

THE CONVERGENCE OF **BLOCKCHAIN & AI** IN DRUG DISCOVERY.

A TECHNICAL WHITE PAPER

innoplexus[™]

ABHIJIT KESKAR
JUN 15TH, 2018



www.innoplexus.com | info@innoplexus.com

Contents

Introduction	1
Going beyond AI	1
Platform architecture	2
The platform	
The author / researcher's environment	
The consumer's environment	
Overall technology stack	4
Key components of architecture	5
Trusted agent	
IPFS	
Ethereum	
Valuation engine	
Experiment Similarity-Detection Engine	
Elasticsearch	
Two key data flows from a publishing and transaction platform	6
Author's submission of experiment reports	
Consumer acquisition of license to unpublished data	
The road ahead	7

Image Index

Fig 1 - Integrated architecture of a blockchain and AI platform	3
Fig 2 - Architecture for blockchain based component	4
Fig 3 - Design of a trusted agent that resides on an author's environment	5
Fig 4 - Author's submission of experiment reports	6
Fig 5 - Consumer acquisition of license to unpublished data	7

Blockchain + AI: The next frontier for drug discovery

Introduction

The pharma industry is at the cusp of a huge transformation, and companies that remain ahead of the curve will rule the roost. Any insight that can help them shorten drug development time will help build competitive advantage. Quicker drug development is good for both pharma companies and citizens at large, since the sooner life-saving drugs can be made available, the more lives can be saved. While providing insights to pharma companies, it is also important that data belonging to the producers, as well as the consumers, is completely secured.

In this paper, we will outline how two cutting-edge technologies, blockchain and AI, can be integrated into a life science research platform to accelerate drug discovery innovation. A blockchain and AI-powered drug discovery platform could include a secured publishing and license transaction platform for unpublished data to provide secured, holistic insights from a drug development standpoint. Three main themes should guide the development of such a platform: **Security** of producer and consumers' data at any cost, **all-round insights** from all possible sources (published, unpublished, restricted, enterprise data, and third-party subscription data) and **fairness** in data valuation and similarity detection through use of artificial intelligence.

Going beyond AI

In recent years, AI and machine learning technologies have been very popular among global pharma giants for deriving insights from published data.

That said, not all of the insights are available in published data. Sometimes the insights are hidden in unpublished data. And, at times, for various reasons, the experiments don't get published. This is usually common when one of the following things happen:

- The experiment has failed to prove the hypothesis (the hypothesis is proven to be false)
- After the experiment, the direction of the organization has changed
- The experiment has been done in an academic setting (e.g. a doctoral thesis)

Insights hidden in such experiments are not available for pharma companies' consumption. This results in pharma companies reinventing the wheel, thereby wasting time and money

However, a blockchain and AI-powered drug discovery platform can solve this problem by launching a secured publishing and license transaction platform based on Ethereum (blockchain) and Artificial Intelligence.

In the context of unpublished data, the key problems that this platform can solve are:

- Secured publishing of unpublished data on blockchain, which makes it immutable and timestamped (thereby establishing authorship)
- Secured exchange of licenses to use the unpublished data (exclusive or shared)
- Fair valuation of the artifacts using Artificial Intelligence in real time
- Extensive check for similar documents and duplications using Artificial Intelligence in real time

Apart from the unpublished data, the platform can also make other types of data searchable. All of the following sources of data can be made searchable for a pharma company in secured way:

1. Published data
2. Unpublished data (that the company has acquired)
3. Restricted data (like the patient forum, for example)
4. Enterprise data (specific to the company)
5. Third-party subscribed data (that the company has subscribed)

The USP here is that unpublished data, enterprise data, and third-party subscribed data never leaves the company's environment yet is still made searchable on the platform. Blockchain and Artificial Intelligence are leveraged to deliver differentiating value to the users.

Platform architecture

A blockchain and AI-powered drug discovery platform can provide insights by leveraging published data, unpublished data, restricted data, enterprise data, and third-party data. Hence the architecture has some components that reside at the data source's environment, while the rest of the components reside on the platform's environment.

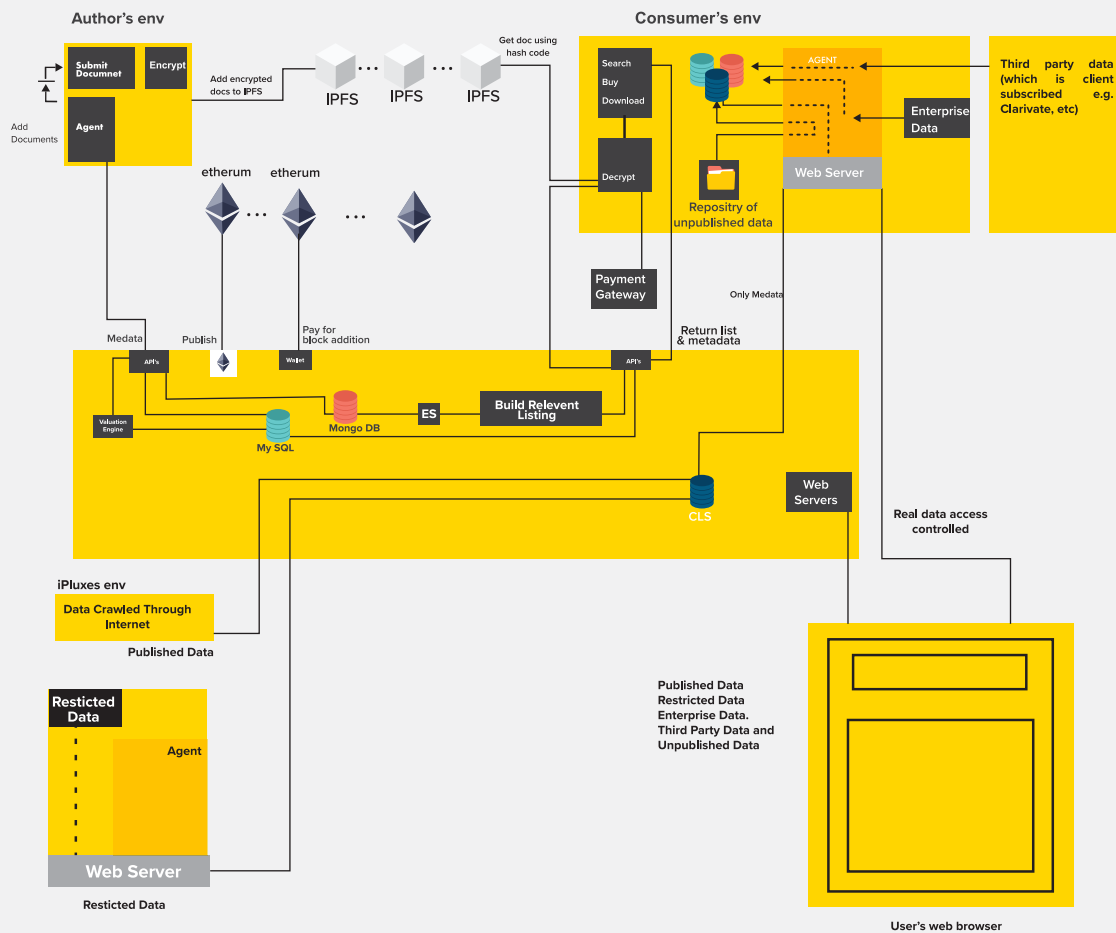


Fig 1 : Integrated architecture of a blockchain and AI platform

The platform

The platform can be responsible for enabling submissions by researchers, listing of unpublished experiment data, license acquisitions, and integration of all five types of metadata into platform's repository. While the platform can provide validations, it can't get the unpublished documents. It only receives metadata. The platform can also ensure real-time valuation (guidance pricing) of the submitted documents purely based on the metadata and keywords from ontology that are present in the document.

The author / researcher's environment

An author can submit unpublished data related to experiments directly to IPFS and then publish it to Ethereum (blockchain), where authorship is established. The author also provides the document key once license acquisition is confirmed. A trusted agent in the author's environment ensures that the actual data doesn't ever get transferred to the platform and that only metadata is transferred.

The consumer's environment

The consumer is typically a big global pharm company who acquires licenses to unpublished data. Upon acquisition, data gets stored locally. Since this unpublished data, enterprise data of the company, and third-party subscription data is the sole property of the company itself, the data never gets transferred. Instead, only the metadata that is extracted by a trusted agent is transferred. Such metadata transferred to the platform remains the sole property of the consumer company and is seen (and made searchable) only to the company. Other users of the platform never see this metadata or actual data. Ownership is established at a data record level in platform's repository.

Overall technology stack

Overall tech stack now involves blockchain and IPFS apart from the usual components, like Elasticsearch, MongoDB and RESTful APIs.

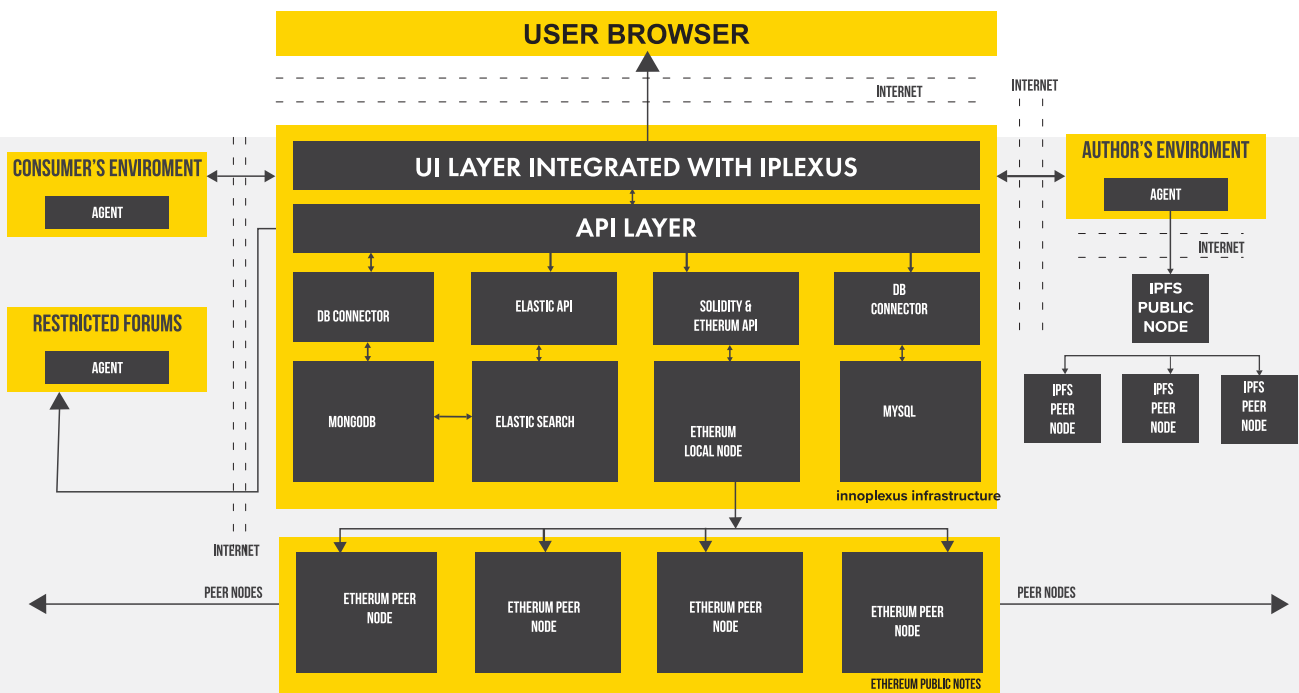


Fig 2 : Architecture for blockchain based component

Key components of architecture

Following are the key components that make up the platform:

Trusted agent

A trusted agent can be designed to be a light-weight component of the platform that resides on the remote environment (either the consumer's, author's, or restricted forum's). It can be configured to generate indexes based on the ontology keywords that appear in the data. It can be a one-time download and install, and it can act in the interest of the owner of data.

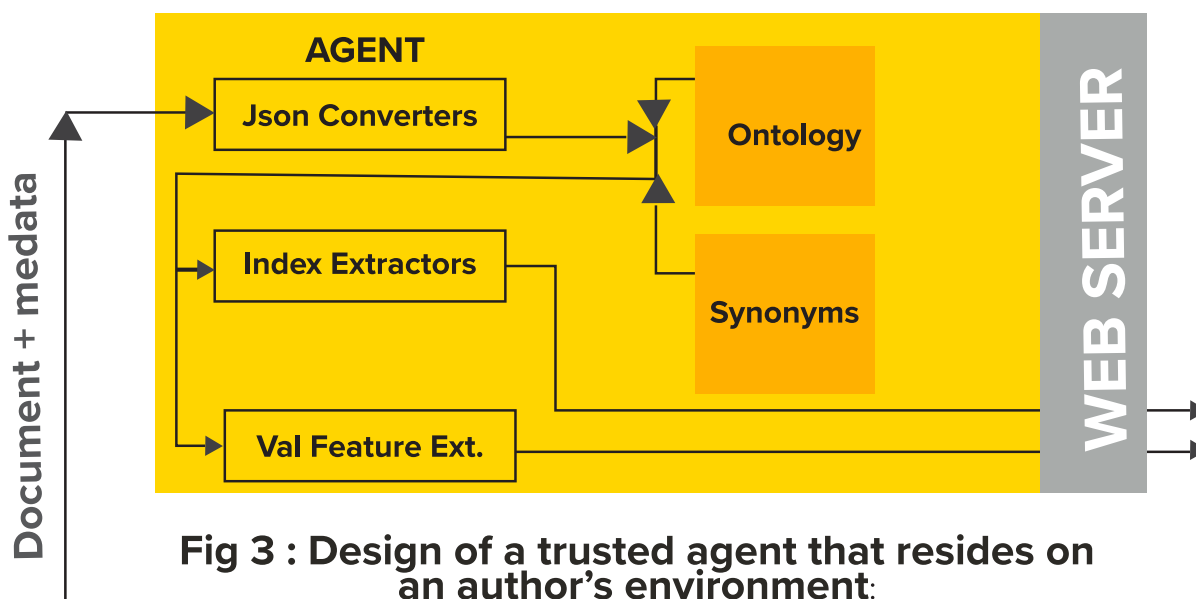


Fig 3 : Design of a trusted agent that resides on an author's environment:

IPFS

Inter Planetary File System (IPFS) is used to store the experiment documents in an encrypted format. IPFS is used because it is distributed (like blockchain) and is very inexpensive as compared to Ethereum. IPFS also helps identify any duplicate submissions from an author. A consumer, upon acquisition of a license, can take the documents directly from IPFS and decrypt them using the document key provided by the author.

Ethereum

Ethereum's public blockchain serves two purposes. It helps in establishing authorship by storing the hashcode of documents in IPFS, an author's id, and a timestamp. It also stores the license transfer transaction for immutability.

Valuation engine

A valuation engine is a patent-pending innovation that uses a machine learning algorithm to generate guidance valuation based solely on the keywords index and metadata. This happens in real time and can bring fairness to the process.

Experiment Similarity-Detection Engine

Engine This is another patent-pending innovation that uses machine learning in real time to detect similar and duplicate submissions, based on IPFS hashcode, keywords index, and world knowledge. It is also useful in recommending similar (relevant) experiments.

Elasticsearch

Elasticsearch provides quick access to a document's metadata using the keywords from ontology and their synonyms. Interestingly, this happens even though the actual document is never sent to the platform.

Two key data flows from a publishing and transaction platform

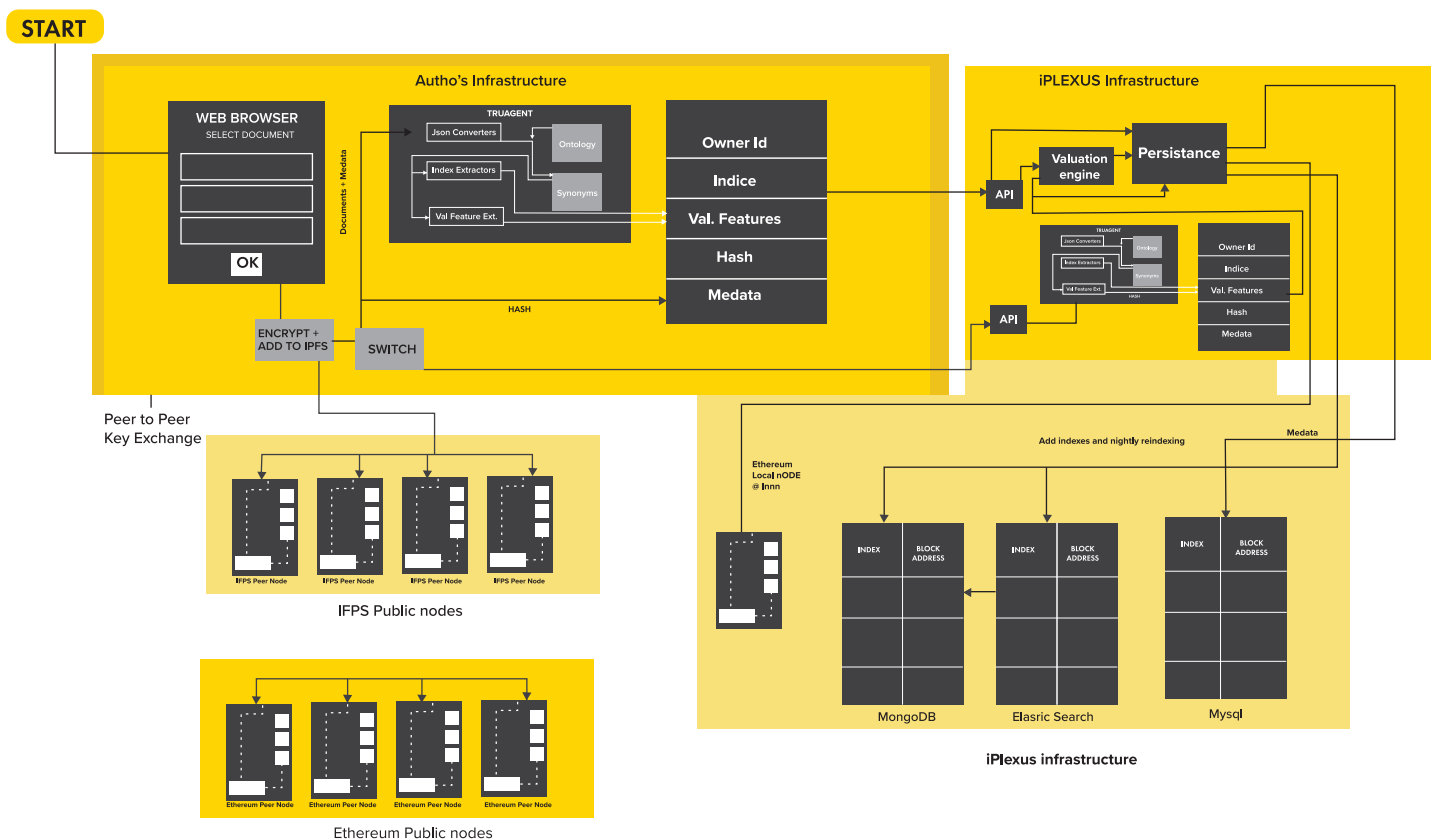


Fig 4 : Author's submission of experiment reports

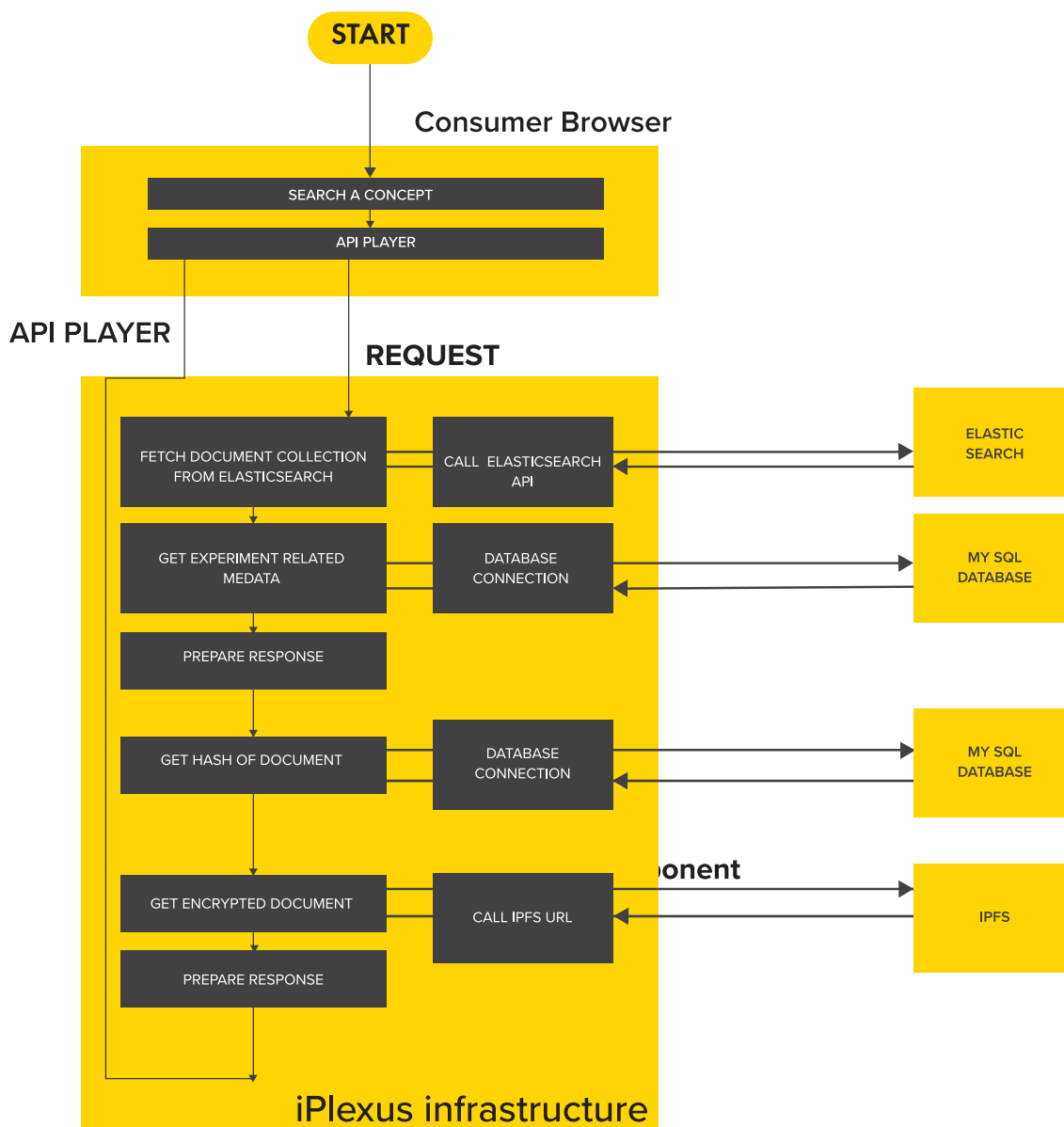


Fig 5 : Consumer acquisition of license to unpublished data

The road ahead

This new architecture is expected to provide a 360-degree view to consumers, thereby reducing the drug development cycle time. Many consumers have shown interest in adopting this type of blockchain and AI-powered drug discovery platform. Additionally, when a trusted agent architecture is used, many newer data sources are expected to become integrated with such a platform. Many new restricted forums are likely to join the platform as well. The road ahead shows significant promise for ground-breaking pharmaceutical innovations.