基于元数据和智能合约技术的政务数据确权与共享研究

李静雯[1]，林绍福[1]，李宏卓[2]，贾晓丰[3]

[1](北京工业大学 软件学院，北京 100124)

[2](北京信息安全测评中心， 北京 100101)

[3](北京市信息资源管理中心,北京 100101)

通讯作者：贾晓丰, E-mail: jiaxf@bjeit.gov.cn

摘要：数据确权与开放共享一直是电子政务领域的挑战性问题，传统的数据共享平台一定程度上解决了数据共享问题，但信息孤岛、数据权责不清等难题依旧存在。针对这些问题，本文利用区块链的安全性、可追溯、去中心化、不可篡改等特点，基于区块链的基础架构，提出了一种新的技术方案。即运用智能合约技术，创新政务信息资源开放共享的基础要素——元数据的管理方式，将元数据存储到区块链上并发布，同时将数据授权服务整合在智能合约中，实现政务数据的确权与智能共享。与传统技术方案相对比，此方案实现了政务数据共享的数据权责界定、可控、可信和可溯源，提高了共享效率，从而能更好地解决数据权责不清和难共享的难题。

关键词：电子政务、区块链、智能合约、元数据、数据确权、智能共享

中图法分类号：TP311

中文引用格式：陈翔,顾庆,刘望舒,刘树龙,倪超.静态软件缺陷预测方法研究.软件学报. http://www.jos.org.cn/1000-9825/0000.htm

英文引用格式: Chen X, Gu Q, Liu WS, Liu SL, Ni C. State-of-the-Art survey of static software defect prediction. Ruan Jian Xue Bao/Journal of Software, 2016 (in Chinese). http://www.jos.org.cn/1000-9825/0000.htm

The Government Data Sharing and Confirmation Based on Metadata and Smart contract

LI Jing-Wen[1]， LIN Shao-Fu[1]， LI Hong-Zhuo[2]， JIA Xiao-Feng[3]

[1](Beijing University of Technology, Software College, Beijing 100124, China)

[2](Beijing Information Security Test and Evaluation Center, Beijing 100101, China)

[3](Beijing Information Resource Management Center，Beijing 100101，China)

Abstract: It is a challenge in the field of e-government to confirm and open data sharing all the time. The traditional data sharing platform solves the problem of data sharing to a certain extent, but the aporia still exist, such as: information islands, unclear data rights and responsibilities. To solve these problems, this paper proposes a new technical scheme based on the block chain infrastructure, which takes advantage of the security, traceability, decentralization, and non-tampering characteristics. That is using smart contract technology to innovate the management mode of metadata, which is the basic element of open sharing of government information resources. Then metadata is stored in the block chain and published. At the same time, the data authorization service is integrated into the smart contract to realize the confirmation and intelligent sharing of government data. Compared with the traditonal technical scheme, this paper realizes the data confirmation, controllability, trustworthiness and traceability of government data sharing, and improves the sharing efficiency, so as to better solve the problem of unclear data rights and responsibilities and hard to share.

Key words: e-government, block chain, smart contract, matadata, data confirming, intelligent data sharing

电子政务发展至今，数据共享的老问题依然存在，突出体现在利益责任难划分、政策约束难执行、共享标准不统一、安全保障不完善等问题上[9]。为解决上述问题，国家出台一系列政策文件并提出到 2020 年，基本实现政府数据部门共享和社会开放。实现政府数据的部门共享与社会开放，离不开政务信息资源目录体系的建立，而政务信息资源目录体系的建设离不开元数据的设计、编制和管理，其完整性、统一性、准确性等将直接影响信息资源的开放共享程度。在学术界、产业界和政务部门，针对政务数据共享、解决信息孤岛问题的研究不少，所提出的建设大数据共享平台、信息上云、设计共享模型等方案一定程度上缓解了部门内和部门间信息孤岛问题，但跨部门、跨层级的数据共享问题依然没有很好解决，其原因主要是尚未形成针对政务数据共享的数据权责界定、可控、可信和可溯源的更有效的解决方案。

近年来兴起的区块链技术的主要特性之一则为在可信的环境下保护数据安全、不被篡改，实现数据的溯源与确权。尽管区块链应用体系目前远未成熟，但区块链技术在社会应用方面的巨大潜力已经凸显。基于区块链的技术特性，针对政务信息资源共享的难点问题，利用智能合约技术，研究创新元数据的编制、存储与发布的解决方案，并应用到数据确权和信息资源共享开放共享方面，对于进一步解决跨部门、跨层级的信息共享问题，具有重要的创新研究意义和应用价值。

1　基于智能合约的数据确权与共享逻辑模型

智能合约的概念可以追溯到 1994 年，几乎与互联网同时出现。由备受广泛赞誉的密码学家尼克.萨博（Nick Szabo）首次提出“智能合约”（Smart contract）这一术语，他对于智能合约的定义是：“一个智能合约是一套以数字形式定义的承诺，包含合约参与方可以在上面执行这些承诺的协议。”

智能合约不完全附属于区块链系统，但其受益于区块链无法篡改的特性，合约代码可以保持完整性与不变性。其可以借助区块链平台回应外界信息并传递其存储的信息等，还可基于区块链的公钥私钥加密体系，直接对合约存储信息的读取权限进行限制[1]。可以说，智能合约是区块链技术的拓展应用。本文利用智能合约可以储存信息、发布信息及可控的读取权限机制等特点，通过对元数据的采编、存储、发布，来实现在区块链上政务数据的确权和智能共享，总体设计模型结构如图 1 所示。其中，元数据是数据确权和数据共享的基础和依据，数据确权是数据共享的前提和保障。1 代表基于智能合约的元数据编译流程；2 代表基于智能合约的数据确权流程；3 代表基于智能合约的智能共享流程。

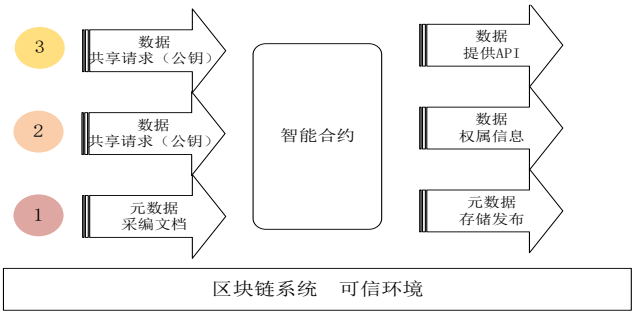


Fig 1　The logic model of data confirmation and sharing based on Smart contract

图 1　基于智能合约的数据确权与共享逻辑模型

2　基于智能合约的元数据存储和发布

元数据通常被定义为关于数据的数据，广义上包括对数据、服务等资源描述的数据，能较好地解决网络信息资源的描述、发现、控制和管理问题。元数据是政务信息资源目录体系的关键要素，在开展目录体系建设中，各政务部门将依据国家政务信息资源核心元数据标准和地方核心元数据标准进行扩充或减少元数据的数据项。本节将介绍国家和北京市核心元数据标准选型，并选取重要元数据项进行基于智能合约的设计。

2.1 核心元数据标准选型

我国政务信息资源核心元数据包括 6 个必选的元数据实体、元数据元素和 6 个可选的元数据实体、元

数据元素。其中，6 个必选的元数据实体、元数据元素分别是信息资源名称、信息资源摘要、信息资源提供方、信息资源分类、信息资源标识符、元数据标识符。

根据国家政务信息资源核心元数据标准，北京市政务信息资源目录体系地方标准定义的政务信息资源核心元数据包括 7 个元数据实体，分别是资源负责方、资源需求方、资源格式信息、关键字说明、时间范围、资源分类和元数据联系方。元数据元素从定义、名称、数据类型、值域、短名、注解 6 个方面描述其主要属性。[13]

以北京市部门政务信息资源目录为例，一条标准的元数据包含内容如表 1 所示。

Table1　Metadata directory

表 1 元数据目录

一级分类	二级分类	数据类型
信息资源分类	类分类（必填）	字符串
	项分类（必填）	字符串
	细目分类	字符串
	分类五级	字符串
	分类六级	字符串
	分类七级	字符串
	分类八级	字符串
	分类九级	字符串
	分类十级	字符串
	分类十一级	字符串
	分类十二级	字符串
信息资源名称（必填）	--	字符串
信息资源代码（选填）	--	整型
信息资源提供方	信息资源提供方（必填）	字符串
	提供方内部部门（选填项）	字符串
资源提供方代码（选填）	--	整型
信息资源摘要（选填）	--	字符串
信息资源格式	格式分类（必填）	字符串
	格式类型（必填）	字符串
	其他类型资源格式描述（选填）	字符串
信息资源普查	数据存储总量（G）（必填）	浮点型
	结构化信息记录总数（万）（必填）	浮点型
	已共享的数据存储量（G）（必填）	浮点型
	已开放的数据存储量（G）（必填）	浮点型
	已开放的结构化记录数（万）（必填）	浮点型
信息项信息	信息项名称（必填）	字符串
	数据类型（必填）	字符串
	数据长度（必填）	整型
共享属性	共享类型（必填）	字符串
	共享条件（必填）	字符串
共享方式	共享方式分类（必填）	字符串
	共享方式类型（必填）	字符串
开放属性	是否向社会开放（必填）	字符串
	开放条件（对外开放时必填项）	字符串

更新周期（必填）	--	时间型
发布日期（必填）	--	时间型
关联及项目名称（选填）	--	字符串
国家系统 ID	--	整型

2.2 元数据的采集与编制

基于智能合约的元数据编制分为元数据采集、元数据描述文档编制、智能合约编制、存储发布智能合约四个步骤。其具体过程为，部门采集元数据信息，形成 XML 文档，并根据文档编制智能合约，并发布存储至区块链系统中，如图 2 所示。

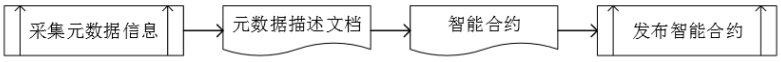


Fig 2　Compilation process

图 2 编制流程

基于上述编制流程，以北京市工商局食品流通许可证的许可证编号为实例，根据 2.1 节中元数据标准，采集智能拓展的元数据项进行编制示范，元数据项条目如表 2 所示。

Table 2　Metadata item

表 2 元数据项条目

类别	代码名称	数据类型	内容
信息资源名称	NameofInformationItem	字符串	Food circulation permit
信息项名称	InformationName	字符串	Certificate number
信息资源代码	InformationResourceCode	整型	000039
提供方	Provider	字符串	BJGSJ
数据类型	DataTper	字符串	String
公开类型	SharedType	字符串	Unconditional
是否向社会公开	SocialOpening	字符串	False

在进行智能合约的开发之前，需对本系统的开发环境进行配置，配置表如表 3 所示。

Table 3　System Development Environment Table

表 3 系统开发配置表

内容	使用环境
系统平台	Win10
开源区块链	以太坊 geth
区块链编程语言	Go 语言
智能合约编程语言	Solidity 语言

根据所采集的元数据项，进行智能合约的编制。当有用户访问该合约时，合约输出信息如表 4 所示。

Table 4 Smart contract output

表 4 合约输出信息表

用户操作	输出信息	
用户输入 ID	NameofInformationItem	Food circulation permit
	InformationName	Certificate number
	InformationResourceCode	000039
	Provider	BJGSJ
	DataTper	String
	SharedType	Unconditional
	SocialOpening	False

2.3 元数据存储与发布

元数据的智能合约编制完成之后，通过目录账户选择发布与存储。合约的存储过程如图 3 所示，在当前系统周期内，各节点记录所有交易信息与智能合约，在达成共识之后，由记录节点记录生成新区块，链接至区块链之上。新发布的智能合约将存储至此区块之中[2]。

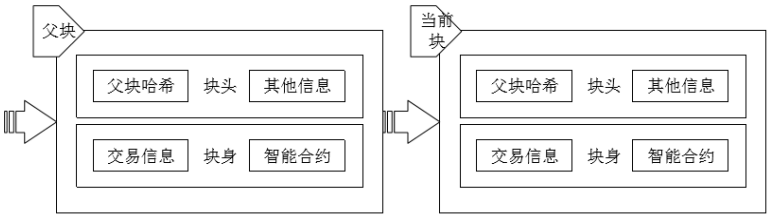


Fig 3　Smart contract storage

图 3 智能合约存储

同时，账户可选择定向发布或广播发布该合约。定向发布过程为制定具体地址，将该合约发送至该地址，其他节点不会收到发布通知。广播发布过程为向全网广播合约信息，所有节点都将收到新合约发布的信息。

3　基于元数据和智能合约技术的政务数据确权与智能共享

基于智能合约技术的元数据设计较大程度上解决了数据字典缺失、数据权责确定不清等问题，其更好地服务于目录的编制，实现信息资源的权责界定与智能共享。

3.1 数据确权

区块链技术作为一种去中心化系统，具有安全、透明等特性，其能够保证记录的数据不被篡改。基于智能合约的元数据在每一次数据交易及数据共享的过程中，都会记录参与方的所有信息，形成数据确权证明，既可以保护数据权属方，也可追溯数据责任方。以身份验证类元数据进行数据确权设计，如图 4 所示，其余元数据确权方式类似。

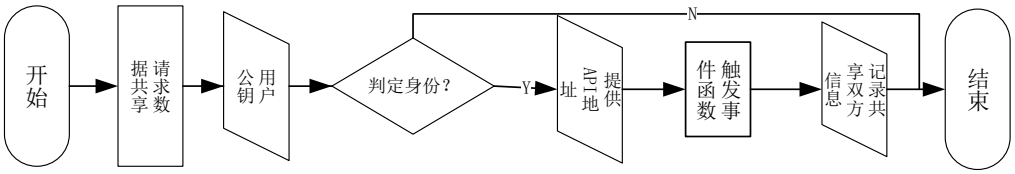


Fig 4　Authentication metadata confirmation flowchart

图 4 身份验证类元数据确权流程图

用户通过合约查看元数据，确定需求、发起访问请求，智能合约收到请求后查询用户公钥，对比合约内所存公钥，若匹配成功则提供 API 地址，并触发事件函数，将双方共享信息记录至区块链内，供双方随时查看，鉴定数据所有权。当需要查询此次共享操作时，输入共享编号，系统将从当前区块开始遍历，若区块中存在相同编号，则定位到共享信息记录的所在位置，即可找到此次共享记录的全部信息，包括数据权属方、责任方、共享内容、共享用途等，从而实现数据共享服务和应用的溯源。

3.2 智能共享

（1）传统共享模式

北京市共享交换平台的核心业务服务包括目录服务、安全服务、共享交换服务。其中，目录服务为共享交换服务提供政务信息资源目录的查询接口，以实现政务信息资源的定位功能，同时目录服务为安全服务提供政务信息资源目录数据的查询接口，实现细粒度化的资源授权服务。传统共享业务流程图如图 5 所示。

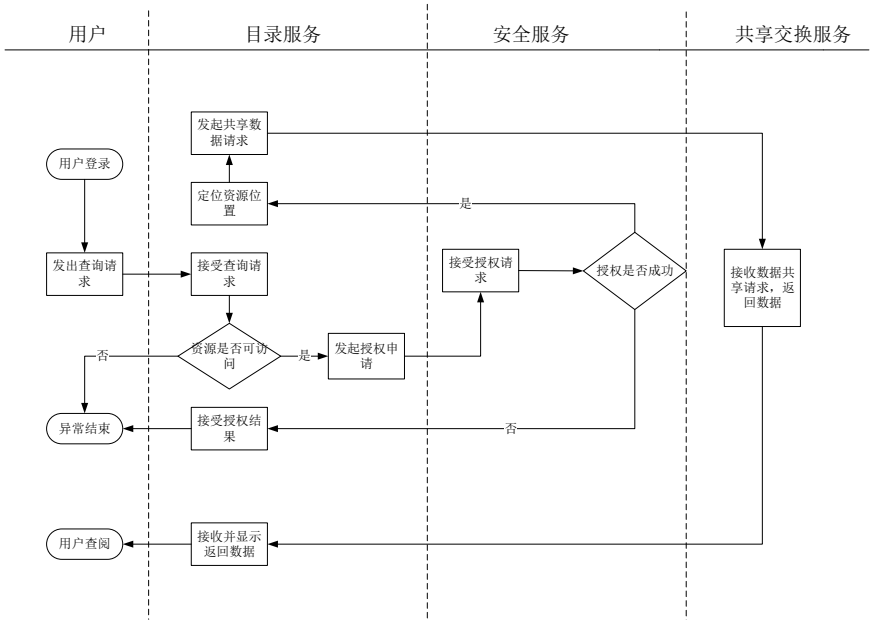


Fig 5　Traditonal shared business flowchart

图 5 传统共享业务流程图

当前北京市共享交换平台建设日趋完善，目录服务体系功能也日趋完整，但如图 5 所示，共享步骤繁多，由于各类信息资源共享开放程度不同，授权服务还需线上线下同时向资源所属部门申请，并等待各部门的审批。这也很大程度的限制了资源共享交换的审批速度，若不同部门资源部门所存数据格式不统一、所用系统不统一，将形成共享瓶颈，造成信息孤岛。为简化共享流程、提高共享效率，下文提出基于区块链架构的智能共享方案。

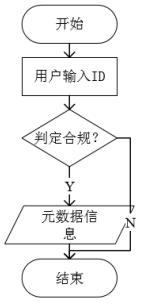
（2）智能共享

基于上节智能合约元数据体系的设计与编制，来实现信息资源的智能共享。智能共享的主要思想是将信息资源共享交换授权服务整合在智能合约中，从而实现共享交换自动化、智能化和去人工化，高速、便捷的共享信息资源。

在元数据描述中，数据的共享开放一般分为三类，分别是无条件共享、有条件共享和不共享。不共享类元数据不涉及此模块，无条件共享及有条共享类元数据智能合约设计如下。

① 无条件共享类元数据设计

通过智能合约元数据实现数据共享交换的方法如图 6、7 所示。



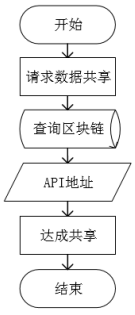


Fig 6　View metadata flowchart　　Fig 7　Data shared flowchart

图 6　查看元数据流程图　　图 7　数据共享流程图

用户通过合约查看元数据信息，决定是否需要对该元数据描述的数据进行访问和存取，若需要，则向该智能合约发起请求。由于是无条件共享数据，智能合约收到请求后，即可发送数据 API 地址给用户，至此，完成数据共享。

② 有条件共享类元数据设计——身份验证类

政务信息资源中，有一部分数据是带有隐私性的，但需要在部门间共享，所以有条件共享类元数据设计是必不可少的。有条件共享类元数据分为身份验证类和协议限制类两大类，身份验证类设计流程图如图8、9所示。

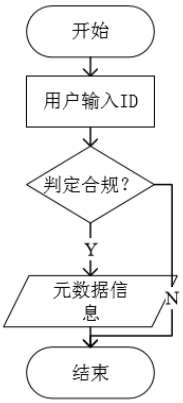


Fig 8　View metadata flowchart

图 8 查看元数据流程图

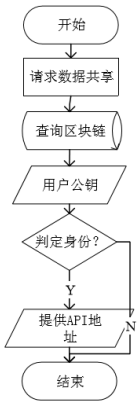


Fig 9 Authentication data shared flowchart

图 9 身份验证类数据共享流程图

在编制合约时，将关系紧密的部门节点公钥写入智能合约中，在用户向智能合约发起访问请求时，合约查询用户公钥，对比合约内部所存公钥，匹配成功则发送数据 API 地址，否则拒绝用户访问请求。

③ 有条件共享类元数据设计——协议限制类

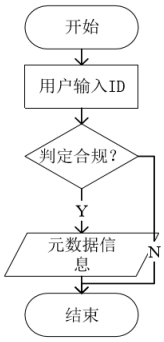


Fig 10　View metadata flowchart

图 10 查看元数据流程图

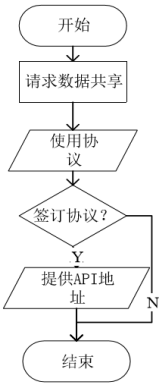


Fig 11　Agreement limits data shared flowchart

图 11 协议限制类数据共享流程图

对于有条件共享类信息资源，在实际应用中，部分带有低敏程度信息的数据在共享交换中需签署相关协议。在设计此类元数据应用时，将协议内容直接存储在智能合约中，用户通过元数据确定自身需求，并向智能合约提出请求，合约将协议发送给用户，并要求其使用公钥签名，签署成功后即可发送数据 API 地址，达成数据共享。

基于智能合约的数据共享能有效地解决共享效率、数据安全等问题。

4　总结

本文从当前政务信息资源共享开放存在的数据权责不清、共享流程繁杂等重点问题出发，结合区块链技术，基于元数据和智能合约技术，研究提出去人工、去中心、自动化的数据确权和智能共享技术方案。经笔者和团队的初步原型实现验证，此方案能实现共享流程的可追溯，统一数据共享标准，简化共享流程，实现从人工授权、签署一纸协议到自动化、智能化、无纸化的转变，能较大程度上提高部门间、跨层级的政务数据共享效率。目前在政务领域对于区块链的应用还处于探索时期，本研究还有待在实践应用中进一步完善。未来，将研究如何利用区块链技术在元数据初始采编阶段，进一步确保元数据的唯一性和精准性，同时进一步研究数据共享开放应用过程中的溯源和绩效评估。

References：

[1] Delmolino K, Arnett M, Kosba A, et al. Step by Step Towards Creating a Safe Smart Contract: Lessons and Insights from a Cryptocurrency Lab[C]// International Conference on Financial Cryptography and Data Security. Springer, Berlin, Heidelberg, 2016:79-94.

[2] Pass R, Seeman L, Shelat A. Analysis of the Blockchain Protocol in Asynchronous Networks[C]// International Conference on the Theory and Applications of Cryptographic Techniques. Springer, Cham, 2017:643-673

[3] Li M, Weng J, Yang AJ, Lu W. CrowdBC: A blockchain-based decentralized framework for crowdsourcing. IACR Cryptology

[4] Yafei Chen. Research and Application of Warehouse Receipt Transaction Based on Smart Contract on the Blockchain[A]. Wuhan Zhicheng Times Cultural Development Co., Ltd.Proceedings of 2nd International Conference on Mechanical, Electronic, Control and Automation Engineering (MECAE 2018)[C].Wuhan Zhicheng Times Cultural Development Co., Ltd:,2018:9.

[5] Mingyang Mao. Blockchain-based Technology for Industrial Control System CyperSecurity[A]. Wuhan Zhicheng Times Cultural Development Co., Ltd.Proceedings of 2018 International Conference on Network, Communication, Computer Engineering (NCCE 2018)[C].Wuhan Zhicheng Times Cultural Development Co., Ltd:,2018:5.

[6] Svein Ølnes,Jolien Ubacht,Marijn Janssen. Blockchain in government: Benefits and implications of distributed ledger technology for information sharing[J]. Government Information Quarterly,2017,34(3).

[7] Eliza Mik. Smart contracts: terminology, technical limitations and real world complexity[J]. Law, Innovation and Technology,2017,9(2).

[8] Kristen N. Griggs,Olya Ossipova,Christopher P. Kohlios,Alessandro N. Baccarini,Emily A. Howson,Thaier Hayajneh. Healthcare Blockchain System Using Smart Contracts for Secure Automated Remote Patient Monitoring[J]. Journal of Medical Systems,2018,42(7).

附中文参考文献:

[9] 孟庆国.基于三权分置的政务数据交换共享与实现机制[J].软件和集成电路,2018(08):30-31.

[10] 蔡维德,郁莲,王荣,刘娜,邓恩艳.基于区块链的应用系统开发方法研究[J].软件学报,2017,28(06):1474-1487.

[11] 张成成.基于智能合约的公钥证书发放方案[J].电子技术与软件工程,2017(23):191-192.

[12] 梅洁,张宝明,王佳茂.基于区块链下的供应链上游采购流程的设计构想[J/OL].电子商务:1-4[2018-09-26].https://doi.org/10.14011/j.cnki.dzsw.20180723.002

[13] 李文生.政务信息资源目录体系理论与实践[J].图书馆学研究,2011(17):10-15+37.

[14] 钱卫宁,邵奇峰,朱燕超,金澈清,周傲英.区块链与可信数据管理:问题与方法[J].软件学报,2018,29(01):150-159.