on the emergence of permissioned, distributed ledger systems. *Report, available online, Apr.*

Szabo, N. (1997). Formalizing and securing relationships on public networks. *First Monday, 2*.

Tötösy de Zepetnek, S., & Jia, J. (2014). Electronic journals, prestige, and the economics of academic journal publishing. *CLCWeb: Comparative Literature and Culture, 16*, 12.

Tan, L. (2019). *Token Economics Framework.* Singapore: Economics Design Limited.

Taylor, P. J., Dargahi, T., & Dehghantanha, A. (2020). A systematic literature review of blockchain cyber security. *Digital Communications and Networks, 6(2)*, 147-156.

Teece, D. J. (1992). Competition, cooperation, and innovation: Organizational arrangements for regimes of rapid technological progress. *Journal of economic behavior & organization, 18*, 1–25.

Teixeira da Silva, J. A., & Dobránszki, J. (2015). Problems with traditional science publishing and finding a wider niche for post-publication peer review. *Accountability in research, 22*, 22–40.

Tennant, J. P., Dugan, J. M., Graziotin, D., Jacques, D. C., Waldner, F., Mietchen, D., . . . Colomb, J. (2017, November). A multi-disciplinary perspective on emergent and future innovations in peer review. *F1000Research, 6*, 1151. doi:10.12688/f1000research.12037.3

Tiebout, C. M. (1956). A pure theory of local expenditures. *J. Polit. Econ. 64, 416–424. doi: 10.1086/257839.*

Tobin, A., & Reed, D. (2016). The Inevitable Rise of Self-Sovereign Identity. *The Sovrin Foundation.*

Treiblmaier, H. (2020). Toward more rigorous blockchain research: Recommendations for writing blockchain case studies. In *Blockchain and Distributed Ledger Technology Use Cases* (pp. 1–31). Springer.

Tsang, P. P., & Wei, V. K. (2005). Short linkable ring signatures for e- voting, e-cash and attestation. *International Conference on Information Security Practice and Experience* (pp. 48-60). Springer.

Turner, F. M. (1980). Public Science in Britain, 1880-1919. *Isis (The University of Chicago Press, The History of Science Society)*, pp. 589-608.

Underwood, S. (2016, October). Blockchain Beyond Bitcoin. *Commun. ACM, 59*, 15–17. doi:10.1145/2994581

Van Noorden, R. (2013). Open access: The true cost of science publishing. *Nature News, 495*, 426.

Van Vugt, M., Biel, A., Snyder, M., & Tyler, T. R. (2000). *Cooperation in modern society: Promoting the welfare of communities, states, and organizations* (Vol. 1). Psychology Press.

Varian, H. R. (1995). Economic Mechanism Design for Computerized Agents. *USENIX workshop on Electronic Commerce*, (pp. 13–21).

Vo, H. T., Mohania, M., Verma, D., & Mehedy, L. (2018). Blockchain-Powered Big Data Analytics Platform. In A. Mondal, H. Gupta, J. Srivastava, P. K. Reddy, & D. V. Somayajulu (Ed.), *Big Data Analytics* (pp. 15–32). Cham: Springer International Publishing.

von Hayek, F. (1991). *The Fatal Conceit.* University of Chicago Press. Retrieved from https://EconPapers.repec.org/RePEc:ucp:bkecon:9780226320663

Walch, A. (2019). Deconstructing'Decentralization': Exploring the Core Claim of Crypto Systems. *Crypto Assets: Legal and Monetary Perspectives (OUP, forthcoming 2019).*

Wang, P. e. (2016). Open peer review in scientific publishing: A Web mining study of PeerJ authors and reviewers. *Journal of Data and Information Science 1, 4*, 60-80.

Wang, X., Liu, C., Mao, W., & Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. *Scientometrics 103 (2)*, 555–564.

Ware, M. (2008). *Peer review: benefits, perceptions and alternatives.* Citeseer.

Wattenhofer, R. (2016). *The science of the blockchain* (First edition ed.). Erscheinungsort nicht ermittelbar: Inverted Forest Publishing.

Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics 62*, 117–131.

Whisler, T. L., & Leavitt, H. J. (1958). Management in the 1980's. *Harvard Bus. Rev. 36*, 41–48.

Wieringa, R. J. (2014). *Design science methodology for information systems and software engineering.* Springer.

Wilkinson, M. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data 3.*

Williams, G. (2018 ). *Kant's Account of Reason.* Retrieved from The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.): https://plato.stanford.edu/archives/sum2018/entries/kant-reason

Willinsky, J. (2005). The unacknowledged convergence of open source, open access, and open science. *First Monday.*

Wilsdon, J., Bar-Ilan, J., Frodeman, R., Lex, E., Peters, I., Wouters, P. F., & others. (2017). Next-Generation Metrics: Reponsible Metrics and Evaluation for Open Science. Report of the European Commission Expert Group on Altmetrics.

Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum Project Yellow Paper, 151.* Retrieved from https://github.com/ethereum/yellowpaper

Yudkowsky, E. (2017). *Inadequate Equilibria: Where and How Civilizations Get Stuck.* Machine Intelligence Research Institute.

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology, 30*, 75–89.

Zuboff, S., & Maxmin, J. (2004). *Zuboff, Shoshana, and James Maxmin. The support economy: Why corporations are failing individuals and the next episode of capitalism.* New York: Penguin.