Blockchain-based MDM for data governance A vaccine supply chain use case

Imane Ezzine
Ecole Mohammadia d'Ingénieurs
Mohammed V University in Rabat
Rabat, Morocco
imaneezzine@research.emi.ac.ma

Laila Benhlima
Ecole Mohammadia d'Ingénieurs
Mohammed V University in Rabat
Rabat, Morocco
benhlima@emi.ac.ma

Abstract*—The effects of COVID-19 have quickly spread around the world, testing the limits of the population and the public health sector. High demand on medical services are offset by disruptions in daily operations as hospitals struggle to function in the face of overcapacity, understaffing and information gaps. Faced with these problems, new technologies are being deployed to fight this pandemic and help medical staff governments to reduce its spread. Among these technologies, we find blockchains Artificial Intelligence which have been used in tracking, prediction applications and others. However, despite the help that these new technologies have provided, they remain limited if the data with which they are fed are not of good quality. In this paper, we highlight some benefits of using Blockchain and AI to deal with this pandemic and some data quality issues that still present challenges to decision making. We present a general Blockchain-based framework for data governance and particularly we propose a MDM solution based on blockchain and AI that aims to ensure ~~a~~ high level of data trust, security, and privacy. Finally, a use case of healthcare supply chain is described to approach our Master Data Management MDM model based on blockchain.

Keywords*—Covid-19, Blockchain, AI, Big Data, Data Quality, data governance, MDM.

I. INTRODUCTION

Coronavirus 19 or COVID-19 has shaken the whole world, not only in the health system, but also in the economy, education, transport, politics, etc.

Researchers, businessmen and innovators from around the world have invested to find a way to stop the spread of this pandemic by using innovative technologies and putting it at the service of governments and authorities in order to track and contain the outbreak.

Recently, many initiatives to set up tracking applications, dashboards on the state of the spread of the virus worldwide, mobile health self-monitoring application and others, have been able to see the day and demonstrated their effectiveness in this fight. These applications are based on innovative technologies such as blockchain, Artificial Intelligence, Big data ..., etc.

In fact, the blockchain has been deployed in patient tracking applications during the confinement, supply and delivery of the vaccine, etc. AI, on one hand, has been able to help in the fight of this pandemic by identifying the symptoms caused by coronaviruses such as detecting fever b using thermal cameras, the use of robots and drones to monitor patients and support the manufacture of vaccines. On the other hand, Big Data found the perfect fields to reveal itself, via analytics, the collection of data from heterogeneous databases and with different formats and even more to combine this data to have dashboards about to spread of the virus in different parts of the world.

Many articles and research have been published online lately discussing these technologies and their roles in the fight against COVID-19. however, a data quality concern that arises and prevents reporting the total the effectiveness of these applications.

Now with the big data emergence, the volume of data is increasing, data is being produced at an increasing velocity, data types and formats have more variety, and data veracity is becoming more uncertain. This is the case in particular for master Data.

Master data is a basic information, fundamental for the activity of the company, that is shared or duplicated in several systems. This business data must be identifiable and recognized everywhere in the organization, regardless of the service responsible for it, the information system, the server or the software that hosts, processes or records it, division or subsidiary that produces it. Master data typically describes business objects such as "customer", "product", "supplier", "localization", "employee", etc.

The goal of Master Data Management is to create a single version of data that serves as a reference for all the other data of the business. Master data are of great importance and ensuring its integrity can be challenging.

To tackle these challenges, we propose a framework based on blockchains to improve data quality. In this framework, we present a Master Data Management (MDM) solution based on blockchain and AI to ensure a high level of data trust, security and privacy.

The rest of the paper is organized as follows. In section II, we will discuss the benefit of using these technologies which are AI, Big data and blockchain to deal with the COVID 19. Then, we discuss some issues related with data quality that researchers faced in their fight against Coronavirus in section III. We present our framework in section IV and detail the proposed solution for the MDM component of the framework. In section V, we present a use case of an MDM model based on blockchain for healthcare supply chain. Finally, we end with a conclusion and future works.

II. BIG DATA, BLOCKCHAIN AND AI TO DEAL WITH THE CORONAVIRUS

A. The mechanism of Blockchain technology

Blockchain is a technology that redefines trust in the new generation systems. It doesn't need a mediator like corporations and governments, almost always come as central entities that receive, process and store the transactions. There are no mediators who are obliged to process the transactions using correct business logic and have full control on data privacy and security so the trust is decentralized. Users just need to trust the system and the smart code that is shared between all the participants. From technical point of view, Blockchain is a distributed database that exists on a P2P network and every node in the network is on the same level as all the other nodes. Although nodes can come in many forms but there is no central node that is an authority. Every node stores a local copy of the Blockchain. If consensus of nodes agrees upon transaction's validity, then the transaction is considered valid. When a transaction is created, it has to go through validation and confirmation stages before it enters the Blockchain and it is broadcasted to the network. P2P nodes share the transaction between themselves almost in a real time. If valid, the node saves the transaction into its transaction pool. If not, it is immediately removed. Some of the nodes are called "miners" which take all the available transactions from the pool and include them in a new candidate block. a proof of work concept is calculating a random hash value using data of the candidate block. The correct hash value must satisfy a defined difficulty target. This number is calculated using all block's metadata including the hash of the previous block. This is the key to blockchain security. If someone tries to change a transaction from the past, the hash value of the block that contains the transaction must be calculated again. All hash values for the blocks that came afterwards must be calculated again too. This is not feasible, unless more than half of the nodes in a network are infected. Once a new block is created, it is broadcasted to the network. All nodes receive the block, validate it and all the transactions in it. If valid, all nodes put it as the next block in their local blockchain. Transactions that are included in the created block, are then removed from the pool [21].

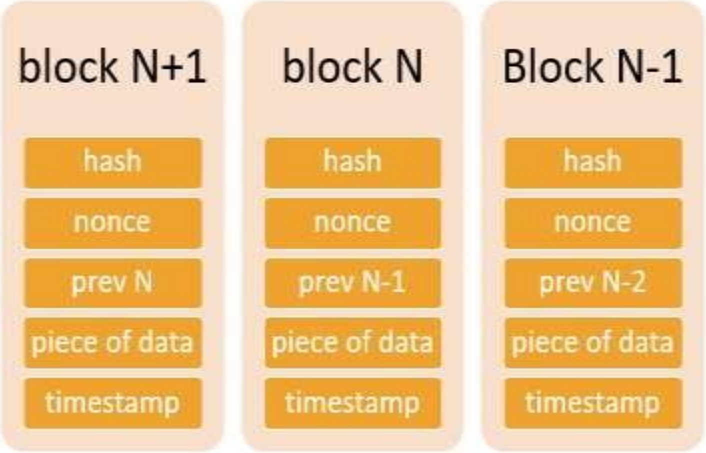

Fig.1: Blockchain proposed structure

B. Blockchain implementations to deal with Covid-19 pandemic

The blockchain technology is a distributed registry that acts as a shared database, keeping all of its copies synchronized and verified. In a recent work [3], the blockchain is formally defined as "of blocks containing messages, proof of work and a reference of the previous block and stored in a shared database, which is capable of carrying out transactions on the P2P network by keeping irreversible historical records and transparency".

In the context of health data management, some researchers have used this technology to reduce the propagation of the COVID-19 pandemic in the World and to guard (be aware) against a future pandemic.

Blockchain technology has the following advantages: the ability of a blockchain's registry to remain unchanged and indelible, to keep authorized users responsible for any transaction and to share data with appropriate authentication without third party intervention. Thus, it was wise to adopt such technology to keep track of health data related to COVID-19 and also to search for and contain infected people while bringing more confidence and integrity to this data that users use and share [4].

Several applications are using blockchain in these pandemic times. Here are some examples from around the world:

- Tracking Infectious Disease Outbreaks:

  The blockchain technology is used to track the surveillance of public health data, especially for epidemics of infectious diseases such as COVID-19. With increased transparency of the blockchain, this will translate into more accurate reporting and effective responses. Blockchain can help quickly develop treatments because they would allow rapid processing of data and therefore early detection of symptoms before their spread to epidemics. In addition, it will allow government agencies to keep track of virus activity, patients, new suspected cases, etc.

- Donation Tracking:

  Blockchain is a solution for Trust in the case of major donation problems. With the help of blockchain capabilities, donors follow the funds and track their donations until they receive verification about their contributions have been received by victims.

- Management Crises

  Blockchains helped by instantly alert the public to the coronavirus by global institutes like the World Health Organization (WHO) and by providing governments with recommendations on how to contain the virus. It provides a platform where all concerned authorities such as governments, health professionals, the media, health organizations, the media and others can keep each other informed of the situation and prevent it from getting worse.

- Securing Medical Supply Chains

  Blockchain has proven effective in managing the supply chain in various industries [8]; similarly, it is beneficial for tracking and tracing medical supply chains. Blockchain-based platforms are useful for examining, recording and tracking the demand, supply and logistics of epidemic prevention equipment without allowing anyone to follow the process.This technology helps streamline medical supply chains, ensuring that doctors and patients have access to the tools when they need them, and preventing contaminated items from reaching stores.

- Fake news

  Blockchains have been used to remedy the problem linked to fake news that is spreading across the globe and which constitutes a new challenge for the media environment, but also for businesses. In fact, quite a few companies have developed blockchain-based solutions to allow companies to guarantee the accuracy of the information they disseminate, such as Wiztrust from a French company, Wiztopi.

  Despite the sophistication of the technologies deployed to fight against the coronavirus, this remains insufficient to have good decisions, because in order for these technologies to achieve their objective for which they are designed for, they need good data quality.

C. Big data against COVID-19

Big Data has been hailed by experts as one of the leading resources in the fight against COVID-19. Commonly, big data is the information asset characterized by such a high volume, velocity and variety that require a specific technology and analytical methods to extract a useful information to serve the decision making.

In COVID 19 context, it refers to the patient data such as physician notes, X-Ray reports, case history, list of doctors and nurses, and information of outbreak areas.

Experts have found uses for Big Data and Big Data analytics platforms for various purposes, such as:

- Containment control and Social Distancing Analysis:

  In addition to strict quarantine measures which kept people in place at the peak of the epidemic, some countries like China have made widespread use of big data to contain the virus.

  The adoption of big data technology consist that citizen must scan a code on his smartphone and enter his information to show where he has been for the past two weeks. The information is added to a database which can be checked to confirm if he has completed quarantine or not.

- Better interconnectivity across national data systems

  In the light of the outbreak of the virus, many countries in the world, tried to intercept their citizens who were traveling and came back. So, they tried to link medical records on the national health insurance database with customs and immigration records to identify and test people who had recently travelled from China, sought medical care, or showed signs of severe respiratory illness [14].

  In these circumstances, the use of Spatial Data Science and location-based data streams is more important than ever. From understanding and predicting the dynamic of the COVID-19 spread over time and space, to empowering our administrations with insights and tools to better plan and respond

  against the dramatic pressure inflicted on our health infrastructure, emergency systems and overall economic system; we now have the opportunity to put forward our technology and efforts to serve this global quest and flatten the curve

- Data visualization to track COVID 19 outbreak

  To track COVID-19's spread in real time, governments, scientific institutions and companies created many "dashboards" for visualization of the disease by making resources available, including funds and the opening of large-volume data repositories, the most frequently used, is that of the John's Hopkins' Center for System Science and Engineering (CSSE)who provides real-time visualization [15].

  In these circumstances, the use of data from heterogeneous sources and location-based data streams is more important than ever and thanks to these dashboards, governments can control, understand and predict the dynamics of distributed COVID-19 in time and space, moreover, it has given authorities perspectives to better plan and respond to the dramatic pressure exerted on health infrastructure, emergency systems and the global economic system

- Big data analytics

  A new trend has appeared, that of computer scientists, biologists' researchers and doctors gathered to collaborate and take advantage of this explosion of data on COVID 19 from hundreds of thousands of medical records from coronavirus patients into effective treatments and predictive analytical tools that could help lessen or end the global pandemic.

  By using the big data analytics on this sources cited before, they could help to implement large-scale COVID-19 investigations, develop comprehensive treatment solutions. This would also help healthcare providers to understand the virus development to find an effective vaccine [16].

D. *AI and its benefits to deal with Coronavirus*

Artificial intelligence (AI) is one of the means or pathways for understanding the virus and developing prevention and control measures. This includes the use of mathematical modeling to understand virus transmission, structural biology to determine the structure of the virus and develop vaccines, computational biology to understand the evolution of the virus, as well as docking studies to screen for drugs and inhibitors [6]

The World Health Organization has declared that the use of AI has been very beneficial in controlling the spread of COVID19.
Indeed, through the use of mathematical modeling, researchers have recently been able to develop algorithms that can be used to predict the underlying dynamics of the evolution of the virus and the immune response to viral changes.

These AI-based models can be developed and trained to analyze huge amounts of data from heterogeneous sources and at incredible speed. Analysis using AI-based methods is more efficient and scalable and supports timely decision-making. As with the coronavirus, AI is likely to play an essential role in the early detection of future epidemics, to stop or limit the spread and save lives. Some of the uses of AI to fight COVID19 include [7]:

- DISEASE SURVEILLANCE AI:

  BlueDot, based in Canada, has used machine learning and natural language processing to track, recognize and report the spread of the virus faster than the World Health Organization and the Center for Disease Control and Prevention (CDC) in the United States. While concerns may exist regarding the potential violation of individuals' civil liberties, the political regulations that other AI applications have faced, will ensure that this technology is used responsibly.

- VIRTUAL HEALTHCARE ASSISTANTS (CHATBOTS)

  Stallion.AI, based in Canada, used its natural language processing capabilities to create a multilingual virtual health worker who can answer COVID-19 questions, provide reliable information and clear directions, recommend actions protective measures, check and monitor symptoms and advise individuals whether they need hospital screening or home self-isolation.

- DIAGNOSTIC AI

  AI improved diagnostic time in the COVID-19 crisis using technology such as that developed by Linking Med, a Beijing-based oncology data platform and a medical data analysis company. There is also an open source AI model that analyzes CT images, identifies lesions, and quantifies in terms of number, volume and proportion. This platform, unprecedented in China, was powered by Paddle Paddle, the open-source deep learning platform from Baidu.

- FACIAL RECOGNITION AND FEVER DETECTOR AI

  Cameras with multisensory AI-based technology have been deployed in airports, hospitals, nursing homes, etc. The technology automatically detects people with fever and tracks their movements, recognizes their face and detects if the person is wearing a face mask.

- INTELLIGENT DRONES & ROBOTS

  The public deployment of drones and robots has been accelerated due to the strict social distancing measures necessary to contain the spread of the virus. To ensure compliance, some drones are used to track individuals who do not use face masks in public, while others are used to disseminate information to a wider audience and also to disinfect public spaces. Patient care, safe for healthcare workers, has also benefited from the fact that robots are used to deliver food and medicine. The role of cleaning and sterilizing isolation rooms has also been fulfilled by robots.

  A team from MIT has developed a machine learning model that uses coronavirus data and a neural network to determine the effectiveness of quarantine measures and predict the spread of the virus.

In the next section, we will present the most encountered data quality issues in this fight against the COVID-19.

III. Data quality challenges

Despite this innovation and the advancement of medical big data, policy makers should approach data with care, as the data in circulation has questionable quality. Indeed, much data is still lacking and the available data may not be exact or reliable and may contain substantial uncertainty, concerning, for example, the precise timing and natural history of the cases [8].

Researchers have found that the big data collected and used in research on COVID 19, presents data quality problems, among others we find:

A. Data privacy

There is growing concern about the way governments are using data to respond to the COVID-19 crisis. With the emergence of new technologies to collect, disseminate and use data to support the fight against COVID-19, the need to ensure that they follow best ethical practices. But while the government’s efforts are directed to slow coronavirus outbreaks, there are also concerns that gathering information about people's geo-location and other personal data to aid management of the pandemic risks infringing on the person privacy more than ever before. In fact, some Governments gave the authorities the right to require telecoms companies to use or access mobile phone location information without user consent to hand over data of people with confirmed infections to track their location. The data has enabled the rapid deployment of a notification system alerting people to the movements of all potentially contagious people in their neighborhoods or buildings. [3][7] and [9]. No one fears "technology for good". But we must not relax the basic privacy requirements, the strategies for maintaining anonymity, encrypting data and preventing our information from ending up in the wrong hands.

B. Data security

At the IT level, data quality and security controls must be ensured. Weaknesses in data integrity, which are common when data from personal digital devices is used, can introduce small errors into one or more factors, which in turn can have a disproportionate effect on large predictive models ladder. In addition, data breaches, insufficient or ineffective anonymization and biases in the data sets can become major causes of distrust of public health services [10] [7].

C. Data Trust

The current COVID-19 pandemic raises important questions about opening, sharing and using data, and highlights the challenges associated with data trust. Without data we cannot understand the pandemic. Only based on good data can we know how the disease is spreading, what impact the pandemic has on the lives of people around the world, and whether the counter measures countries are taking are successful or not.

In one hand, Social media has become a conduit for spreading rumors, deliberate misinformation and wrong data, many perpetrators are deploying sites such as Facebook, Twitter, YouTube, and WhatsApp to create a sense of panic fake news and confusion in such circumstances of coronavirus. The pressing issue is fake news spread more rapidly in social media than the ones from reliable sources and damages the authenticity balance of news ecosystem and eliminating trust in the data and the information [11]. In the other hand, there is Data source providers witch could be sensor nodes or agents that produce permantely a large quantity of data items. These data items describe the properties of certain entities or events for example check of fever in coronavirus with thermometer. Data users are the final information consumers who expect to receive trustworthy data. Due to the possible presence of malicious source providers and inaccurate knowledge generated by intermediate agents, the information provided to the data users could be wrong or misleading [12]. Relying on trustworthy sources is always good advice, but now it is an absolute must. It is a raw material whose reliability is a prerequisite for a precise analysis [13].

In the next section, we present some related works to data governance, then we present our framework for big data governance based on blockchain technology and we present our MDM component based on blockchain solution.

IV. MDM BASED ON BLOCKCHAIN AND AI IN DATA GOVERNANCE TO IMPROVE DATA TRUST

Data quality and data governance are two related domains, but in the same time they are two separate disciplines. Many organizations spend a lot of money in a data quality tool hoping that it will solve their issues with data accuracy and trust. However, organizations need data governance first to create the foundation for enterprise-scale data quality.

A. Related works

Data governance quickly gained popularity and it is now considered an emerging field [19]. It provides a mastery of data management includes other concepts and practices and helps companies to improve and maintain the quality of data and their use [17].

Today, Big Data has brought many challenges to organizations; for example, confidentiality and security in terms of personal information leakage and monitoring of customer privacy [6]. Big Data governance is of crucial importance for each organization because data has become a very important currency.

Soares (2013) defines big data governance in a clear and comprehensive manner as follows: Big data governance is part of a broader information governance program that formulates policy relating to the optimization, privacy, and monetization of big data by aligning the objectives of multiple functions [20].

However, working with Big Data raises new challenges and risks, not only the integrity and data quality threatened, but also working with semi-structured and unstructured data in real time, and how to guarantee secure access to data. it is for these reasons that there is too little attention given to the governance of Big Data.

Unfortunately, most big data technologies do not offer data governance functionality. It is necessary to establish big data governance frameworks in companies for decision-making. Thus, a solid Big Data governance framework is essential for the success of all Big Data projects and the management of this data.

There are a very few studies on regulatory issues and big data governance; most studies focus on Big Data and analysis, the cloud, the Internet of Things, mobility or social media, algorithms and architecture. The data governance framework proposed by A. Al-Badia and other authors in [17] contains five interrelated decision domains such as data principles, data quality, metadata, data access and the data life cycle.

Effective data governance has always played a critical role in master data and MDM systems, as this data follows rules with clear denotations regarding metadata, ownership, authority and quality [23]. Master data management (MDM) primarily revolves around the creation of a trusted source of highly structured data throughout an organization [24]. Redundancies and inaccuracies are common inconsistencies that can appear in systems but thanks to MDM hubs they are largely taken care of and removed from the systems, so ideally there is one version of trusted data [26]. However, and even with Master data management advantages, some researches [29,30] have presented many challenges to master data, such as;

The difficulty to define master data because data definition differs between companies.

Data is often stored in multiple information systems and databases, as data has been developed and evolved in silos over the past decades [16].

Traditional data integration techniques work well for structured data. However, and with the transition to Big Data Incorporation into Organizations [25], Traditional data integration can handle some of the characteristics of Big Data, but it has failed to handle semi-structured data and unstructured data on a large scale, which creates challenges for enterprise data management practices, causing data quality issues that are very common in businesses nowadays. Hence the role of Data Master needs to evolve from a simple collection of the most useful structured data to a tool for leveraging governance standards models for semi-structured and unstructured data [22].

In our research, we think that it is wise to consider Big data governance as potential solution in order to face these problems, and specially, improve the Master Data Management.

Master data helps businesses and their information systems to identify the different components involved in their day-to-day operations. This data must be recorded and processed to keep a record of truth for all business critical processes. Master data management helps to make quick decisions based on objectively accurate information.

B. *Our Framework for big data quality*

The big data governance framework can present a way to improve the quality of data. It is based on timely, reliable, significant and sufficient data, while respecting the rules and processes compatible with Big data. In addition to the quality level of Big data, data governance will improve the strategy for the protection the private information like personal data and disclosure data alongside data security by setting up mechanisms against attacks.

In our research and in order to improve the quality of big data, we opted to set up a Big data governance framework. This framework (Fig.2) will have as a plus the use of data from a database build from blockchains to guarantee data trust.

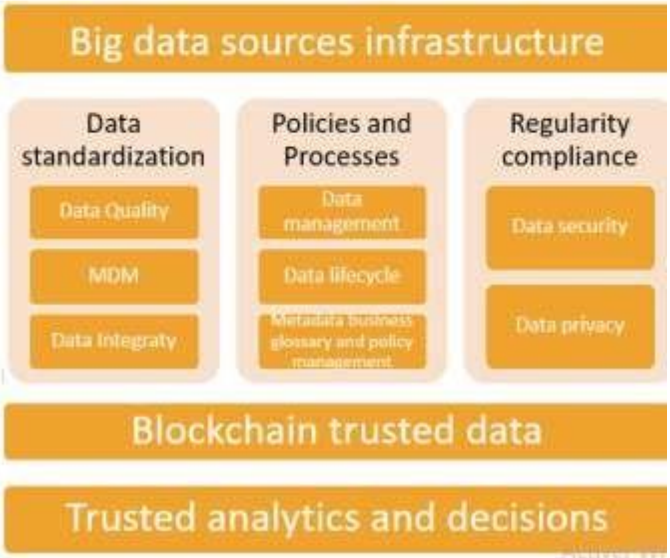

Fig.2: Our framework for Big Data governance

This framwork consists of the following layers:

- Big data sources infrastructure: The exponential growth in the number of data sources will continue to make it difficult to collect reliable information. It includes transaction data, data from social networks, content and machine data. This can lead to uncertainty, such as the origin, quality, source and accuracy of the data. The role of data governance is to mask this complexity and to make managing a complex architecture as simple as managing a single database.
- Data standardization: This layer includes:
  - Data quality: Defining, monitoring and improving data accuracy, completeness and timeliness. It has the ability to parse, standardize, validate and match enterprise data.
  - Master data management: It manages multiple data domains, including customer, product, account, location, reference data and more. IT handles any domain or style and provides the flexibility to define custom domains as required. Also, it helps in managing shared data to reduce redundancy and ensuring better data quality through standardized definition and use of data values. It also plays a prominent role in ensuring that trusted data is accurately maintained.
- Processes and policies: This layer includes:

Data management: is an operational concept focused on the implementation and coordination of policies and procedures. Data managers manage essential data resources, including making data decisions, making recommendations, and developing policies.

- Metadata, business glossary and policy management: define both metadata and governance policies with a common component used by all integration and governance engines. it contains capabilities for data discovery, metadata management, business glossary of terms and definitions, governance policy definition and management and governance project blueprint design.
- Data lifecycle management: it manages the existence of the data from its reception or creation until it makes more sense to keep these massive volumes and therefore

the deletion and archiving of this data from the business system.

- Regularity compliance: These include privacy and security, this involves hiding data in applications to protect sensitive data, monitoring repositories to prevent data breaches, protecting data from external or internal attack and take compliance. It must also be taken into account that appropriate policies and procedures must be followed (created and defined in the Processes and policies component) to prevent the misuse of Big data, taking into account regulatory and legal risks when managing social media, geolocation, biometrics and other forms of personally identifiable information. At this level, we are considering the use of mechanisms based on artificial intelligence and machine learning to allow high levels of cybersecurity and to detect fakes news.
- Trusted blockchain data: this layer represents a data lake that contain the blockchains created from trusted data collected. In fact, each data has been checked in all processes and rules established to realize the big data governance, now it's going to be stored in a
- blockchain. We have to precise that if we store all the data in the blockchain, we will end up with a problem of storage. So we have chosen in our solution, to store a piece of the data, and that piece will indicate the original data when we need it. In addition, an automatized mechanism is going to be in charge of the creation of this blockchains, because there is always a risk of error occurring, as long as the human factor is involved. In fact, the mean goal is that these blockchains will serve as a database, and all the incoming data has to be of high quality to create it and grantee the trustworthiness of the data. All occurring events are registered with accuracy, so the trustworthiness of the stored data will be good.
- Data analysis and decision making: Analytical applications rely on a Big Data platform to process and analyze information. In turn, the analytics engines of the Big Data platform rely on a reliable and certain database in order to return precise and usable results and to integrate this information into other business systems.in our case, the database is build up from blockchains that contain the data processed with the different processes developed in the other components

The Framework may provide visible benefits such as:

- Reinforced security, obtained by locating critical data, identifying owners and users of data, assessing and correcting risks relating to critical data
- Better data quality, allowing better decisions for the company
- Greater operational efficiency, thanks to processes and procedures allowing faster and easier data management
- Reduced data management and storage costs
- A decrease in the number of security breaches, thanks to better training on data resource management

C. Master Data Management:

Master data helps businesses and their information systems to identify the different components involved in their day-to-day operations. This data must be recorded and processed to keep a record of truth for all business critical processes. Master data management helps to make quick decisions based on objectively accurate information.

Master Data Management refers to all the methods, tools, concepts and processes to ensure that master data is correctly identified, of good quality, free of errors and usable without any risk. This set of techniques makes it possible to constitute a single repository, also called a master file. It thus makes it possible to rationalize the sharing of data between the various departments and employees of the company.

Among the techniques that are part of Master Data Management are data cleansing, consistency, elimination of duplicates, consolidation, updating, and the establishment of data descriptions.

D. The proposed MDM component

Master Data Management makes it possible to maintain consistency in the data used to deploy all of the company's actions. Therefore, it must provide reliable and accurate data to users. This is the reason why we use blockchain technology to store MDM (figure 3). Thanks to this we can have a structured form of data to secure it and dispatch it in the network

It's a way to organize MDM in a secure, decentralized and flexible way: no central server, no risk of hacking, corruption or loss of data, and no dependence on a cloud service.

Indeed, the blockchain allows the visibility and traceability of the data stored in its blocks. For example, companies can trace and track documentation and financial transaction. Compared to traditional manual and error-prone approaches, shared information is much more accessible. The digitally extended enterprise can use all the parts, products, suppliers, warehouses, inventory, documentation, tracing and financial transactions stored on the blockchain to function as an efficient and optimized pipeline.

To have a single view, the common MDM stored in the blockchain must be uniform with the other MDMs of participants and therefore the artificial intelligence algorithms will be implemented in such a way as to detect potential attributes of the MDM, to match them with the existing attributes in the MDM. This guarantees efficiency, autonomy and agility. AI will be used to:

clean the data ensure that the necessary data is accurate and complete

Collect additional information related to the master data, thus minimizing the need for manual data entry and, therefore, less inefficiency and inaccuracies.

Ensure that established master data management standards and practices are met and followed to minimize the need to manually perform data governance activities

Enable automated keyword extraction by providing support to avoid costs and simultaneously increase data quality since it can read and extract relevant keywords from the product information supplied and automatically assign them

- Automate data management, i.e. manage and maintain data to make it easily accessible when needed. This automation will save time and resources.
- Match and merge master data to avoid duplicate item master data: With an automated query processing during product onboarding, the data manager can be notified of the existing item and duplicate item master data can

thus be avoided.

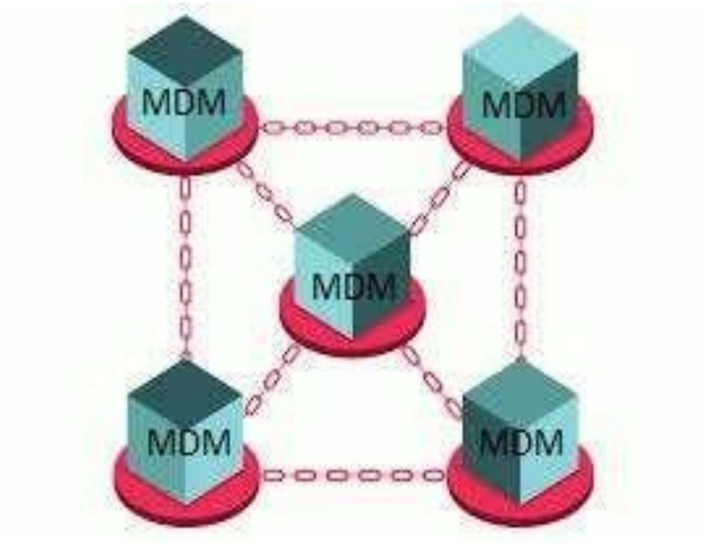

Fig.3: The proposed MDM Model based on Blockchain

In our blockchain, the first block is called the genesis block. Each block has its hash as a unique ID that includes the hash of the previous block (fig. 4). In this way, a chronological chain is formed. Usually, a block stores a set of timestamped transactions that are validated by stakeholders in the network. Once it gains consensus, the block is accepted and stored by all parties in the blockchain and can no longer be modified. Therefore, trust in and transparency of transactions between the stakeholders of the network are significantly improved.

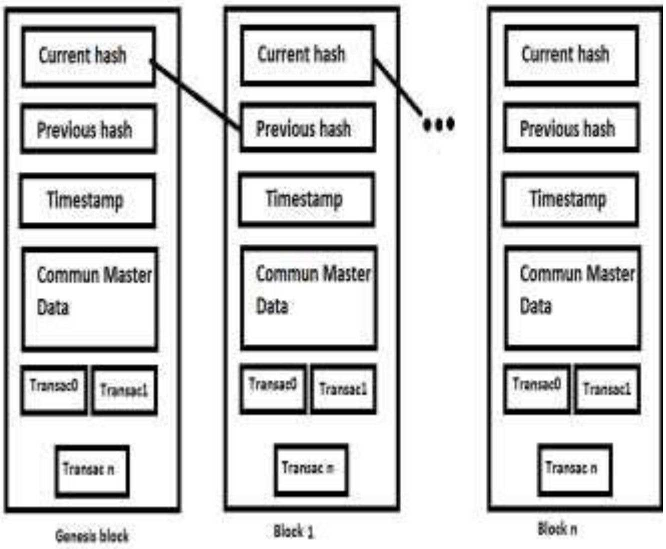

Fig.4: Our bockchain components [updated from [32]]

In the next section, a use case of vaccine supply chain based on blockchain in described with a focus on our conceptual MDM component.

V. Vaccine supply chain Use case

Traditional Supply Chain Management mechanisms usually suffer from a wide range of issues such as lack of information sharing, long data retrieval times and unreliable product tracing. This is the case for vaccines and especially for covid- 19 vaccines where tracing is of high importance.

A blockchain-based vaccine supply chain depends on a blockchain-driven network of trust. The required entities or nodes of the blockchain network are the manufacturer, distributors, transport agencies and hospitals (Fig. 5). A supply chain must be created including all stakeholders such as suppliers, vendors, distributors and healthcare provider.

The model can be described as follows: The manufacturer sends the vaccine to the wholesaler using a transport agency according to their requirement. The wholesaler sends vaccine to hospitals and clinics again through the transport agencies. If the hospital or clinic's requirement is very large, it can directly order from the manufacturer through transport agency, or it can collect vaccine from the wholesaler directly.

The main problem with existing supply chains is that each stakeholder owns his own Master data, since manufacturer deploys its own separate solution and has its own MDM. Participants along the chain who wish to work with that manufacturer must integrate their platforms to the manufacturer's one. However, two manufacturers can use different MDM, meaning that each distributor and healthcare providers need to deal with different Masters data and it is the same with every new agent they work with, increasing the overall complexity of the entire system.

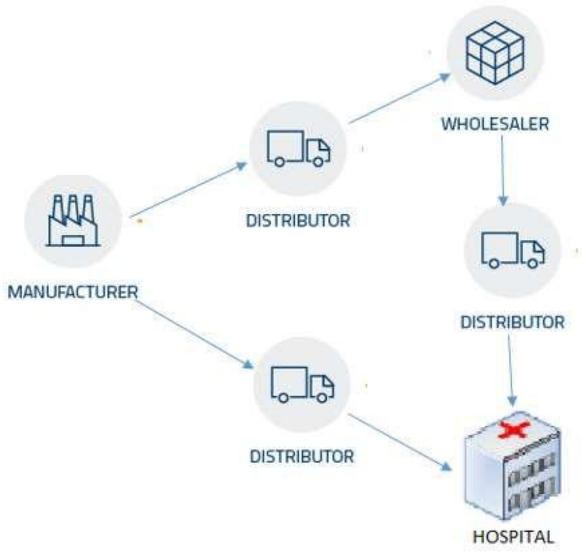

Fig.5: Example of a Supply chain schema

A. A conceptual MDM component based on blockchain

To deal with this issue, we propose to create a common Master data for all the participants in the healthcare supply chain (fig. 6). In fact, our solution is based on storing a uniform Master data in the healthcare supply chain blockchain. Hence, every node participating in the MDM blockchain will need a complete copy of the common Master Data to start with. Once a participating node receives a new block in the blockchain, it will take the new transactions in the block and simply use the functions either create/update/delete its own copy of the Master Data.

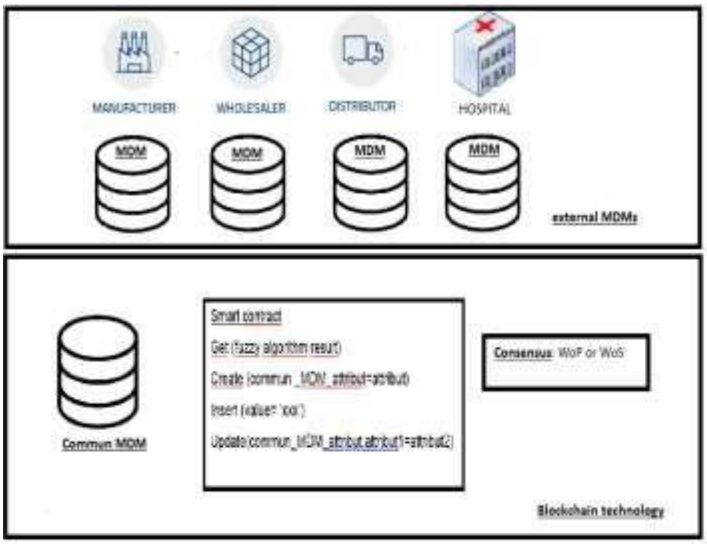

Fig.6: The common MDM Model based on Blockchain

The main consideration in creating and maintaining common master data is the mapping and merging of master data that has been created by different stakeholders. A common MDM system will include functionality to automatically merge similar masters data as much as possible on the basis of consensus. In addition, a common MDM system will provide functionality to determine the best possible master data.

For the Master data of Covid 19 vaccine, each participant to the blockchain network has its own Master Data: MD1, MD2….,MDn. If in each Master Data, the attribute vaccine_name is written in different ways, by using Fuzzy algorithm for matching, we can detect the master record which will be hashed and its hash will be stored in the common Master Data in the blockchain. Figure 7 shows the process of matching records between Masters Data to select the common Master Data.

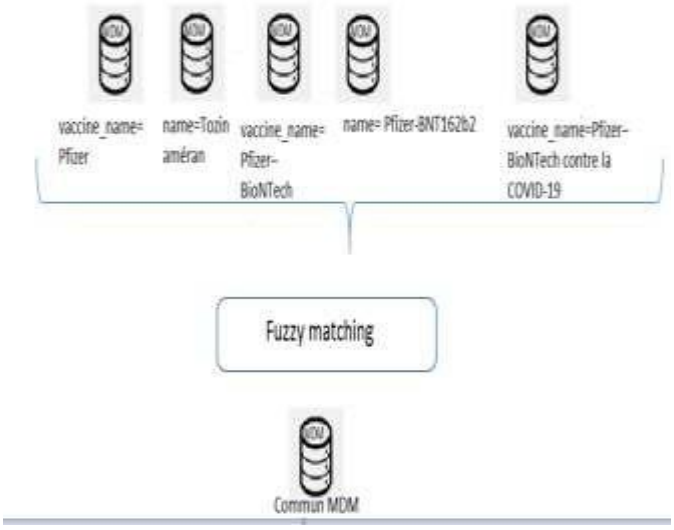

Fig.7: example of AI algorithm to update the commun MDM

If the data already exist, no modification is done in the common MDM, otherwise, the Master Data with the highest score will be taken into consideration and will become the common Master Data.

In the following part, we present an example of the supply chain of a vaccine with the various processing based on blockchain technology and with our MDM component.

An example of vaccine supply chain blockchain solution

Throughout the supply chain, the use of blockchain technology has enabled many advantages such as: secure sharing of information, facilitates product quality and operations monitoring, data acquisition in real time, transparency and visibility. Vaccines belong to a production lot and therefore must have a lot number and be labeled with serial numbers. Vaccines packaged for transport and packaging should also contain the serial numbers of the medicines inside. The entire path of the drug must be made visible to all entities of the blockchain. Manufacturing inputs such as chemical ingredients and other parameters can be updated and linked to the serial number of the products which is the id of the vaccine [31] (Fig.8).

Figure 8 presents the vaccine supply chain diagram. In fact, all participants registered in the blockchain must have a unique identifier which will be the combination of a private key and a public key. When transferring from manufacturers to distributor, both parties must digitally sign using their private keys in the distributed ledger, and the transaction is added to the block. This transaction stored in the blockchain will also store other information such as: the order number, the date of dispatch, the low code of the package and the temperature. Other entities on the blockchain should check the validity of the transaction before adding another transaction so that no one can deny or tamper with that transaction in the future.

The main purpose of the distributor is to distribute the vaccine in accordance with the requirements of Hospitals and clinics through the transport agencies. Once the distributor receives the vaccine from the manufacturer, all parts of the blockchain will know that so this transaction will be stored in the block with the following information: date of receipt, package barcode and temperature. The transport agency will then transport the vaccine from the distributor to hospitals and clinics and it is another transaction that will be added to the other information (fig. 8)

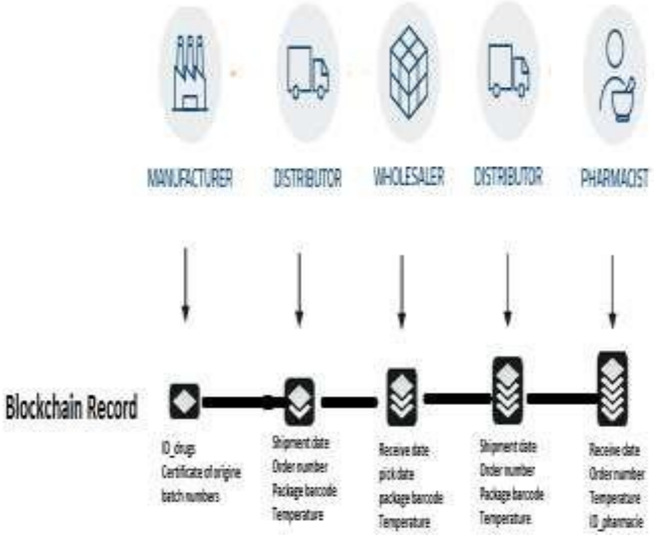

Fig.8: example of a vaccine supply chain based on Blockchain

VI. CONCLUSION AND FUTURE WORKS

As the world increases in its fast adoption of new technologies in the fight against the COVID-19, the use of Big Data cannot replace the development of efficient health infrastructures and the development of strict systems and protocols for examination and monitoring. However, AI and blockchains based applications have been used with success in some countries. Nevertheless, remains the biggest challenge of ensuring security, privacy and trust in data. In the first part of this article we presented a set of different perspectives on key design elements, challenges, opportunities, and best practices for big data, AI and blockchain technologies.

While the proposed framework is not exhaustive, it does provide a basis to improve data quality in the field of Big Data as a starting point in this framework. We are focusing to implement the blockchain and AI in the MDM component,

since the need of master data management will remain challenging in most companies.

In the second part of the paper, we presented how blockchain can be a solution to some MDM challenges and prevent the issues in a transparent and secure manner so that we can have a single trusted view. In the proposed use case, we showcase how blockchain can be used also to add traceability and visibility to the drug supply chain to prevent the issues of drug counterfeiting and especially for vaccines. A model of decentralized blockchain architecture has also been presented, which will also make the drug supply more robust, transparent and trustworthy.

As future work, we will work on integration and implementation of more AI methods in the blockchain to enhance the MDM component since AI and blockchain are now permanent building blocks to consider for businesses and organizations that want to successfully integrate the latest master data technologies to have better compliance conditions and a smoother order of operations.

References

[1] R. A. Addi, A. Benksim, M. Amine, M. Cherkaoui, ” COVID-19 Outbreak and Perspective in Morocco” Electron J Gen Med. 2020;17(4):em204.

[2] D. C. Nguyen, M. Ding, P. N. Pathirana,and A. Seneviratne,, “Blockchain and AI-based Solutions to Combat Coronavirus (COVID- 19)-like Epidemics: A Survey,” preprint.

[3] A. Azim, M. N. Islam and P. E. Spranger, “Blockchain and novel coronavirus: Towards preventing COVID-19 and future pandemics,” published.

[4] T. P. Mashamba-Thompson and E. D. Crayton,” Blockchain and Artificial Intelligence Technology for Novel Coronavirus Disease 2019 Self-Testing”, published.

[5] H. M. Yassineaand Z. Shah, How could artificial intelligence aid in the fight against coronavirus?, EXPERT REVIEW OF ANTI- INFECTIVE THERAPY2020, VOL. 18, NO. 6, 493–49.

[6] Z. Allam and D. S. Jones , “On the Coronavirus (COVID-19) Outbreak and the Smart City Network: Universal Data Sharing Standards Coupled with Artificial Intelligence (AI) to Benefit Urban Health Monitoring and Management ,” published

[7] B.Tang, N. L. Bragazz, Q. Li, S. Tang, Y. Xiao and J. Wu “An updated estimation of the risk of transmission of the novel coronavirus (2019- nCov)”Infectious Disease Modelling,Volume 5, 2020, Pages 248-255

[8] S. Saberi, M. Kouhizadeh, J. Sarkis and L. Shen, ‘Blockchain technology and its relationships to sustainable supply chain management, InternationalJournal of Production Research(2019), 57:7, 2117-2135,

[9] M. Ienca and E. Vayena,”On the responsible use of digital data to tackle the COVID-19 pandemic”, Department of Health Sciences & Technology, Swiss Federal Institute of Technology in Zurich, Zurich, Switzerland.

[10] S. Tasnim, M. M. Hossain and H. Mazumder, “Impact of Rumors and Misinformation on COVID-19 in Social Media “ J Prev Med Public Health. 2020;53 (3): 171-174

[11] C. Dai, D. Lin, E. Bertino and M.t Kantarcioglu, « An Approach to Evaluate Data Trustworthiness Based on Data Provenance”,Secure Data Management, 2008, Volume 5159

[12] C. M. Pulido, B. Villarejo-Carballido, G. Redondo-Sama and A. Gómez,« COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information” ,International Sociology2020, Vol. 35(4) 377 –392Q. Pham, D. C. Nguyen, T. Huynh-The, W. Hwang, and P. N. Pathirana, “Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts,” IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE, April 2020,Preprint

[13] D. Buhalisa and R. Leungb, “Smart hospitality—Interconnectivity and interoperability towards an ecosystem”,International Journal of Hospitality Management Volume 71, April 2018, Pages 41-50

[14] N. Naudé, “Artificial intelligence vs COVID-19: limitations, constraints and pitfalls”, AI & Soc (2020), published.

[15] C. J. Wang, C. Y. Ng, R. H. Brook, “Response to COVID-19 in TaiwanBig Data Analytics, New Technology, and Proactive Testing“,JAMA. 2020;

[16] A. Al-Badia, A. Tarhinia, A. Islam Khan, Exploring Big Data Governance Frameworks, Procedia Computer Science 141 (2018),

[17] Pages 271–277

[18] Alhassan, I., Sammon, D. and Daly, M., Data governance activities: an analysis of the literature, Journal of Decision Systems, 2016

[19] J. Hagmann, Information governance–beyond the buzz, Records Management Journal, vol. 23 (3) 2013, pp. 228-240.

[20] V. Morabito, Big Data Governance. Big Data and Analytics 2015,Pages 83–104.

[21] J. Wang, M. Li, Y. He, H. Li, K Xiao, & C. Wang, A Blockchain Based Privacy-Preserving Incentive Mechanism in Crowdsensing Applications. IEEE Access ,2018,, 6, 17545–17556.

[22] T. K. Das, M. R. Mishra. A Study on Challenges and Opportunities in Master Data Management. International Journal of Database Management Systems · May 2011

[23] R. Silvola, O. Jaaskelainen, H. Kropsu-Vehkapera, H. Haapasalo. Managing one master data –challenges and preconditions. Industrial Management & Data Systems, Vol. 111 Iss 1 pp.146 – 162. 2011

[24] P. Rishartati A. Adetia, N. D. Rahayuningtyas, Y. Ruldeviyani, J. Maulina. Maturity Assessment and Strategy to Improve Master Data Management of Geospatial Data Case Study: Statistics Indonesia. 5th International Conference on Science and Technology (ICST),

[25] Yogyakarta, Indonesia. 2019

[26] S. T. Ng, F. J. Xu, Y.Yang, M. Lu. A master data management solution to unlock the value of big infrastructure data for smart, sustainable and resilient city planning.Creative Construction Conference 2017, CCC 2017, 19-22 June 2017, Primosten, Croatia

[27] Banerjee, Arnab (2018). [Advances in Computers] || Blockchain Technology: Supply Chain Insights from ERP.

[28] Wu, H., Cao, J., Yang, Y., Tung, C. L., Jiang, S., Tang, B., … Deng,

[29] Y. (2019). Data Management in Supply Chain Using

Blockchain: Challenges and a Case Study. 2019 28th International Conference on Computer Communication and Networks (ICCCN).

[30] Silvola, Risto; Jaaskelainen, Olli; Kropsu-Vehkapera, Hanna; Haapasalo, Harri Managing one master data – challenges and preconditions. Industrial Management & Data Systems, (2011). 111(1), 146–162.

[31] Vilminko-Heikkinen, Riikka; Pekkola, Samuli . Establishing an Organization's Master Data Management Function: A Stepwise Approach. , [IEEE 2013 46th Hawaii International Conference on System Sciences (HICSS) - Wailea, HI, USA (2013.01.7-2013.01.10)] 2013 46th Hawaii International Conference on System Sciences - 4719–4728.

[32] G. Knolmayer, and M. Rothlin ,“Quality of material master data and its effect on theusefulness of distributed ERP systems”, (2006), Lecture Notes in Computer Science, Vol. 4231,pp. 362-71.

[33] M. Sahoo, S. S. Singhar and S. S. Sahoo. A Blockchain Based Model to EliminateDrug Counterfeiting. Springer, (2020)

[34] Thomas, Ciza; Fraga-Lamas, Paula; M. FernÃ¡ndez-CaramÃ©s, Tiago (2020). Computer Security Threats || Deploying Blockchain Technology in the Supply Chain. , 10.5772/intechopen.83233(Chapter 5)