the innovation hub for
precision medicine

DayOne.swiss Blockchain Project Report 2017/2018

Integration of Blockchain Technology into existing Genome Data Management and Clinical Trials applications

Authors: Daniel Burgwinkel, Thomas Brenzikofer, Mike Gault, Ivo Lohmus, Timo Kanninen, Wolfgang Weiss, Mimmo Garribo, Beat E. Widler.

Contact: dayone@blockchain.jetzt

3th July 2018

1 Summary

This report gives an overview of the activities of the DayOne.swiss Blockchain project in Basel, which were initiated in 2016 and officially started in 2017. As two projects for the integration of Blockchain Technology into existing Genome Data Management and Clinical Trials applications were successfully conducted in Basel.

1.1 Project partners

DayOne - the innovation hub for precision medicine has a vision to "Create a world leading precision medicine hub respected for its impact and collaboration across silos." The initiative is managed by BaselArea.swiss in close collaboration with the Canton of Basel-Stadt. www.dayone.swiss

Guardtime was founded in 2007 in Estonia and has now a team of over 150 cryptographers, developers and security architects, with decades of experience defending networks from nation-state attack. In Estonia all E-Health and E-Government records are secured by Guardtime Blockchain Technology. The cryptographic algorithms of the Guardtime are approved by the NIST (US National Institute of Standards and Technology [1]) and are applied in regulated environments. www.guardtime.com

BC Platforms is a swiss-finnish bioinformatics software company established in 1997 with the headquarter in Basel and R&D in Helsinki. It started as a Spin-off from a unique genome project with MIT and provides Clinical genomics software for healthcare and software solutions for biobanks and genomic research. BC Platforms is runs a Data Science Business with a global network of biobanks with clinical and genome data and conducts data science projects with pharmaceutical industry. BC Platforms partners with Microsoft, which provides a Genomics Service Platform [1].
www.bcplatforms.com

Ethical was founded in Basel in 2014. Ethical is operating in the eClinical Software and Clinical Data Management sector with a special focus on Endpoint Adjudication software services. The Cloud Software Service Ethical eAdjudication® supports Endpoint Adjudication Committees' operations and data management in a quality controlled environment.
www.ethicalclinical.com

1.2 Goals of DayOne Blockchain initiative

The DayOne.swiss Blockchain initative explores how Blockchain Technology can support Precision Medicine to ensure data integrity, compliance and secure data exchange in context of Life Sciences research and Digital Health. The initiative brings together experts from different fields of Life Sciences research, Blockchain technology and healthcare practitioners. The Blockchain initiative started in 2017 with following activities:

- Workshops to identify Blockchain use case and business Models which are relevant to the region of Basel.
- Blockchain Proof-of-concepts in context of Genome Data Management and Clinical trials were conducted.
- Reviews with experts from the academia and industry.
- From 2016-2018 five Blockchain seminars at University of Applied Science (FHNW) were conducted by Prof. Dettling and Dr. Burgwinkel [2].
- Presentations at scientific conferences:
  - BASEL LIFE 2016, 2017, 2018 (www.basellife.org)
  - DayOne conference 2017 (www.dayone.swiss)
  - Decentralized.com 2017 in Cyprus

2 What is Blockchain Technology and how can it be applied in Precision Medicine

This chapter provides a short introduction to the basic terms of blockchain technology and illustrates how blockchain platforms can be used in Life Sciences.

The subject of blockchains is currently being discussed intensively by both business and IT managers as well as regulatory authorities [3]. Business managers see new disruptive business models, and the technology fascinates IT professionals who have had the chance to collect preliminary experience from the digital currency Bitcoin. FDA is looking at blockchain e.g. in the Information Exchange and Data Transformation (INFORMED) initiative: "…emerging technologies such as blockchain to enable secure exchange of health data at scale”[4].

The term blockchain describes a technological concept which stores data not in a central database, but rather distributed among the systems of users with the help of cryptographic protocols. The word “blockchain” was chosen because the data is stored in individual blocks which are then distributed and filed among the systems of the network participants and the order of the blocks is documented by means of a chain [5].

Even though it is only a technical concept, experts believe that this approach will revolutionize business models in various fields. If one wishes to use this technology for a given area of activity, the following questions arise:

- Which applications and use cases in Life Sciences can be realized on the basis of blockchains?
- What data can be sensibly stored in blockchains?
- Which transactions can be supported by blockchains?
- What technical restrictions exist?

A plethora of articles explaining the functional principles of blockchains can be found in the press and on the internet. For a basic understanding it is important to distinguish between the following terms:

- Blockchain as a **technical concept** in IT, which uses cryptographic methods like hash-values and hash-trees that have been known for more than thirty years.
- Blockchain **software** that provides programming code for cryptographic operations. In 2018 there are more than fifty different commercial as well as open source software products available.

- Blockchain **applications** for the implementation of a certain use case. Typically, these applications are operated with the help of blockchain software or on a blockchain platform.
- **Blockchain platforms** which use a certain software and are operated on the internet as a service, e.g. as an open peer-to-peer network or as a commercial service.
- **Blockchain as a Service** which provides the necessary software and services in a cloud. In such cases, a chosen blockchain solution can be operated on virtual computers in the cloud.

Depending on the concept, business-relevant data can be stored in the blockchain (e.g. transaction data) or the data in the blockchain can contain references to external data, e.g. because the data requires a high storage volume or is confidential.

Blockchains can be employed in many areas and offer different functionalities. From a high-level view we can classify blockchain use cases in three categories.

- Blockchains for proof of data integrity
- Blockchains for registration and certification
- Blockchains for the settlement of transactions

In this report we will focus on the first class of use cases and describe the Use Case and Proof of Concept that was conducted in the DayOne Initiative.

Proof of data integrity can be provided using the blockchain, i.e. it can be verified that data has not been subsequently changed.

Until now, similar functions have been implemented with the use of digital signatures or storage media with procedures for integrity protection. In comparison to using digital signatures or hardware-based integrity protection, a blockchain has the following advantages:

- Blockchain technology can prove both the integrity and the completeness of a set of data, as well as the chronological sequence.
- When using digital signatures, data is signed with the key of a person or organization. Thus management of public and private keys is necessary and can be very cumbersome in the long run.
- Blockchain technology is primarily a software-based procedure and is thus independent of the hardware used. So even for data stored in the cloud, data integrity can be verified.

The use of a blockchain for proof of data integrity is based on the following procedure:

- Data is generated outside of the blockchain, e.g. a document or data set.
- The proof of integrity is generated by means of a hash algorithm and filed in the blockchain.
- Data is checked for integrity at regular intervals. The time interval is chosen based on the protection requirement. Thus, in the context of cybersecurity, important data is checked at short intervals in order to detect any manipulation of the data by an attacker. For long-term archiving, longer intervals are typically chosen for monitoring.
- In addition to periodic checking of the entire data pool, individual documents can be checked as required, e.g. an external auditor can check the authenticity of a document.

These functions are of interest in application areas where proof that data hasn't been retrospectively manipulated is particularly important. Examples include research data for medical products, diagnoses in health care or the configuration of machinery.

3 Secure Genome Data Management with Blockchain Technology

3.1 Abstract

In 2017 the first DayOne project was started with the title “Genome Data Management - Using Blockchain technology to ensure provenance of genome sequence data in cross-organizational workflows in cloud environments”. Following project partners collaborated:

- Guardtime.com – Blockchain Technology provider
- BCplatforms.com - Genome Data Management solutions

3.2 Context of the project

Next Generation Sequencing (NGS) is technology for extracting genomic data sequence from DNA. Process includes multiple steps, including DNA extraction from sample (typically blood or spit), production of raw data using NGS device and processing raw data to generate final sequence (secondary analysis). The application BC Platforms “BC|PIPE NGS” is an automated workflow for performing secondary analysis, monitoring data quality and ensuring data governance throughout the process.

Challenge: processing of large data files

Instead of producing whole genome sequencing (3 billion basepairs in humans) directly, current NGS devices produce overlapping short sequence reads (some hundreds of basepairs). To ensure proper coverage over whole genome, sequence from one sample is read typically at least 30 times, often hundreds of times (coverage). As a result, hundreds of millions short sequence reads are produced for one sample. Size of the raw data for whole genome is roughly 250 Gb / sample, depending on the coverage.

To build actual genome sequence, these short sequence reads are first matched to reference genome (alignment) and then called (variant calling). This process requires significant amount of computer resources. The whole process is often called as secondary analysis, while actual sequence data analysis and interpretation is called as tertiary analysis.

Trend: Use of Cloud Computing

Sequence data production is volume business. With expensive sequencing devices, data production costs / sample are lower. As a consequence, raw NGS data production is consolidating to larger units (University sequencing centers), or even national genome centers. **Data processing is often performed using Cloud** or external calculation environment, using secondary analysis software tools from different software vendors.

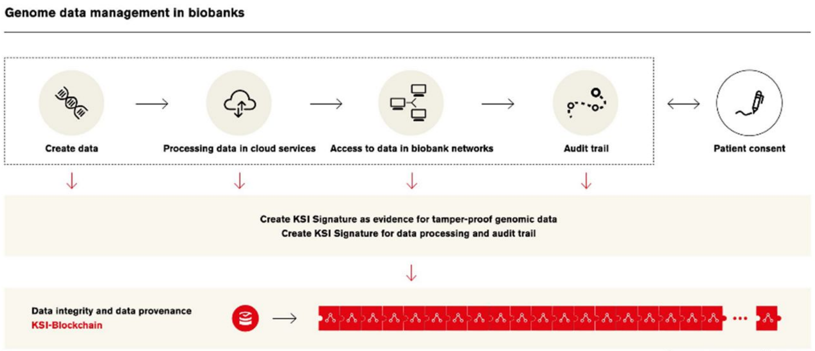
At present NGS data is mainly used for cancer, or rare disease diagnostics in healthcare. Constantly dropping prices of NGS data are driving its use also for complex diseases, and research use (secondary use of healthcare data), increasing data volumes significantly.

The application BC Platforms “BC|PIPE NGS” is designed to

- scale to population scale whole genome sequencing by automating secondary analysis process and eliminating manual work,
- implement continuous data quality monitoring to detect immediate problems, and predict future problems, and
- in NGS production process involving many steps and parties ensure data governance throughout the whole process.

3.3 Solution

In this use case we want to show how to ensure data integrity and data provenance for genome data management can be realized with blockchain technology.


Genome data management in biobanks
Create data
Processing data in cloud services
Access to data in biobank networks
Audit trail
Patient consent
Create KSI Signature as evidence for tamper-proof genomic data
Create KSI Signature for data processing and audit trail
Data integrity and data provenance
KSI-Blockchain

Source: www.guardtime.com

The solution provides following benefits:

- Data integrity is ensured in all process steps and in all environments
- Proof of data integrity, audit trails and compliance with patient consent can be combined

The applications for Genome Data Management use the KSI Blockchain of Guardtime [6], which is e.g. used for securing the eHealth system of Estonia. The researcher Buldas and Saarepera from Guardtime were the first cryptographers to give a formal security proof in 2003 i.e. what properties do you need for hash-functions and data structures in order to build a formally verifiable security proof [7] . The KSI Blockchain is in production since 2008 and uses in context of e.g eHealth, eGoverment services and in Defence and Aerospace industry.

Which kind of data is stored in the Blockchain?

- Only cryptographic evidence is stored in the Guardtime KSI Blockchain.
- No genome data or personally identifiable information is stored in the KSI Blockchain.

How is the genome data managed?

Genome data does not leave the borders of the research organization. In addition to the blockchain a file with cryptographic evidence (KSI signature) is created and stored within the organization.

How do you proof data integrity in the overall process of genome data management?

Today typically every research organization runs a digital archive and there is no overall method to proof data integrity.

With the new solution all involved research organizations and cloud providers use the KSI Signatures.

How can you proof that you are compliant with the patient consent?

In the current health care system, a patient has to trust that all parties act according to his consent. With the solution an evidence of data and process integrity in all processing steps can created and matched with audit trails. In this way the research organization can proof that the processing of the patient data was compliant with the patient consent.

4 Integrating Blockchain Technology in Clinical Trials Cloud services

In a series of DayOne workshops the concept for the Integration of the Guardtime KSI Blockchain with the clinical trials software of ethicaclinical.com was worked out.

4.1 Context of the project: Clinical Endpoint Adjudication

Is crucial for reducing patients' risks and study bias when assessing and comparing an endpoint at baseline vs. study end in a multi-center clinical trial when subjective – such as image-based – endpoints represent a pivotal efficacy and/or safety parameter.

In fact, where endpoint assessment is complex, includes reader's driven components – aka subjective assessments - and/or cannot be blinded, a central assessment of efficacy or safety events, made by a panel of independent experts following a blinded standardized procedure increases accuracy, transparence, and homogeneity of judgments.

Challenge: Data integrity in Clinical Trials

Since the implementation of GxP standards the ALCOA and more recently the ALCOAC principles have been a basic requirement of quality and compliance when interacting with regulatory authorities in the field of medicinal and device products. ALCOA stands for:

- Attributable
- Legible
- Contemporaneous
- Original
- Authentic
- Complete

For any regulator such as EMA, FDA, JPMA or any Notified Body one basic requirement of any electronic system used in managing data related to a pharmaceutical or device product is the ability to produce an indelible Audit Trail. In this context the term indelible is key. Indeed, regulators must rule out with 100% certainty that audit trail data - i.e., the evidence about who recorded any data for the first time and when, who modified it and when, and finally who released it as genuine data – that such Audit Trail data has been tampered with – i.e., modified without leaving any trace of the change - after the original recording of the data in the system's user database. In the “old days” carbon copies – or with a bit more modern technology so-called NCR paper – was used to provide the generator of the data – e.g., a trial investigator – with his personal contemporaneous copy. In case of suspicion of any wrong doing by the sponsor of a trial – e.g., the sponsor tampering with the audit trail by accessing this data through an “administrator’s back-door” – the contemporaneous copy left with the investigator would have allowed a regulatory inspector to detect such fraud. With the move to fully electronic data capture systems the contemporaneous copy became in fact a copy on what got stored on the electronic system that is typically under the direct or indirect (when hosted by a CRO/vendor) control of the sponsor – aka the potential suspect. Not surprisingly this makes inspectors and Health Authorities uncomfortable as they can no longer produce bulletproof evidence that audit trail data is genuine and hence trial data is genuine. Recent repeated cases where CMOs tampered with batch records to make out of specifications batches “acceptable”, have spurn regulators concerns. Solutions presented to rebuild trust in audit trail data are typically paper-based “solutions”: print the records and have them signed and dated by the investigator or other accountable person. A bit more elegant but still saddled with practical issues: generate a contemporaneous PDF copy on the local computer of the data generator (beside the fact that this generates a plethora of files that need to be managed for future retrieval many organizations do not allow importation of alien files for obvious cybersecurity reasons.

4.2 Solution

The Ethical eAdjudication ® Solution was integrated with the Guardtime KSI Blockchain, which is e.g. used in the eHealth System of Estonia. Every Time an operation is performed in the eAdjudication System a new Audit Trail Record is generated in its database. For every record and for every audit trail information a unique KSI Signature is created. The KSI Signature serves as proof-of-evidence that data integrity, time and authorship were not tampered. The KSI Signature is stored in the Adjudication® System and the evidence is registered in the Guardtime KSI-Blockchain.

At any time after, the Audit Trail Records present in the Adjudication ® System database can be matched with the corresponding KSI Signature. In addition, the record and signature can be validated against the KSI Blockchain to ensure that no data whatsoever has been altered or even deleted since the time of its first registration.

This feature provides irrefutable evidence that any data point entered by an adjudicator or other contributor has never been modified without a GxP audit trail stating reasons, time and originator of the change.

Only cryptographic information (hashvalues) is send to the KSI Blockchain while the actual sensitive data generated by the system are only recorded in the original database and, therefore, ensuring compliance with all applicable data privacy and protection laws and regulations such as the GDPR and HIPAA.

5 Outlook

The DayOne project demonstrated that cryptographic technologies can be applied to Life Sciences research processes without creating new data silos. As the technology is used in the Estonian eHealth system since 2012 it shows that it is production ready. For more information on the Day One Blockchain activities please visit www.dayone.swiss

6 References

[1] "BC Platforms Launches an End-to-End Solution for Precision Medicine powered by the Microsoft Genomics Service," *BC Platforms*, 06-Mar-2018. [Online]. Available: http://www.bcplatforms.com/news/bc-platforms-launches-end-end-solution-precision-medicine-powered-microsoft-genomics-service/. [Accessed: 28-Mar-2018].

[2] "Seminarreihe Digitale Geschäftsmodelle mit Blockchaintechnologie," *FHNW - Fachchochschule Nordwestschweiz*. [Online]. Available: https://www.fhnw.ch/de/weiterbildung/wirtschaft/seminar-digitale-geschaeftsmodelle. [Accessed: 26-Apr-2018].

[3] "Digital Footprints in Drug Development: A Perspective from within the FDA," *Digit. Biomark.*, vol. 1, no. 2, pp. 101–105, 2017.

[4] O. of the Commissioner, "Oncology Center of Excellence - Information Exchange and Data Transformation (INFORMED)." [Online]. Available: https://www.fda.gov/aboutfda/centersoffices/officeofmedicalproductsandtobacco/oce/ucm543768.htm. [Accessed: 07-Feb-2018].

[5] D. Burgwinkel, Ed., *Blockchain Technology*. Berlin ; Boston: De Gruyter, 2017.

[6] "KSI Technology | Industrial Scale Blockchain | Guardtime." [Online]. Available: https://guardtime.com/technology. [Accessed: 31-May-2018].

[7] A. Buldas and M. Saarepera, "On provably secure time-stamping schemes," in *In Advances in Cryptology — ASIACRYPT 2004*, 2004, pp. 500–514.