

# 政务数据标识技术研究进展 及下一代政务数据标识体系

王昀<sup>1,2</sup>, 郭毅峰<sup>2</sup>, 苏晓亮<sup>2,3</sup>, 周武爱<sup>2</sup>, 张皖哲<sup>2</sup>, 许大虎<sup>2</sup>, 周强<sup>2</sup>, 冯建华<sup>1</sup>

1. 清华大学计算机科学与技术系, 北京 100084;
2. 中移信息系统集成有限公司, 北京 100032;
3. 清华大学航天航空学院, 北京 100084

## 摘要

政务数据标识是建设全国一体化政务大数据体系的一项基础性工作。对数据标识技术的研究进展进行了总结, 比较了不同数据标识技术编码规则的异同, 并进一步总结了政务数据标识及应用进展。结合政务数据所具有的权责明确、安全性要求高、兼容性需求强等特点, 提出了下一代政务数据标识体系Gcode。Gcode由外部码、内部码和安全码3个部分组成。其中, 外部码兼容了统一社会信用代码, 内部码建立了“机构-部门-系统-数据”的关联关系, 安全码通过引入区块链技术实现防伪验真。Gcode具有权责明确、兼容性强、安全性高等特点, 能够支持政务数据跨层级、跨地域、跨系统、跨部门、跨业务共享, 可有力推动实现政务数据“一数一源”。

## 关键词

政务数据; 数据标识; 数据治理

中图分类号: D035

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2024004

## *Research progress of government data identification technology and the next generation government data identification system*

WANG Yun<sup>1,2</sup>, GUO Yifeng<sup>2</sup>, SU Xiaoliang<sup>2,3</sup>, ZHOU Wuai<sup>2</sup>, ZHANG Wanzhe<sup>2</sup>, XU Dahu<sup>2</sup>, ZHOU Qiang<sup>2</sup>, FENG Jianhua<sup>1</sup>

1. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
2. China Mobile Information System Integration Co., Ltd., Beijing 100032, China
3. School of Aerospace Engineering, Tsinghua University, Beijing 100084, China

## *Abstract*

Government data identification is a fundamental work in building a national integrated government big data system. This article reviewed the research progress of data identification technology, compared the similarities and differences in coding rules of different data identification technologies, and further reviewed the progress of government data identification

and its applications. Based on the clear rights and responsibilities, high security requirements, and strong compatibility requirements of government data, the next generation government data identification system Gcode was proposed. Gcode consists of three parts: external code, internal code, and security code. Among them, the external code was compatible with the Code for Unified Social Information, the internal code established an association relationship of "institution-department-system-data", and the security code achieved anti-counterfeiting verification by introducing blockchain technology. Gcode has clear rights and responsibilities, strong compatibility and high security, and can support cross-level, cross-region, cross-system, cross-department, and cross-business sharing of government data, effectively promoting the implementation of "one data, one source" of government data.

### *Key words*

government data, data identification, data governance

## 0 引言

2020年4月,中共中央、国务院印发《关于构建更加完善的要素市场化配置体制机制的意见》,强调要“加快培育数据要素市场”。2021年3月通过的《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》(以下简称“十四五”规划)中明确提出“建立数据资源产权、交易流通、跨境传输和安全保护等基础制度和标准规范”,对数据要素市场培育工作做出更加明确的战略性部署。2022年10月,国务院办公厅印发的《全国一体化政务大数据体系建设指南》(以下简称《指南》)提出,到2025年,政务数据质量显著提升,“一数一源、多源校核”等数据治理机制基本形成。随着政务大数据体系建设的持续推进,覆盖的领域逐渐拓宽,政务数据标识作为政务数据生产、分配、流通等各环节的基础性支撑工具,日益凸显其重要性。

政务数据标识技术大多由数据标识技术发展而来。国际上广泛使用的标识体系有OID<sup>[1]</sup>、Handle<sup>[1]</sup>、GS1<sup>[2]</sup>和EPC (electronic product code, 电子产品代码)<sup>[1]</sup>。OID编码标识体系是由ISO/IEC、ITU-T

联合提出的标识机制,用于对包括实体、虚拟、复合对象等在内的任何类型数据进行唯一命名。OID编码采用树状结构,不同层次之间的分隔符为“.”,而且层数无限制。Handle编码标识体系由美国国家研究计划机构(The Corporation for National Research Initiatives, CNRI)提出,用于唯一标识互联网上的数字对象,为数字对象提供标识、动态解析和安全管理等服务。Handle编码由前缀、分隔符“/”、后缀组成,其一般形式为“前缀/后缀”。GS1<sup>[3]</sup>编码标识体系起源于美国超级市场的UPC码,继而由欧洲国家发展出EAN码,主要面向商品流通领域。GS1码由厂商识别代码/前缀码、商品项目代码、校验码3个部分组成。EPC编码标识体系是运行在GS1上的编码标准,用于对供应链中的对象(包括物品、货箱、货盘、位置等)进行全球唯一的标识。国内自主研发的标识体系有Ecode<sup>[4]</sup>、CSTR<sup>[5]</sup>、ISLI<sup>[6]</sup>、NIOT<sup>[7]</sup>等。Ecode编码标识体系是我国自主创新研究出来的一整套物联网标识体系标准,用于标识抽象、虚拟对象。与OID和Handle编码标准在网络标识方面不同的是,Ecode支持网络节点位置可变。此外,数据标识技术在科技资源领域的应用是除了物联网领域之外的另一大领域。例如,CSTR编码由中国科技资源代号(China science

and technology resource, CSTR)、科技资源标识注册机构代码、科技资源类型代码和内部标识符4个部分组成。国际标准关联标识符(international standard link identifier, ISLI)规定了信息与文献领域中可被唯一识别的实体之间的关联。ISLI编码由十进制数字构成,包括服务字段、关联字段和校验字段3个部分。其中,关联字段包含定义源和目标之间关系的长度可选编码。国家物联网标识管理公共服务平台NIOT编码由协议标识符、异构标识区和对象编码3个部分组成,分隔符为“:”。国内外数据标识技术比较见表1。

此外,学术界也对构建更加合理、有效的标识体系提出了构想。窦悦等人<sup>[9]</sup>提出的国家“数联网”根服务体系拟建立包含5个层级的统一标识规则体系,分别是国家“数联网”根服务体系国家级节点;聚焦人、企、车、物、地等不同本体对象的二级节点;聚焦区域或行业的三级节点;聚焦各类组织机构的四级节点和聚焦某一独立个体的五级节点。5个层级节点会逐级分配并下发标识规则,每级节点均会继承由上级节点下发的固定标识串码和自定义标识长度,其中,自定义标识是指各级节点可以结合自身实际需求,建立对应本级别别的标识编码规则,从上级继承的固定标识串码连同制定完成的本级标识编码也将成为下一级节点的固定标识串码。此外,能够实现标识互联的技术也已有诸多研究,例如IDMapping<sup>[10]</sup>、Cookie Mapping<sup>[11]</sup>等。

可以看出,主流数据标识技术主要应用于物联网、信息资源检索等领域,其编码规则包括网络标识、对象标识以及校验码。网络标识代码包括根节点、各级区域/行业节点、企业端节点,赋予各级节点的网络存储设备一个前缀码;对象标识实现对的产品标识代码结合使用。其技术路线是

根据网络标识代码找到其节点数字地址,从而找到对象信息存储位置,并获悉对象的具体信息。随着政务数据在评估大数据发展及政府治理和产业发展状况<sup>[12]</sup>等领域或方向得到越来越多的应用,下一代政务数据标识体系应当具备的权责明确、高安全性、强兼容性等特点逐渐受到关注。尤其是,2022年12月2日印发的《中共中央国务院关于构建数据基础制度更好发挥数据要素作用的意见》(以下简称“数据二十条”)明确指出,“根据数据来源和数据生成特征,分别界定数据生产、流通、使用过程中各参与方享有的合法权利,建立数据资源持有权、数据加工使用权、数据产品经营权等分置的产权运行机制……”。这对发展能够充分释放数据要素价值的下一代政务数据标识体系提出了更高的要求。同时,面向数据的体系结构(data-oriented architecture, DOA)所秉承的“面向数据和以数据为核心”思想也为数据要素价值的释放奠定了基础模式<sup>[13]</sup>。为此,本文首先介绍了政务数据标识技术的研究进展及不足;进一步地,针对政务数据标识技术研究存在的不足,提出了下一代政务数据标识体系Gcode,以期发展为下一代政务数据标识技术提供参考。

## 1 政务数据标识技术研究的进展及不足

### 1.1 研究进展

政务数据的标识主要依赖于政府侧推动。早在2005年,《政务信息资源目录体系》的第5部分“政务信息资源标识编码规则”规定政务信息资源标识符的编码由5位前段码和不定长的后段码构成,前后段之间用“/”隔开。其中,5位前段码由

表1 国内外数据标识技术对比

数据标识技术	应用场景	编码规则	特点	不足	文献
OID	任何类型的对象	树状结构,不同层之间用“.”分隔,层数无限制。OID名称包括数字和字母数字两种形式,数字名称的值是一个大于0、小于16 000 000的正整数,字母数字名称形式的值是一个不小于1个字符且不大于100个字符的可变长字符	应用广泛,兼容性强,拓展性强	长标识会增加系统复杂性;存在单点故障与负载过重的问题	[1]
Handle/DOI	数字对象	由前缀、分隔符“/”、后缀三部分组成,即“前缀/后缀”。前缀由若干个数字串和分隔符“:”组成,后缀由任意个UTF-8字符组成的字符串,可自行定义	应用广泛,有利于数据互联互通、数据确权、数据管理及数据保护	缺少政务领域应用案例。没有解决节点解析服务权利不对等的问 题	[1]
GS1	商品流通	由厂商识别代码/前缀码、商品项目代码、校验码3个部分组成。按照代码的位数,又分为13位代码、8位代码和12位代码3种情况。对于13位代码情况,其厂商识别代码由7~10位数字组成,商品项目代码由2~5位数字组成,校验码由1位数字组成。对于8位代码情况,其前缀码有3位,商品项目代码由4位数字组成,校验码由1位数字组成。对于12位代码情况,其厂商识别代码由6~10位数字组成,商品项目代码由1~5位数字组成,校验码由1位数字组成	系统性、科学性、全球统一性和可扩展性	不适用于政务数据量级快速增长、多变的情况	[2]
EPC	物联网体系内物品	由版本号、域名管理者、对象分类、序列号组成的一组数字,分别用于标识EPC的长度或类型、生产厂商的信息、类型、货物单品	科学性、全面性、兼容性、合理性	较为昂贵	[1]
Ecode	抽象的实体、虚拟对象	由版本V、编码体系标识NSI和主码MD三部分构成。其中,V和NSI定义了MD的结构和长度,由Ecode管理机构统一分配,MD为标识对象代码,由标识对象管理方分配	从应用出发,具备后发优势	存在单点故障与负载过重问题,用户体量小	[4]
CSTR	科技资源	由中国科技资源代号(CSTR)、科技资源标识注册机构代码、科技资源类型代码和内部标识符4个部分组成。CSTR和科技资源标识注册机构代码之间用半角符号“:”分隔,其他各部分之间用半角符号“-”分隔	侧重于科技资源的标识	用户体量小,缺少政务领域应用案例	[5]
ISLI	信息与文献及其关系	由十进制数字构成,包括服务字段、关联字段和校验字段3个部分。其中,服务字段包含6位数字构成的服务编码,关联字段包含定义源和目标之间关系的长度可选编码,校验字段是1位数值,用于防止出错	侧重于标识实体之间的关系	不适用于标识对象	[6]
NIOT	物联网对象	由协议标识符、异构标识区和对象编码3个部分组成,分隔符为“:”。其中协议标识符是国家平台标识服务协议代码“niot”。异构标识区是子平台标识代码和上级平台代码组成,用“-”分隔。对象编码由对象所有者或者授权商定义并分配	兼容性强	具有单点失效以及易遭受攻击的风险	[7]
DID	人员、组织和事物等	由Scheme、DID Method和DID Method Specific String三部分组成,用“:”分隔	不依赖于中心化注册机构,可自我控制	不适用于体量庞大的政务、公共数据	[8]

10个阿拉伯数字(0~9)和26个大写英文字母(A~Z)构成,由政务信息资源前段码管理中心统一管理和分配,前段码分配给电子政务网络中拥有政务信息资源管理权的各级目录管理者。不定长的后段码采用《信息技术、信息交换用汉字编码字符集、基本集的扩充》中规定的任意字符构成,由拥有前段码的目录管理者或行政部门自行管理和分配,必须保证在本部门范围内的唯一性。后段码原则上采用顺序号,由程序自动生成<sup>[14-15]</sup>。

在实际应用中,北京市的政务信息资源标识符编码分为前段码和后段码两部分,用“/”连接。其中前段码表示产生或提供信息资源的机构编码,共6位,由10个阿拉伯数字(0~9)和24个大写英文字符(除I和O之外的其他A~Z的字符)组成。前两位“11”代表北京;第三位“0”代表市级,字母代表区县级;后三位代表政务机构或乡镇、街道,由各级前段码的管理机构参考政府各级编办的组织机构代码编写。后段码分为业务类编码、信息资源类编码和其他编码。后段码的第1位表示编码类型,其中业务类编码用Y表示,信息资源类编码用Z表示,其他的编码类型根据需要扩展。考虑到编码规则的兼容性,在编码规则中规定,各国家机关既可以采用已有的业务编码方案和信息资源编码方案,也可以制定本部门的业务编码方案和信息资源编码方案;在缺省情况下,可采用6位无意义顺序码<sup>[16]</sup>。黑龙江省的政务信息资源代码由6位国家代码、不定长的“类”“项”“目”“细目”,以及6位信息资源顺序码构成<sup>[17]</sup>,能够从基础信息、主题、部门等角度对政务信息资源分类和编码。

此外,贾东琴等人<sup>[18]</sup>利用数字对象唯一标识符(digital object unique identifier, DOI)对政府信息资源进行标识,进而系统规范政府信息资源登记管理、实现政

府信息资源有效定位和整合、完善政府信息公开监督和保障制度,保证政府信息能够全面真实地提供给公民。在政府信息资源登记管理方面,可将DOI作为政府信息资源元数据中的应用元素,并规范应用元素的标准体系,进而规范政府数字资源登记管理。运用元数据机制对离散式分布的各类电子政务信息资源进行统一描述和集中化登记管理。在政府信息资源有效定位方面,DOI为用户提供了对数字资源的永久性访问,避免了由资源地址的改变造成用户链接失效的问题。在政府信息资源有效整合方面,将DOI引入政府信息管理中,并通过开放链接功能,实现政府信息资源的有效整合、共享,便于公民对上下级或同级部门的相似或相关信息的发现和查找。

综上所述,研究人员已经明显地意识到政务数据的标识应当关注政务数据自身而非政务数据的物理地址。概括来看,现有政务数据标识技术有两种方案。一种是直接使用较为成熟的数据标识技术,例如基于Handle系统的DOI;另一种是延续现有数据标识技术的思想,构建政务领域的数据标识技术,例如,通过发挥类似网络标识功能的前段码找到数据所在的政府部门,通过发挥类似对象标识功能的后段码找到政府部门内部具体的数据。此外,对政务元数据进行标识的萌芽也逐渐显露。这对于政务数据的统一描述、访问和管理具有重要意义。

## 1.2 存在的不足

不同于物联网<sup>[1]</sup>、工业互联网<sup>[19]</sup>等数据标识技术研究已有较为深厚积累的领域,政务数据标识技术研究仍然欠缺,暂无专门针对政务数据的原创性标识技术。政务数据标识研究面临着明权责性、高安全性、强兼容性的挑战。

在明权责性方面,“数据二十条”针对政务数据各个流通环节的数据权利与职责提出了明确要求,能够明确权责的政务数据标识技术是充分发挥政务数据要素市场价值的基础性工作。以统一资源定位系统(uniform resource locator, URL)<sup>[20]</sup>为代表的信息资源标识技术落脚于标识信息资源的物理位置。一旦信息资源的物理位置发生改变,URL无法正常工作。URL的这种性质决定了其无法对数字对象进行长期、稳定的定位,无法明确政务数据的所属机构、部门和业务系统,从而对政务数据权责明确造成障碍。

在高安全性方面,政务领域的安全性要求就是责任要求,政务数据标识技术必须具备安全性技术和制度保障。政务数据有别于普通数据之处在于其包含着个人或主体的隐私信息,因而在共享或开放过程中需要保障数据安全。现有数据标识技术在安全性要求方面普遍无法满足,尤其缺少对数据生产、流通、使用全过程的动态安全保障措施。

在强兼容性方面,政务数据标识技术能够兼容现有的不同地区、不同行业甚至不同部门已建数据标识体系是顺利完成升级改造的关键。缺少统一规范的信息资源描述和管理模式将导致不同标识技术之间兼容性不够。从现有的研究进展情况来看,各级地方政府和部门采用的标准各不相同,存在多种多样的存储格式,没有对资源进行统一描述和管理,造成政府部门间数据难以共享,政务数据服务性不强。

## 2 Gcode:下一代政务数据标识体系

### 2.1 体系构成及动态性

针对现有政务数据标识技术研究存在

的不足,本文提出了由外部码、内部码、安全码3个部分构成的下一代政务数据标识体系Gcode,其结构和各个组成单元的含义如图1所示。

码体系标识指的是政务“码”的体系标识,设定为默认值“Gcode”。字母“G”表示的是“Governance”。为了避免计算机程序处理过程中遇到特殊字符的转义问题,码体系标识和组织/机构编码之间不设置连接符。

组织/机构编码指的是法人和其他组织的统一社会信用代码,其编码规则遵循统一社会信用代码的生成规则。依据《法人和其他组织统一社会信用代码编码规则》的规定,统一代码由18位的阿拉伯数字或大写英文字母(不使用I、O、Z、S、V)组成,包括第1位登记管理部门代码、第2位机构类别代码、第3~8位登记管理机关行政区划码、第9~17位主体识别码(组织机构代码)、第18位校验码5个部分<sup>[21]</sup>。其中,第3~8位的登记管理机关行政区划码依据《中华人民共和国行政区划代码》(GB/T 2260—2007)给出,由6位数字组成,精确到县区级别,故统一社会信用代码包含了地域信息。

部门编码指的是组织/机构内某个部门的编码。组织/机构在设立时,一般会由上级部门出台相应的“三定方案”,规定该组织/机构的职能配置、内设机构和人员编制。根据内设机构设置组织/机构内部各个部门的编码。内设机构的数量通常为十几个至几十个。因此,部门编码的编码规则为:由不少于3的 $M$ 位阿拉伯数字组成,例如:若 $M=3$ ,则从001开始,至999结束,依次类推。

业务系统编码指的是部门内某个业务系统的编码。《指南》指出,要全面摸清政务数据资源底数。为此,通常需要开展系统普查工作,即通过调研并收集组织/机

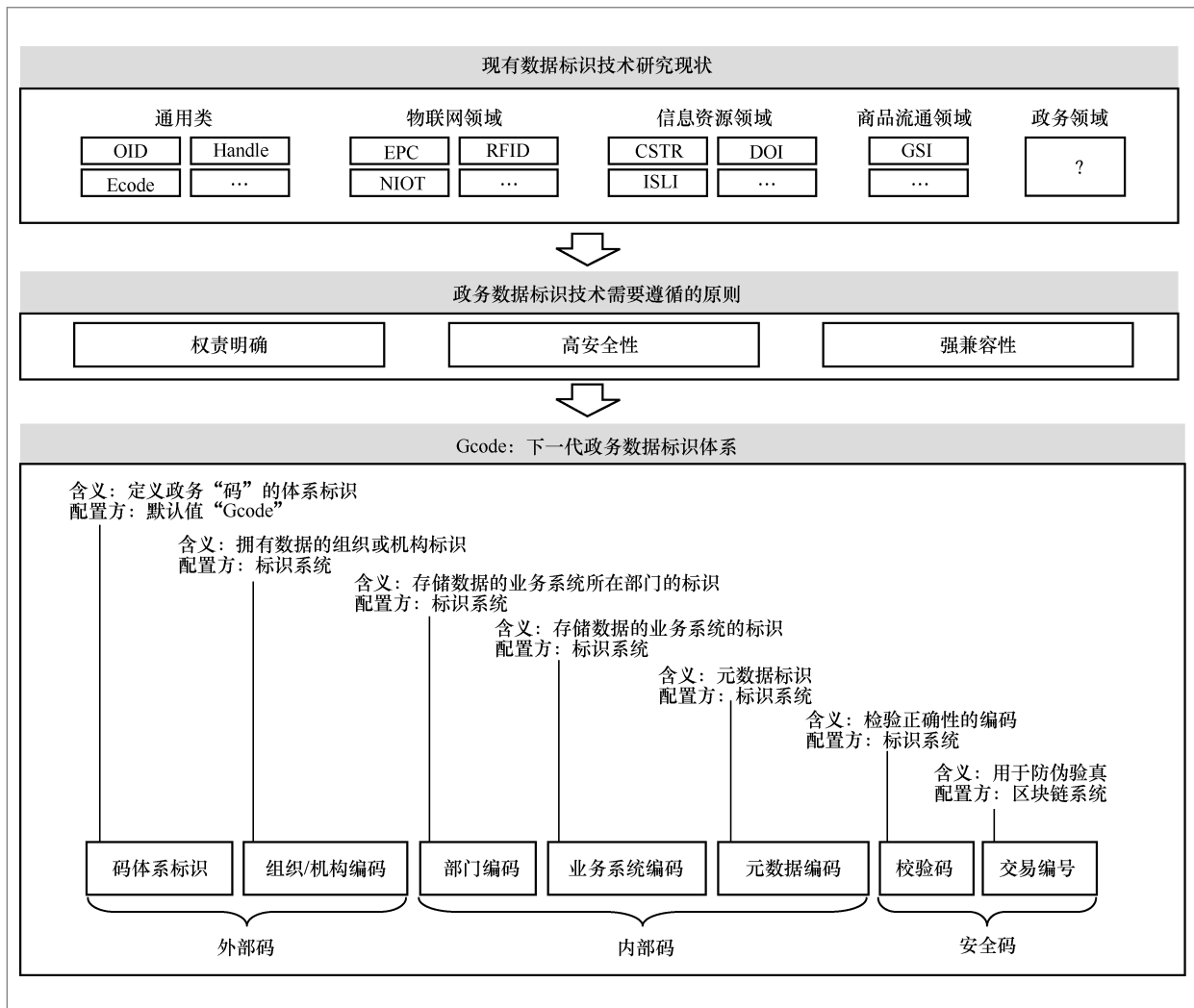


图1 Gcode的提出背景、原则、组成单元及各个部分的含义

构各个部门所管理的业务系统信息，厘清政务数据资源家底。在此过程中，每个业务系统将获得编码，该编码可以为系统普查过程中自行产生的编码，也可以为国家发展和改革委员会等部门批复的项目编号。不同部门业务系统的数量不一，大的部门业务多，系统也相应较多，小的部门业务少，系统也相应较少。业务系统编码的编码规则为：由不少于4的N位阿拉伯数字组成，例如：若N=4，则从0001开始，至9999结束，以此类推。

元数据编码指的是部门的某个元数据

的编码。元数据是描述数据的数据。一般而言，元数据的概念比较宽泛，既可以是数据的配置信息，也可以是数据内容的描述。政务场景下，元数据概念应用较多的是数据项名，例如，数据库表中的某个字段名。元数据由业务系统产生，与部门行政职权相关。一般来说，元数据的量相对较大。元数据编码的编码规则为：由不少于6的S位阿拉伯数字组成，例如：若S=6，则从000001开始，至999999结束，以此类推。

校验码一般为1位字符，该字符根据前述所有编码，依据特定的计算规则产生。

《法人和其他组织统一社会信用代码编码规则》中使用的校验码参照GB/T 17710<sup>[22]</sup>生成。ISLI中的校验码的生成规则为：在服务字段和关联字段从右往左依次交叉分配加权因子1和2；各位数字和加权因子相乘，相乘结果大于10，则将乘积所含数据相加；将所有乘积相加，总和除以10取余，再用10减余数；所得结果为1~9，则为校验码，为10，则校验码为0。GS1编码中的校验码的生成规则为：将包括校验码在内的字符从右往左依次编序号；将从序号2开始的所有偶数位代码求和后乘以3；将从序号3开始的所有奇数位代码求和；将偶数位和与奇数位和相加；用大于或等于上步所得结果且为10的整数倍的最小数减去上步所得结果，差值即校验码。可以看出，校验码的生成已经具有较多参考依据。Gcode体系中的校验码编码规则采用比较成熟的GB/T 17710中提供的方法。

交易编号指的是该标识上链之后产生的交易编号。区块链技术具有难以篡改、可追溯、去中心化、透明可信等特点，能够保证信息的完整与可靠<sup>[23-24]</sup>。区块链技术能够化解数据共享面临的安全与效率矛盾，在政务数据共享，例如，在不动产区块链信息共享等场景中，展现出巨大的应用潜力<sup>[25]</sup>。将区块链技术引入数据标识体系能够实现标识的防伪验真，增强政务数据在使用和流通过程中的安全性<sup>[26]</sup>。交易编号的编码规则为：采用国密SM3算法对Gcode中除了交易编号之外的内容进行加密计算，长度设置为定长64位，只包含英文小写字母和阿拉伯数字。

Gcode也充分考虑到了标识的动态更新机制。未来，需要为Gcode建立统一的标识管理机构，对Gcode的生成、分配、维护等进行全方位管理。其中，外部码由统一的标识管理机构提供并管理，内部码由拥有数据的单位提供并由统一的标识管理机构

管理，安全码同样由统一的标识管理机构提供并管理。

## 2.2 体系特点及创新性

Gcode在政务数据标识技术的研究基础之上，充分考虑了政务数据的特点。总体来说，Gcode具有权责明确、安全性高、兼容性强的特点，这也是Gcode的创新之处。

在权责明确方面，政务数据一般归属于特定责任部门，因而政务数据的生产、申请、使用等环节需要明确谁能够干什么。为了明确权责，标识技术需要建立“机构-部门-系统-数据”的关联关系，推动实现“一数一源”。为此，Gcode的标识对象设置为元数据，而不是数据的物理地址。因此，当数据的物理地址发生变化后，标识结果不受影响。对元数据标识的基础是建立“机构-部门-系统-数据”关联关系，这可有力推动“一数一源”，便于明确权责。

在安全性高方面，政务数据中包含着大量个人隐私信息，甚至包含涉密信息，因而对安全性的要求较高。也正因如此，各种面向政务应用场景的方法研究都努力地聚焦于如何避开操作私域数据。为此，Gcode设置校验码，能够对自身编码的正确性进行静态验证；Gcode还引入了区块链技术，将数据标识结果上链，达到动态防伪验真的效果。

在兼容性强方面，如前文所述，部分地区已经开展了政务数据标识技术的研究和推广应用。统一社会信用代码等已经能够实现对政务组织/机构的唯一标识。新的政务数据标识技术的发展需要最大限度地满足利旧原则，尽可能减轻系统改造的工作量，减少对现有业务系统的影响。为此，Gcode采用了类似于现有OID、

Handle/DOI的架构,使其可以兼容现有主流的数据标识体系;Gcode纳入了统一社会信息用代码,能够兼容现有的组织/机构的标识体系,减少了开发和系统改造的工作量。

### 3 结束语

本文提出了由外部码、内部码和安全码3个部分组成的下一代政务数据标识体系Gcode,具备权责明确、安全性高、兼容性强的特点。但是,Gcode存在进一步优化的空间。首先,Gcode致力于建立严格的“机构-部门-系统-数据”关联关系,由此推动实现“一数之源”。然而,繁杂交错的机构、部门、系统、数据关系可能导致实际情况要远远比设想的规则复杂。例如同一个业务系统可能由多个部门共同管理、同一个数据可能出现在多个机构中、“机构-部门-系统-数据”关联关系动态更新机制需要建立等。此外,Gcode在实际应用中也可能面临着诸多困难。例如,需要耗费精力开展部门梳理、系统普查、目录编制等工作,区块链技术的引入可能引发数据失控的风险等。以上问题将为进一步优化Gcode提供重要研究方向。

总之,Gcode权责明确、安全性高、兼容性强,具有支持“跨层级、跨地域、跨系统、跨部门、跨业务”的能力,未来的实际应用中势必为政务数据标识提供强有力的技术支撑,有力地推动“一数之源”的实现。

### 参考文献:

[1] 张秋霞,田润.物联网标识体系与Ecode编码标准比较研究[J].中国自动识别技术,2021(1):57-66.

ZHANG Q X, TIAN R. Comparative study on Internet of Things identification system and ecode coding standard[J]. China Auto-ID, 2021(1): 57-66.

[2] 国家质量监督检验检疫总局,中国国家标准化管理委员会.商品条码零售商品编码与条码表示:GB 12904—2008[S].北京:中国标准出版社,2009.

General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Bar code for commodity - Retail commodity numbering and bar code marking: GB 12904—2008[S]. Beijing: Standards Press of China, 2009.

[3] Information technology - Automatic identification and data capture techniques - GS1 application identifiers and ASC MH10 data identifiers and maintenance[Z]. 2016.

[4] 国家质量监督检验检疫总局,中国国家标准化管理委员会.物联网标识体系物品编码Ecode:GB/T 31866—2015[S].北京:中国标准出版社,2016.

General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Identification system for Internet of Things—entity code: GB/T 31866—2015[S]. Beijing: Standards Press of China, 2016.

[5] 国家质量监督检验检疫总局,中国国家标准化管理委员会.科技资源标识:GB/T 32843—2016[S].北京:中国标准出版社,2017.

General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Science and

- technology resource identification: GB/T 32843—2016[S]. Beijing: Standards Press of China, 2017.
- [6] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 中国标准关联标识符: GB/T 32867—2016[S]. 北京: 中国标准出版社, 2016. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Chinese Standard Associated Identifier: GB/T 32867—2016[S]. Beijing: Standards Press of China, 2016.
- [7] 中国科学院计算机网络信息中心. 国家物联网标识管理公共服务平台技术规范第4部分: 对象标识编码结构要求: Q/NIOT 004—2016[S]. 2016. Computer Network Information Center, Chinese Academy of Sciences. Technical specification for national iot identification management public service platform part 4: requirements for object identification coding structure: Q/NIOT 004—2016[S]. 2016.
- [8] A primer for decentralized identifiers[Z]. 2019.
- [9] 窦悦, 易成岐, 黄倩倩, 等. 打造面向全国统一数据要素市场体系的国家数据要素流通共性基础设施平台: 构建国家“数联网”根服务体系的技术路径与若干思考[J]. 数据分析与知识发现, 2022, 6(1): 2—12. DOU Y, YI C Q, HUANG Q Q, et al. Constructing a common data circulation infrastructure platform for the national unified data factor market: technical path and policy thinking of constructing the national “data networking” root service system[J]. Data Analysis and Knowledge Discovery, 2022, 6(1): 2—12.
- [10] JUNG M, KIM W S, CHUNG S H, et al. A blockchain-based ID/IP mapping and user-friendly fog computing for hyper-connected IoT architecture[J]. IJICTDC, 2018, 2(2): 12—19.
- [11] GHOSH A, MAHDIAN M, MCAFEE P, et al. To match or not to match: economics of cookie matching in online advertising[J]. ACM Transactions on Economics and Computation, 3(2): 1—18.
- [12] 郭明军, 陈沁, 安小米, 等. 我国大数据发展指数构建及实践应用: 从政务数据与社会数据融合的视角[J]. 大数据, 2022, 8(2): 182—192. GUO M J, CHEN Q, AN X M, et al. Construction and practical application of big data development index in China—from the government data and social data fusion perspective[J]. Big Data Research, 2022, 8(2): 182—192.
- [13] CABRERA C, PALEYES A, THODOROFF P, et al. Real-world machine learning systems: a survey from a data-oriented architecture perspective[EB]. arXiv preprint, 2023, arXiv:2302.04810.
- [14] 中国标准化研究院, 大唐电信集团, 北京信息资源管理中心, 等. 政务信息资源目录体系第5部分: 政务信息资源标识符编码方案: GB/T 21063.5—2007[S]. 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会, 2005. China Institute of Standardization, Datang Telecom Group, Beijing Information Resource Management Center, et al. Government information resource catalog system part 5: government information resource identifier coding scheme: GB/T 21063.5—2007[S]. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, National Standardization Administration of China, 2005.
- [15] 李青元, 孙晓利. 电子政务标准解读及对电子政务中空间信息基础设施标准的思考[J]. 电子政务, 2006(11): 18—22. LI Q Y, SUN X L. Interpretation of

- e-government standards and reflections on spatial information infrastructure standards in e-government[J]. E-Government, 2006(11): 18-22.
- [16] 王薇. 政务信息资源标识符编码方案研究[J]. 信息化建设, 2010(2): 21-22.  
WANG W. Research on the coding scheme of government information resource identifier[J]. Informatization Construction, 2010(2): 21-22.
- [17] 黑龙江省市场监督管理局. 政务信息资源目录第1部分: 编码规则: DB23/T 2839—2021[S]. 2021.  
Market Supervision Administration of Heilongjiang Province. Government information resource catalog part 1: coding rules: DB23/T 2839-2021[S]. 2021.
- [18] 贾东琴, 董伟. DOI在政府信息公开中的应用初探[J]. 图书情报工作, 2009, 53(15): 138-141.  
JIA D Q, DONG W. Exploration on the application of DOI in freedom of government information[J]. Library and Information Service, 2009, 53(15): 138-141.
- [19] 谢家贵, 齐超, 朱佳佳. 工业互联网标识解析体系架构及部署进展[J]. 信息通信技术与政策, 2020(10): 10-17.  
XIE J G, QI C, ZHU J J. Architecture and deployment progress of identifier resolution system for Industrial Internet[J]. Telecommunications Network Technology, 2020(10): 10-17.
- [20] BERNERS-LEE T, MASINTER L, MCCA HILL M. Uniform resource locators (URL)[J]. RFC, 1994, 1738: 1-25.
- [21] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 法人和其他组织统一社会信用代码编码规则: GB 32100—2015[S]. 北京: 中国标准出版社, 2015.  
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Coding rule of the unified social credit identifier for legal entities and other organizations: GB 32100—2015[S]. Beijing: Standards Press of China, 2015.
- [22] 国家标准化管理委员会. 信息技术安全技术校验字符系统: GB/T 17710—2008[S]. 北京: 中国标准出版社, 2008.  
Standardization Administration of the People's Republic of China. Information technology-Security technique - Check character systems: GB/T 17710—2008[S]. Beijing: Standards Press of China, 2008.
- [23] 李燕, 马海英, 王占君. 区块链关键技术的研究进展[J]. 计算机工程与应用, 2019, 55(20): 13-23, 100.  
LI Y, MA H Y, WANG Z J. Research progress on key technologies of blockchain[J]. Computer Engineering and Applications, 2019, 55(20): 13-23, 100.
- [24] 曾诗钦, 霍如, 黄韬, 等. 区块链技术研究综述: 原理、进展与应用[J]. 通信学报, 2020, 41(1): 134-151.  
ZENG S Q, HUO R, HUANG T, et al. Survey of blockchain: principle, progress and application[J]. Journal on Communications, 2020, 41(1): 134-151.
- [25] 王鹏, 魏必, 王聪. 区块链技术在政务数据共享中的应用[J]. 大数据, 2020, 6(4): 105-114.  
WANG P, WEI B, WANG C. Application of blockchain technology in government data sharing[J]. Big Data Research, 2020, 6(4): 105-114.
- [26] 汪靖伟, 郑臻哲, 吴帆, 等. 基于区块链的数据市场[J]. 大数据, 2020, 6(3): 21-35.  
WANG J W, ZHENG Z Z, WU F, et al. Blockchain based data marketplace[J]. Big Data Research, 2020, 6(3): 21-35.

## 作者简介



王昀(1979- ),男,中移信息系统集成有限公司高级工程师、总经理,主要研究方向为数字政府。



郭毅峰(1978- ),男,博士,中移信息系统集成有限公司数字政府事业部副总经理、首席专家,主要研究方向为数字政府、政务数据治理。



苏晓亮(1979- ),男,中移信息系统集成有限公司运维服务部高级工程师、总经理,主要研究方向为政务信息化建设。



周武爱(1993- ),男,博士,中移信息系统集成有限公司规划技术部工程师,主要研究方向为政务数据治理、人工智能、自然语言处理。



张皖哲(1995- ),男,博士,中移信息系统集成有限公司规划技术部工程师,主要研究方向为政务数据治理。



许大虎(1979- ),男,中移信息系统集成有限公司数字政府事业部副总经理,主要研究方向为政务服务。



周强(1986-),男,中移信息系统集成有限公司数字政府事业部工程师、总监,主要研究方向为政务服务。



冯建华(1967-),男,博士,清华大学计算机科学与技术系教授、副主任,主要研究方向为数据库管理系统、数据安全与隐私保护、信息检索。

收稿日期: 2023-09-11

基金项目: 中移信息系统集成有限公司自立项目海算平台项目(第1期)(No.R2310AJE)

Foundation Item: China Mobile Information System Integration Co., Ltd.'s Independent Project-Haisuan Platform Project (Phase 1) (No. R2310AJE)