

OPEN SCIENCE ORGANIZATION

---

## OSO: An Idea Platform

---



contact@oso.network  
November 15, 2018  
v 0.3

## Disclaimer

This is a work in progress.

DRAFT

### **Authors/Contributors**

Gajendra Katuwal, Kirk Gosik, Abinash Koirala, Sandip Pandey, [you ?]

### **Acknowledgements**

Aaditya Prakash  
Andrew Kang  
Noman Land  
Prof. Stanislaw Radziszowski

## Abstract

The current system has been proven to be ill-suited for the modern scientific research. The current problems in research such as irreproducibility, high publication and subscription fees, publish or perish mentality, funding-crisis, etc. stem from the current inefficient system. So, the research ecosystem needs a complete overhaul and with the right use and adoption of technology it is not impossible. Here, at Open Science Organization (OSO) [1] we are developing a community-managed system to make research open, unbiased, and efficient. We present OSO platform which is a *decentralized Idea platform* and can be a research ecosystem in itself; a single place for all the activities in a research cycle i.e. funding to publishing. The platform will contain all the components required to manage (own, create, publish, review, revise, fund) an idea. The proposed platform will significantly reduce the cost of scientific research and inherently incentivizes openness and cross-pollination of ideas. In our platform, idea creators (e.g. researchers) will be able to receive funding from anyone and anywhere in the world and the general public and investors can fund any ideas.

Over the last few centuries, scientific research has transitioned from intellectual pursuits to a commercial group endeavor. With time, several intermediaries were introduced that provided value to the scientific research. However, with the passage of time, research ecosystem has transformed into a complex network of intermediaries and as a result, a system that is highly unfit for modern scientific research, has been created. Due to this ill-suited system, scientific research has been plagued by several problems. Here, we will briefly mention the major problems in the current research ecosystem but please refer to our first white paper for the detailed information [1].

### 0.1 Closed and expensive dissemination of scientific publications

- The cost of publishing can range from \$2000 to \$10000 per publication [2]. To give a comparison, Google One (<https://one.google.com/about>) provides 2TB for \$9.99 per month and with this rate, all the scientific articles published in 2017 ( 2M articles) can be hosted as public Google Drive files for \$9.99 a month! *So why do we pay \$5000 to publish a paper?*
- Researchers (and universities) pay a heavy amount to subscribe to journals for accessing their own publications. For example, Harvard University paid more than \$16 million as subscription fees in 2012 alone [3, 4]
- *“The peer-review — the most critical process behind the value creation of the current scientific publishing industry — is done for free by the scientific community.”* And in return, they have to pay hefty amount to publish the papers and access them.

## 0.2 Closed and Slow Review Process

- The rationale behind the decision to accept or reject a publication is not disclosed to the public
- Richard Smith, former editor of the British Medical Journal, says — “*All journals are bought — or at least cleverly used — by the pharmaceutical industry*” [5]
- The median review time i.e. the time between submission and acceptance of a publication is on average 100 days. The median online publication time i.e. the time between acceptance and online debut of a publication is on average 25 days (<http://blog.dhimmel.com/history-of-delays/>). [6]
- There is no mechanism to promote the replicability of the studies under the traditional review model.
- 70% of researchers have tried and failed to reproduce the experiments of other scientists and more than 50% have failed to reproduce their own experiments [7]

## 0.3 Funding Crisis

- Funding biases and undemocratic decision-making on research directions.

### Imbalance in demand and supply of funding

- Due to the scarcity of funding, a hyper-competitive atmosphere which has been created and is detrimental to the organic growth of scientific research
- Scientists are spending more time writing grants than doing actual research
- It promotes short-term gain vs long-term gain in science; it suppresses the creativity, original thinking, and risk taking and promotes safe research
- It promotes quantity over quality of scientific publications and encourages the scientists to take the short-cuts and exaggerate their results due to the pressure of publishing
- Federal funding (% of GDP) in basic research which is the foundational layer of the human knowledge is decreasing [8]. Robbert Dijkgraaf, a mathematical physicist and string theorist writes “*But driven by decreasing funding, against a background of economic uncertainty, global political turmoil, and ever-shortening time cycles, research criteria are becoming dangerously skewed towards short-term goals that may address more immediate problems, but miss out on the huge advances that human imagination can bring in the long term.*” [9]

In our first white paper [1], we discussed the general problems in current scientific ecosystem, presented modern day science as a community endeavor, made a case why we need a community-owned decentralized platform to manage this collective endeavor, and presented a simple web3 based solution i.e. a decentralized platform. In this paper, we present an idea platform which, in long-term, may become the main foundation of the new and better research ecosystem.

Note: See 8 for use cases.

## 1 OSO -an Idea platform

Imagine everything that happens in the current research ecosystem is transferred to one decentralized platform which is collectively managed and owned by a community; OSO -an Idea platform is trying to be one. OSO platform is a decentralized *idea platform*: a unique combination of modular web3 components required *to own, create, iterate, review, store, publish, and fund ideas*. It strives to be a common platform for research ecosystem where all the activities in a research cycle – fund → research → review → publish – can be performed in a common platform. Goal of OSO is to be such a platform.

### OSO Idea Platform is:

- **A decentralized idea platform.**
  - Just like Ethereum [10] is a smart contract platform i.e. a set of protocols to deploy self-executing code in a decentralized fashion, OSO platform is a set of protocols to deploy (own, host, revise, share, etc.) your ideas. OSO network is a global network of ideas just like Internet is a global network of computers.
- **A global decentralized research lab.**
  - Just like Github is a popular site to develop and share open-source softwares. OSO platform is a place to develop and share ideas. It strives to be a single place that will host the evolution of global human knowledge.

See Section 8 for example use cases.

The fundamental characteristics of the OSO platform are:

- **Open:** The platform is open-source in nature; everything related to the platform (e.g. protocols, software, etc.) are made public. It is designed to incentivize rapid and open sharing of ideas while ensuring that the value created by an idea is appropriately shared to its creator(s).
- **Decentralized:** The platform is made up of modular web3 components to avoid central point of ownership and failure.
- **Community-based:** The platform is collectively owned and managed by a community of users and interested stake holders.
- **Collaborative:** The platform incentivizes collaboration by design. It incentivizes the cross-pollination among the ideas, thereby encouraging collaboration among the Idea creators (e.g. researchers). The end goal of OSO platform is to transform research from the current "publish or perish" mentality into harmonious collective endeavor.
- **Flexible:** The platform will be flexible in design to allow different needs of the users. It will contain the core protocols to manage ideas in a decentralized way and others can build applications on top of the platform to provide services to the users.

The **flexible design of OSO platform** will be able to serve the different needs of the users. The core OSO platform will have technical components required to own, create, iterate, review, store, publish, and fund ideas. Within the OSO network, users can customize the use of protocols to form *sub-networks* within the OSO-network and the sub-networks can overlap with each other. See Figure 1. Just like a group of users with similar interests can create a sub-Reddit in the Reddit platform to suit their topic of interest and make their customized rules to govern the sub-reddit, users (e.g. researchers) can create sub-networks inside the OSO network to suit their custom needs. For example, a group of researchers can create a sub-network for their research domain. There can be multiple sub-networks for a subject domain. Users will have the complete freedom to create sub-networks and switch in-between them. We expect this design will create a healthy competition and collaboration among sub-networks.

Within a sub-network created for a research domain, researchers can create idea-sharing or publishing *channels*. Channels are themed publication tracks with custom focus and rules (similar to traditional journals). There can be multiple channels in a sub-network and channels can span multiple sub-networks. Channels will have their own custom rules (e.g. specialization track, selectivity, review policy, etc.) Sub-networks can have their own rules about the creation and management of the channels.

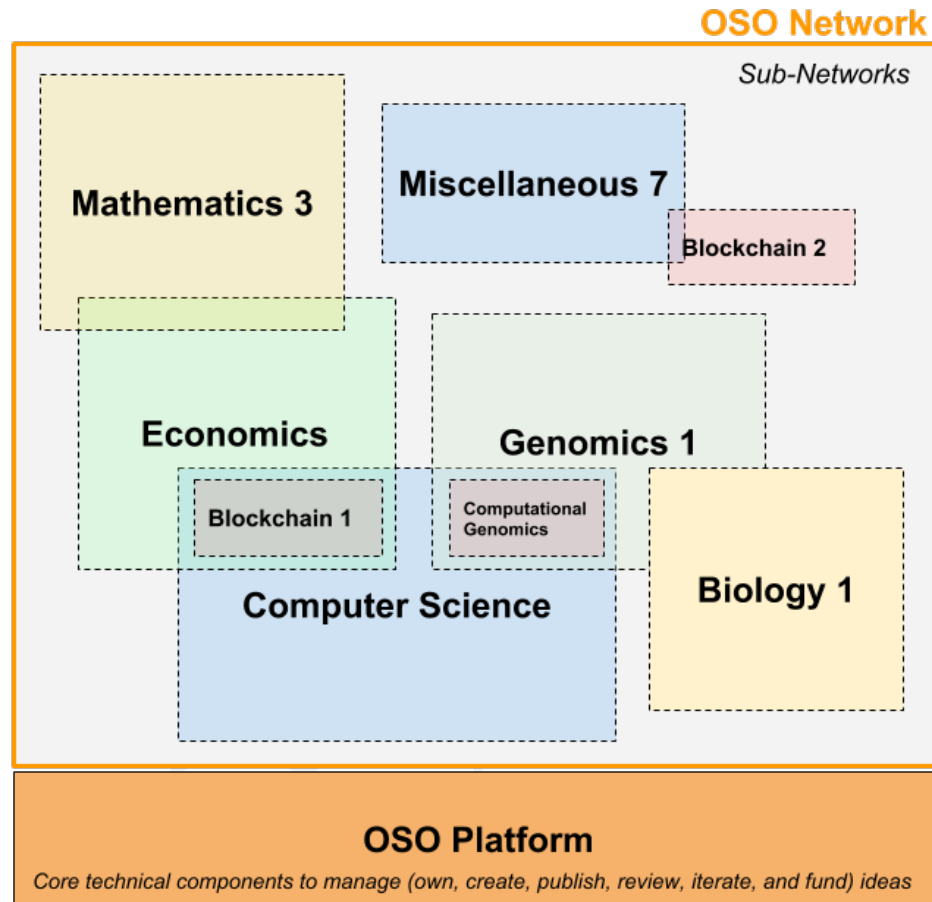


Figure 1: **Sub-networks.** The core OSO platform will have technical components required to own, create, iterate, review, store, publish, and fund ideas in a decentralized way. Users have choice on how they want to use the OSO platform. A group of users with similar interests can create a sub-network for a topic of interest with custom rules. There can be multiple sub-networks for a subject domain which will create a healthy competition and collaboration among sub-networks. Within a sub-network, researchers can create idea-sharing or publishing *channels*. Channels are themed publication tracks with custom focus and rules (similar to traditional journals).



## 2 Value Proposition

We are creating a decentralized idea platform that serves two markets: academic publishing and crowd funding market.

### Current Market size and profit margin

#### Publishing Market

Wiley's gross profit margin was 73% for fiscal year 2018 [11]. Based on recent reports of annual profits, the biggest players—Elsevier, Wiley, and Springer Nature (merged since May 2015) collected 913 million Euros (\$1274m), \$687m, and nearly 600m Euros (\$714m), respectively. *Who foots the bill?* Academic institutions, largely. \*And this bill is likely transferred to education cost\*. Approximately 7.6 Euros (\$8.3) billion goes into journal access every year. After paying these fees, even the wealthiest institutions gain access to only a fragment of the scientific literature; the less wealthy—a smidgen. [11] The variance in prices has lead the community to question the efficacy of current publication model. [12]

#### The cost of publishing

“Data from the consulting firm Outsell in Burlingame, California, suggest that the science-publishing industry generated \$9.4 billion in revenue in 2011 and published around 1.8 million English-language articles — an average revenue per article of roughly \$5,000. Analysts estimate profit margins at 20–30% for the industry, so the average cost to the publisher of producing an article is likely to be around \$3,500–4,000.” (direct excerpt from [12])

“Most open-access publishers charge fees that are much lower than the industry's average revenue, although there is a wide scatter among journals. The largest open-access publishers — BioMed Central and PLoS — charge \$1,350–2,250 to publish peer-reviewed articles in many of their journals, although their most selective offerings charge \$2,700–2,900. In a survey published last year, economist Bo-Christer Björk of the Hanken School of Economics in Helsinki and psychologist David Solomon of Michigan State University in East Lansing looked at 100,697 articles published in 1,370 fee-charging open-access journals active in 2010 (about 40% of the fully open-access articles in that year), and found that charges ranged from \$8 to \$3,900. Higher charges tend to be found in 'hybrid' journals, in which publishers offer to make individual articles free in a publication that is otherwise pay-walled (see 'Price of prestige'). Outsell estimates that the average per-article charge for open-access publishers in 2011 was \$660.” (direct excerpt from [12])

## Crowd Funding Market

**GofundMe** The website GoFundMe.com, which allows individuals to collect donations for personal needs or projects, raised \$1.09 billion over the past 12 months, according to a new report by the giving portal. [13]

**For Profit Company:** <https://experiment.com/> – Crowd Funding Scientific Research  
Experiment has a 45.37% success rate.

- 822 funded projects
- 43,532 backers
- \$8,006,559 pledged

## Value creation by OSO Idea platform

### 2.1 For Researchers

- Cheap and fast validation and publication of their ideas. The cost is estimated to go down by a factor of 10 and the publication process will speed up by a factor of 5 to 10.
- Single point of access and convenient to follow global research
- Convenient place to receive crowd funding for their research ideas. With successful crowd funding, they can do the actual research than spending their majority of the time fighting for grants.
- Common protocol enables efficient meta-research. Summarizing the results of several findings is an important scientific process. Efficient literature review or background research.
- An ideal place for cross-pollination of the ideas and hence the improvement in collaboration
- Flexibility to create their own sub-network that fits their needs
- Researchers can use their metrics in the platform such as reputation and expertise scores as a proxy of an academic achievement

### 2.2 For Investors

- A convenient single place to explore new and original ideas

- Investors get chance to invest in basic-research. The success rate is lower than traditional investments but the yield rate can be exceptionally high.
- Small but highly-informed investors, for example, researchers themselves, can conveniently invest in the high potential research ideas. OSO platform provides provides a convenient way to monetize their specialized knowledge through investing.
- For philanthropists, it will be a convenient place to donate or do benevolent investing to validated research ideas

## 2.3 For Entrepreneurs

- A convenient single place to explore and track the development of new and original ideas
- Entrepreneurs can use the platform to connect with the idea creators, potential advisors, and investors

## 2.4 For Policy Makers & Organizations

- A common platform to explore global research in a macroscopic level and understand global trends and pain points
- Common protocol means easy meta-research. Efficient research surveys and reviews.
- Easy for the government and other agencies to create their policies and funding strategies

# 3 Conceptual Components of OSO Platform

OSO platform is essentially an *idea management platform*. There are three main entities in OSO platform: 1) **Idea** which is an abstraction of an intellectual contribution 2) **User** which creates and utilizes the idea, and 3) **Cryptographic tokens** that incentivizes the fair and healthy interaction among users and ideas.

## 3.1 Idea

Idea is an entity which generates new information at the time of its creation. It is an abstraction of an intellectual contribution and is the main commodity of the OSO platform. An idea can be anything physical or virtual object that contains some useful information, for example, publication, book, website, blog, data, legal document, code, review, etc.

For more info, please see Generalized Idea Protocol (GIP) 4.

### 3.2 User

A user is an entity that creates and utilizes the ideas. It can be a human being (e.g. researcher, engineer, student, investor, etc.) or an organization (e.g. research institute, government, company, etc.). A user has three main elements: 1) Identity, 2) Reputation, and 3) Expertise

#### Identity

Each user has an asymmetric cryptography based identity. There are three main components of an identity: 1) private-public key pair, 2) human readable username that is uniquely mapped to the public key, and 3) real-world identity. The private key is owned by the user and is used to authenticate its identity in the platform.

Note: We acknowledge that the mapping between the cryptography-based identity and the real-world identity is an extremely difficult problem and has not been solved fully yet. As an initial solution, we plan to use existing databases to bind these two identities. For example, university databases can be used to verify the real-world identities of the students and professors.

#### Expertise Score

The level of expertise of a user in a subject domain  $D$  is quantified by a metric called *expertise score*  $e_D$ .

$$e^D = \mathcal{E}^D(V^D, \downarrow e^D) \quad (1)$$

where:

$\mathcal{E}^D$  is an *expertise function* of a suitable form. The exact form of the expertise function can be different in different sub-networks and subject domains and is decided by the users of the respective sub-networks and domains.

$V_D$  is the total value created by the user that is related to the domain  $D$  and is defined as:

$$V^D = \sum_i v_i * s_i^D \quad (2)$$

where  $v_i$  is the total value created by an idea  $I_i$  in the OSO platform (i.e. total no. of OSO tokens attracted by the idea) and  $s_i^D$  is the similarity of the  $i^{th}$  idea with the domain  $D$ . The similarity function can be different for different research domains.

$\downarrow e^D$  is the aggregated expertise rating **received** (hence the down arrow) by the user from other users for the domain  $D$ . It is defined as:

$$\downarrow e^D = \frac{1}{N} \sum_i e_i^D * \downarrow e_i^D \quad (3)$$

$e_i$  is the expertise score of  $i^{th}$  peer in domain  $D$  and  $\downarrow e_i^D$  is the expertise rating **received** (hence the down arrow) by the user from the peer.  $N$  is the number of expertise ratings received by the user.

## Reputation Score

A user with a high expertise score is expected to act fairly but which might not happen all the time in practice. This requires a separate score called *reputation score* to quantify the collective assessment of the acts of a user. The reputation of a user  $r^S$  in a sub-network  $S$  is quantified as:

$$r^S = \mathcal{R}^S(e^S, \downarrow r^S) \quad (4)$$

where:

$\mathcal{R}$  is an *reputation function* of a suitable form that is unique for each sub-network.

$\downarrow e^D$  is the aggregated peer reputation rating of the user given by other users in sub-network  $S$ .

It is defined as:

$$\downarrow r^S = \frac{1}{N} \sum_i r_i^S * \downarrow r_i^S \quad (5)$$

$r_i$  is the reputation score of  $i^{th}$  peer in sub-network  $S$  and  $\downarrow r_i^S$  is the reputation rating received by the user from the peer.  $N$  is the number of reputation ratings received by the user.

## 3.3 Tokens

Crypto tokens facilitate the quantification of intellectual contributions and provide necessary incentives for sustainable operation of the platform. OSO platform has two type of tokens: 1) OSO utility token which facilitate the healthy operation of the platform and 2) IDEA tokens to quantify the value of the ideas.

### 3.3.1 OSO utility token

OSO utility token is like the fuel of the platform. It provides incentive to the researchers to provide quality intellectual contributions to the platform. It acts as form of value exchange where the value is the *intellectual contribution* i.e. idea which is the commodity of the OSO platform. OSO tokens can be earned by providing intellectual contributions such as peer-review, advising, bug hunting, etc. They can be spent to receive intellectual contributions from other users in exchange.

### 3.3.2 IDEA tokens - Unique Idea token for each Idea

Each idea can have its own unique IDEA tokens and the owner of the idea has the sole authority to generate its IDEA tokens. IDEA tokens facilitate the divisible ownership in the *promise of something* from an idea. Our platform only provides the portal to interact with external value (e.g. receive external funding), we leave it to the idea owners (e.g. researchers) on how they want to use the portal. Depending upon the agreement between the owner and investors, the *promise of something* can mean several different things such as share in the value creation from the idea, ownership of the

idea, governance of the idea, etc.

Note: Only OSO tokens interact with the environment outside the OSO platform. IDEA tokens are only meant to facilitate divisible ownership in ideas. They are automatically converted to OSO tokens while interacting with the external environment.

## 4 Generalized Idea Protocol (GIP)

Research or any intellectual process can be fully described by three basic components: 1) Definition of an intellectual contribution (i.e. Idea object) 2) Relationship among idea objects, 3) State change of the idea objects in response to an action by an agent or group of agents (i.e. users). Generalized Idea Protocol (GIP) <https://github.com/open-science-org/GIP> describes these three components. GIP fully defines an idea, its relationship with other ideas, and its state change in response to a user (e.g. researcher).

### Representing all Human Knowledge by a Graph

With the representation of the ideas by GIP, the collection of the ideas will form a directed graph (digraph) that grows over time. A directed graph [14] is an ordered pair  $G = (V, E)$  where  $V$  is a set of elements, called vertices or nodes and  $E$  is a set of ordered pairs of vertices, called directed edges or simply edges. In OSO platform, a node in the digraph will represent an idea (say  $I_0$ ) and will contain enough information to describe the idea; see Figure 2 The edge between two nodes (representing ideas  $I_0$  and  $I_1$ ) will describe the relationship (magnitude and direction of idea and value flow) between  $I_0$  and  $I_1$ . The graph shows that the idea  $I_0$  is related to the ideas  $I_1$ ,  $I_2$ , and  $I_3$ .

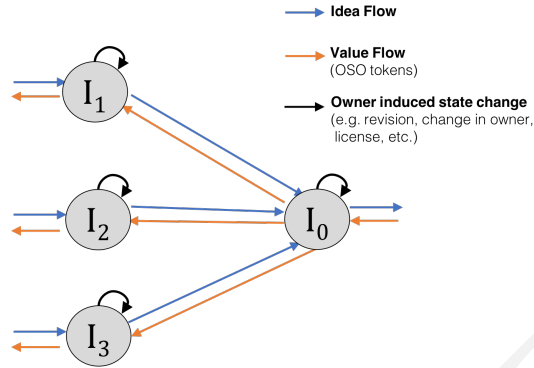


Figure 2: **Representation of Human Knowledge by a Digraph.** A node in the digraph will represent an idea (say  $I_0$ ) and will contain enough information to describe the idea. The edge between two nodes (representing ideas  $I_0$  and  $I_1$ ) will describe the relationship (magnitude and direction of idea and value flow) between  $I_0$  and  $I_1$ . Here, according to the digraph, idea  $I_0$  is related to the ideas  $I_1$ ,  $I_2$ , and  $I_3$

## 4.1 Definition of an Idea

**Idea** is an entity which generates new information at the time of its creation. It is a conceptual object encapsulating an intellectual contribution and its properties. It can be a publication, book, website, blog, data, legal document, or anything physical/virtual object or concept. An idea object is completely defined by its state i.e. idea state and is represented by a node in the DAG. An idea state contains meta data of the idea and the pointers to the actual content of the idea (e.g. code, data, publication, etc. A basic idea state can look like:

```
{
  ID: hash of the state + content
  Name: "some cool idea"
  Ownership: [Author1: %, Author2: %, Organization1: %, ...]
  Wallet: controlled by the owners
  License: defines how other users and ideas can use it
  Relationship : with other ideas (parents & siblings)
  Content: pointers to the idea contents
  Reputation:
  Reviews: a list of idea objects
  ...
}
```

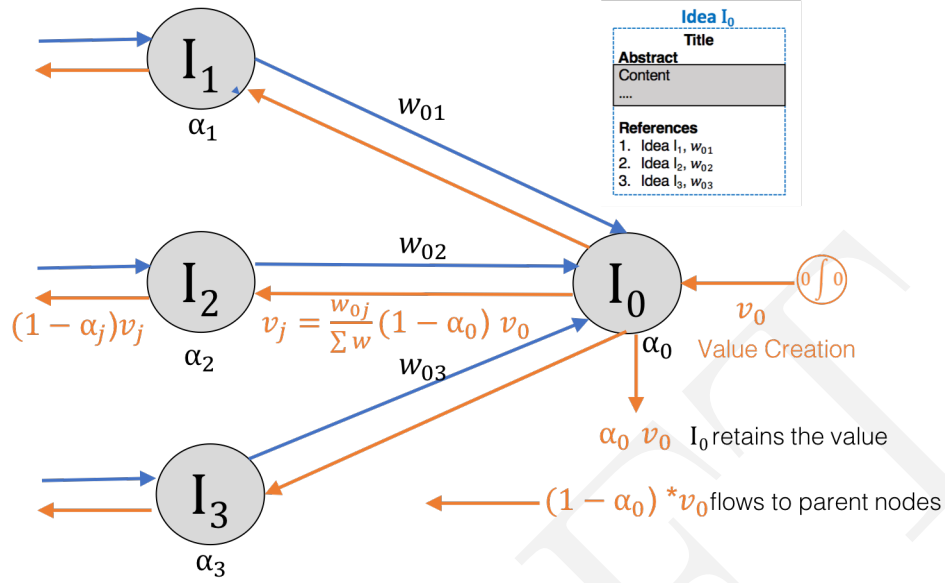


Figure 3:  $v_j$  = Value generated by idea  $I_j$ ,  $w_{ij}$  = association strength between Idea  $I_i$  and  $I_j$ ,  $\alpha_j$  = absorption coefficient of Idea  $I_j$ ;  $0 < \alpha < 1$ . **Idea Flow.** Idea  $I_0$  (publication here) uses (cites) ideas  $I_1$ ,  $I_2$ , and  $I_3$ , so there is a flow of idea from  $I_1$ ,  $I_2$ , and  $I_3$  to the idea  $I_0$ . **Value Flow.** Idea  $I_0$  generates a value of  $v_0$  (OSO tokens). The idea  $I_0$  retains the value of  $\alpha_0 v_0$  and passes  $(1 - \alpha_0) v_0$  to the parent ideas as the "cost of idea". A  $j^{th}$  parent node will receive the value of  $v_j = \frac{w_{0j}}{\sum_j w} (1 - \alpha_0) v_0$

## 4.2 Relationship Among Ideas

Any new idea always has some level of dependency on previous ideas. For example, an idea for a successful cancer drug is a novel as it can be but it still depends on several other ideas to build the foundation for that idea. So, it is important to define the relationship among ideas and GIP makes an effort towards that.

In GIP, the relationship between two ideas will define:

### 4.2.1 Idea Flow: the association strength among the ideas

The idea flow describes the intellectual association among the ideas. Whenever a new idea utilizes previous ideas (e.g. a publication cites previous publications), it is in essence borrowing the intellectual contributions from these ideas. In Figure 3, idea  $I_0$  uses (e.g. a publication citation) ideas  $I_1$ ,  $I_2$ , and  $I_3$ , so there is an *idea flow* from  $I_1$ ,  $I_2$ , and  $I_3$  to the idea  $I_0$ . The weights  $w_{ij}$  of the edges between two ideas  $I_i$  and  $I_j$  denote the strength of association between them. Theoretically,



the strength of association between the idea  $I_0$  and its  $j^{th}$  parent idea node  $I_j$  is defined as:

$$w_{0j} = \frac{\mathbf{I}(I_0, I_j)}{\sum_j \mathbf{I}(I_0, I_j)} \quad (6)$$

where  $\mathbf{I}(I_0, I_j)$  is the mutual information between  $I_0$  and  $I_j$ . In practice, we leave it to the users (market) to estimate these weights. Our understanding is that since everything is open by default and users are incentivized to increase their reputation scores, users will collaborate for fair estimation of these weights.

#### 4.2.2 Value Flow: the value sharing agreement among the owners of the ideas

A mechanism that evaluates the borrowed intellectual contributions from the ideas and rewards it to the authors of the ideas can introduce several positive effects in the research ecosystem. In OSO platform, we have a similar mechanism and we call it the *Value Flow*. The value flow can be thought as a *cost of idea* i.e. the value created by an idea has to be shared with the ideas it was based upon. In OSO platform, for example, if a new idea is a scientific publication, the author has to pay some OSO tokens to the cited publications and the cost related to it is called the cost of idea. The nature and amount of the cost of idea are different for different sub-networks and subject domains. It can be nil if the authors of the used ideas chose to do so.

**Value Generation:** In Figure 3, idea  $I_0$  generates a value of  $v_0$  (OSO tokens). The value generation can be from several sources such as:

- **Use:** A new idea uses the idea  $I_0$  and pays  $v_0$  as the *cost of idea*. In general, the idea  $I_0$  will generate some value whenever one of its child idea generates some value.
- **Investment:** An investor invests a value of  $v_0$  in the idea  $I_0$
- **Gift:** A philanthropist gifts a value of  $v_0$  to the idea  $I_0$

**Value Retention:** The idea  $I_0$  retains the value of  $\alpha_0 v_0$  out of the generated value of  $v_0$  where  $\alpha_0$  is its absorption coefficient. Loosely speaking,  $\alpha_0$  is the proxy measure of the intellectual independence of the idea  $I_0$ . We acknowledge that there is no algorithmic way to estimate  $\alpha_0$  and we leave it to the market for its fair estimation. For example, it can be set by the agreement among the authors of the ideas.

**Value Flow Back:** The total value of  $(1 - \alpha_0)v_0$  out of the generated value of  $v_0$  flows back to the parent ideas as a payment to the cost of idea because the intellectual contributions from the parent ideas were used to create the idea  $I_0$  and hence the value created by it. The value  $(1 - \alpha_0)v_0$  that flows upstream is shared by the parent nodes in proportion to their degree of association  $w_{0j}$  with their common child  $I_0$ . A  $j^{th}$  parent node will receive the value of:

$$v_j = \frac{w_{0j}}{\sum_j w} (1 - \alpha_0) v_0 \quad (7)$$

Out of the received value of  $v_j$ , the  $j^{th}$  parent node will retain the value of  $\alpha_j v_j$  and will pass the value of  $(1 - \alpha_j) v_j$ . In theory, the value can flow all the way back to the root idea. However, it is practical to continue the value flow process only if the cost of the process is less than the passed value of  $(1 - \alpha_j) v_j$  to the parent nodes.

### 4.3 Importance of idea node

(work in progress)

The importance of a node (centrality) can have various forms depending upon the context. <https://en.wikipedia.org/wiki/Centrality>  
This concept will be crucial to create a search layer on top of GIP. Let's say it "Idea Search Engine".

Centrality of an idea node can be considered as a potential energy to capture the value flow back and investment in an idea (node) is like buying a percentage of the potential energy.

### Different actions in an Idea object:

(TO DO:)

**4.3.1 Action: An entity (e.g. researcher) publishes an idea**

**4.3.2 Action: An entity reviews an idea**

**4.3.3 Action: An entity opens its published idea for investment**

### Merits of GIP

- Each unique idea can be identified and its evolution over the time can be tracked. It provides a formal framework to evaluate the relationship between two different ideas
- It facilitates the efficient sharing of the ideas and prevent the community from reinventing wheels, thereby accelerating the pace of the scientific research.
- Meta-research: a good visualization of GIP objects will help the scientific community to identify the pain points of scientific research. It can be used to find the optimal distribution of available resources (e.g. fund allocation) required in order to maximize scientific throughput.
- GIP can be considered as a way to store the collective thought process of the whole scientific community.
- GIP stores the evolution of global human knowledge.

## 5 Decentralized Peer-Review

Peer-review can be considered as a process for validation of an idea, similar to a quality control step. The OSO platform will provide the necessary features for a decentralized peer-review process where we consider the review of an idea as an idea itself. Different sub-networks and publishing channels can have different peer-review process that suit their specialties. However, the peer-review process in the sub-networks have to follow these fundamental characteristics:

- Peer-review process is open by default
- Selection of reviewers is fair and transparent
- Perpetual Review. In addition to an initial review required to publish in a channel, an idea goes through a perpetual review process i.e. other users can provide a review and rating to the idea.
- The review decision doesn't necessarily has to be hard reject/accept. Even if an idea is rejected, it has to be published in a separate channel (meant for rejected ideas) or in the same channel with a tag 'rejected' alongside with the full history of review process.

## 6 Technical Summary

### 6.1 Architecture Components

OSO idea platform is designed with openness and extensibility in mind to create a platform where people can contribute new ideas and extend the system. Figure 4 shows the high-level technical architecture of the whole system. We briefly discuss each of the components in this section.

#### 6.1.1 Smart Contracts Platform

Idea protocols are written as smart contracts which executes on a smart contracts platform like Ethereum. Our protocols are platform agnostic but as for the first implementation, we selected Ethereum as the platform blockchain and Solidity as the smart contracts language for the implementation. Blockchain is used as a settlement layer and anchor for building and maintaining trust, therefore, non-scalability and low transaction throughput would not affect the platform.

#### 6.1.2 Decentralized Storage System

Every idea on the idea platform has some associated data, research papers, and reviews which are essential to the idea itself. Therefore, we need a storage system that can handle the storage of

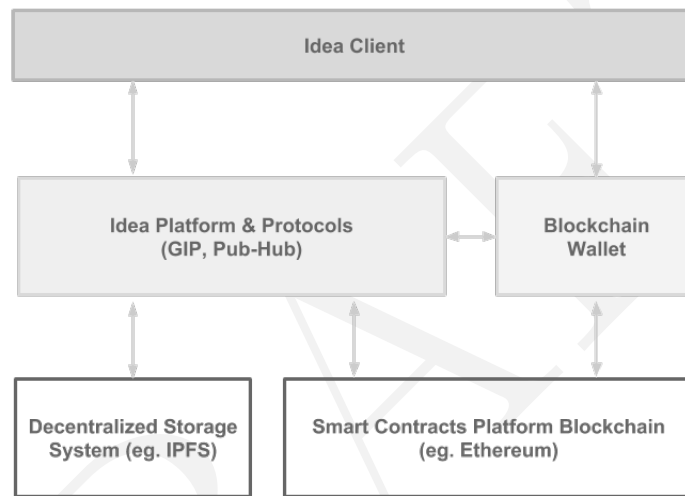


Figure 4: **High-level Technical Stack of OSO platform.** Idea platform interacts with Blockchain and Decentralized storage platform along with wallet for smart contract transactions. See Section 6 for detailed information.

data associated with the ideas on the platform that is highly available and distributed. For this, we selected Inter Planetary File System (IPFS) [15, 16]. IPFS is a well known distributed file system that is scalable to handle huge amount of data with the ability to track and retrieve any file with its own unique URI. We use this unique URIs to point to resources in the idea contracts.

### **6.1.3 Idea Platform and Protocols**

Like GIP and Idea-Hub discussed in subsequent chapters. At the core of the idea platform lies a distributed overlay network that segregates different protocols while utilizing the same underlying distributed network architecture. The overlay network can be better understood in terms of communities. Each protocol constitutes a community of people interested in the idea. One such community and the protocol is the GIP. The platform is open and extensible to encourage people to create new protocols and communities on top of it.

### **6.1.4 Idea Client**

Idea client is the end user interface that allows users to interact with the Idea platform and the protocols. For example, an user can submit the idea or upload papers or review an idea in GIP using the client. The client also includes an built-in blockchain wallet which can hold and transact Idea tokens. As for the first implementation, the wallet is an ERC20 / ERC721 wallet that can hold any compatible token along with Ether itself. To invoke any action on the Idea platform that updates the state of the idea requires transaction fee which is paid with our token. The wallet handles all the interaction with the idea platform interactions and the blockchain.

## **6.2 Blockchain agnostic & Minimum use of blockchain**

Blockchain is the anchor point for building and maintaining trust within the system. Within Idea platform, blockchain is used for identity, idea and token ownership, and execution of smart contract transactions. Proof of authenticity of the proposed idea based on the registration timestamp, idea signature and the immutability of the blockchain records is an important strength of the platform.

OSO Idea platform is designed as a decentralized platform using modular components including web3 technology and blockchain is a missing puzzle for trust. Our careful consideration from the beginning is to be blockchain agnostic. We started the reference implementation with smart contracts pioneer and currently one of the best smart platform Ethereum but we welcome the implementations of Idea protocols with other platforms. We always strive to support compatibility of rest of the components of the Idea platform to work seamlessly with any blockchain implementation of the protocols.

### 6.3 OSO utility token

OSO utility token facilitates the transactions within Idea protocols. The token can be exchanged to ether and used as gas for transactions. It is a standard ERC20/ERC721 token by design.

### 6.4 Identity

We have an in-house solution called Unique Researcher Identity (URI) <https://github.com/open-science-org/URI> that provides unique cryptography based identity to the researchers. However, we acknowledge that decentralized identity is a challenging problem that has not been fully solved yet. We are constantly updating our knowledge about this topic and we are open to other better solutions.

## 7 Idea-Hub: Decentralized Publication System

Where the ideas live (and stored)!

A decentralized storage system for all of scientific ideas and materials that will instantly be hashed and linked together upon submission to the system. The information would be considered a node in the network and could range from a hypothesis to on going work on a project all the way up to a manuscript. The storage system will use web3 and blockchain based ideas. It would include a Merkle tree of material all connected into a mesh of ideas. By implementing this system everything entered into it will automatically be indexed and referenced as part of a system of ideas. It will easily be searched and find connections between each node in the Merkle tree. When information is updated any node depending on this information will automatically be updated and/or flagged that information in the chain has changed. This could help with retraction watch of publications and also when updated information gets presented around a subject.

#### Features

- Idea storage
- Merkle Tree of Ideas/Citations
- Store weighting system for attribution that was created by GIP

Idea-hub has two major subsystems:

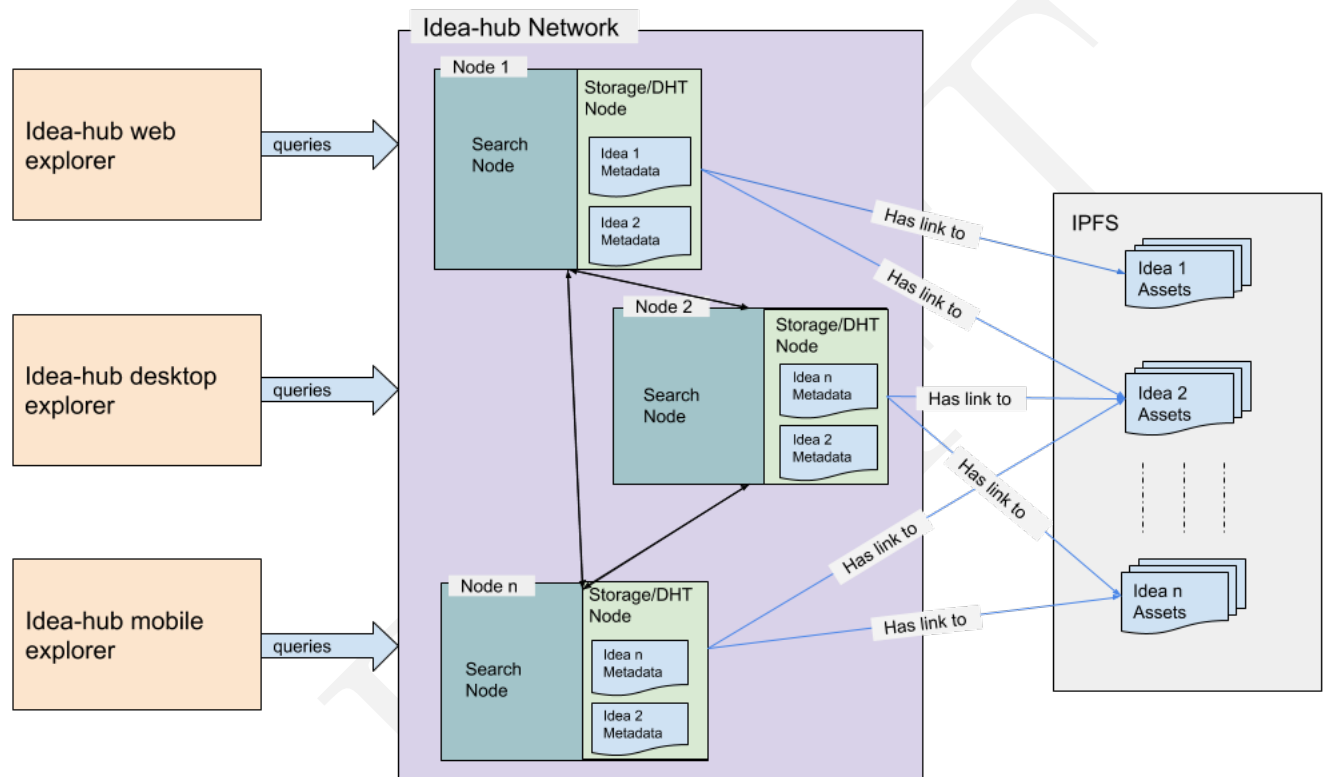


Figure 5: **Idea-Hub**. One place to store and interact with the ideas (e.g. publications). Two main components: **1. Assets Storage Network:** A decentralized file storage network for storing the files and assets related to ideas (uses IPFS in the current design). **2. Idea-Hub Network:** A decentralized network of Search and Storage nodes which store the metadata info of ideas along with IPFS links to the related files in the Idea Assets Storage Network.

### 7.0.1 Assets Storage Network

A decentralized file storage network for storing the files and assets related to ideas. Currently designed to use IPFS for storing the actual files.

### 7.0.2 Idea-Hub Network

A decentralized network of Search and Storage nodes. These nodes store the metadata info of ideas along with IPFS links to the related files in the Idea Assets Storage Network. Each node can act as either a Search node or a Storage node or both. Search nodes are responsible for indexing and providing a streamlined query interface for the Idea-hub clients. Storage nodes are responsible for storing the Idea metadata and related IPFS file links in a decentralized and fault tolerant way. The storage nodes are essentially nodes of a highly available Distributed Hash Table of Idea Metadata.

#### Advantages

- Automatic Indexing of content
- Dependencies hash linked together to ideas
- this could include peer reviews, citations or other types of references
- Retractions will be inherently found when information is updated
- Updates to ideas will be readily visible by changes in links and hashes of content attached to it. (such as peer reviews)

## 8 Use Cases

Below we describe two simple example use cases of the OSO idea platform. There can be several other use cases which we have not mentioned here or even not imagined at this point.

### 8.1 Academic Publishing

Ramanujan is a mathematical researcher and he wants to publish his idea in Channel Blue of Mathematics Sub-network. Channels are publishing venues with their own custom rules(e.g. specialization track, selectivity, review policy, etc.). He submits his idea  $I_{1729}$  to the channel blue and creates a bounty of 60 OSO tokens as the "*cost of review*" to pay reviewers for their intellectual contributions i.e. reviews. The channel blue has its own policy. Three mathematical researchers Gauss, Turing, and Aryabhata meet the requirements for reviewers and are interested to review, so they claim the OSO tokens and provide their reviews. In OSO platform, a review itself is an idea and the reviews



are submitted as idea objects to the channel blue's policy (can be a smart contract). As per the policy, based on the majority voting among the reviewers, there are possible four outcomes: 1) Accept and publish in channel blue, 2) Suggest to publish in channel yellow and provide a rationale of the decision 3) Reject with rationale of the decision and automatically publish in channel red, and 4) Send a revision signal to the author (researcher Ramanujan) alongside with their reviews.

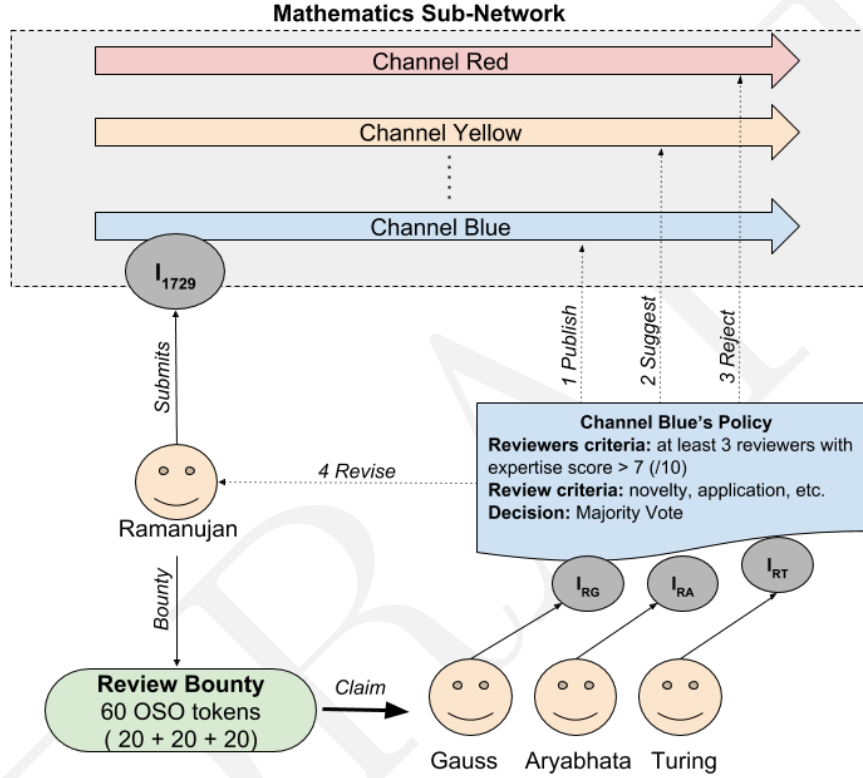


Figure 6: **Academic Publishing Example.** Ramanujan, a mathematical researcher submits his idea  $I_{1729}$  to the channel blue of Mathematics Sub-network and creates a bounty of 60 OSO tokens as the *cost of review* to pay reviewers for their intellectual contributions i.e. review. Three researchers Gauss, Turing, and Aryabhata meet the requirements and they claim provide their reviews as idea objects to the channel blue's policy and receive OSO tokens. Based on the majority voting among the reviewers, there are 4 possible outcomes: 1) Accept and publish in channel blue, 2) Suggest to publish in channel yellow 3) Reject and automatically publish in channel red, and 4) revision.



Once the drug is developed and is commercially available, the profits from the sales will be distributed proportionally to the staked IDEA tokens that were created earlier. Alice, Bob and Charlie can use these funds to continue their research, Biotech D can use these funds to grow there company and Investor E can enjoy a return on their investment for helping such a noble cause come to fruition.

Everyone in the network that is contributing is incentivized in this scenario. Alice, Bob and Charlie all receive funding by doing the research they love. The better and more impact their research is to society the more funding will be generated back to them to continue their research. Biotech company D will save time and money by having validated basic science research to help develop further into therapies all while being able to advertise and obtain necessary funding from investors. An investor E, will have a less risky investment opportunity by having access to high quality validated research that companies want to actively develop. Agreements will be made up front and the network will handle all distribution of ideas and funding appropriately. This could also alleviate companies and/or investors of some of the management time and costs that comes with such interactions. By having a decentralized platform that helps coordinate and distribute the necessary information, it will help facilitate better operations. In this scenario, everyone that contributes is incentivized. Higher the impact of their research, more the funding will be generated for their research. It makes the interactions more transparent and efficient and moves science forward in a meaningful direction.

## **9 TO DO (in next version)**

### **9.1 Why IPFS over DAT, Swarm, etc.?**

- DAT
- IPFS
- Swarm
- IPFS vs. Swarm
- IPFS vs. DAT (response by Juan Benet)

### **9.2 Use of IPLD**

We feel the key questions is not what we store the data on but ultimately the idea object that we store. We think at the base level either will be great for storing the underlying data/files. We think the key is defining the idea object in such a way that we can use IPLD (<https://ipld.io/>) to index and know how to search/link ideas together. This can be used regardless of which system we store the data on.

### **9.3 Choice of Smart Contract Language**

- Vyper vs. Solidity

### **9.4 Economic incentives for IPFS (or DAT) nodes.**

- File Coin vs. Custom solution
- Universities acting as OSO nodes

### **9.5 Attack Vectors**

TO DO:

## References

- [1] G. J. Katuwal, “Open science organization- a decentralized autonomous organization for better scientific ecosystem,” 2017. [https://github.com/open-science-org/wiki/blob/master/OSO\\_white\\_paper.pdf](https://github.com/open-science-org/wiki/blob/master/OSO_white_paper.pdf).
- [2] M. Ware and M. Mabe, “The STM Report: An overview of scientific and scholarly journal publishing,” *The STM Report*, 2015.
- [3] T. C. Bergstrom, P. N. Courant, R. P. McAfee, and M. A. Williams, “Evaluating big deal journal bundles,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 26, pp. 9425–30, 2014.
- [4] T. Phan, L. C. Hardesty, J. Hug, and C. L. Sheckells, “Documentation for the Academic Libraries Survey (ALS) Public Use Data File,” 2011.
- [5] “Flacking for big pharma.” <https://theamericanscholar.org/flacking-for-big-pharma/#.W4nDupNKj5Y/>.
- [6] “The history of publishing delays.” <https://blog.dhimmel.com/history-of-delays/>.
- [7] M. Baker, “Is there a reproducibility crisis?,” *Nature*, vol. 533, pp. 452–454, 2016.
- [8] “Data check: U.s. government share of basic research funding falls below 50” <http://www.sciencemag.org/news/2017/03/data-check-us-government-share-basic-research-funding-falls-below-50>.
- [9] “Knowledge is a kind of infrastructure.” <https://blogs.scientificamerican.com/guest-blog/knowledge-is-a-kind-of-infrastructure/><https://blogs.scientificamerican.com/guest-blog/knowledge-is-a-kind-of-infrastructure/>.
- [10] G. Wood, “Ethereum: a secure decentralised generalised transaction ledger,” *Ethereum Project Yellow Paper*, pp. 1–32, 2014.
- [11] R. T. Thibault, A. MacPherson, S. Harnad, and A. Raza, “The rent’s too high: Self-archive for fair online publication costs,” *arXiv preprint arXiv:1808.06130*, 2018.
- [12] R. Van Noorden *et al.*, “The true cost of science publishing,” *Nature*, vol. 495, no. 7442, pp. 426–429, 2013.
- [13] “Gofundme says its users raised \$1 billion last year.” <https://www.philanthropy.com/article/GoFundMe-Says-Its-Users-Raised/234430>.

- [14] J. Bang-Jensen and G. Z. Gutin, *Digraphs: Theory, Algorithms and Applications*. Springer Publishing Company, Incorporated, 2nd ed., 2008.
- [15] I. Baumgart and S. Mies, “S/Kademlia: A practicable approach towards secure key-based routing,” *Proceedings of the International Conference on Parallel and Distributed Systems - ICPADS*, vol. 2, 2007.
- [16] J. Benet, “IPFS - Content Addressed, Versioned, P2P File System (DRAFT 3),” no. Draft 3.