SIMULACRUM 1.0

Incentivizing and Crowdsourcing Availability of Research Outputs: Simulacrum 1.0

Vikram Dhillon*
*Electronic mail: dhillonv10 at gmail dot com

Unbaised dissemination of clinical research and fostering further development are of critical importance in life sciences where the lives of patients depend on the quality of treatment administered. However, positive results are generally twice as more likely to get published as negative results and very often as a result of negative data missing in action, the studies become unreplicable because the rare working positive conditions are published. Streamlining published research and encouragement of publishing negative data from experiments in the context of the methodology and not the proposed hypothesis of researchers allows platforms like figshare to emerge as leaders in data sharing and collaborative research. A common criticism with the figshare model that has recently surfaced is the lack of motivation and significant return for adoption of the platform amongst principal investigators (PIs). We propose a solution to incentivize, promote sharing and collaboration of data through the figshare platform by attaching a bitcoin-based cryptocurrency as a reward for uploading to figshare. Creating value through the adoption of a cryptocurrency platform to assist research sharing also allows for our solution to be used as a non-profit or a Distributed Autonomous Community (DAC) entity. We also discuss to extend the described protocol and methods to provide significance to negative results and clinical trials by making them easier to publish and share through the figshare platform.

1. OVERVIEW

Negative data is very difficult to publish and often just remains in lab notebooks where it is of no use. We release now, more so the importance of releasing as much data as possible when we hear about scientific papers being retracted because of misconduct or that the results produced were not replicable. In addition to that, this is even more critical for clinical trials for new drugs where almost all of the positive data is presented however only a fraction of the negative results are made available, even if the negative results show that the new drug is nearly as effective as placebo. For doctors who follow the studies, they are misled in prescribing those new drugs to patients where they might have harmful side-effects or worse yet no effects at all. This is our first-attempt to incentivize massive research outputs with Simulacum 1.0. Our goal with this platform is to create a sustainable incentive for researchers and scientists to release as much data as possible through known and functional channels and make it publically available.We will be using Figshare as our data repository simply because it was created to help disseminate research data and it does so very well. Using a strong data-backend like figshare has numerous advantages in that we will not have to create our own backend, figshare is easy to use and many researchers either already use it or are familiar with it. The ideal goal for our platform is to augment peer review and make the studies independently replicated by providing more data from the experiments conducted.

This paper is organized as follows: First we briefly go over the core of our platform which is a fork of Datacoin, then we describe what our motivations behind Simulacrum are, afterwards we discuss the

support and the storage modules that make up our platform. Finally after a lengthy discussion of our platform, we provide two extensions that result from our discussion and show the real power of what can be accomplished with Simulacrum.

2. TECHNICAL BACKGROUND: DATACOIN

Datacoin is a primecoin derivative for evolution of blockchain into a metadata storage service based on decentralized currency that supports the storage. In the datacoin model, a user pays for the storage of their files with the currency they mine using the miner and the files can never be removed from the blockchain. Currently fee is 0.05 DTC per 1Kb of data. In the datacoin payment support model the miners who mine the blocks keeping the blockchain alive are rewarded from this fee. The reward model is the traditional currency associated with mining the block and each block has an associated DTC reward which decreases after a certain number of blocks have been mined. The computational resources to mine the blocks increase as the number of miners increases, and datacoin being a primecoin derivative uses the same Proof of Work (PoW) thereby also generating new Cunningham prime-chains and verifying them which hold great significance in number theory research. In this manner, datacoin not only creates a business model for storage but also generates useful byproduct as prime chains.

Datacoin is highly modular containing separate upload scripts and a datacoin browser which is currently in development. This modularity will allow us to create more addons to the infrastructure to the Bitcoin 2.0 models where the core transaction-mining network remains unchanged however more layers of interpretation can be added without any disruptions or latency in the network as they operate separately. Generalizations to the datacoin platform allow us to as discussed in the following sections will allow us to develop a general infrastructure for the development of highly-specific metadata storage applications automatizing and decentralizing current services. In most technical aspects such as calculating difficult of blocks, datacoin inherits from primecoin, also to be noted here: While calculating difficulty primecoin also manages to implement an analogue of Kimono gravity well which retarget after every N number of block and adjusts very quickly keeping fairness in the mining process.

3. INTRODUCTION TO SIMULACRUM 1.0

High reproducibility of scientific experiments has always been a cornerstone in research. Lately, the number of retractions has gone up even for the results published in high impact journals because they can't be replicated properly. To reduce the publication of the flukes that worked as results, the availability of an easy to use and widely adoptable data-platform/repository is pertinent. To meet these requirements, the trade-offs from the two parties involved need to be satisfied:

—From PIs - The trade-off for their time in terms of a reward or reward points that count towards enrollment in a program

—For the costs of the reward program to be dispersed in a fashion such that no one entity bears the load of supporting it

These two requirements are critical for the success of our platform, one of the main objections towards making more and more results available in a public repository is that why would the PIs spend their valuable time in doing so. As long as a journal accepts their results, the scientific community considers the results valid. However, often after the acceptance and publication of the papers we notice that the results can't be replicated leading to retractions. The second requirement breaks the high-barrier of entry because of shortage of cash-flow from research institutes towards project that might seem extraneous: A research institute has any number of researchers submitting their papers for publication and most of them get accepted. The institute is more likely to hold to the standards of a journal

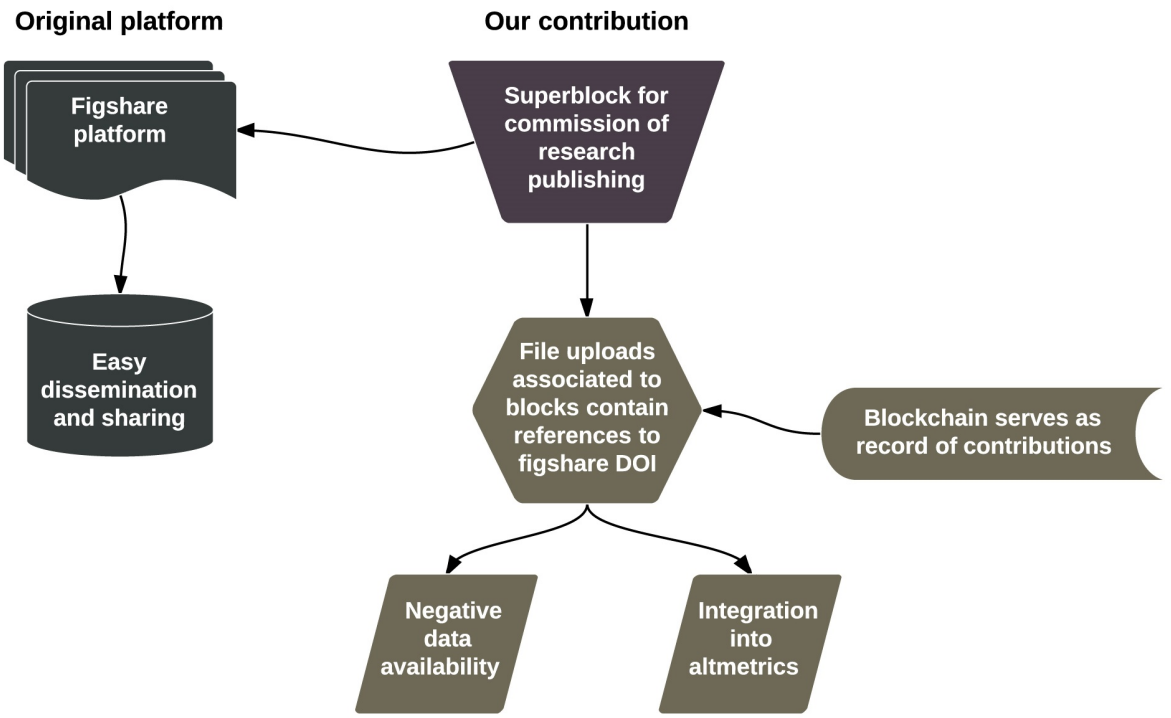

Fig. 1. Overview of our platform

that have been established, instead of having to go through publishing more data particularly if it's negative data. This extra data is not required by the journal and therefore the scientific community would not object to an institute that does not provide it.

Similarly, institutes are less likely to create and promote initiatives to publish as much data as possible because as long as majority of their researchers get positively accepted, this measure seems extraneous and adding them leads to added weight (in the form of extra financial support which creates that high-barrier of adoption for our platform). One possible solution to the second requirement is to create a stream of capital that supports our platform and at the same time does not interfere with grants or financial support by an intuition therefore making it very low-friction for an institute. The added weight of providing new data can be offset and distributed in a crowd-sourcing model of a cryptocurrency. The financial value provided such a cryptocurrency would be completely separate from institutional support and to use it they would only need to the electricity costs for a computer to act as a miner. This will be discussed at length in the upcoming sections.

To satisfy the two requirements mentioned earlier, our platform tries to answer them in terms of two modules: A storage module that satisfies the ease of use and easy adoption into an institute and also a support module.

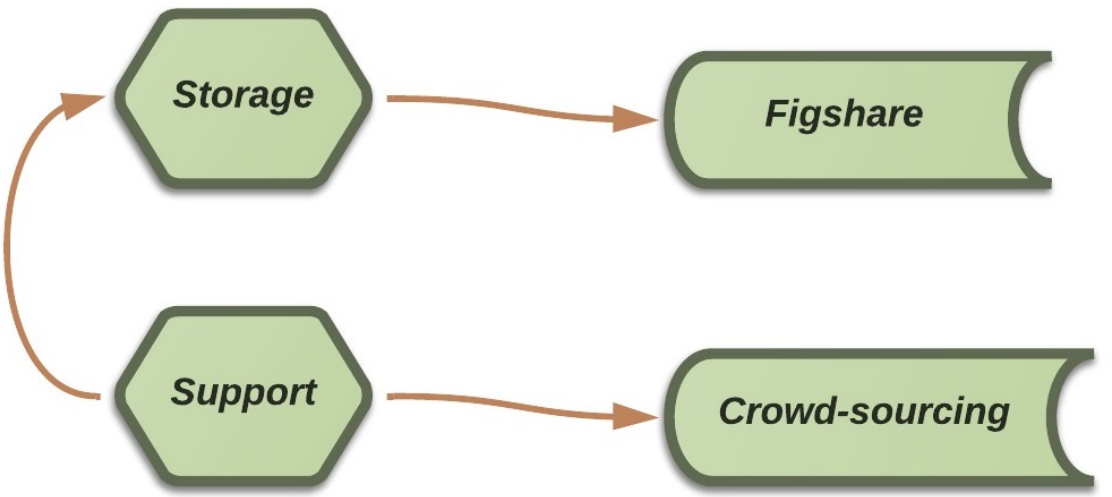

Fig. 2. Organization of simulacrum

4. CROWD-SOURCING MODEL: SUPPORT MODULE

Our platform relies heavily on balancing the sharing of data and generating an support the sharing, and thus in order to make this model self-sustaining in a limited sense we propose to rely on a crowd-sourcing model which will integrate itself very well within the extensions. In our model, we have two core entities: Miners and Assignees. Both have intertwined, but they are well defined roles are as follows:

(1) Miners:
   —Miners comprise of anyone interested in mining the cryptocurrency and using it to make a contribution to an institute to see the impact their work has on influencing research and making more results publicly available.

(2) Assignees
   —Assignees play a very interesting role in our model, they are similar to reviewers or editors for a journal. They assign the resources that either the institute has mined or were donated by miners and allocate them accordingly to the researchers. The aim here is that once after a study has already been published, any data (negative or related experiments) that the researchers wish to release can now be done through Figshare and the researcher will send the assignee a DOI from Figshare. The assignee will then curate the DOIs and upload them in the metadata descriptor to the blockchain for every release window.
   —The Assignees within any institute should have a publicly verifiable identity that can be encoded by a PGP-key that they will use to sign the uploading document every release cycle.

In addition to the core entities that comprise our network, we can foresee at least two more entities to emerge from our model:

(1) Non-profit gateways:
   —A non-profit organization is an interesting prototypical entity that can adopt this model. This organization exists on the blockchain so to speak, and will mine the cryptocurrency just as miners. Also just like organizations they can receive funding as contributions from miners and distribute them as previously decided by their publicly available agenda.
   —Another function that non-profits can serve is as an exchange gateway where a non-miner can contribute to a non-profit via their already established channels and the non-profit can pay forward to an institute in crypto.
(2) Patrons:
   —Patrons are similar to $miners$ in that they make contributions however they don't do so directly with their own mined crypto, but instead they trade for the crypto with fiat through an exchange gateway and pay forward with that.
   —Most institutional researchers acknowledge their funding sources in publications and for our miners one method of rewarding their efforts will by the assignees to recognize addresses that have contributed to the Figshare DOI upload as a Patron. The Patrons can choose to be identified or not as they wish but this allows another prototype entity to exist in our model.

This dual role of assignees in our model is very similar to what maintainers in the linux-kernel community have. They are employed by a hardware company that uses the linux-kernel so they maintain they daily jobs however since the kernel is open source, the maintains also have a very strong presence to the community where they curate and review patches.

5. FIGSHARE PLATFORM: STORAGE MODULE

An easy way to understand the involvement of Figshare in our schematic is to think of Fighare to us is as imgur is to Reddit, it is being utilized as a data or research output storage platform. UNIX philosophy in software integration is to let one tool perform one function that it can do very efficiently. In accordance to that, we treat figshare as a storage platform with relation to our development modules. This allows the user/researcher to have no further knowledge of the protocol behind it that we have proposed, they will simply use figshare to upload the research they have worked on and submit the final DOI to the $assignee$ The assignee can then take the DOI or multiple DOIs, compile them into a metadata descriptor text file and upload that to the blockchain. We propose the following metadata descriptor format to be used and universalized to promote standard descriptors:

Parameters	Descriptor
time_stamp	Time stamp on the uploaded file
file_hash	A hash applied on the metadata file
uploader_name	Name of the assignee
Patrons	Supporters of the conducted research
public_key	Public key associated with the uploader
uploads	An array for associated DOIs on figshare, for instance $'DOI'_1,' DOI'_2, ...'DOI'_n$

After having discussed the storage and the support modules, we believe that in Simulcrum a cryptocurrency is the correct path to take because creating an open, transparent and decentralized platform can benefit greatly from leveraging cryptocurrencies as they were inherently created based on these principles. The funding or resources that an $assignee$ allocates to the researchers can be used almost as purchase power for programs that become part of the platform. One example we can envision is that of a faculty-wellness program, where the researchers who have contributed DOIs to the $assignee$ get enrolled in this program using the reward points they earned for releasing the new data.

6. EXTENSIONS: CLINICAL TRIALS

The proposed extensions to the core-platform are the most exciting aspects for us since these extensions are capable of creating the sustainable model for us accomplishing the purpose of Simulacrum.

6.1 OTP Index: Open Trials Published

With the storage module as extensive as Figshare, any experimental data can be deposited and the first target we want to reach out to is data from clinical trials. Often a very high amount of data is collected out of which only selective small portion of the negative data get published. In part, this is a major reason why flukes get published as targets that can't be replicated. We hope that with the PIs getting an incentive to release more experimental data than was published, we can improve the general quality of research and enhance the peer review process from simply being peer review to independently replicable.

Retraction of papers has disastrous consequences for an institute and all the more when it's from a high profile journal. One consequence of using our platform is the creation of an index that quantifies how much data in trials or just from experiments an institute has made available in public that can be used by anyone with the expertise to check or replicate the results. We define this index as Open Trials Published (OTP), as an institute adopts this platform and start generating more data, we will start to see a certain level of trust build upon the fact that their results can be independently verified. Naturally as a consequence of this, the institute can claim a higher level of confidence that studies published by them and their OTP index number goes up demonstrating that their studies are independently replicable. NIH can do also do this as an overall grant program requirement, having an high OTP index can be a positive indicator for the grant reviewers that the grant resources being provided are being used properly

6.2 Financial Derivatives

One consequence from OTP can be used in a feedback loop into the platform. The idea is that for an institute with a high OTP, they can release a financial derivative based on IPO. We define the RIPO (Research Initial Public Offering) as a monetary evaluation of the research being done based on the criteria of the OTP index and the funding recieved from NIH or other agencies (all publicily reported metrics) to support a particular program. After the RIPO is done, the research program will rise in value as the stocks are traded and they will keep a portion of the stocks. As they rise in value, this allows for more resources to be diverted towards the programs that are generating high quality research with an equally high OTP index.

This need not be done inside an institute where other administrative constrains prevent RIPO from occurring. Any entity on the blockchain can act as a RIPO gateway through an $assignee$ to start the process and the gateway is also required to disclose all the funding to the $assignee$ who can then include it in the release window as contributions from a $patron$. In this fashion, the RIPO stays outside the institute while keeping a sense of transparency of research being done and generating revenue for the platform to be fed back into the cycle. High OTP indexes and particularly financial derivatives based upon them can also serve another more important function: Attracting the bright young scientists in training to these programs. If the financial derivative program works out for an institute adopting our platform, they can present their success and OTP in a yet another fashion which is their stocks.

The two extensions shown here have only scratched the surface of what is possible with entities reading off the blockchain for release windows by $assignee$. In near future, we hope to find and create even more entities that can benefit from our platform and detail in a frictionless, low-barrier to adoption techniques to incorporate our platform in their work flow. In some sense, this is not a huge step away from researchers who post on their own blogs.

7. NAMING CHOICE

So why did we call it Simulacrum? Jean Baudrillard wrote a book $Simulacra and Simulation$ and in it, he comments on the post-modern society and that our current society has replaced all reality and meaning with symbols and signs, and that human experience is of a simulation of reality. To us, these symbols that exist in our mind hold more meaning since they are ideal and perfect. We don't have to worry about fixing anything wrong with them, since nothing can go wrong with them. Moreover, these simulacra simply hide that anything like reality is relevant to our current understanding of our lives and replace them with idealistic versions where we seem to find everything in harmony and a state of functioning. The simulacra hide the ugly and rotten reality of our world and we become prisoners to our own minds failing to look past the idealism.

This we believe is the current state of medicine - We often think of it as all-powerful, miraculous science and our simulacra give us this experience. However at it's core, there is a lot of work to be done and this paper is just the beginning.