



# 中华人民共和国国家标准

GB/T XXXXX—XXXX

## 数据基础设施 标识要求

Data infrastructure—Identification requirements

(点击此处添加与国际标准一致性程度的标识)

(征求意见稿)

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX – XX – XX 发布

XXXX – XX – XX 实施

国家市场监督管理总局  
国家标准化管理委员会 发布



目 次

前 言 ..... III

1 范围 ..... 1

2 规范性引用文件 ..... 1

3 术语和定义 ..... 1

4 标识编码规则 ..... 1

    4.1 主体标识码 ..... 1

    4.2 平台、接入连接器标识码 ..... 1

    4.3 数据产品、数据资源与高质量数据集标识码 ..... 2

    4.4 智能体数据标识码 ..... 3

5 标识赋码要求 ..... 3

    5.1 重要功能设施代码标识赋码要求 ..... 3

    5.2 业务节点标识赋码要求 ..... 3

    5.3 接入连接器标识赋码要求 ..... 3

    5.4 数据标识赋码要求 ..... 3

    5.5 智能体数据赋码要求 ..... 4

6 标识解析要求 ..... 4

7 标识安全要求 ..... 4

附 录 A      （资料性） 标识码查询流程 ..... 5

附 录 B      （资料性） 校验码 ..... 6

参 考 文 献 ..... 8



## 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。

本文件由全国数据标准化技术委员会（SAC/TC609）提出并归口。

本文件起草单位：中国电子技术标准化研究院、国家信息中心、中国信息通信研究院、全国组织机构代码统一社会信用代码数据服务中心、国家数据发展研究院、中国工业互联网研究院（工业和信息化部密码应用研究中心）、国家工业信息安全发展研究中心、盐城市盐都区数据局、湖北省公安厅、北京大学、北京邮电大学、浙江大学、上海财经大学、北京大数据先进技术研究院、北京泰尔英福科技有限公司、湖北省标准化与质量研究院、成都市标准化研究院、贵州大数据产业集团有限公司、中国移动通信集团有限公司、中国移动通信有限公司研究院、中移动信息技术有限公司、中国联合网络通信集团有限公司、联通数据智能有限公司、中国稀土集团有限公司、华为云计算技术有限公司、浪潮软件科技有限公司、国家电网有限公司信息通信中心（大数据中心）、中电数据产业集团有限公司、中国南方电网有限责任公司、南方电网能源发展研究院有限责任公司、广州数据交易所有限公司、深圳数据交易所有限公司、山西数据交易中心有限公司、上海数据集团有限公司、中电云计算技术有限公司、长三角数链（上海）网络基础设施有限公司、北京观盟信息技术有限公司、中果优品（北京）科技有限公司、北京中数睿智科技有限公司、数据空间研究院、上海市数字证书认证中心有限公司、数据易（北京）信息技术有限公司、上海零数众合信息科技有限公司、中关村工信二维码技术研究院、浙江科正电子信息产品检验有限公司、浙江蚂蚁密算科技有限公司、蚂蚁科技集团股份有限公司、杭州铭崴信息科技有限公司、杭州润州光电技术有限公司、杭州安恒信息技术股份有限公司、武汉现代精工机械股份有限公司、煤炭科学研究总院有限公司、深圳市尚数网科技有限公司、深圳数鑫科技有限公司、洞见科技（雄安）有限公司、云基华海信息技术股份有限公司、全球数源中心（广州南沙）有限公司、中科斯欧（合肥）科技股份有限公司、高颂数科（厦门）智能技术有限公司负责制定。

本文件主要起草人：徐冻、许紫媛、胡琳、陈海、司宏伟、徐春学、马英、王亦澎、景越、袁博、罗乐、金旭冉、邱朝旭、王献琪、袁军、罗松、张钰雯、刘巍、池程、孔吴为、袁羚匀、田野、赵菁华、何帅、纪守领、赵彬彬、胡延平、罗超然、柳熠、张发振、王帅、张鸿冉、田伯成、李姝婧、谢秋琪、刘莎、白玲玉、姚昕金、王仕品、茹志强、张帆、喻炜、李征、彭巍、王萍、贾子萌、金洁华、李冠洲、伊然、杨瑞、王宇、王一飞、杨瞩远、王项男、谢卫军、黄玮泽、林志强、李华杰、王鑫、赵丽丽、王冠军、辛卫民、郭经、国丽、胡成盛、杨秋勇、陈丰、马一宁、柴雁欣、谭天怡、郑彬彬、赵亮、李榕、李中、王儒、周强、魏涛、王博、李鑫，任重远、倪采、侯宁、韩涵、庾朝富、张璐杰、吕培培、杨晶、崔久强、宾军志、兰春嘉、杨珍、牛小兵、孙璐、欧阳小珊、张晓蒙、潘无穷、韦韬，昌文婷、李帜、孙琪、李岳政、田丽丹、王吾冰、詹智勇、武光城、张建中、吴桂荣、廖炳才、张艺贝、李博、姚明、鲁胜强、张丹丹、张乐思、宣秀芳、于万钦、许杰焜。



# 数据基础设施 标识要求

## 1 范围

本文件规范了数据基础设施标识的术语和定义、标识编码、赋码、解析及安全要求等。

本文件适用于数场、可信数据空间、数联网、数据元件、隐私保护计算、区块链等技术体系支撑的各类层级的数据基础设施建设，包括区域、城市、行业、企业、个人等数据基础设施。

## 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 32100-2015 法人和其他组织统一社会信用代码编码规则

GB 11643-1999 公民身份号码

GB/T 4754-2017 国民经济行业分类

GB/T 17710-2008 信息技术 安全技术 校验字符系统

GB/T 2260-2007 中华人民共和国行政区划代码

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**标识对象 identification object**

数据基础设施涉及的特定物理或逻辑对象，包括机构、人、物品、数据产品、数据资源、平台系统等。

### 3.2

**标识码 identification code**

标识码是对标识对象的唯一编码，本文件所述标识指标识码。

## 4 标识编码规则

### 4.1 主体标识码

主体标识码固定为18位，主体类型及其标识码编码规则如下：

- 法人和非法人组织。主体标识码为统一社会信用代码，其编码规则应符合 GB 32100 的要求。
- 自然人。主体标识码为中华人民共和国居民身份证号码，其编码规则应符合 GB 11643 的要求。
- 其他类型主体。主体标识码可向标识码登记注册平台申请，由平台分配 18 位数字或字母作为其主体标识码。

### 4.2 平台、接入连接器标识码

平台标识码共32位，由类型码（1位）、主体标识码（18位）、重要功能设施代码（4位）、随机码（8位）、校验码（1位）组成，编码结构见图1：

- 底座编码。类型码为“1”、主体标识码（18位）、重要功能设施代码（4位）、随机码（8位）、校验码（1位）。
- 重要功能设施代码编码。类型码为“2或3”、主体标识码（18位）、重要功能设施代码（4位）、随机码（8位）、校验码（1位）。

- c) 业务节点编码。类型码为“4”、主体标识码（18位）、重要功能设施代码（4位，业务节点编码由重要功能设施分配时用重要功能设施代码）、随机码（8位）、校验码（1位）。
- d) 接入连接器编码。类型码为“5”、主体标识码（18位）、重要功能设施代码（4位，接入连接器编码由重要功能设施分配时用重要功能设施的代码）、随机码（8位）、校验码（1位）。
- 注：主体标识码。编码规则参见4.1。

重要功能设施按照区域分类时，类型码为“2”，参照GB/T 2260 第5章中的前4位数字作为本文件区域代码；按照行业分类时，类型码为“3”。其中1~2位固定为数字“00”，3~4位对应GB/T 4754第5章中规定2~3位作为本文件行业代码。

随机码：当用户申请多个标识码出现代码重复时，随机码进行区分，确保标识码唯一性。

校验码：校验码采用GB/T 17710的有关规定，见附录B。

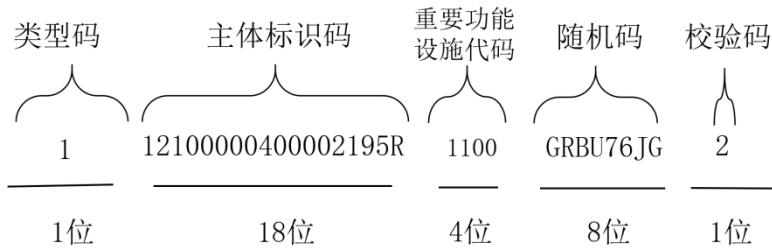


图1 底座标识码示例图

4.3 数据产品或数据资源标识码

数据产品或数据资源编码共32~65位，由类型码（1位）、主体标识码（18位）、重要功能设施代码（4位）、随机码（8位）、校验码（1位）、扩展码（最多32位）组成，如选择扩展码，扩展码之前加“•”（1位）分隔：

- a) 类型码。数据产品类型编码为“6”，数据资源类型编码为“7”。
- b) 主体标识码。编码规则参见4.1。
- c) 重要功能设施代码。与受理平台的重要功能设施代码一致。如第1位是大写字母“A”，则标识为数据集。
- d) 随机码。当用户申请多个标识码出现代码重复时，随机码进行区分，确保标识码唯一性。
- e) 校验码。校验码采用GB/T 17710的有关规定，见附录B。
- f) 扩展码。为可选项，编码长度为1~32位，可兼容其他标识编码，不选时省略。

如某数据产品分配的标识码为：

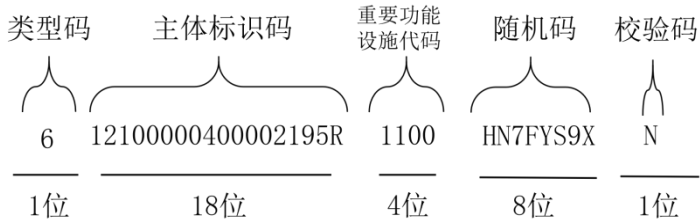


图2 数据产品标识码示例图

如某平台登记的数据资源码为：12（登记类型）15（行业分类）20250228（日期码）1BH95JZY（随机码），为其分配的标识码为：



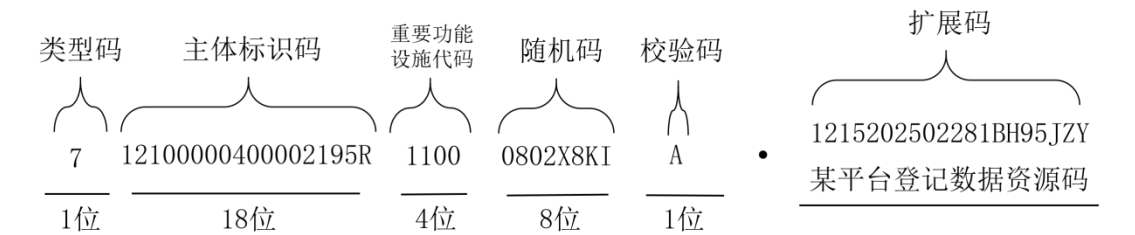


图 3 数据资源标识码（含扩展码）示例图

4.4 智能体数据标识码

智能体数据编码共32~65位，由类型码（1位）、智能体主体标识码（18位）、重要功能设施代码（4位）、随机码（8位）、校验码（1位）、扩展码（最多32位）组成，如选择扩展码，扩展码之前加“.”（1位）分隔：

- a) 类型码。智能体数据编码为“8”。
- b) 智能体主体标识码。智能体主体标识码可向标识码登记注册平台申请，由平台分配 18 位数字或字母作为其主体标识码。
- c) 重要功能设施代码。与受理的重要功能设施代码一致。
- d) 随机码。当智能体申请多个标识码出现代码重复时，随机码进行区分，确保标识码唯一性。
- e) 校验码。校验码采用 GB/T 17710 的有关规定，见附录 B。
- f) 扩展码。为可选项，编码长度为 1~32 位，可兼容智能体生成编码，不选时省略。

如某平台登记的智能体数据编码为：6K8P2X7M9N3Q5Z1W4V0T8（扩展码），为其分配的标识码为：

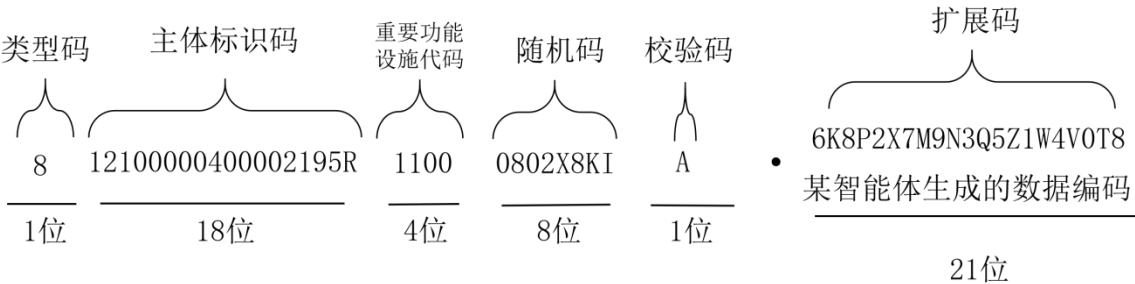


图 4 智能体标识码示例图

5 标识赋码要求

5.1 重要功能设施代码标识赋码要求

重要功能设施代码标识码由底座分配，标识码一旦分配不可变更。

5.2 业务节点标识赋码要求

业务节点向重要功能设施代码注册时，由重要功能设施向业务节点分配标识码，标识码一旦分配不可变更。

5.3 接入连接器标识赋码要求

数据主体向重要功能设施、业务节点注册接入连接器时，由重要功能设施分配接入连接器标识码。

5.4 数据标识赋码要求

数据标识含数据产品标识或数据资源标识，赋码要求如下：

- a) 用户通过业务节点或重要功能设施登记数据产品/数据资源时，由重要功能设施分配标识码。
- b) 用户通过接入连接器首次登记某一数据产品/数据资源时，由重要功能设施为该数据产品/数据资源分配标识码。

- c) 其他登记平台可通过重要功能设施提供的登记接口进行数据资源或数据产品登记，重要功能设施为其分配相应标识码，原有标识可作为扩展码保留，完成原有标识与新分配标识码转换。
- d) 如登记的数据产品为数据集时，由底座统一分配标识码。

### 5.5 智能体数据赋码要求

依托智能体生产的数据，赋码要求如下：

- a) 智能体数据赋码的责任主体为机构、组织或自然人。
- b) 用户通过业务节点或重要功能设施登记智能体数据时，由重要功能设施登记、备案标识码。
- c) 其他登记平台可通过重要功能设施提供的登记接口进行智能体数据登记，由平台向重要功能设施登记、备案标识码，完成原有标识与新分配标识码转换。

## 6 标识解析要求

底座、重要功能设施、业务节点应配置相关解析模块，共同构成标识解析体系。标识解析要求如下：

- a) 底座、重要功能设施、业务节点应支持使用标识码跨域查询，提供相关节点查询服务目录，查询流程见附录 A。
- b) 底座应提供标识编码、赋码、解析、寻址、维护等能力，维护标识信息库。
- c) 应支持使用标识码作为解析标识符，使用标识码可解析出底座、重要功能设施、业务节点、接入连接器、数据产品、数据资源等标识对象的位置信息、状态信息或者其他相关信息。
- d) 底座、重要功能设施、业务节点需配置标识解析模块，完善标识与 IPv6 等网络标识的映射关系，可通过 DNS 解析等机制实现标识解析。

## 7 标识安全要求

标识码是身份确认、节点平台互联的关键，应满足标识安全性要求：

- a) 标识使用环境安全：涉及到标识的传输、存储时应符合国家信息安全等级保护、商用密码应用安全性评估等要求。
- b) 标识管理设施安全：涉及标识赋码、查询或解析的平台节点，宜建立标识码冗余检测与纠错机制、唯一性校验、高并发处理、全链路日志审计及容灾恢复等处理方案。

附 录 A  
(资料性)  
标识码查询流程

使用标识跨域查询位置的流程描述如下：

- a) 用户在业务节点或重要功能设施对数据标识发起查询请求。
- b) 业务节点查询本地标识数据库并返回结果，查询完成。
- c) 如未查询到该数据标识，业务节点将查询请求发至其对接的重要功能设施，重要功能设施查询本地标识数据库并返回结果，查询完成。
- d) 如重要功能设施未查询到该数据标识则将查询请求发至底座，由底座查询标识数据库并返回结果（该标识所在的重要功能设施或业务节点地址），查询完成。

附录 B  
(资料性)  
校验码

1 校验码

校验码采用GB/T 17710 2008, MOD 37-36校验码系统。

1.1 校验公式

标识码中各个位置上的号码字符值应满足下列公式的校验：

$$((\dots(((M + a_n) \parallel M \cdot 2) \{M+1\} + a_{n-1}) \parallel M \cdot 2) \{M+1\} + \dots + a_1) \parallel M = 1$$

式中：

n是包含校验字符的串的总字符数；

i是从右侧数的字符位置（最右端i = 1，空格/分隔符不算）；

a<sub>i</sub>是i位置对应的字符值（表1）；

M是字符集的字符数目，是偶数；M和M + 1是两个模数；

||M是模M取余，如果余数是0，就用M代替；

|M + 1是模 M + 1 取余，这里余数不会是0。

表 B.1 标识码中a<sub>i</sub>数值

字符	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G	H
值	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
字符	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
值	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35

1.2 混合系统递归计算

生成校验字符a<sub>1</sub>的计算法则的描述为：用j=1… (n-1)，n为包含校验字符在内的串中字符的数目。

当j=1时，定义P<sub>j</sub>=M，公式如下：

$$S_j = P_j \parallel M+1 + a_{n-j+1}$$

$$P_{j-1} = S_j \parallel M \cdot 2$$

式中：

||<sub>M</sub>——除以M后的整余数，如果为0，则M值应被替换；

||<sub>M+1</sub>——除以（M+1）后的整余数，在经过上述处理后该余数不会为0；

a<sub>n-j+1</sub>——字符值。

所以下一个a<sub>1</sub>：

$$P_n + a_1 \equiv 1(\text{mod } M) \text{ 或 } a_1 = -(1 - P_n)(\text{mod } M)$$

验证效验字符a<sub>1</sub>的算法则描述为：用j=1, …, (n-1)，n为包含校验字符在内的串中字符的数目。

当j=1时，定义P<sub>j</sub>=0。公示如下：

$$S_j = P_j \parallel M+1 + a_{n-j+1}$$

$$P_{j+1} := S_j \|_M \cdot 2$$

如果 $S_n \equiv 1(\text{mod } M)$ 则串正确。

参 考 文 献

- [1] 《国家数据基础设施建设指引》
-