

THE DESIGN AND DEVELOPMENT OF A BLOCKCHAIN-BASED EPRO SYSTEM FOR COLLECTING CLINICAL DATA

KWANGSOO JANG¹, OOK LEE*

¹ Department of Information Systems, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, 04763
Seoul, South Korea.

* Department of Information Systems, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, 04763
Seoul, South Korea.

E-mail: ljks8605@hanyang.ac.kr, ooklee@hanyang.ac.kr

ABSTRACT

Objectives: The present study discussed the design and development of an ePRO system that records and manages clinical trial data on a blockchain network so as to prevent any arbitrary modification, forgery, or falsification and improve the reliability of the data. **Methods:** The overall system design and development of ePRO consists of (1) advancement and development of blockchain technology, (2) development of a blockchain clinical trial system model, (3) development of a clinical trial data collecting tool using blockchain, and (4) technological verification of the blockchain-based ePRO system. **Results:** The results of this study suggest a potential to save human resources and costs for CDM operation by increasing reliability through the implementation of an on-chain clinical trial system. Additionally, the on-chain clinical trial system is predicted to help prevent harm from serious abnormal reactions, including patients' death, subsequent side effects, and birth defects caused by clinical trial errors. **Conclusions:** The ePRO system can serve as a platform for medical and clinical services to cope with expanding clinical data fields. When the system developed in this study is combined in the near future with state-of-the-art technologies such as IoT and AI, a new paradigm for clinical data repositories and management systems is predicted to develop. Based on the clinical data collected by this ePRO system, when a clinical data warehouse is established and linked to medical information systems, a servitization model that can accelerate innovation in the medical service market is expected to emerge.

Keywords: *Blockchain; Clinical Trials as Topic; Electronic Health Records; Mobile Applications; Pilot Projects*

1. INTRODUCTION

1.1 Background and Necessity of the Study

As of 2018, the global market for clinical trials reached USD 36.38 billion. With an anticipated average annual growth rate of 8.76 percent, it is projected to grow as large as USD 65.39 billion by 2025 [1]. The South Korean clinical trial market has received favorable evaluations at the international level, being described as a next-generation frontrunner for clinical trials. In 2014, South Korea stood in seventh place in the global market for clinical trials, a remarkable leap from 19th place in 2007 [2]. Critically, the quantitative increase and diversification of clinical trials have highlighted the importance of the monitoring, management, storage, and sharing of clinical data [3]. Clinical data consisting of diagnosis data, treatment data,

prescription data, and lab data is directly related to patients' health. With clinical data, any omission, forgery, falsification, or error may lead to erroneous prescription or treatment, thereby producing serious consequences. In addition, any qualitative error that may occur in clinical trial data may lead to errors that seriously affect the safety of patients or the interpretation of study results [4]. Furthermore, there have been cases of revocation after a medicine entered the market due to serious side effects, including death, birth defects, and cardiovascular side effects, brought on by inappropriate management and interpretation of clinical trial data and clinical results [5,6].

The reliability of clinical data requires proper clinical data management (CDM). CDM refers to the process of managing and collecting clinical data in accordance with clinical regulatory standards.

The major objective of the CDM process is to collect data in order to allow the provision and interpretation of high-quality data while minimizing the possibility of error or omission [7]. Recently, an argument has been raised that blockchain technology, characterized by its data inalterability and assurance, may proactively prevent the possibility of deliberate omission, forgery, or falsification of clinical data. For instance, Wong et al. [8] has developed a system that can ensure constancy in clinical trial data by means of blockchain technology. This system allows the tracking of clinical data, thus enhancing confidence and the ease of data management by regulatory institutions. Global pharmaceutical companies are known to have taken notice of the effectiveness of blockchain technology for CDM. The Pistoia Alliance, a nonprofit organization established in 2007 by pharmaceutical firms including Pfizer in the U.S. and GlaxoSmithKline in the U.K., decided to employ blockchain technology as a means to verify sources of clinical data and ensure their integrity [9]. Blockchain technology does not manage clinical data on the servers of hospitals and medical institutions, but rather saves the data on a distributed ledger shared by interested parties [10]. In addition, individual keys that allow data identification and confirmation are generated separately, meaning the ownership of patients' health information can be clearly imparted to the entity who generated it. Consequently, blockchain technology is expected to facilitate a patient-oriented Personal Health Record (PHR) healthcare ecosystem.

1.2 Purpose of the Study

The purpose of this study is to discuss the design and development of an ePRO system that records and manages clinical trial data on a blockchain network so as to prevent any arbitrary modification, forgery, or falsification and improve the reliability of the data. Current clinical data systems (CDS) were developed to be used by accessing a related database. Clinical data is centralized, and the actual ownership of the data belongs to the owner of the server in which the database is stored. Conventionally, the data has not been managed, tracked or monitored by a proper system but rather by hand, which incurs high costs. The key issue is that when a CDS manager accesses the database and modifies clinical data, the modified data is still recognized as original data, which allows forgery or falsification of clinical data. The ePRO system suggested in this study uses blockchain technology to prevent any modification of clinical data or

falsification by an exterior entity. Any modification of clinical data, even partial, invariably alters the value of the consecutive hash chain of the blockchain, making it impossible to invisibly change the data.

1.3 Literature Review

1.3.1 Industrial Trends in Clinical Trials

The aging of the global population has dramatically increased the demand for care for patients with geriatric diseases, including dementia and Parkinson's disease. Chronic diseases such as hypertension, diabetes mellitus, and arthritis are anticipated to eventually make up more than 70 percent of the disease suffered by the world population [11]. New clinical treatments and the development of new pharmaceuticals represent a high value-added business and a major future growth engine. Clinical trials in particular account for 65 percent of the cost of new medicine development, determine more than 70 percent of success probability, and occupy the essential position in the development of new pharmaceuticals [12].

According to ClinicalTrials.gov, a database for registering clinical trial operated by the National Institutes of Health (NIH) in the U.S., the cumulative number of clinical trials registered worldwide between 2000 and early 2019 totaled 299,634. The statistics published in 2018 by the Korea National Enterprise for Clinical Trials indicate that the total number of approved clinical trials for medicines increased by 21 cases from 658 in 2017 to 679 in 2018 (3.2 percent), and by 30 cases from 628 in 2016 to 658 in 2017 (4.8 percent), showing how the number of approved clinical trials for medicines has continued to increase over the last three years.

Clinical trials are a high-value-added knowledge-based business that create considerable wealth in the industry. According to a 2018 announcement by the Korea National Enterprise for Clinical Trials [13], it was forecast that an investment of KRW 400 billion in clinical trials would bring as much as KRW 18 trillion in socio-economic profit over a decade. Emerged as a future core industry and new growth engine, clinical trials have developed into a strategic industry through the expansion of the related industrial scale and continuous increases in R&D investment.

1.3.2 Blockchain Technology

Blockchain has emerged as a technology that allows the comprehensive management of medical data and its use by third parties other than patients [14]. Since Satoshi Nakamoto [15] first introduced the Bitcoin blockchain that employs hashing, public-key cryptography, and P2P and ledger technology, a number of blockchains have appeared. Within a blockchain, a participant makes a digital signature on data including information, messages, and transactions, before delivering the data to nodes connected via a P2P network. The delivered data is approved by a majority of nodes using a Proof-of-Work (PoW) consensus mechanism, and the approved data is stored on a block. Blocks include hash values of input data that are displayed in a Merkle tree structure [16]. The blocks generated onwards include information on the previous block, meaning newly generated blocks are connected in chains, thus constituting a blockchain [17].

In the case of medical data, it is of great importance to verify the identity of the data generator and sender, as well as whether there have been any arbitrary modifications (changes) in the original data – in other words, to confirm the integrity, identification, and authenticity of the data [18]. Blockchain can ensure the integrity and authenticity of data, so it can play an essential role in the comprehensive management and use of medical data. Blockchain uses public-key cryptography and hash functions to enable a data sender to create a digital signature with a private key before providing the public key to a receiver so that the receiver can use the public key to verify the identity of the data sender and the data's integrity.

2. METHODS

This study designed and developed an ePRO system as a clinical trial data collection system operated on the basis of blockchain. The overall system design and development largely consists of (1) advancement and development of blockchain technology; (2) development of a blockchain clinical trial system model; (3) development of a clinical trial data collecting tool using blockchain; and (4) technological verification of the blockchain-based ePRO system.

It was discovered that it was impossible to write entire clinical trial data on Bitcoin blockchain (seven transactions per second (TPS)) or Ethereum blockchain (15 TPS). Consequently, this study advanced blockchain technology by developing an algorithm called Proof-of Formulation (PoF) as a

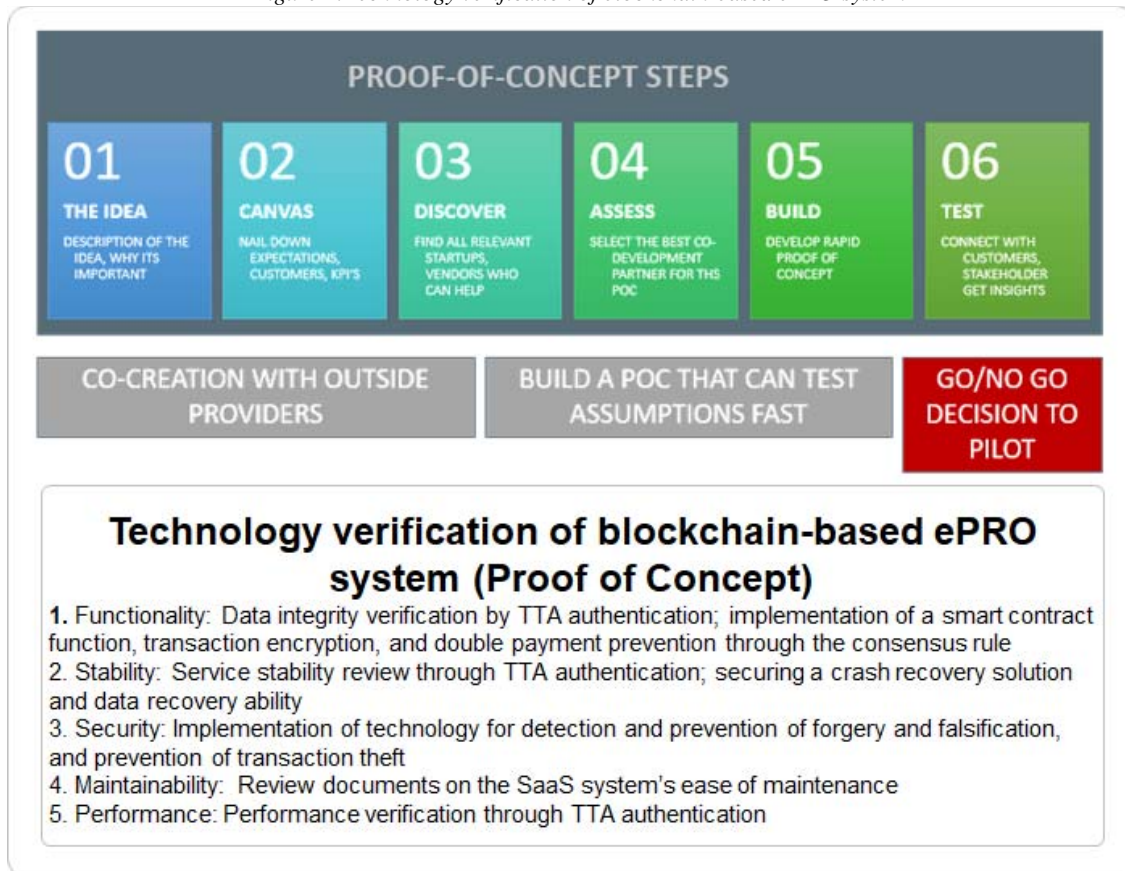
new consensus structure and then developed a mainnet with PoF and applied it to the clinical trial system.

This study proposes to implement a service-based system (SaaS: Software as a Service), for which we standardized a clinical trial system and developed a SaaS-type standard data model so that researchers can easily establish and operate a system on their own. We applied this standard data model in order to develop Excel input forms and rules. Accordingly, researchers can easily input the contents of a CRF (case report form) into Excel, through which the on-chain clinical trial system was developed without a web programming process. In addition, we sampled existing CRFs to test the clinical trial system and created two Excel files. These were used to determine whether the clinical trial system could in practice be effectively operated.

After completing the aforementioned processes, we developed a block-explorer interface through which SaaS type service provision screens, admin data verification and editing screens, and subject diary screens can be provided. Researchers and monitoring personnel are able to directly create audit trails of on-chain clinical trial data.

Lastly, for technological verification of the system, a clinical trial led by a researcher was conducted on 50 subjects with diabetes mellitus, hypertension, and dyslipidemia at Seoul St. Mary's Hospital. In addition, the Telecommunications Technology Association (TTA) was asked to verify the blockchain technology applied to the system (Figure 1).

Figure 1: Technology verification of blockchain-based ePRO system



2.1 System Design

Here, we aim to discuss the on-chain based system architecture along with PoF (Proof-of Formulation), which is the major algorithm for the ePRO system.

2.1.1 Proof-of Formulation

In order to overcome the inefficiency and low performance in existing blockchain, we developed PoF, a new consensus algorithm, and applied it to this system. The PoF consensus is designed to identify blocks and transactions in real-time and prevent forks. As a single consensus mechanism, this structure can overcome the disadvantages of existing consensus structures such as the energy consumption of PoW (Proof-of Work), security failures of PoS (Proof-of Stake), and centralization of DPoS (Delegated Proof-of Stake). Mining and block production by the PoF algorithm designed for this study are carried out differently from existing blockchain platforms. On the PoF blockchain, a formulator node serves as a block generator. The observer nodes check a created block in real-time to

prevent double-spending. Formulators play a pivotal role in the PoF algorithm. The priority order of block production is calculated based on scores according to the following formula:

$$\text{Score: uint64(Phase)} \ll 32 + \text{uint64(binary.LittleEndian.Uint32(hash[:4]))}$$

The PoF algorithm does not require extraordinary computational performance like the PoW algorithm, and has a structure to elect a formulator acting as a block generator in an unpredictable order which differs from the DPoS algorithm in which only elected delegators can participate in mining. Since a formulator node is designed to take part in mining only once per step according to the designated priority, the PoF has a higher level of distributed structure than does DPoS. The ranking is randomly set with no particular rules and the formula is changed at every step, making it impossible to predict the ranking. Through this structure, errors displayed on the platform according to the PoS system can be prevented. Also, mining is carried out at high speed for four seconds with a short

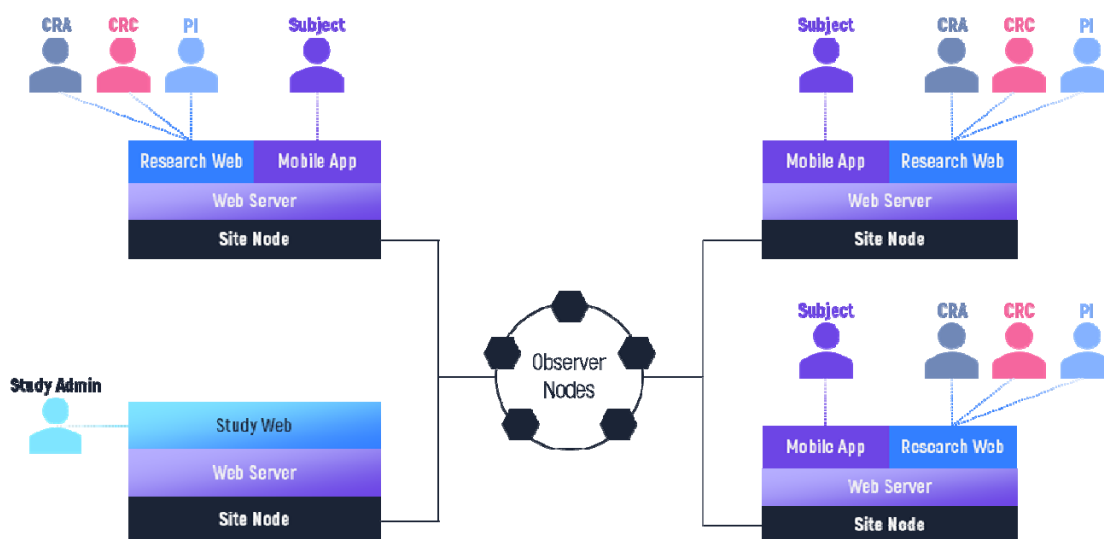
block time of only 0.5 seconds, and the observer nodes immediately check and confirm created blocks. In this manner, blocks can be rapidly produced. It is designed to quickly distribute blocks after validity verification by at least three among the five observer nodes. The PoF algorithm was designed as a new consensus algorithm that not only maintains heightened performance in the chains but also improves security. The blockchain algorithm developed in this study can implement an

on-chain clinical trial system through which clinical trial data is written on the chains.

2.1.2 On-chain Clinical Trial System Architecture

Figure 2 shows the flow of information between a clinical trial process and the participants on blockchain within the ePRO system developed in this study.

Figure 2 Flow diagram of the clinical trial process and stakeholder transactions



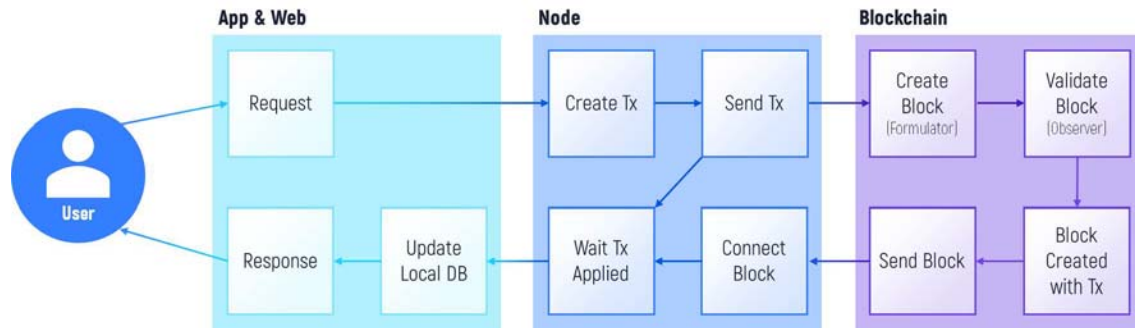
When blockchain is operating under consensus through the PoF algorithm, the system encrypts and decodes data for use only in the approved nodes. Various processes, such as Study, User, Subject, Visit, and Query, which serve to change the blockchain data for performing clinical trial data management functions in blockchain, define and manage transactions. Furthermore, by transmitting the processed results to the eCRF Web Provider Server, it composes data sets for offering web services. By adding an authentication module taking charge of user authentication, including in the login process, only authenticated users are allowed to access and use the system, from the function of the blockchain to the web service, which is the users' endpoint. The PoF algorithm consists of five observer nodes to verify blocks and the formulators to produce blocks. The Study Node, which clinical trial organizers (for example, sponsors of pharmaceutical companies or researchers, or contract research organizations (CROs)) can directly operate and add institutions to and manage, and the Site Node, which is installed at

individual participating institutions (for example, hospitals), take part in the blockchain operation as formulators. Figure 3 shows the usage cycle from data transmission, transaction issuance, and block production to issuance of extracted transactions and data extraction on blockchain within the ePRO system.

A typical web system is made up of analysis of requests from users and data composition; database command execution and results composition; and the sending of results. The web system combined with the blockchain developed for this system operates differently, however. For instance, when a user request is input, the system analyzes the information, produces a transaction, signs, sends the information to blockchain, and stands by for the inputting of transaction results. If the relevant transaction is verified and reflected in the chain, and the reflected block is transmitted, the system delivers the results to a location waiting for transaction results. Through this process, it updates the local database for web service and delivers results to the user.

Figure 3A flow diagram of a transaction through app and web, node, and blockchain in the ePRO system

a clinical trial. For this work to be performed, after



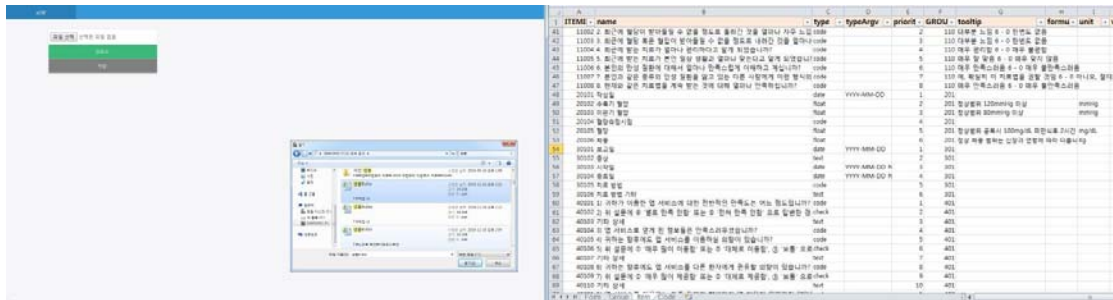
3. RESULTS

This system is characterized by a SaaS-type structure. Researchers may access the entry screen and create screens on which they can easily design

login to the admin page “Study (clinical trial),” a researcher, without the need for web coding, inputs a form for a clinical trial that is pre-set into Excel and uploads this Excel file in the admin page to create clinical trial pages (Figure 4).

Figure 4 Examples of Excel form upload screens and definition documents according to clinical trial rules

The design of a clinical trial process must be the admin modules. It is possible to identify, modify,



carried out based on fixed forms and rules. The method provided by the ePRO system developed for this study allows researchers to create an eCRF (electronic case report form) without any special knowledge of programming. The created eCRFs are all operated as an on-chain based clinical trial system using blockchain. Figure 5 refers to an eCRF system admin page. The eCRF created by the SaaS-type eCRF generator can be managed through

and delete subjects’ data , but all data change history remains on the blockchain to allow an audit trail (history management). The relevant detailed screens can be confirmed through block explore.

Figure 5: The eCRF admin page screen:

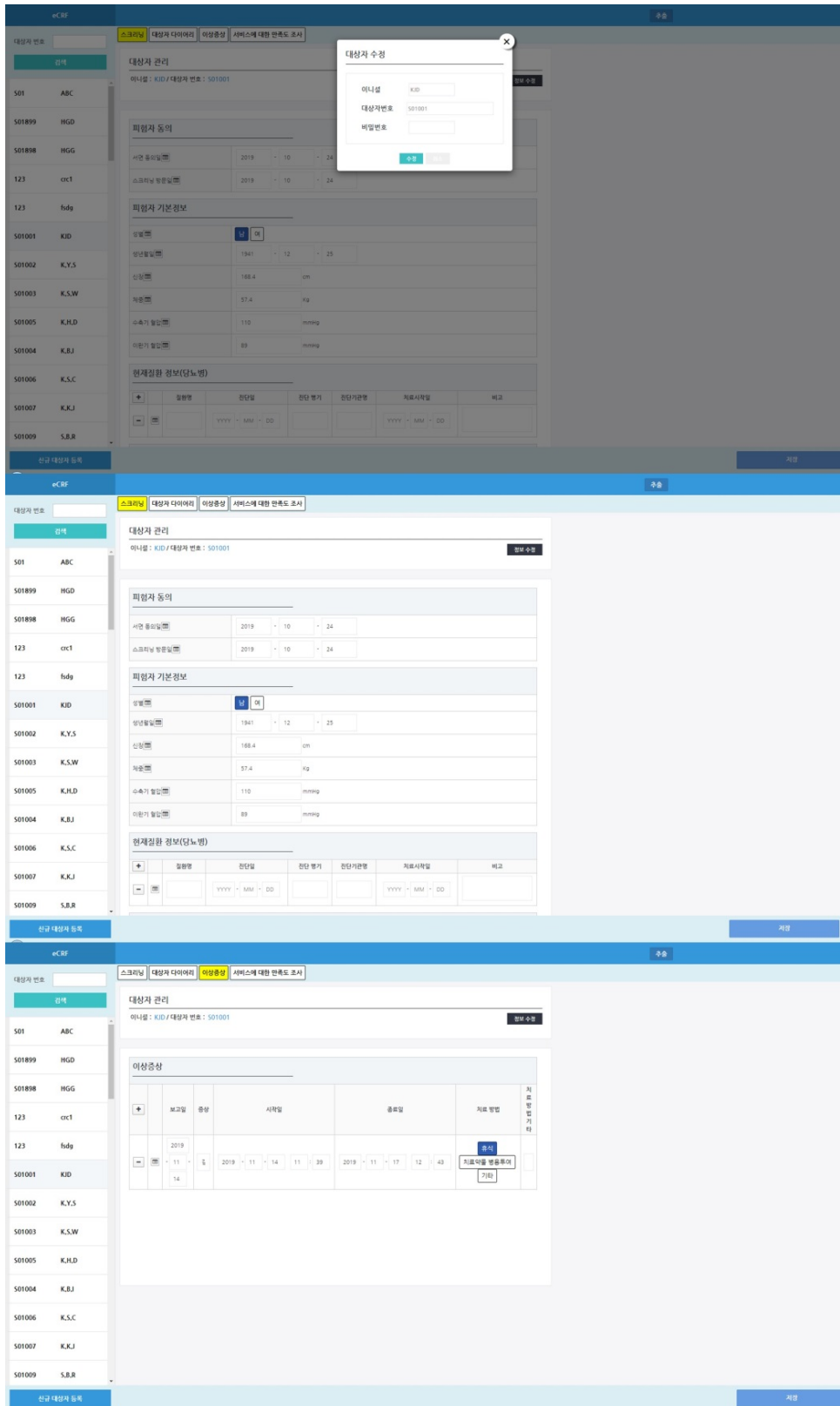


Figure 5 The eCRF admin page screen
 Figure 6 The mobile input tool for subjects

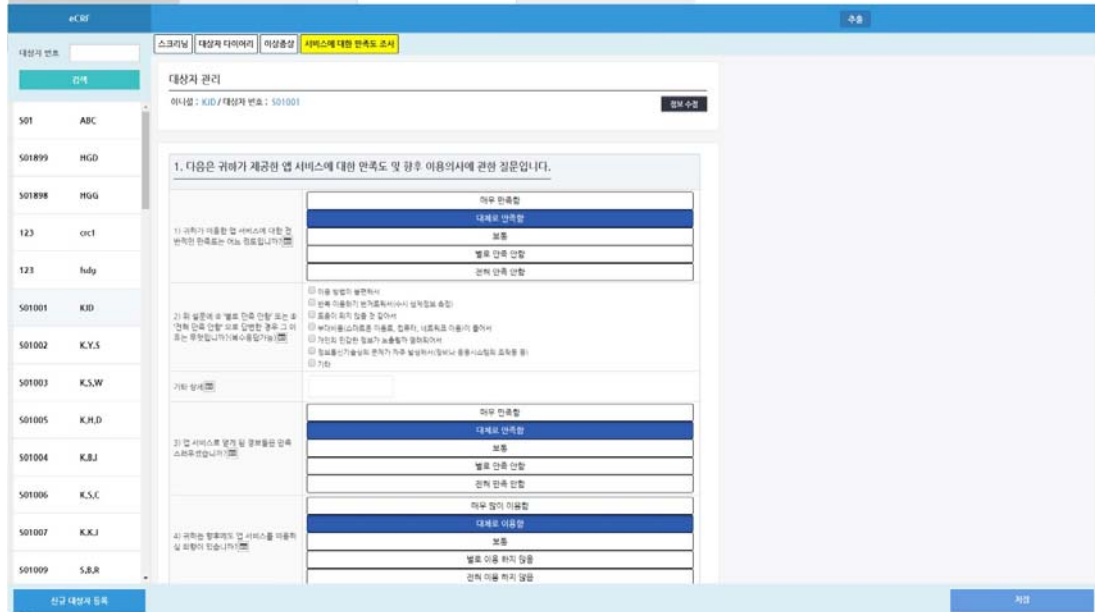


Figure 6 presents the mobile input tool screens provided to the clinical trial subjects. The mobile input tool for subjects is also a SaaS-type generator and creates an eCRF, and the application developed as an app is provided to subjects. When subjects set up the app on their cellular phone and input biometric information, including height, weight, blood pressure and blood sugar level, the information is transmitted to researchers. When their period of participation in a clinical trial ends, a satisfaction survey screen pops up on subjects' mobile phones, which allows the evaluation of satisfaction. Figure 6 shows the screen for checking the results of satisfaction surveys completed by subjects.

Figure 7 shows the block explorer screen for managing data generated and written on blockchain.

The three provided data tables include “Height,” showing the height of block or the number of blocks created so far; “Formulators,” referring to the number of formulators, or nodes maintaining blockchain; and “Transactions,” which lists the number of activities on blockchain such as data input, changes, and modifications occurring in a clinical trial system. In addition, the following three are provided: “Transactions tables,” indicating lists of activities on blockchain; “Block tables,” showing information about tables and blocks actually created; and “Transaction type per second graphs,” indicating the frequency of transaction activities per

second and their trend. The table displaying server portion of the screen. bandwidths in numbers is placed at the bottommost

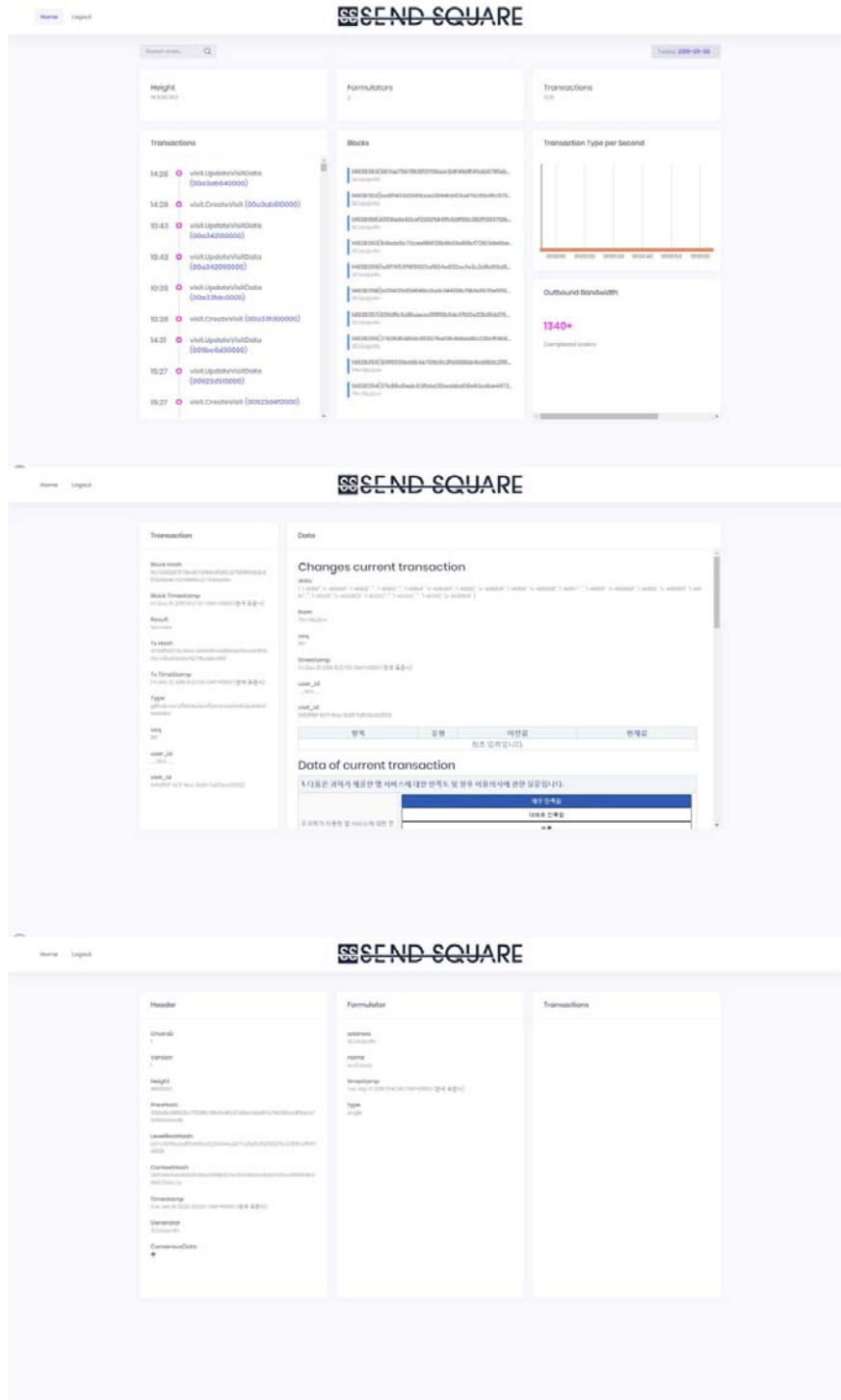


Figure 7 The block explorer screen

By clicking the detailed transaction information in the block explorer, it can be confirmed that everything is recorded on blockchain from the hash information of a block containing a relevant transaction and block creation time to transaction results, transaction hash value, transaction creation time, transaction type, and user ID and even visitor identification information, as shown in Figure 7. In

addition, by offering visualized information on entered, modified, and changed data items, the system provides monitoring personnel and researchers with an interface that allows them to more easily manage the history without understanding the technical workings of the system.

Lastly, as all block information contained on the blockchain is given by the explorer, it is possible to confirm current chain status and verify whether the records/history of data input, modification and changes have been written on blockchain. The block information screen displays block header information, information on the formulators that have produced the blocks, and information on the transactions contained on a block.

4. DISCUSSION

This study discussed the design and development of a blockchain technology-based ePRO system that can prevent the modification or manipulation of clinical trial data and enhance its reliability. A researcher may intervene to eliminate specific data or alter the normal range in order to draw desired results from data analysis. This may not only affect the licensing of medicines, but also seriously undermine the safety and reliability of the product, and moreover, public health. When using the ePRO system, however, even cases where the constancy of blockchain data is ignored, data is manipulated, or clinical trials are passed deceptively can be exposed afterwards by reloading the blockchain data. This can be used as evidence for legal recourses. Within the ePRO system, every generated chain is recorded, and the chain's header is recorded on a regular basis as well, which can prevent the falsifying of an entire chain to alter the data. In addition, clinical data is distributed on the blockchain, so the actual ownership of the data can be imparted to its generators, including medical personnel and patients. Such an ePRO system designed around blockchain technology would allow an enhanced clinical trial process compared to conventional CDS. The results of this study thus suggest a potential to save human resources and costs for CDM operation by increasing reliability through the implementation of an on-chain clinical trial system. Additionally, the on-chain clinical trial system is predicted to help prevent harm from serious abnormal reactions, including patients' death, subsequent side effects, and birth defects caused by clinical trial errors.

One of the most important tasks in the blockchain and clinical data convergence industry market is to discover a use case model. If relevant institutions, including clinical trial centers, corporate research institutes, pharmaceutical companies, and university research centers, use the ePRO system developed in this study, a successful use case model is expected to be presented.

The results of this study consist of not only a blockchain platform for clinical trials, but also modules that can serve as a platform for medical and clinical services to cope with expanding clinical data fields. The medical services of today, in contrast to the "medical treatments" of the past, emphasize the importance of prevention, diagnosis, and rehabilitation based on patient data. When the system developed in this study is combined in the near future with state-of-the-art technologies such as IoT and AI, a new paradigm for clinical data repositories and management systems is predicted to develop. Lastly, based on the clinical data collected by this ePRO system, when a clinical data warehouse is established and linked to medical information systems, a servitization model that can accelerate innovation in the medical service market is expected to emerge.

It is our intention that the ePRO system and data modeling will be enhanced to accommodate other clinical areas in the future work. It may be particularly advantageous to divide and manage chains to accommodate data from different specifications or standards. The independent multi-chain structure of this software can be used properly, and the company plans to expand from diabetes and chronic diseases to endocrine diseases such as adrenal, pituitary, hyperlipidemia and thyroid diseases and further to other models of diseases.

5. CONCLUSION

The results of this study consist of not only a blockchain platform for clinical trials, but also modules that can serve as a platform for medical and clinical services to cope with expanding clinical data fields. The medical services of today, in contrast to the "medical treatments" of the past, emphasize the importance of prevention, diagnosis, and rehabilitation based on patient data. When the system developed in this study is combined in the near future with state-of-the-art technologies such as IoT and AI, a new paradigm for clinical data repositories and management systems is predicted

to develop. Lastly, based on the clinical data collected by this ePRO system, when a clinical data warehouse is established and linked to medical information systems, a servitization model that can accelerate innovation in the medical service market is expected to emerge.

It is our intention that the ePRO system and data modeling will be enhanced to accommodate other clinical areas in the future work. It may be particularly advantageous to divide and manage chains to accommodate data from different specifications or standards. The independent multi-chain structure of this software can be used properly, and the company plans to expand from diabetes and chronic diseases to endocrine diseases such as adrenal, pituitary, hyperlipidemia and thyroid diseases and further to other models of diseases.

Acknowledgement

This research was supported by the MSIT(The Ministry of Science and ICT), Korea, under Fostering blockchain company supervised by the NIPA(National IT Industry Promotion Agency)(NIPA-2019-S0834-19-1021)

REFERENCES:

- [1] Medgadget. Global Clinical Trial Market [Internet]. Medgadget; 2019 [cited at 2019 Nov 23]. Available from <https://www.medgadget.com/2019/11/global-clinical-trial-market-is-anticipated-to-reach-65385-million-us-by-2025-at-a-cagr-of-8-76.html>.
- [2] Korean Ministry of Health and Welfare. 2020 Year announcement of global competitiveness of clinical trial for the advancement of clinical trial into the top 5 countries [Internet]. MHW; 2019 [cited at 2019 Nov 25]. Available from: http://www.mohw.go.kr/react/al/sal0301vw.jsp?PAR_MENU_ID=04&MENU_ID=0403&CON_T_SEQ=325129&page=1.
- [3] H.S. Song and B. S. Park, “Invited Clinical Trials : Biocapital, Ethical Variability, and the Industrialization of Clinical Trial in Korea”, *Korean Journal of Clinical Pharmacy*, Vol. 26, No. 2, 2016, pp. 181-186.
- [4] D. Knepper, C. Fenske, P. Nadolny, A. Bedding , E. Gribkova, J. Polzer, J. Neumann, B. Wilson, J. Benedict and A. Lawton. “Detecting data quality issues in clinical trials: Current practices and recommendation”, *Therapeutic Innovation & Regulatory Science*, Vol. 50, No. 1, 2016, pp. 15-21.
- [5] A. Gupta, “Fraud and misconduct in clinical research: A concern”, *Perspectives in Clinical Research*, Vol. 4, No. 2, 2013, pp. 144-147.
- [6] ER. Pryor, B. Babermann and ME. Broome, “Scientific misconduct from the perspective of research coordinators: a national survey”, *Journal of Medical Ethics*, Vol. 33, No. 6, 2007, pp.365-369.
- [7] N. Malik, N. Jain and U. Nagaich, “Clinical data management: Tools and regulations”, *International Journal of Pharmacy & Life Sciences*, Vol. 9, No. 3, 2018, pp. 5767-5771.
- [8] DR. Wong, S. Bhattacharya and A. Butte, “Prototype of running clinical trials in an untrustworthy environment using blockchain”, *Nature Communications*, Vol. 10, No. 1, 2019, pp. 1-8.
- [9] Cointelegraph. Medical R&D alliance expands blockchain project to include data sharing [Internet]. Cointelegraph; 2019 [cited at 2019 Nov 25]. Available from: <https://cointelegraph.com/news/medical-rd-alliance-expands-blockchain-project-to-include-data-sharing>.
- [10] Y. Chen, S. Ding, Z. Xu, H. Zheng, and S. Yang, “Blockchain-based medical records secure storage and medical service framework”, *Journal of Medical Systems*, Vol. 43, No. 1, 2018, pp.1.
- [11] Korea National Enterprise for Clinical Trials. Year Korean Clinical Trials Industry Year Book [Internet]. KNECT; 2019 [cited at 2019 Nov 25]. Available from: <https://www.konect.or.kr/kr/board/press/boardView.do?bbsIdx=5589>.
- [12] G. R. Kim and D. H. Lee, “A study on the protection, integrated management and uses of medical data through blockchain technologies”, *Korea Commercial Law Association*, Vol. 37, No. 4, 2019, pp. 279-327.
- [13] Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System [Internet]. 2019 [cited at 2019 Nov 24]. Available from <http://bitcoin.org/bitcoin.pdf>.
- [14] B. Sharma, C. N. Sekharan and F. Zuo, “Merkle-tree based approach for ensuring integrity of electronic medical records”, *Proceedings of 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, IEEE (New York City, USA), November 8-10, 2018, pp.983-987.
- [15] J. Y. Kim, “A comparative study of block chain: Bitcoin, Namecoin, MediBlock”, *Journal of*



Science and Technology Studies, Vol. 18, No. 3,
2018, pp. 217-254.

- [16]D. Glover and J. Hermans, “Improving the traceability of the clinical trial supply chain”, *Applied Clinical Trials*, Vol. 26, No. 11/12, 2017, pp. 36-38.