

团体标准

T/GBA xxx—2023

区块链 基因数据应用审计实施规范

Blockchain Genetic Data Application Audit Implementation Standards



20XX-XX-XX 发布

20XX-XX-XX 实施

粤港澳大湾区标准创新联盟 发布

目 次

前言 1

引言 2

1 范围 3

2 规范性引用文件 3

3 术语和定义 3

4 缩略语 4

5 实施原则 4

6 系统架构 5

7 功能要求 5

参考文献 11



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XXXXXX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：

本文件起草人：

本文件为首次发布。



引 言

鉴于分布式商业应用的需求日益增长，分布式技术凭借其卓越的弹性、高效的经济性、对等网络与容错能力逐步得到广泛应用，其中，区块链技术尤为突出。这项技术融合了分布式架构、存储、块链数据结构、点对点网络、共识算法、隐私保护及智能合约等先进信息技术，预示着信息技术的创新变革。

“区块链”与“分布式账本技术”两词在实践中常常交替使用。然而，区块链技术在严格意义上描述了一个在对等网络环境中，通过明确、可靠的规则，建立不可篡改、可追踪的块链数据结构来管理信任数据的方法。另一方面，分布式账本技术的范围更为广大，其重点在于通过复制和共享的方式处理事务，而不特定于某种数据结构。因此，实践中常见这两技术的混合应用，为了本文的简洁性，我们使用“区块链”来代指两者。

在生命科学领域的基因数据应用中，由于其特殊的业务需求与监管要求，单一实体控制的私有链或公有链都无法完全满足。为此，业界采取了将传统分布式技术与生物信息技术相结合的方法，进一步探索联盟链技术。鉴于各机构在区块链技术研发上的差异，制定区块链技术标准化规范显得尤为重要。



区块链 基因数据应用审计实施规范

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、系统框架、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，也适用于主管监管部门、第三方评估机构对基因数据应用业务进行监督、管理、评估时参考。

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。

GBT 25069-2022 信息安全技术 术语
GB/T 35890-2018 高通量测序数据序列格式规范
GB/T 42752-2023 区块链和分布式记账技术 参考架构
ISO 20387-2018 生物技术生物建库生物建库一般规则

3 术语和定义

GBT 25069-2022、GB/T 35890-2018、GB/T 42752-2023、ISO 20387-2018 界定的以及下列术语和定义适用于本文件

3.1

区块链 blockchain

一种在对等网络环境下，通过透明和可信规则，构建不可伪造、不可篡改和可追溯的块链式数据结构，实现和管理事务处理的模式。注：事务处理包括但不限于可信数据的产生、存取和使用等。

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

以数字形式定义的能够自动执行条款的合约。

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 genomic data

与生物先天遗传或后天获得的基因特征相关的数据，一般通过对该生物体样本的基因测序得到，能提供关于该生物体生理或健康方面独特的信息。[GDPR, 4.13]

3.10

关联数据 associated data

任何与生物样本及基因数据（3.11）有关的信息，包括但不限于表型、临床、处理过程等信息。[ISO 20387:2018, 3.3]

3.11

数据获取 acquisition

取得生物样本、基因数据（3.11）、关联数据（3.14）。[ISO 20387:2018, 3.2]

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: 一种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、加密和密钥交换中小企业数字化成熟度等级

VCF: 变异识别格式 (Variant Call Format)

5 实施原则

5.1 合法合规原则

应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

所有业务和操作均应有详实记录，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，同样，区块链的各节点间数据也应保持统一。

5.4 分级授权原则

应实施分级授权制度，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

应确保用户的隐私权益，防止任何隐私泄露事件。

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

基因行业区块链平台整合并描述其多样化的功能组件。在架构中，各种特定的功能被有序地归类于不同的层次，这些层次之间的模块和组件通过精确定义的接口进行互动。区块链平台功能组件的分层架构示意图（图1），架构总体上分为五个层次：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。具体如下：

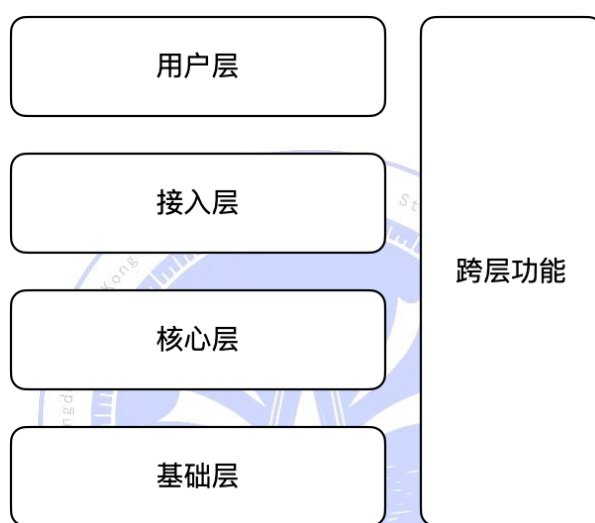


图1 基因数据区块链平台功能组件分层框架

6.1 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。

6.2 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。

6.3 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。

6.4 基础层

应提供所有必要的运行环境和基本组件，如数据存储、执行容器和通讯网络等。

6.5 跨层功能

应支持能在多个层次之间发挥作用的功能组件。

7 功能要求

功能组件通用集（图2），展示了基因行业区块链平台的各功能组件在具体的参考架构实现中的实现。



图2 基因数据区块链功能组件通用集

7.1 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。它在大多数使用场景下，主要涉及基础资源、链的操作和智能合约的处理。具体的技术规范包括：

7.1.1 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。

7.1.2 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。

7.1.3 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的多种服务功能。

7.2 接入层功能要求

接入层功能主要涉及接入管理、协议管理和链上节点管理三个方面。

7.2.1 接入管理

接入管理为终端和用户的核心服务，包括账户信息查询、事务详情查看，以及如何安全、高效地管理接入请求。

- a) 接入管理功能为终端应用和用户层提供与核心层的接入服务。
- b) 应提供以下服务：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。
- c) 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）。

7.2.2 协议管理

协议定义加入网络的设备所需遵循的技术规范，涉及硬件、软件和端口等关键技术参数。

7.2.3 链上节点管理

节点是区块链网络中的基础实体，由安装特定区块链软件、能够连接至互联网、拥有独立IP地址，并能为外部提供服务的物理或虚拟服务器组成。

链上节点管理功能不仅提供对基因数据审计节点的信息查询，还进行相关管理操作。其技术规范应满足：

- 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。

7.3 核心层功能要求

核心层的功能组件包括：共识机制、账本管理、隐私保护、数据加密、摘要生成和数字签名。此外，还可考虑引入时间同步服务和智能合约处理。

7.3.1 共识机制

根据业务需求，可以选择合适的共识算法来构建共识机制，如POW、POS和PBFT等。共识机制组件应具备：

- 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。

7.3.2 账本管理

账本是区块链中用于记录和管理分布式数据的存储结构。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。账本记录功能组件应具备：

- 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。

7.3.3 隐私保护

隐私保护是为确保区块链应用中的用户身份和交易详情等敏感信息不被非法获取或泄露。只有经过充分的授权，这些信息才能被获取。基因数据审计的隐私保护主要通过数据加密和访问控制实现。隐私保护功能组件应支持以下策略：

- 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。

7.3.4 数据加密

加密组件为数据提供保护，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范。

7.3.5 摘要生成

摘要组件应确保：

- a) 数据完整性保护；
- b) 根据给定的明文和摘要值，验证数据是否遭到篡改。

7.3.6 数字签名

数字签名保证数据的完整性和其来源的不可否认性。它结合了非对称加密和数字摘要技术。具体的操作包括创建数字签名和验证签名。根据不同的非对称加密算法，常见的签名算法有RSA、ECDSA和SM2。

数字签名组件应：

- 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。

7.3.7 时序服务

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。时序服务功能组件应当：

- 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。

7.3.8 智能合约

根据业务需要，基因数据区块链系统可以引入智能合约功能。该功能组件需满足：

- 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。

7.4 基础层功能要求

基础层包含如下关键功能组件：存储、计算、对等网络。

7.4.1 存储

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。常用的解决方案包括关系型数据库、键值对数据库和文件数据库等。

存储组件需求如下：

- 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。

7.4.2 计算

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。

具体需求为：

- 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。

7.4.3 对等网络

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。网络传输组件需满足：

- 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

7.5 跨层功能要求

7.5.1 开发组件

开发功能组件应涵盖：开发环境管理、构建管理、以及测试管理。

7.5.1.1 开发环境管理

开发环境管理应包含元数据生成和服务配置脚本的定制功能，具体要求为：

- 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。

7.5.1.2 构建管理

构建管理宜包括自动化构建、错误提示、审核以及多环境支持方面的核心功能和要求。

- a) 提供自动化软件构建和编译功能。
- 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。

7.5.1.3 测试管理

测试管理应包含测试流程、报告自动生成、环境隔离和测试资源管理方面的主要要求与功能。

- 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。

7.5.2 运营组件

运营功能组件宜包括：服务目录及交付管理、运营策略、监控及日志管理、跨链服务管理。

7.5.2.1 服务目录及交付管理

对服务目录及交付管理的详细记录要求。

- a) 详细记录基因数据区块链服务相关的技术和流程信息。

7.5.2.2 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。

- a) 策略包括业务、技术、安全、隐私和认证等方面。

7.5.2.3 监控与日志管理

监控与日志管理是区块链网络节点状态的监控以及节点运行日志的存储与分析的关键功能。

- 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。

7.5.2.4 跨链服务管理

连接相关 FBSP 的运营、业务及管理系统。

7.5.3 安全组件

面对各种安全威胁，如身份、业务、网络、终端、基础和管理类威胁，组件应提供如下功能体系：

- 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

7.5.4 审计组件

确保基因数据审计区块链服务满足审计要求，防止基因数据资源的非法行为。至少应包括：审计支持、易审计、可追溯功能。

7.5.4.1. 审计支持

审计支持宜包含数据不可篡改性、治理规则、数据保存、实时审查、证据保存和系统对接方面的核心要求和功能。

- 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。



参 考 文 献

- [1] GB/T 5271.18-2008 信息技术词汇 第 18 部分:分布式数据处理
- [2] GB/T 11457--2006 信息技术, 软件工程术语
- [3] GB/T 25069-2022 信息安全技术 术语
- [4] GB/T 32399-2015 信息技术 云计算 参考架构
- [5] GB/T 35890-2018 高通量测序数据序列格式规范
- [6] YY/T 1723-2020 高通量基因测序仪

