ICS 35.240.99
CCS L70.84

团 体 标 准

T/GBA 014—2023

区块链 基因数据应用实施规范

Blockchain genetic data application implementation specification

2023-12-25 发布

2023-12-30 实施

粤港澳大湾区标准创新联盟 发 布

目　　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由深圳华大智造科技股份有限公司提出。

本文件由粤港澳大湾区标准创新联盟归口。

本文件授权粤港澳大湾区标准创新联盟组织伙伴和所有成员单位使用，联盟组织伙伴需等同采用转化为自身团体标准，并在全国团体标准信息平台上公开标准基本信息。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件起草单位：深圳华大智造科技股份有限公司、香港科技大学、澳门城市大学、深圳致星科技有限公司、广州金域医学检验中心有限公司、北京市朝阳区三里屯社区卫生服务中心、深圳市标准化协会、深圳市星创数字经济研究中心、深圳市前海智慧版权创新发展研究院。

本文件起草人：李士森、杨梦、陈凯、应作斌、张骏雪、柴迪、杨自飞、陈芳、颜妙丽、李映华、林垂旭、赵辰梅、但丹、冯奕翡、周宏、郝汉。

本文件为首次发布。

引　言

随着信息技术的飞速发展，特别是分布式技术的广泛应用，区块链技术作为其显著代表，已经在多个行业中显示出巨大潜力。区块链技术不仅仅是分布式账本技术的一个分支，还特别强调在对等网络中以块链数据结构管理信任数据的能力。这种结构确保了数据的不可篡改性和可追溯性，是处理敏感信息，如基因数据的理想选择。在生命科学领域，尤其是基因数据处理方面，区块链技术的应用提供了一种独特的解决方案，能够满足该领域特有的业务需求和监管要求。考虑到基因数据的敏感性和复杂性，联盟链技术，一种结合了传统分布式技术与生物信息技术的方法—显得尤为关键。然而，鉴于不同机构在区块链技术研发上的多样性，制定一个统一的技术标准化规范变得至关重要。这不仅有助于确保技术的兼容性和安全性，而且促进了行业内的健康发展和创新。

本文件旨在提供一套明确的指导原则和实施规范，以便在保障数据安全和隐私的同时，有效利用区块链技术来处理和共享基因数据。通过这种方式，能够更好地利用这一突破性技术，为生命科学领域带来更广泛的应用前景和价值。

区块链 基因数据应用实施规范

1 范围

本文件规定了区块链与分布式账本技术在基因数据审计中的系统功能和规范要求，内容涵盖了实施原则，系统架构，功能要求。

本文件适用于基因数据审计服务提供者规范业务活动，可作为主管监管部门，第三方评估机构对基因数据应用业务进行监督管理及评估时的依据。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 25069 信息安全技术 术语

GB/T 35890 高通量测序数据序列格式规范

GB/T 42752 区块链和分布式记账技术 参考架构

ISO 20387 生物技术—生物样本保藏—生物样本库通用要求

3 术语和定义

GB/T 25069、GB/T 35890、GB/T 42752、ISO 20387界定的以及下列术语和定义适用于本文件。

3.1

区块链 blockchain

使用密码技术链接将共识确认过的区块按顺序追加形成的分布式账本。

3.2

链式数据结构 chained-block data structure

一段时间内发生的事务处理以区块为单位进行存储，并以密码学算法将区块按时间顺序连接成链条的一种数据结构。

3.3

共识算法 consensus algorithm

区块链系统中各分布的节点对事务或状态的验证、记录等行为达成一致确认的方法。

3.4

智能合约 smart contract

存储在分布式记账技术系统中的计算机程序。

3.5

摘要生成 digest generation

又称摘要函数（或称Hash函数），通常通过将任意长度的消息输入变成固定长度的短消息输出来保障数据的完整性。

3.6

联盟链 consortium blockchain

由一组利益相关的参与者使用，仅有授权节点可接入，接入节点可按规则参与共识和读写数据的一类区块链部署模型。

3.7

区块链存证 blockchain proof of existence

为了保证存证信息（电子数据）的完整性和真实性，采用区块链技术实现多节点共识的存证服务。

3.8

基因 gene

控制生物性状的遗传信息，通常由 DNA 序列来承载，亦即一段具有功能性的 DNA 或 RNA 序列。

3.9

基因数据 genetic data

与生物先天遗传或后天获得的基因特征相关的数据，一般通过对该生物样本的基因检测得到，能提供关于该生物体生理或健康方面独特的信息。

3.10 **关联数据 associated data**

任何与基因数据有关的信息。

注：包括但不限于表型、临床、处理过程信息。

[来源：ISO 20387:2018，3.3]

4 缩略语

下列缩略语适用于本文件。

API：应用程序接口（Application Programming Interface）

BaaS：区块链即服务（Blockchain as a Service）

CA：认证授权（Certification Authority）

DLT：分布式账本技术（Distributed Ledger Technology）

ECC：椭圆曲线加密（Elliptic Curve Cryptography）

ECDSA：基于椭圆曲线数学的数字签名算法（Elliptic Curve Digital Signature Algorithm）

PBFT：实用拜占庭容错共识机制（Practical Byzantine Fault Tolerance）

POS：权益证明共识机制（Proof of Stake）

POW：工作量证明共识机制（Proof of Work）

RSA：一种公钥密码学算法，用于加密和数字签名（Rivest－Shamir－Adleman）

SM2：一种公钥密码学算法，广泛用于数字签名、加密和密钥交换

5 实施原则

5.1 合法合规原则

严格遵守国家相关法规及生物信息安全监管要求，并为监管审计提供必要的技术支持。

5.2 可追溯原则

所有业务和操作均有详实记录，以保证可追踪性和可审计性。

5.3 数据一致性原则

链上链下的数据应保持一致性，区块链的各节点间数据保持统一。

5.4 分级授权原则

实施分级授权制度，明确不同角色的权限，确保数据访问的安全性。

5.5 隐私保护原则

确保用户的隐私权益，防止隐私泄露事件。

5.6 安全原则

采取所有必要措施确保链上链下数据的安全性。

5.7 业务导向原则

以业务需求为导向进行技术研发，优先考虑实际的业务应用场景。

6 系统架构

6.1 架构图

基因数据区块链平台功能组件分层架构见图1，架构分为五个层次：用户层，接入层，核心层，基础层和一个横跨各层的跨层功能集合。

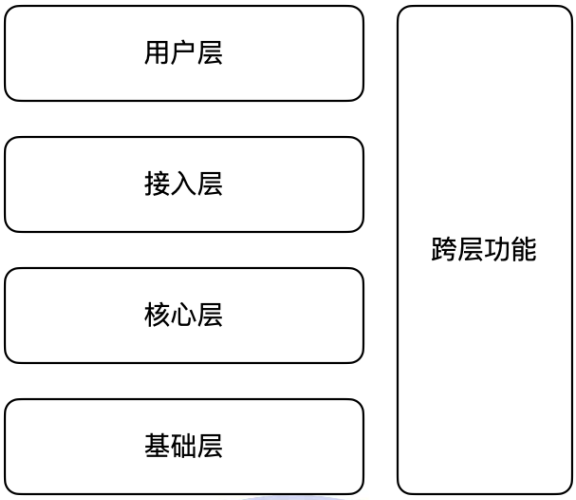

图1 基因数据区块链平台功能组件分层架构

6.2 用户层

为使用方提供与基因数据区块链服务的交互入口，通过此层进行相关的管理功能操作和维护，并有能力提供跨层的服务支持。

6.3 接入层

针对用户层或终端应用，它提供高效、可靠和通用的访问能力，这包括封装核心层的功能、提供高效的缓存和负载均衡，以及标准化的接入协议。

6.4 核心层

集成节点间的共识机制、隐私保护、加密和数字签名等模块，确保系统的完整性、安全性和不可篡改性，并能根据不同业务场景定制智能合约和统一服务。

6.5 基础层

基础层提供所有必要的运行环境和基本组件。

注：如数据存储、执行容器和通讯网络等。

6.6 跨层功能

跨层功能提供能在多个层次之间发挥作用的功能组件。

7 功能要求

7.1 功能组件通用集

基因数据区块链功能组件通用集见图2。

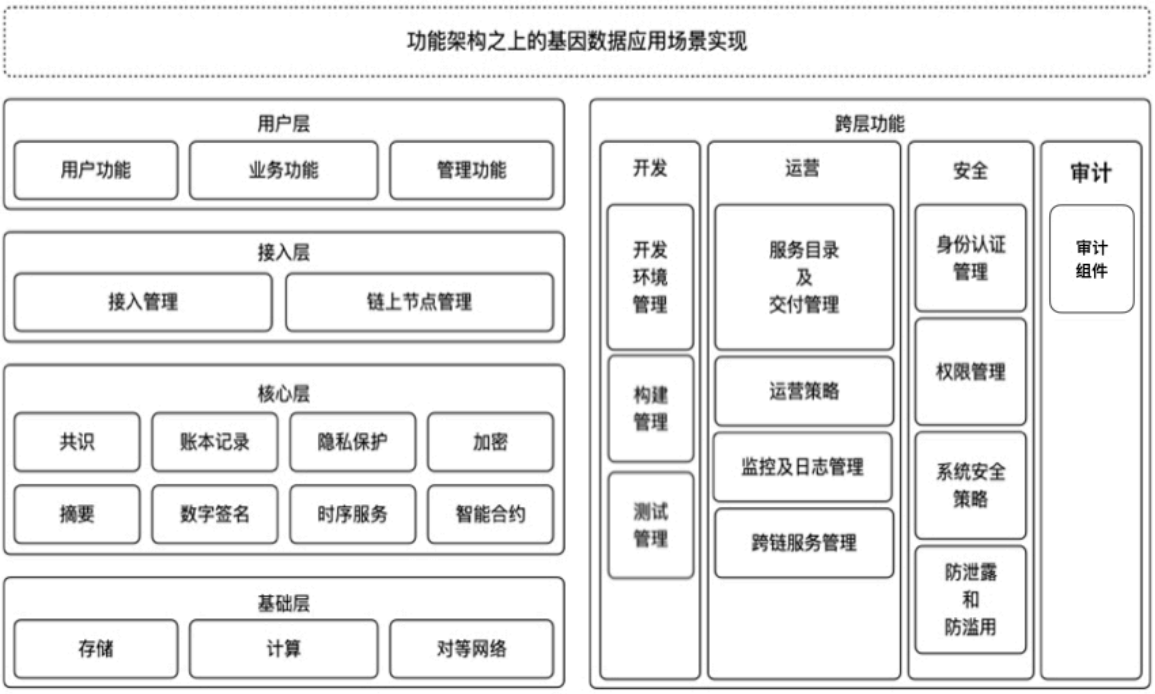

图2　基因数据区块链功能组件通用集

7.2　用户层功能要求

用户功能组件为使用方提供对基因数据审计服务的访问和使用，主要涉及基础资源，链的操作和智能合约的处理。

7.2.1　用户功能

包括命令行界面、图形用户界面和应用程序接口等多种形式。

7.2.2　业务功能

允许使用方提交特定的事务请求（如查询、更新）到基因数据审计网络。

7.2.3　管理功能

管理功能组件为使用方提供包括成员管理、服务活动监控、事件处理、问题报告和安全管理在内的多种服务功能。

7.3　接入层功能要求

7.3.1　接入管理

接入管理功能应包括：

a)　应提供使用方账户的基本信息查询、基因区块链的区块和事务详情查询、及将使用方的特定事务请求提交到基因数据审计网络；

b)　接入管理功能应考虑：接口服务的能力管理（如接口调用频率和查询缓存设置）、接口的访问权限配置（如为不同用户设置不同权限）和接口的通讯安全（如通讯报文加密）。

7.3.2　协议管理

管理加入网络的设备所应遵循的技术规范，涉及硬件、软件和端口等关键技术参数。

7.3.3　链上节点管理

链上节点管理功能应提供对基因数据审计节点的信息查询及进行相关管理操作。其功能应包括：

a) 查询节点服务器的实时状态信息；
b) 控制节点服务的启动和关闭；
c) 配置节点服务的能力参数；
d) 监控节点的网络连接状态；
e) 管理节点的授权和权限配置。

7.4 核心层功能要求

7.4.1 共识机制

共识机制应选择合适的共识算法来构建共识机制，其功能应包括：
a) 支持多节点共同参与共识过程并完成确认；
b) 允许独立节点对提交到基因数据审计网络的信息进行合法性验证；
c) 阻止单一节点在未经其他共识节点确认的情况下进行信息更改；
d) 具有容错能力，能够处理包括节点的物理故障、网络连接问题、非法控制等各种可能的错误和异常。

7.4.2 账本管理

账本管理功能应包括：
a) 提供持久化的账本存储机制；
b) 保证多个节点都能保存完整的区块链存证记录；
c) 仅向授权的参与者提供真实且完整的数据查询；
d) 确保各节点间相同账本数据的完全一致性；
e) 通过多节点的共同维护和记录，实现数据的公开、防篡改和可靠存储。

7.4.3 隐私保护

隐私保护功能应包括：
a) 利用认证机构代表用户进行交易，使用户信息和行为细节不在区块链网络中公开；
b) 替代全网广播，仅在授权的节点之间传输数据；
c) 使用密钥控制对用户数据的访问，只有持有密钥的用户才能解密；
d) 采用高级隐私技术如零知识证明、环签名或同态加密来进一步确保隐私。

7.4.4 数据加密

数据加密组件为数据提供保护，应包括加密和解密操作。所采用的加密算法应符合国家密码管理局的规定。

7.4.5 摘要生成

摘要生产功能应包括：
a) 通过摘要以确保数据的完整性校验；
b) 根据给定的明文和摘要值，验证数据是否遭到篡改。

7.4.6 数字签名

数字签名功能应包括：
a) 提供数字签名和验证功能，确保信息的机密性、完整性和不可否认性；
b) 与权威的第三方CA机构签发的数字证书集成。

7.4.7 时序服务

时序服务功能应包括：
a) 为账本记录提供统一的时序；
b) 具备容错能力来应对可能的时序错误；
c) 能够集成如国家授时中心等的可信第三方时序服务；

d) 时序在基因数据审计区块链系统中扮演着重要的角色，确保行为或数据的时间一致性。系统可以选择引入特定的时序工具或机制。

7.4.8 智能合约

智能合约功能应满足：

a) 提供合适的编程语言与开发环境；
b) 支持对合约内容的静态与动态检查；
c) 有能力在虚拟机等运行载体上执行合约；
d) 保障合约内容的完整性并抵御篡改；
e) 在多方共识的背景下支持合约的升级；
f) 与外部数据交互的智能合约，其影响范围应仅限于合约内部，避免对整体系统造成影响。

7.5 基础层功能要求

7.5.1 存储

存储功能应包括：

a) 在点对点网络中，每个节点都能够部署并使用；
b) 高效、安全、稳定地提供数据服务；
c) 对于采用分库分表的策略，能够实现数据的分片与路由处理；
d) 存储组件负责处理区块链生成的数据，例如账本和交易信息，并提供写入与查询功能。

7.5.2 计算

计算功能应包括：

a) 为区块链系统提供运行环境；
b) 在点对点网络中，每个节点都能够采用此组件；
c) 计算组件支持区块链系统的计算需求，涵盖容器技术、虚拟机技术、云计算等。

7.5.3 对等网络

对等网络功能应满足：

a) 提供高效、安全的点对点通信；
b) 基于点对点通信的多播能力；
c) 支持节点动态的增减和识别；
d) 区块链系统基于分布式对等网络的底层结构，利用对等网络协议组织各个节点。节点间通过点对点协议交换信息，以支撑上层功能。

7.6 跨层功能要求

7.6.1 开发组件

7.6.1.1 开发环境管理

开发环境管理功能应包括：

a) 提供配置元数据的生成支持；
b) 支持服务配置脚本及组件的定制与生成；
c) 开发环境管理应应包含元数据生成和服务配置脚本的定制功能。

7.6.1.2 构建管理

构建管理功能应包括：

a) 提供自动化构建和编译功能；
b) 显示编译错误与提示；
c) 实施构建过程审核流程；
d) 支持多语言和多平台。

7.6.1.3 测试管理

测试管理功能应包括：

a) 管理测试计划、方案、报告及用例；
b) 自动产出测试报告；
c) 确保测试不影响生产环境；
d) 支持测试流程自动化；
e) 管理测试用例库和测试数据库。

7.6.2 运营组件

7.6.2.1 服务目录与交付管理

服务目录与交付管理应详细记录基因数据区块链服务相关的技术和流程信息。

7.6.2.2 运营策略

运营策略应考虑业务、技术、安全、隐私及认证等关键领域的策略。

7.6.2.3 监控与日志管理

监控与日志管理功能应包括：

a) 监控区块链网络节点进程、通信状态和共识效率；
b) 存储和分析节点运行日志。

7.6.2.4 跨链服务管理

连接相关 FBSP 的运营、业务及管理系统。

7.6.3 安全组件

安全组件功能应包括：

a) 身份认证管理；
b) 权限管理；
c) 系统安全策略；
d) 支持多种身份验证方法；
e) 认证备案、签名和验签功能；
f) 设置访问和资源使用权限；
g) 通信加密、数据加密存储；
h) 节点主机安全、智能合约验证；
i) 综合安全管理组件；
j) 数据获取合规。

7.6.4 审计组件

审计组件功能应包括：

a) 实现不可篡改、可追溯、可审查的记录；
b) 明确审计治理规则；
c) 保存服务、资源、性能相关数据；
d) 允许审计机构加入网络实施实时审计；
e) 保存与审计相关数据和证据；
f) 允许审计方实时访问或定期获取数据；
g) 实时核查并用作审计证据；
h) 与其他系统对接与数据提取；
i) 确保基因数据区块链服务满足审计要求，防止基因数据资源的非法行为。

参 考 文 献

[1] GB/T 5271.18 信息技术词汇 第 18 部分：分布式数据处理
[2] GB/T 11457 信息技术 软件工程术语
[3] GB/T 32399 信息技术 云计算 参考架构
[4] YY/T 1723 高通量基因测序仪