ICS 35.240.99
CCS L70.84

Social Organization Standard

T/GBA 014—2023

Blockchain genetic data application implementation specification

区块链 基因数据应用实施规范

(English Translation)

Issue date: 2023-12-25 Implementation date: 2023-12-30

Issued by Guangdong-Hong Kong-Macao Greater Bay Area Standards Innovation Alliance

Contents

Foreword

This stabdard is drafted in accordance with the rules set forth in GB/T 1.1-2020 *Directives for Standardization -- Part 1: Rules for the Structure and Drafting of Standardizing Documents.*

This standard is proposed by MGI Tech Co., Ltd.

This standard is prepared by the Guangdong-Hong Kong-Macao Greater Bay Area Standards Innovation Alliance.

This standard is authorized for use by organizational partners and all members of the Guangdong-Hong Kong-Macao Greater Bay Area Standards Innovation Alliance. These members are required to adopt the same transformation into their own group standards and publicize the standard basic information on the National Group Standards Information Platform.

Please note that some of the contents of this document may involve patent rights. The issuing body of this document shall not be held responsible for identifying any or all such patent rights.

This is the first release.

Introduction

With the rapid development of information technology, particularly the widespread application of distributed technology, blockchain technology, as a significant representative, has shown immense potential in multiple industries. Blockchain is not just a branch of distributed ledger technology but also particularly emphasizes the ability to manage trust data in peer-to-peer networks using a blockchain data structure. This structure ensures the immutability and traceability of data, making it an ideal choice for handling sensitive information, such as genetic data. In the field of life sciences, particularly in the processing of genetic data, the application of blockchain technology offers a unique solution that meets the specific business needs and regulatory requirements of this field. Given the sensitivity and complexity of genetic data, the technology of consortium chains, which combines traditional distributed technology with bioinformatics, becomes particularly crucial. However, given the diversity of different institutions in the development of blockchain technology, it is vital to establish a unified technical standardization specification. This not only helps ensure the compatibility and security of the technology but also promotes healthy development and innovation within the industry.

This document aims to provide a set of clear guiding principles and implementation specifications to effectively utilize blockchain technology for processing and sharing genetic data while ensuring data security and privacy. Through this approach, it is possible to better leverage this breakthrough technology, bringing broader application prospects and value to the field of life sciences.

Blockchain genetic data application implementation specification

1 Scope

This document specifies the system functions and normative requirements of blockchain and distributed ledger technology in genetic data auditing, covering implementation principles, system architecture, and functional requirements.

This document is applicable to genetic data auditing service providers for standardizing business activities. It can also serve as a basis for supervisory authorities and third-party assessment organizations to oversee, manage, and evaluate genetic data application businesses.

2 Normative References

The contents of the documents listed below form essential parts of this document through normative references within the text. For referenced documents with specified dates, only the versions corresponding to those dates are applicable to this document. For undated referenced documents, the most recent version, including all amendments, is applicable to this document.

GB/T 25069 *Information Security Technology - Terminology*

GB/T 35890 *High Throughput Sequencing Data Sequence Format Specification*

GB/T 42752 *Blockchain and Distributed Ledger Technologies - Reference Architecture*

ISO 20387 *Biotechnology - Biobanking - General Requirements for Biobanking*

3 Terms and Definitions

The terms and definitions defined in GB/T 25069, GB/T 35890, GB/T 42752, ISO 20387, and the following terms and definitions apply to this document.

3.1 Blockchain

A distributed ledger formed by sequentially appending blocks confirmed through consensus, linked using cryptographic techniques.

3.2 Chained-block data structure

A data structure where transactions occurring over a period are stored in units called blocks, which are then connected in chronological order into a chain using cryptographic algorithms.

3.3 Consensus algorithm

The method by which distributed nodes in a blockchain system reach a consensus on the verification and recording of transactions or states.

3.4 Smart contract

A computer program stored in a distributed ledger technology system.

3.5 Digest generation

Also known as a digest function (or Hash function), it typically works by converting an input message of any length into a fixed-length short message output to ensure data integrity.

3.6 Consortium blockchain

A type of blockchain deployment model used by a group of stakeholders, where only authorized nodes can access the network. These nodes can participate in consensus and read/write data according to the rules.

3.7 Blockchain proof of existence

To ensure the integrity and authenticity of evidential information (electronic data), a notarization service is implemented using blockchain technology, which achieves consensus among multiple nodes.

3.8 Gene

Genetic information that controls biological traits, usually carried by DNA sequences, i.e., a functional segment of DNA or RNA sequence.

3.9 Genetic data

Data related to the innate or acquired genetic characteristics of an organism, generally obtained through genetic testing of biological samples. This data can provide unique information about the physiology or health of the organism.

3.10 Associated data

Any information related to genetic data.

Note: This includes, but is not limited to, phenotypic, clinical, and process information.

[Source: ISO 20387:2018, 3.3]

4 Abbreviations

The following abbreviations apply to this document:

API: Application Programming Interface

BaaS: Blockchain as a Service

CA: Certification Authority

DLT: Distributed Ledger Technology

ECC: Elliptic Curve Cryptography

ECDSA: Elliptic Curve Digital Signature Algorithm

PBFT: Practical Byzantine Fault Tolerance

POS: Proof of Stake

POW: Proof of Work

RSA: A public-key cryptography algorithm used for encryption and digital signatures (Rivest–Shamir–Adleman)

SM2: A public-key cryptography algorithm widely used for digital signatures, encryption, and key exchange

5 Implementation Principles

5.1 Principle of legality and compliance

Strictly comply with relevant national regulations and bioinformatics security supervision requirements, and provide necessary technical support for regulatory audits.

5.2 Principle of traceability

All business activities and operations are thoroughly recorded to ensure traceability and auditability.

5.3 Principle of data consistency

The data on-chain and off-chain should maintain consistency, and the data among various nodes of the blockchain should be uniform.

5.4 Principle of hierarchical authorization

Implement a hierarchical authorization system, clarify the permissions of different roles, and ensure the security of data access.

5.5 Principle of privacy protection

Ensure the protection of users' privacy rights and prevent incidents of privacy breaches.

5.6 Principle of security

Take all necessary measures to ensure the security of data both on-chain and off-chain.

5.7 Business-oriented principle

Conduct technological research and development guided by business needs, prioritizing actual business application scenarios.

6 System Architecture

6.1 Layered architecture

The layered architecture of the genetic data blockchain platform's functional components is shown in Figure 1. The architecture is divided into five levels: user layer, access layer, core layer, base layer, and a aross-layer function collection that spans across these layers.

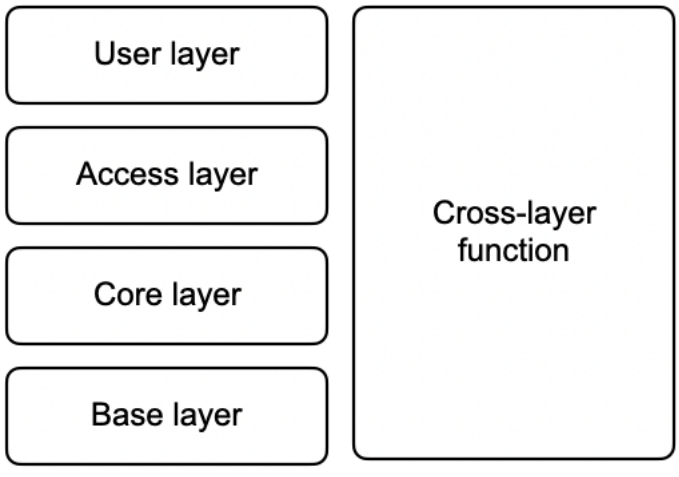

Figure 1: Layered Architecture of Functional Components in the Genetic Data Blockchain Platform

6.2 User layer

Provides an interactive entrance for users to access the genetic data blockchain services. Through this layer, users can perform related management functions and maintenance, and it has the capability to provide cross-layer service support.

6.3 Access layer

Targeted towards the User Layer or terminal applications, it provides efficient, reliable, and universal access capabilities. This includes encapsulating the functions of the Core Layer, offering efficient caching and load balancing, as well as standardized access protocols.

6.4 Core layer

Integrates modules such as consensus mechanisms among nodes, privacy protection, encryption, and digital signatures, ensuring the system's integrity, security, and immutability. It also enables customization of smart contracts and unified services according to different business scenarios.

6.5 Base layer

The Base Layer provides all the necessary operational environments and basic components.

Note: This includes data storage, execution containers, communication networks, etc.

6.6 Cross-layer functions

Cross-layer functions provide functional components that can operate across multiple levels of the system.

7 Functional Requirements

7.1 General set of functional components

See Figure 2 for the general set of functional components of the genetic data blockchain.

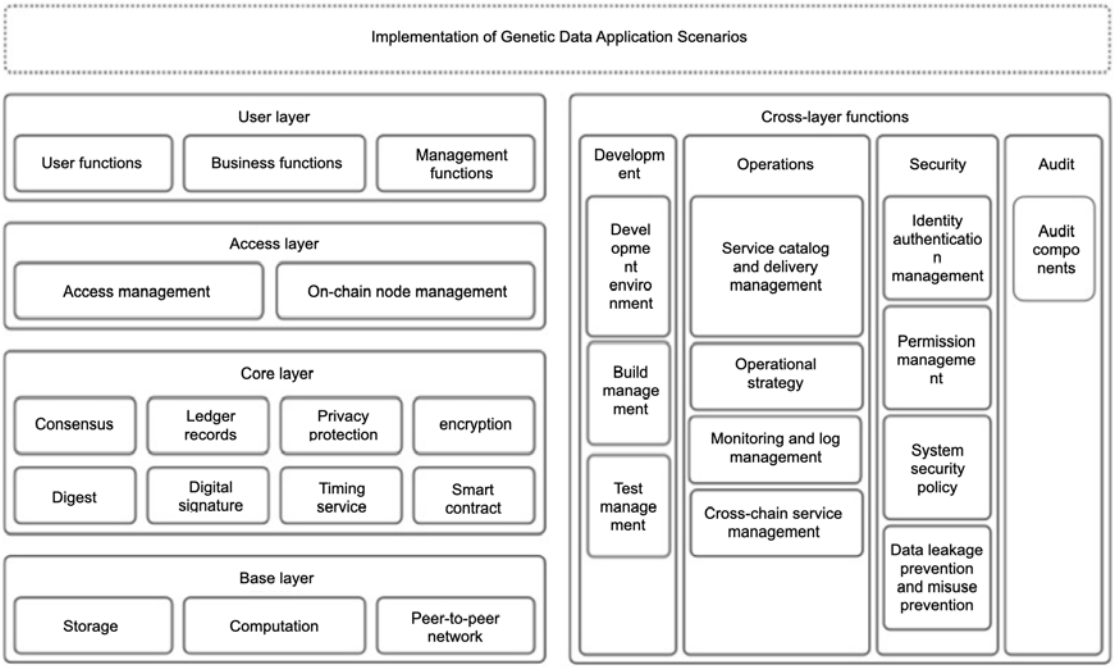

Figure 2: General Set of Functional Components of the Genetic Data Blockchain

7.2 User layer functional requirements

The User Functional Component provides users with access to and use of genetic data auditing services, primarily involving basic resources, chain operations, and smart contract processing.

7.2.1 User functions

Includes various forms such as command-line interfaces, graphical user interfaces, and application programming interfaces.

7.2.2 Business functions

Allows users to submit specific transaction requests (such as queries, updates) to the genetic data auditing network.

7.2.3 Management functions

The Management Function Component provides users with a variety of service functions, including member management, service activity monitoring, event handling, issue reporting, and security management.

7.3 Access Layer Functional Requirements

7.3.1 Access management

The access management function should include:

a) It should provide basic information queries for user accounts, block and transaction detail queries for the genetic blockchain, and the submission of specific transaction requests from users to the genetic data auditing network;

b) Access management functions should consider: management of interface service capabilities (such as API call frequency and query caching settings), configuration of access permissions for the interface (such as setting different permissions for different users), and communication security of the interface (such as message encryption).

7.3.2 Protocol management

Manages the technical specifications that devices joining the network should follow, involving key technical parameters such as hardware, software, and ports.

7.3.3 On-Chain node management

The On-Chain Node Management function should provide information queries and related management operations for genetic data auditing nodes. Its functionalities should include:

a) Querying the real-time status information of node servers;

b) Controlling the start and shutdown of node services;

c) Configuring the capability parameters of node services;

d) Monitoring the network connection status of nodes;

e) Managing the authorization and permission settings of nodes.

7.4 Core layer functional requirements

7.4.1 Consensus mechanism

The Consensus Mechanism should select an appropriate consensus algorithm to establish the mechanism. Its functionalities should include:

a) Supporting multiple nodes to jointly participate in the consensus process and complete confirmation;
b) Allowing independent nodes to validate the legitimacy of information submitted to the genetic data auditing network;
c) Preventing any single node from making information changes without the confirmation of other consensus nodes;
d) Possessing fault tolerance capabilities to handle various potential errors and anomalies, including physical failures of nodes, network connectivity issues, and illegal control.

7.4.2 Ledger management

Ledger management function should include:

a) Providing a persistent ledger storage mechanism;
b) Ensuring that multiple nodes can maintain a complete record of the blockchain notarization;
c) Providing authentic and complete data queries only to authorized participants;
d) Ensuring complete consistency of the same ledger data across various nodes;
e) Achieving public, tamper-proof, and reliable data storage through the joint maintenance and recording by multiple nodes.

7.4.3 Privacy Protection

Privacy Protection function should include:

a) Utilizing certification authorities to represent users in transactions, ensuring that user information and behavioral details are not publicly disclosed in the blockchain network;
b) Replacing broad network broadcasting with data transmission only among authorized nodes;
c) Using encryption keys to control access to user data, where only users with the keys can decrypt the data;
d) Employing advanced privacy technologies such as zero-knowledge proofs, ring signatures, or homomorphic encryption to further ensure privacy.

7.4.4 Data Encryption

The Data Encryption component should provide protection for data, including encryption and decryption operations. The encryption algorithms used should comply with the regulations of the National Cryptography Administration.

7.4.5 Digest Generation

The Digest Generation function should include:

a) Using digests to ensure data integrity verification;
b) Verifying whether data has been tampered with, based on given plaintext and digest values.

7.4.6 Digital Signature

The Digital Signature function should include:

a) Providing digital signing and verification functionalities to ensure the confidentiality, integrity, and non-repudiation of information;

b) Integration with digital certificates issued by authoritative third-party Certification Authorities (CAs).

7.4.7 Timing Service

The Timing Service function should include:

a) Providing a unified timing for ledger records;

b) Having fault tolerance capabilities to deal with potential timing errors;

c) Being able to integrate trusted third-party timing services, such as those provided by national time service centers;

d) Timing plays an important role in the genetic data auditing blockchain system, ensuring the temporal consistency of actions or data. The system can opt to incorporate specific timing tools or mechanisms.

7.4.8 Smart Contract

The Smart Contract function should satisfy:

a) Providing appropriate programming languages and development environments;

b) Supporting static and dynamic checks of contract contents;

c) Having the capability to execute contracts on run-time carriers such as virtual machines;

d) Ensuring the integrity of contract contents and resisting tampering;

e) Supporting contract upgrades in the context of multi-party consensus;

f) For smart contracts that interact with external data, their impact should be limited to within the contract to avoid affecting the overall system.

7.5 Base layer functional requirements

7.5.1 Storage

The Storage function should include:

a) Being deployable and usable by each node in a peer-to-peer network;

b) Providing data services efficiently, securely, and stably;

c) In the case of using a strategy of database and table partitioning, being capable of data sharding and routing;

d) The storage component is responsible for processing data generated by the blockchain, such as ledgers and transaction information, and providing writing and querying functionalities.

7.5.2 Computation

The Computation function should include:

a) Providing a running environment for the blockchain system;

b) Being adoptable by each node in a peer-to-peer network;

c) The computation component supports the computational needs of the blockchain system, covering container technology, virtual machine technology, cloud computing, etc.

7.5.3 Peer-to-Peer Network

The Peer-to-Peer Network function should satisfy:

a) Providing efficient and secure peer-to-peer communication;

b) Multicast capabilities based on peer-to-peer communication;

c) Supporting the dynamic addition, removal, and identification of nodes;

d) The blockchain system is based on a distributed peer-to-peer network infrastructure, organizing each node using peer-to-peer network protocols. Nodes exchange information through peer-to-peer protocols to support upper-layer functionalities.

7.6 Cross-layer functional requirements

7.6.1 Development Components

7.6.1.1 Development Environment Management

The Development Environment Management function should include:

a) Providing support for generating configuration metadata;

b) Supporting the customization and generation of service configuration scripts and components;

c) Development Environment Management should include functionalities for metadata generation and customization of service configuration scripts.

7.6.1.2 Build Management

The Build Management function should include:

a) Providing automated build and compilation functionalities;

b) Displaying compilation errors and prompts;

c) Implementing an audit process for the build procedure;

d) Supporting multiple languages and platforms.

7.6.1.3 Test Management

The Test Management function should include:

a) Managing test plans, strategies, reports, and cases;

b) Automatically generating test reports;

c) Ensuring that testing does not affect the production environment;

d) Supporting automation of the testing process;

e) Managing the test case repository and test databases.

7.6.2 Operational Components

7.6.2.1 Service Catalog and Delivery Management

Service Catalog and Delivery Management should meticulously record technical and procedural information related to genetic data blockchain services.

7.6.2.2 Operational Strategy

Operational Strategy should consider strategies in key areas such as business, technology, security, privacy, and certification.

7.6.2.3 Monitoring and Log Management

The Monitoring and Log Management function should include:

a) Monitoring the processes,communication status, and consensus efficiency of blockchain network nodes;

b) Storing and analyzing operational logs of the nodes.

7.6.2.4 Cross-Chain Service Management

Connects the operational, business, and management systems of the relevant Fabric Blockchain Service Providers (FBSP).

7.6.3 Security Components

The Security Components function should include:

a) Identity authentication management;
b) Permission management;
c) System security policy;
d) Supporting various methods of identity verification;
e) Authentication recordation, signature, and verification functionalities;
f) Setting access and resource usage permissions;
g) Communication encryption, encrypted data storage;
h) Node host security, smart contract verification;
i) Comprehensive security management components;
j) Compliance in data acquisition.

7.6.4 Audit Components

The Audit Components function should include:

a) Implementing tamper-proof, traceable, and auditable records;
b) Clearly defining audit governance rules;
c) Preserving data related to services, resources, and performance;
d) Allowing auditing organizations to join the network for real-time auditing;
e) Preserving audit-related data and evidence;
f) Permitting auditors to access data in real-time or obtain it periodically;
g) Conducting real-time verification for use as audit evidence;
h) nterfacing with other systems for data extraction;
i) Ensuring that the genetic data blockchain services meet audit requirements and prevent illegal activities related to genetic data resources.

Bibliography

[1] GB/T 5271.18 Information Technology Vocabulary Part 18: Distributed Data Processing

[2] GB/T 11457 Information Technology, Software Engineering Terminology

[3] GB/T 32399 Information Technology - Cloud Computing - Reference Architecture

[4] YY/T 1723 High-Throughput Gene Sequencer