



运营技术广角

# 大数据交换信息链

刘楠, 魏进武, 刘露

(中国联合网络通信有限公司研究院, 北京 100032)

**摘要:** 数据流动和共享是大数据产业发展的基础, 但由于数据本身的特殊性和当前数据交易模式机制不健全等原因, 当前数据交易进展缓慢。基于对区块链技术的研究, 设计了一种大数据交换信息链, 能够促进交易数据在交易双方之间的安全、有效流转, 且保证数据提供者的合法权益, 实现了对流动数据的管理和监控。

**关键词:** 大数据; 区块链; 数据交易; 权限控制

**中图分类号:** TP311

**文献标识码:** A

**doi:** 10.11959/j.issn.1000-0801.2016270

## Big data exchange based on information chain

LIU Nan, WEI Jinwu, LIU Lu

China Unicom Research Institute, Beijing 100032, China

**Abstract:** The flow and sharing of data is the foundation of the development of big data industry. However, facing with the specification of data itself and many problems that data trading modes have, the development of data flow and sharing moves slowly. Based on the block chain technology, a new decentralized data exchange platform was designed, which would secure the legal rights of data providers and realize the goal of safe circulation between the only mutual parties of the data exchange, strengthen the regulation and monitoring of flows data.

**Key words:** big data, block chain, data exchange, access control

### 1 引言

随着大数据、云计算、物联网等新技术、新应用向商业世界及公众生活的渗透, 数据体量迎来了爆炸式的增长。数据的共享和交易是大势所趋和市场的必然需求, 也是数据资源价值体现的前提和基础, 但当前数据交易形式严峻, 工业和信息化部电信研究院 (现为中国信息通信研究院)《中国大数据发展调查报告(2015年)》显示, 数据资源短缺仍是制约大数据发展的主要因素, 企业迫切希望政府开放更多的公共信息资源(64.7%)和促进数据流通交易(63.6%)<sup>[1]</sup>。

2015年9月, 国务院印发了《促进大数据发展行动纲

要的通知》, 2016年3月两会期间, 大数据作为国家战略被纳入备受关注的“十三五”计划纲要(草案)中, 其明确提出了“加快推动数据资源共享开放和开发应用, 助力产业转型升级和社会治理创新”的理念。全国各地也纷纷建立大数据交易特区, 2015年4月14日, 全国首家大数据交易所落户贵阳; 2016年4月1日, 上海数据交易中心挂牌成立数据特区; 广东专门成立省大数据管理局; 北京大数据交易服务平台、华中大数据交易所也纷纷投入运营。然而, 大数据特区的出现, 并没特别有效地促进数据的交易和流通, 一方面多数企业出于数据安全考虑, 不愿意将数据通过数据特区进行交易; 另一方面, 数据特区多采取“粗

收稿日期: 2016-08-08; 修回日期: 2016-10-10

2016270-1

矿”式的交易方式,导致交易的数据多为初级、信息价值不大的类型。

针对上述数据交易在方式、模式上的不健全,本文基于区块链技术的研究,设计了一种新型的大数据交换信息链,该方案不仅能够促进交易数据在交易双方的安全流转,且有效解决了数据所有权、数据质量、数据管控诸多方面的问题,能够切实保证数据提供者的合法权益。

## 2 数据流动的意义

纵观大数据产业链,数据源是当之无愧的第一个环节,是大数据产业发展的基础。相较于国外的数据源区块,中国大数据产业具有一些独有特征:由于中国大数据流通在全社会还未形成规模,目前数据源区块主要集中在政府管理部门、互联网巨头、移动通信企业等手中,具体有政府数据源提供商、行业数据源提供商、企业数据源提供商、互联网数据源提供商、物联网数据提供商、移动通信数据提供商、第三方数据服务企业等,这些公司企业机构汇聚了存量、质量好、增长速度快、与社会公众关系密切的海量数据资源,然而因多种原因所致,除了部分自用和信息公开外,大部分没有充分发挥数据资源作为“生产要素、无形资产和社会财富”的应有作用。同时,以数据为导向的企业单位又面临着数据来源匮乏的问题,创意、项目等得不到充足数据资源的支撑,所以大数据产业形成了一个比较怪的现象,数据源拥有者紧守自家海量数据资源,真正的数据需求者“无米下锅”,没有形成一个良好的合作基础平台,不仅难以实现供需双方合作双赢的局面,多行业信息融合协作式发展更是举步维艰<sup>[2]</sup>。现就上述问题产生的原因做如下讨论分析。

首先是认识上的问题,企事业单位或拥有大量数据资源的机构对数据共享和流通的意义不具有清晰准确的认识,把自己所掌握和获取的数据资源看成是私有财产,不愿意共享和开放,甚至在不同部门或同一部门之间都难以实现数据的开放共享。数据开放共享价值不明确、市场不健全、动力不充足,缺乏相应的制度标准,缺乏考核管理体系。

其次,我国当前尚缺乏严格规范的数据开放共享的法规制度,数据的开放共享可能会引起信息安全问题,数据泄密和失控都会加深相关企业的恐惧感,更加不愿、不敢把自己掌握的数据对外界开放共享。

最后,数据共享开放是一个高度专业化的工作,需要

分级分类、收放结合、科学把握。该共享开放而不共享开放会引发数据隔离与封闭、价值损耗、信息孤岛等一系列问题;相反,不该共享开放而共享开放,或者不该大范围共享开放而大范围共享开放,也可能带来更大的损失,甚至危及国家安全。

## 3 数据交易流通现状

推动数据开放和流通在发达国家已成为共识,欧美多国通过国家战略为数据开放背书。在国际大数据交互方面,以德国 Fraunhofer 协会的成果最为突出。Fraunhofer 的 IAIS 研究所(智能分析和信息系统)已在欧洲地区实现数据集市,并拥有多项成功经验,如欧洲工业数据空间(IDS)。IDS 将分散的工业数据转换为一个可信的数据网络空间,目前已经得到欧洲 30 多个重点工业企业支持。在这个空间中,当用户需要数据提供增值服务时,数据可以在被认证的合作伙伴之间共享。

国内各地区数据交易所、数据特区纷纷成立。全国首家大数据交易所落户贵阳;上海数据交易中心挂牌成立数据特区;广东专门成立省大数据管理局;北京大数据交易服务平台、华中大数据交易所也纷纷投入运营。多种数据共享模式的出现,使得多公司、跨行业的数据、大数据应用合作越来越丰富。

数据开放共享是大势所趋,信息使用的边际收益是递增的,信息流动和分享的范围越大,创造的价值就越高,而线上、线下数据化和数据开放正是信息大范围流动的两大前提,自从“互联网+”上升为我国战略后,中央也不断加大力度推动数据开放,但目前我国尚未出台相应的数据共享开放原则、数据格式等相关规范要求,造成对拥有大量数据资源的企业数据共享开放能力不强、水平不高、质量不佳,限制了大数据作为基础性战略资源的开发应用和价值释放<sup>[3]</sup>。

中国联通在大数据交换领域具有得天独厚的优势,电信行业丰富的数据源及基于现网数据的大数据应用为中国联通提供了数据资源基础,并且中国联通在 3 年多的时间里在数据集中、平台建设、对外开放等方面积累了一定的经验。但是,出于数据扩散、信息安全以及数据价值量化等多方面考虑,在数据对外合作方面一直持谨慎保守的态度,使得自身的数据价值实现并没有得到最大化,且没有充分享受到使用对方数据的机会。故应尽快依托大数据技术,兼顾全量数据处理的成本与效率,充分发挥大数据的核心价值与边际效应,实现数据

合作上共享共赢的愿望。

## 4 数据交换方式

通常情况下商品交易和交换,是商品所有权的转移过程,在所有权转移之前,任何一方观看、托管甚至试用都不会真正拥有商品,这些过程也不会对商品提供者造成实质性利益损害。相对于普通商品,数据因本身独有的特征,使得在数据交换过程面临更多的问题:数据具有复制的无差别性,复制具有完全无差异性,在效用上也无差异;数据看过即拥有,数据商品没有了传统所有权的概念,拥有数据商品也变得更为简单、成本更低,看过即拥有了数据商品,就能获得效用。数据的这些特征造成了数据共享、交换技术上的困难。

在数据流通交换过程中,数据拥有者对数据资源的整个生命周期中的权限控制不能得到有效解决,当前数据流动交换主要分为以下3种场景。

(1) 数据拥有者和用户之间的数据交换是直接进行的,不需要第三方中介结构的参与,但在这种方式下,数据交换的效率大打折扣,数据拥有者和用户双方在数据交换中都受到不同程度的限制,数据的价值不能得到充分的发挥。

(2) 最终的用户与数据中间商或其他中介结构进行交换,但由于复制后的数据与原始数据的无差别特性,造成实际数据源拥有者丧失对数据的把控权利,数据资产权益不能得到保障。

(3) 数据流动、交换是以数据拥有者为发送端的 mesh 互联方式。

根据上述3种数据交换流动场景的分析,可以清晰地看出数据交换面临的核心问题:数据流动之后,无法保证不再发生未经许可的数据“二次”流动;数据的安全、隐私无法保证,数据源拥有者一旦让脱敏的数据流动,数据责权无法正确匹配;以数据拥有者为中心的集中控制无法满足数据应用的及时响应。面对不同交易场景下的各种问题,建立满足数据特性的信息交易平台,通过技术机制而不是仅凭承诺来保障数据的安全和数据源拥有者的权益,做到让数据交易双方真正放心,才能加速数据的顺畅流动。

## 5 大数据交换信息链方案设计

### 5.1 方案技术要素

#### (1) 区块链技术

该技术的发展为数据流动带来了新的契机,它是指通

过去中心化和去信任的方式集体维护一个可靠数据库的技术方案。该技术方案让参与系统中的任意多个节点,把一段时间系统内全部信息交流的数据,通过密码学算法计算和记录到一个数据块,并且生成数据块的“指纹”,用于链接下个数据块和实现校验功能,系统所有参与节点来共同认定记录是否为真。它是基于 P2P 网络协议、分布式事务数据库、密码学等技术,以计算共识的方式方法解决 mesh 网络传输、无中心的多方交易的可信可控问题的技术体系和模式,具有完整性、不可篡改性、多方参与和监督的特点。

区块链技术的4个核心技术特征位:共享账本,通过跨商业网络共享式,实现不可更改的分布式交易记录;智能合约,交易条款和交易状态内嵌在区块链系统中,驱动交易执行;隐私保护,保证共享账本,适当的可见性,保障交易是真实和可验证性的;共识机制,所有参与者一致同意才意味交易在网络中通过验证。区块链的核心是在没有中心化中介介入的情况下实现数据的安全性和高可靠性。区块链技术结构如图1所示。

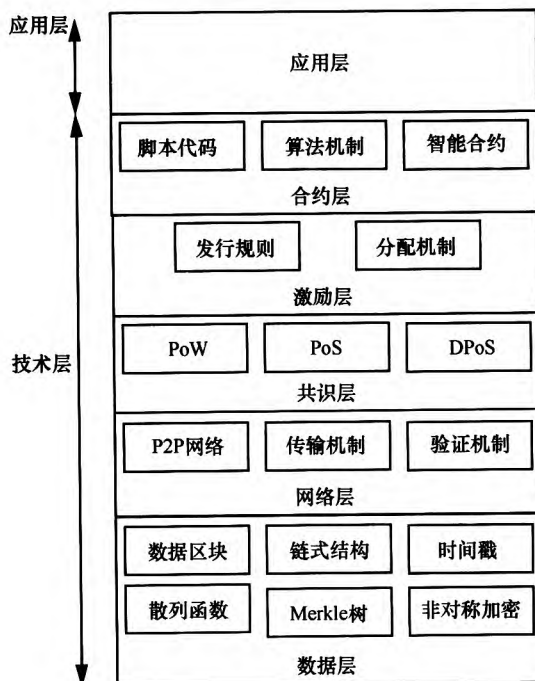


图1 区块链技术结构

区块链技术让中心节点透明化,记账者成为随时可监督、可抛弃、可替换的服务者。各方机构都参与到数据交换的监督过程中,共同维护数据交易的记录,此过程不仅保证了数据拥有者对数据流向的把控,同时也有助于用户对

各部分数据资源交换使用情况的掌握,结合自身需求,选取最优的数据交易方式。

## 5.2 方案设计原则

针对解决数据交易面临的各种问题,首先,数据交易历史较短,数据中介的环节极度缺乏,而这正是大数据流通和交易的中枢所在。其次,可共享数据信息渠道的不畅通,大量数据源未被激活,大量数据拥有者没有数据价值外化的路径。此外,需求端以互联网企业为主,覆盖面不广。最后,在数据流通的安全方面,法律法规的不健全和高技术手段支撑度不足,造成数据拥有者的权益不能得到根本性的保证,使得越来越少的数据资源能够真正进入流通环节。

从解决以上问题为出发点,本文中大数据交换信息链建立原则如下:注重数据共享流通渠道的开放;数据交易流程合法、严密;交易数据展示精准化;交易数据的定价透明化;交易过程有保障,交易记录可追溯、可查询、可申诉;

数据拥有者具有对交易数据把控和监管的权利。

## 5.3 方案建立

### 5.3.1 系统组成

整个数据交易系统主要分为以下两部分。

#### (1) 交易平台

主要作用是搭建基础通信架构,实现成员之间的连接和通信隐藏,成员认证和接入管理;接收、汇总记录请求及交易记录,生成区块、维护公共记录数据索引链和信息交易链;协助交易方完成数据的追溯维权。该交易平台既不作为数据存储中心,也不存留交易数据。

#### (2) 交易成员区

主要指数据交易双方,同时也是业务层面的参与方。主要保存区块链的备份,监督公共区的区块链记录的正确性;存储、维护可共享的隐私数据,并向平台提供可分享数据的描述;发起查询,获取外部数据。

大数据交换方案示意如图2所示。

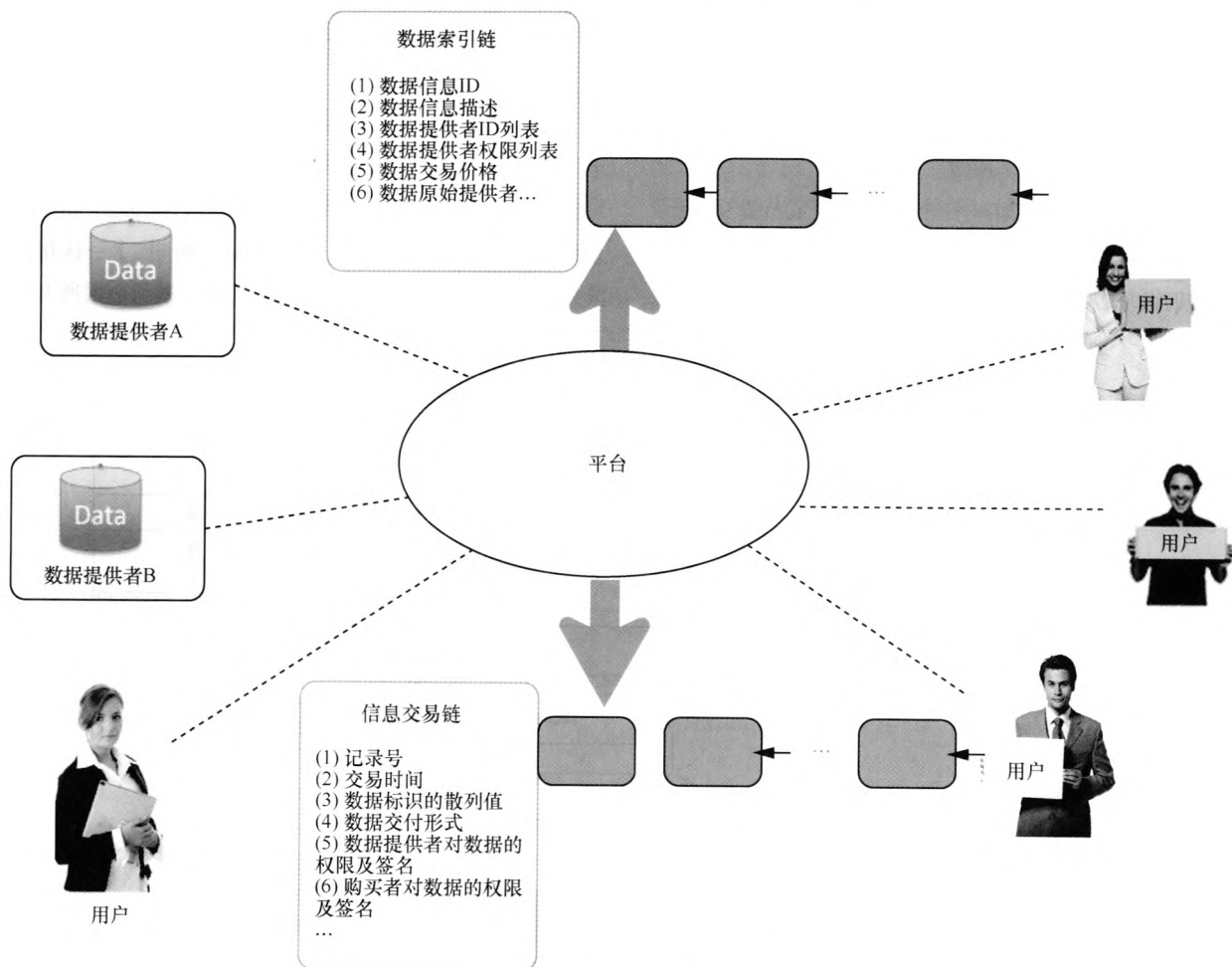


图2 大数据交换方案示意



### 5.3.2 方案初始化

#### (1) 成员加入及认证

成员申请加入,交易平台审核确认成员身份,允许成员加入后,成员向交易平台发送第三方 CA 发放的数字证书或自产生的数字证书,作为成员的身份证书,交易平台将此数字证书与成员真实身份对应。

#### (2) 数据索引链建立

各方数据拥有者(包括数据原始提供者)向平台单方面公布可共享数据的相关描述,包括数据标识 ID、数据字段描述、交换权限、交易价格以及数据原始提供者的相关信息。

交易平台对数据描述进行汇总分类,对同一数据标识的可共享数据合并整理。截止到某一时间点,将汇集的可供交易的数据信息打包形成区块,随后将区块广播到全网所有节点进行审核(后文中提到的交易信息链将为此处审核提供依据),审核通过的区块,链接到数据索引链的末端。数据索引链中的内容无法修改和删除,当可共享数据记录发生变更时,生成新的区块附加在链条末端。数据索引链可供交易系统的所有参与者查询、下载。图 3 为数据索引链示意。

数据索引链特征如下所示。

(1) 交易平台定期(如每小时)按照各个成员提交的新索引、更新索引以及数据交易链中记录的交易详情汇总生成当天的区块,加入链条中。

(2) 链中的记录内容无法更改和删除,只能新增记录。

(3) 每个成员都可以对索引链进行下载和检索,确认自己提供的索引是否在索引链中,形成对交易平台的监督。

### 5.3.3 数据交易流程

#### (1) 申请交易及权限许可

当数据需求方通过数据索引链查询到平台具有所需数据资源时,根据数据提供者 ID 列表、可获取权限及价格明细,确定交易对象。根据交易对象的身份类别的不同,分为以下两种方式:

- 交易对象为数据原始提供者时:用户直接向数据原始提供者发送数据请求,数据原始提供者将交易信息发送给交易平台。
- 交易对象为现有数据持有者时:用户依次向数据持有者发送交易请求,向原始提供者发送交易请求审核,请其对该次交易的数据及数据持有者进行鉴权和许可操作,若审核通过,视为合法交易,数据原始提供者将交易信息发送给交易平台。

#### (2) 交易信息链建立

交易平台对一定时间范围内的交易请求进行汇总,其中关于交易的时间、数据标识的散列值、数据交付方式及定价信息记录为区块的主体部分。此外,区块链头相关信息包括:根据所有交易信息计算 Merkle 树根、上一区块的散列值以及时间戳。随后将生成的区块广播到全网所有节

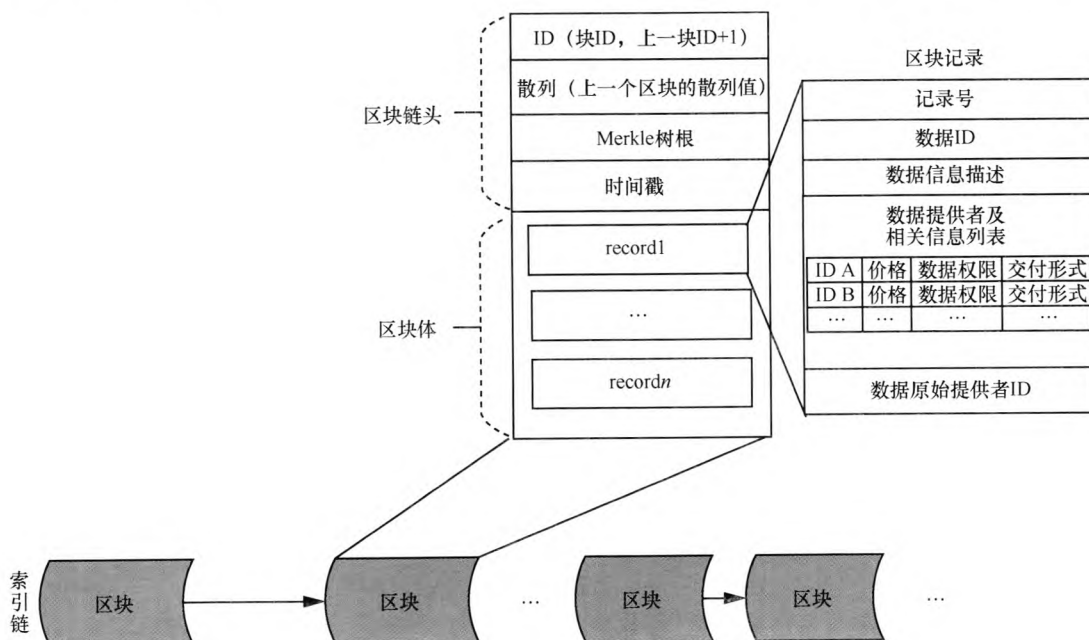


图 3 数据索引链示意

点进行审核操作(数据索引链中对数据持有者的记录将为交易双方提供权限证明)。经所有节点审核通过后,区块链录入数据交易链中。图4为信息交易链示意。

交易信息链特征为:数据索引链中对数据持有者的记录将为交易双方提供权限证明;每个成员都可以对交易链进行下载和检索,确认自己交易是否在交易链中,即形成对交易平台的监督,也为后续的交易追溯和申诉提供证据。

数据索引链和信息交易链同时由交易平台记录更新,

各参加机构维护审核。经过数据交易链审核通过后,交易双方进行实际的数据交付。现列出可能的交付方式:通过API访问的方式、用户侧的App到数据提供者处运行等方式。

获得的数据与描述不相符或获得的信息经证实为虚假信息时,可以向交易平台提交投诉申请。

数据检索链和信息交易链将共同给投诉申请提供证据支撑。

图5为以区块链技术为基础的大数据交换流程描述。

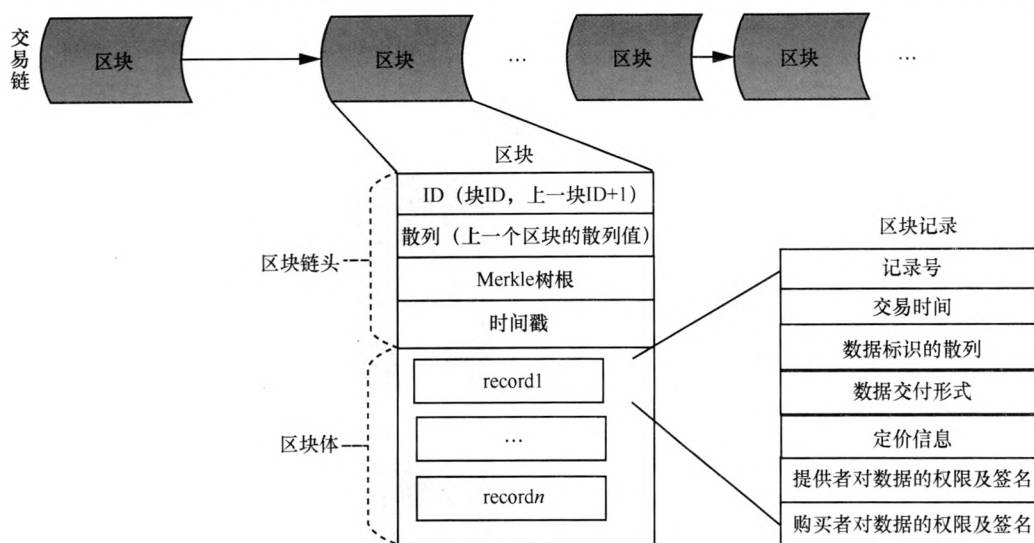


图4 信息交易链示意

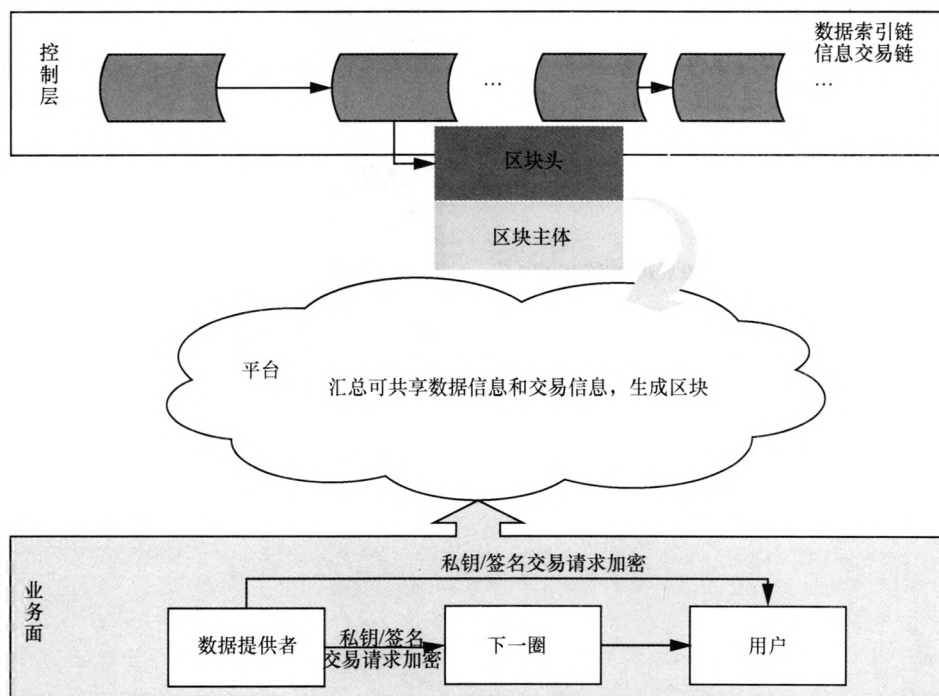


图5 基于区块链技术的数据交换示意

应用流程如下所示。

(1)数据交易方申请加入交易系统,并经系统审核认证。

(2)各方数据提供者向平台提供对外共享或交易的数据描述、交换权限及交换价格等基本信息。

(3)交易平台将收到的所有成员发送的记录汇总形成区块,经审核后被链接到数据索引链末端。

(4)当数据需求方通过数据索引链查询到平台具有所需数据资源时,都需经过数据原始提供者的鉴权操作、许可操作。若审核通过,视为合法交易,数据原始提供者将交易信息发送给交易平台。

(5)交易平台对交易信息进行记录,形成区块,并广播到全网进行审核。若审核通过,视为合法交易,区块附加在信息交易链末端。

(6)交易双方按照约定的数据交付方式进行交易。

(7)发生投诉时,数据检索连和信息交易链为其维权提供证据支撑。

## 6 结束语

大数据交换信息链,保障数据源提供者的权益,使数据在流动过程中可管、可控,加快数据交换和流通,是对数据交易模式一种重要创新。大数据交换信息链的建立对数据交换过程的业务层面和控制层面的联系起到了关键作用,下一阶段将继续深化对大数据交换信息链的研究,完善细化其在整个数据交易流程中的作用,在数据提供者、中间商、用户等参与者间真正实现确权、许可、兑汇、核算的目标。

## 参考文献:

- [1] 中国信息通信研究院. 2015年中国大数据发展调查报告[EB/OL]. (2015-05-29)[2016-07-20]. <http://www.jssme.com/smenews/service/view/71200>.

China Academy of Information and Communications Technology. A survey of China big data development in 2015 [EB/OL]. (2015-05-29)[2016-07-20]. <http://www.jssme.com/smenews/service/view/71200>.

- [2] 李骥宇. 大数据交易模式的探讨[J]. 移动通信, 2016(5): 41-44.

LI J Y. The discussion of big data trading patterns [J]. Mobile Communications, 2016(5): 41-44.

- [3] 杨琪, 龚南宁. 我国大数据交易的主要问题及建议 [J]. 大数据, 2015(2): 38-48.

YANG Q, GONG N N. Reflections on big data exchange of China[J]. Big Data Research, 2015(2): 38-48.

## [作者简介]



刘楠(1991-),女,现就职于中国联合网络通信有限公司研究院软件与系统实验室,主要研究方向为大数据技术。



魏进武(1978-),男,博士,中国联合网络通信有限公司研究院软件与系统实验室主任,主要研究方向为大数据、云计算以及电信IT系统等的设计及研发。



刘露(1984-),女,博士后,现就职于中国联合网络通信有限公司研究院软件与系统实验室,主要从事大数据平台和云计算等架构设计和技术研究工作。