

我国大数据交易的主要问题及建议

杨 琪, 龚南宁

数据堂(北京)科技股份有限公司 北京 100190

摘要

数据的开放和流通是数据资源价值体现的前提和基础,我国数据交易市场仍处在发展的初级阶段,大数据在社会管理和经济发展中远未发挥应起到的作用。为此,对国内外大数据流通和交易市场的现状及特点进行了分析,提出当前我国大数据流通交易环节面临的主要问题是价值链的不完整和对数据资源流通过程中商业秘密和个人隐私泄露的恐惧。最后,从数据商品化、社会认知建立和市场主体权益保护3个方面提出了对于推动数据资源流通的建议。

关键词

大数据;交易;商品化;数据开放;隐私;安全

doi: 10.11959/j.issn.2096-0271.2015017

Reflections on Big Data Exchange of China

Yang Qi, Gong Nanning

Datatang(Beijing) Technology Co., Ltd., Beijing 100190, China

Abstract

Data circulation is the decisive factor realizing big data value. The benefits of big data in social management and economic development are highly restricted because of immature of data exchange market of China. Based on the contrast analysis of the data circulation market, awareness was proposed that the data exchange was mainly hindered by the integrity of the value chain as well as the fear of business secret and personal privacy leakage. Finally, solutions were advised from the following aspects: the commercialization of data, the establishment of social cognition, the protection of the rights and interests of market entities.

Key words

big data, exchange, commoditization, open data, privacy, security

1 引言

大数据的出现,引发了全球范围内深刻的技术与商业变革,已经成为全球发展的趋势以及国家和企业间的竞争焦点,直接关系到国家安全、社会稳定、经济发展和民生幸福等诸多方面。数据资源日益成为人类社会的生产要素和战略资产,而数据的开放和流通是其价值体现的前提和基础。据Gartner Group公司预测,到2017年,约2/3的大数据整合项目将是企业防火墙之外(外部数据)的整合;根据Teradata天睿公司的调查显示,很多公司仅存储了15%与其业务相关的数据,其他85%的数据驻留在其他公司或网站上。麦肯锡全球研究院预测,开放数据在全球的教育、交通运输、消费市场、电力、石油/天然气、医疗健康、消费金融(包括银行、保险和房地产)7个领域可以撬动3.2万亿~5.4万亿美元的经济价值,如图1所示。

由于多源头、跨领域的关联分析才有可能形成更完整的知识和更深刻的智能,数据资源的流通日益成为普遍认知和客观需求。在Bloomberg Ventures创始人Matt

Turck的研究中,专门将与数据流通直接相关的数据生产者、汇集平台以及数据集市(大意如此,原图在各区块内部的细分逻辑上稍显凌乱)归纳为产业链上的数据供给(data sources)环节,如图2所示。

2 数据交易市场概况

针对任何一种商品交易,都可以从供需两端入手进行分析,以形成市场概貌。根据图2的大数据产业区块划分,能够最为直观地反映数据交易流通概貌的就是对产业链条两端的分析,即对数据源和应用区块的分析。

(1)数据源机构或企业处于整个大数据市场的供给端,其主要特征是向用户直接交付数据产品或服务,包括原始数据、加工处理后的数据以及由多份数据整合后的新数据,分别对应数据生产者、数据加工者和数据整合者的角色。

(2)应用端则反映了数据交易的客观需求,是大数据交易市场的主要推动力,而且还带动了分析区块内的企业进入数据流通市场。应用端的发展程度决定了大数据交易的规模上限。

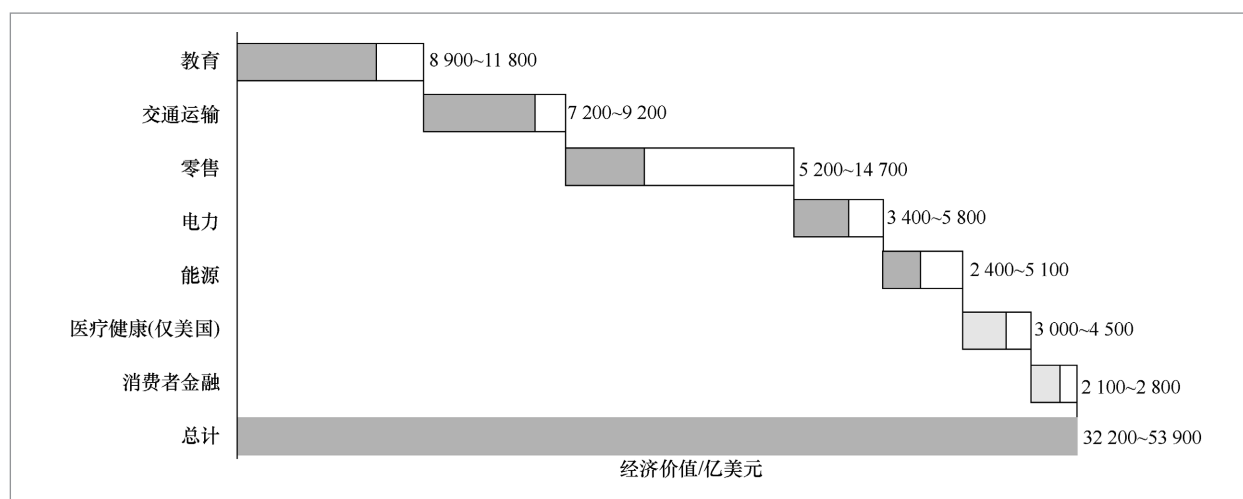


图1 全球数据开放价值(来源于麦肯锡全球研究院)

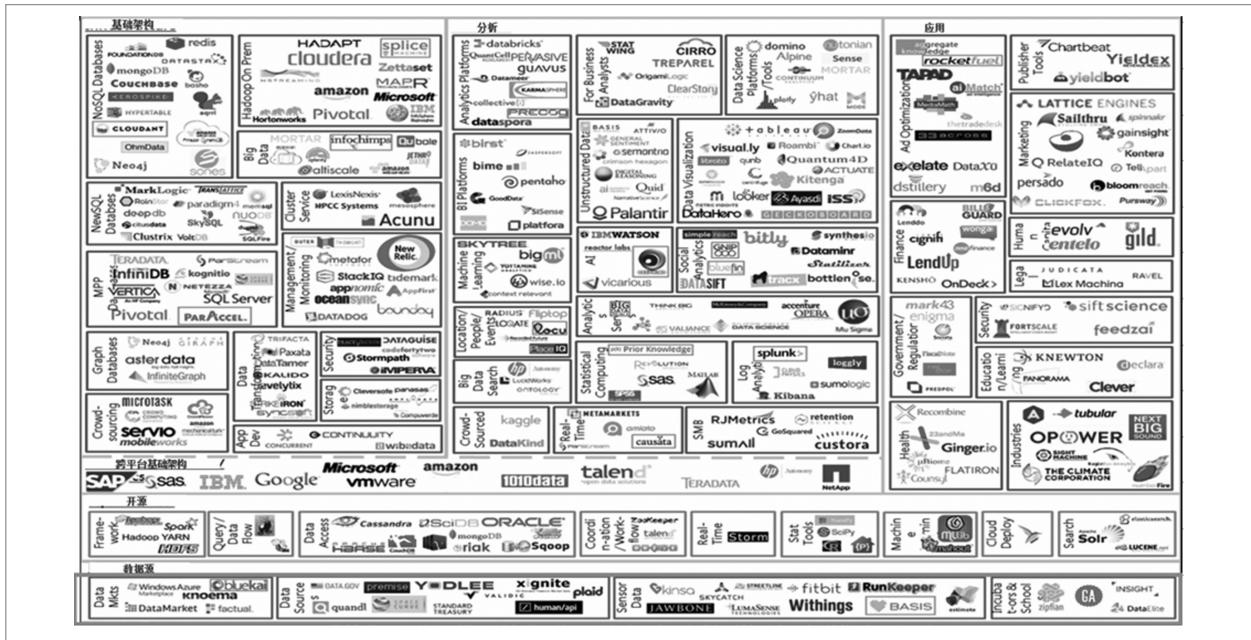


图 2 大数据产业链分布

2.1 国外数据流通和交易现状及特点

在数据的流通和交易方面，欧美发达国家尤其是美国已经走在了前面，图3为国外数据市场的概貌。

如图3所示，数据中介通过政府、公开和商业渠道，从数据源头处获取各类信息，进而向用户直接交付数据产品或服务。其中，数据源头、数据中介和最终用户构成了数据流通和交易的主体。

数据源头和中介环节共同构成了大数据资源的供给端，其中比较典型的案例见表1。

Twitter将自身数据授权给公司Gnip、DataSift和NTT DATA进行售卖；Acxiom等公司通过各种手段收集、汇聚关于企业和个人的信息；Sermo.com和Inrix等公司则通过网络和传感器直接从公众采集数据，获得了传统上单个企业难以采集的海量、实时数据。

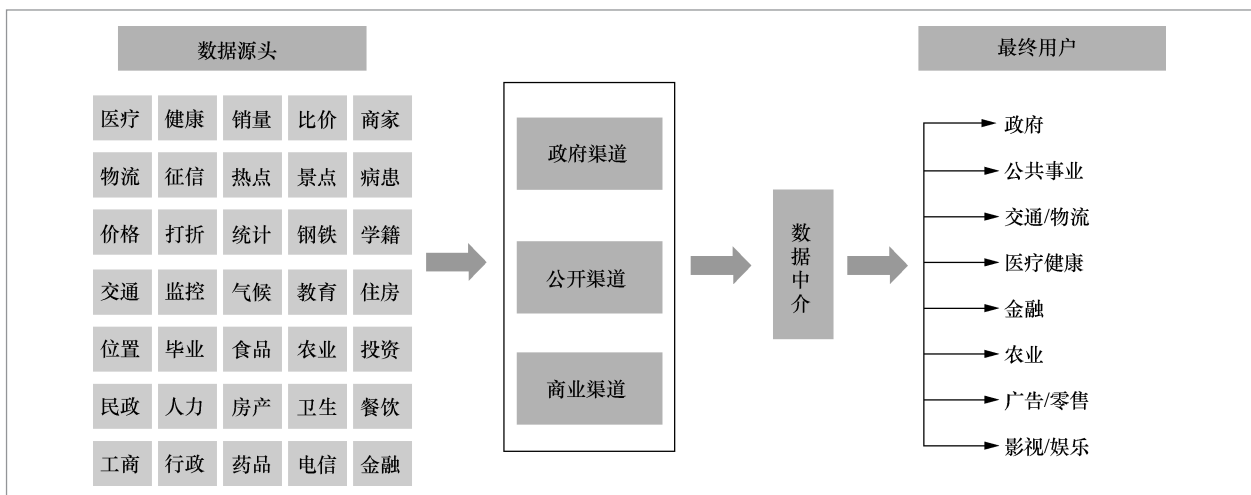


图 3 国外数据市场概貌

表1 数据市场供给端典型案例

数据源头	获取数据方式	数据中介
Twitter	授权	Gnip、Datasift、NTT DATA
各类数据	汇集	Acxiom、Experian
经济数据、金融数据	汇集	彭博、路透、Timetric、1010data、Datamarket
上网行为	汇集	BlueKai、Lotame、RapLeaf
医疗记录、信用信息	汇集	Opera solutions、SureScripts
地理位置	汇集	Factual、InfoChimps
医疗信息	众包	Sermo.com
路况信息	众包	Inrix

数据市场的另一端是数据的需求方,包括各类数据分析服务商和行业用户,涉及政府决策、公共服务、影视娱乐、交通物流、医疗健康、金融、电信、人力资源、零售、广告营销、农业、能源等领域。表2为各行业或领域内的数据用户(即引入外部数据支撑自身产品或服务的企业或机构)总结。

根据数据供给端和需求端的概况,可以总结出以下特点。

(1) 数据中介大多以采集和聚合为主

数据中介层在整个大数据流通和交易中地位突出。根据美国参议院商务、科学与运输委员会发布的报告,全美数据中介市场2012年的总规模已达1 500亿美元,相当于当年美国情报总预算的两倍。相反,数据生产者很少直接面向最终用户,大多通过中介渠道实现自身数据的变现。例如Twitter这

样的海量数据拥有者,也仅是通过授予公司Gnip、DataSift和NTT DATA数据转售权实现数据资源的变现。

(2) 集市类的形态逐渐弱化

全球范围内,相关平台都最终在数据类型上有所侧重,不再以“综合性”为主要策略。Datamarket公司以国民经济与工业相关的数据集为主;InfoChimps公司在地理位置、社交网络、网络信息等方面的数据更为突出,且逐渐转型为PaaS平台;Factual公司从提供全范围的数据交易平台转为专注于提供地理位置相关的数据集。更为极端的例子是致力于社交网络数据的数据市场Kasabi(英国)在2012年7月关闭,仅经营了一年多。

(3) 用户端需求广泛、应用典型

在大数据当前主要的应用领域内,都有

表2 典型数据用户总结

领域	企业或机构
影视/娱乐	Netflix、Rentrak、Pandora、Caesars、United TalentAgency
交通/物流	UPS、Amazon、Uber、Flyontime.us
医疗/健康	Flatiron、Carolinas HealthCare System、Foundation Medicine、IBM&BlueCross
金融	挪威DNB、Zestfinace、BCG、MarketPsych、DerwentCapital Markets、Kabbage
广告/营销	Locately、AirSage、NTT、Route、Gainsight、WGBH-TV
人力资源	Kiran Analytic、Glassdoor、Identified、ResumUp
零售	Walmart、Target、Bloomreach、RelatelQ
农业	Climate、Solum
能源	Vestas Wind Systems、The California ISO、Energy Hub
政府决策公共服务	Fujitsu、WideNoise

比较典型的、引入外部数据支撑自身业务的案例。Rentrak公司基于机顶盒数据,监测各种屏幕上的媒体消费情况,为影视制作公司和广告公司提供咨询服务;Carolinas HealthCare System公司采集200多万客户的消费数据,识别其中高风险的患者;SAP公司从运营商处收集智能手机使用信息和位置数据,并销售给市场营销机构;

表3 国内典型数据供给企业

领域	企业或机构
综合性平台	数据堂、数海、贵阳和武汉大数据交易所等
建筑类	广联达(销售建筑材料价格信息)
金融类	万得信息、九次方、通联数据等
气象类	华风气象
地理/环境类	长地万方、凯立德、北京城际高科、科菱航睿等
生活类	电话邦(号码数据)、无忧停车(停车场信息)等
车联网类	北京汇通天下(销售汽车发动机数据)

表4 国内潜在的大数据供给源

领域	企业或机构
影视/娱乐	克顿、星红橙、猫眼电影、粤科软件
旅游业	汇通天下、中长石基
医疗/健康	慈铭体检、爱康国宾
金融	恒生电子、银联智惠
房地产	易邀中国
航空	中航信
能源	海博思创、国网信通
农业/环保	太谷雨田、云洲智能
车联网	腾讯路宝、元征科技、Autobot、博泰电子
个人/企业	易达讯、上海万达

表5 国内数据市场的用户端典型案例

领域	企业或机构
语音服务	科大讯飞、云知声、思必驰
广告/营销	亿赞普、中长石基
医疗/健康	百度问诊
金融	宜信、元宝铺、数联铭品
房地产	万科地产、万达地产
交通/物流	飞常准、途志、快逸行
地理信息	百度地图、腾讯地图、高德地图

Kabbage公司使用来自于亚马逊公司、UPS公司和Intuit公司的信用评级模型数据,评估中小企业的风险等级;Glassdoor和ResumUp等公司通过各种方式收集企业信息,为求职者提供企业评估服务;Climate公司从200多万个采集点获取天气数据,向农民提供保险服务;Energy Hub公司与传感器网络厂商Earth Networks合作,利用气象数据提高能源供给设施的效率。

2.2 国内数据流通和交易现状及特点

在具体的领域或行业内,我国普遍未形成成型的数据采集、加工、分析和应用链条,以数据为主要服务或产品的案例较少,国内典型数据供给企业见表3^[1]。

尽管案例较少,但实际上我国有大量掌握海量、高价值数据的企业或机构,具有成为社会化数据商品源头的潜质,见表4。

除了上述企业外,特定地区和行业内部也有基本的数据流通机制,比如行业内的数据共享联盟。而在政府端,各地政府也已逐步推动数据开放的工作,其中比较典型的是北京市政务数据资源网和上海市公共信用信息服务平台。

我国数据市场中,引入外部数据支撑业务的典型案例见表5。

结合我国数据市场的供给端和用户端,可以得出以下主要特点。

(1) 大量数据源未被激活。大多数数据拥有者没有数据价值外化的路径。比如,各医疗健康类应用收集了大量的数据,但没有像Sermo.com那样面向医药公司售卖数据。

(2) 需求端以互联网企业为主,覆盖面不广。尤其在O2O趋势下,大型互联网厂商积极引入外部数据支撑金融、生活、语音、旅游、健康和教育等多种服务。但是,与国外相比覆盖面偏窄,我国的政府、公共

服务、农业应用基本缺位，而电信和银行业缺少与外部数据的碰撞。

(3) 诸多企业基于自身的海量高价值数据对外提供服务。比如阿里巴巴网络技术有限公司开始做金融行业的数据分析应用，中国民航信息网络股份有限公司通过“航旅纵横”应用提供航班信息服务。

3 国内数据流通环节的问题

数据源活性不够、应用覆盖面较窄等问题的根本结症在于我国大数据产业发展还处于非常初级的阶段。除了语音和地理信息类数据之外，其他数据的价值实现链条尚未成型。而且，在面对数据资源流通的问题上，数据拥有者普遍存在恐惧心理或没有足够的动力。

3.1 大数据价值实现链条的缺位

参照国外市场的发展现状，数据市场的成熟离不开供给、中介（汇集、清洗和加工）和需求端的协调发展，这3个环节构成了大数据价值实现链条的基本要素。

国内外典型的数据价值链如图4所示。一个典型的国外大数据价值链案例：Sprint公司将设备位置信息提供给数据集成与分析商Locately，由后者在汇总分析之后将结果提供给市场营销公司HAVAS和Mobext，最终Whole Foods、Sears、Target和沃尔玛等连锁超市获得相应的市场调研报告及咨询建议。与之相对应，我国较为成型的数据价值链主要体现在语音和图像领域，比如数据堂（北京）科技股份有限公司采集众客的语音数据，在经过加工标注之后提供给科大讯飞股份有限公司等企业。

相对于其他商品，数据作为商品被用来交易的历史十分短暂。尤其在我国的，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。

大数据的本质并未脱离传统的数据分析概念，已有的专业化分工模式不会因为大数据带来的新特征而消亡，相反，进一步强化了对各环节专业化水平的要求。

(1) 多源数据的汇集。由于数据来自于政府机构、企事业单位、科研院所及社会大众等多个源头，集成整合的技术方法还处于摸索阶段。

(2) 非结构化数据的处理。音频、视频、图像和文本等非结构化数据迅猛增长，使得传统的数据挖掘技术和工具无法满足大数据时代的需求。比如，对于纸质票据，要经过自动化标注的环节转化为结构化的数据，才能为后续的分析挖掘环节所用，纸质票据的处理过程如图5所示。

(3) 应用场景的探索。应用环节是大数据价值实现的落足点，数据在不同行业的利用离不开特定领域的知识，需要大量传统行业信息技术和业务人员的积极参与。

上述制约因素严重妨碍了数据的分析和挖掘工作，使得数据商品的价值无法充分体现，导致数据供给和需求之间存在

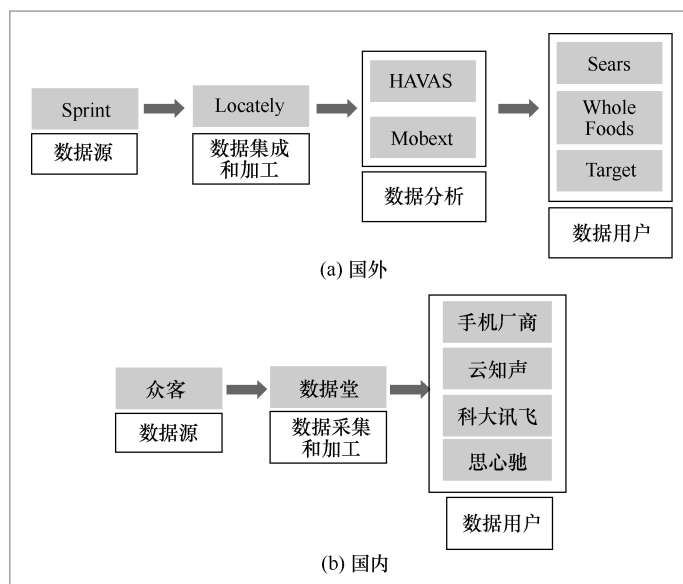


图4 国内外典型的数据价值链

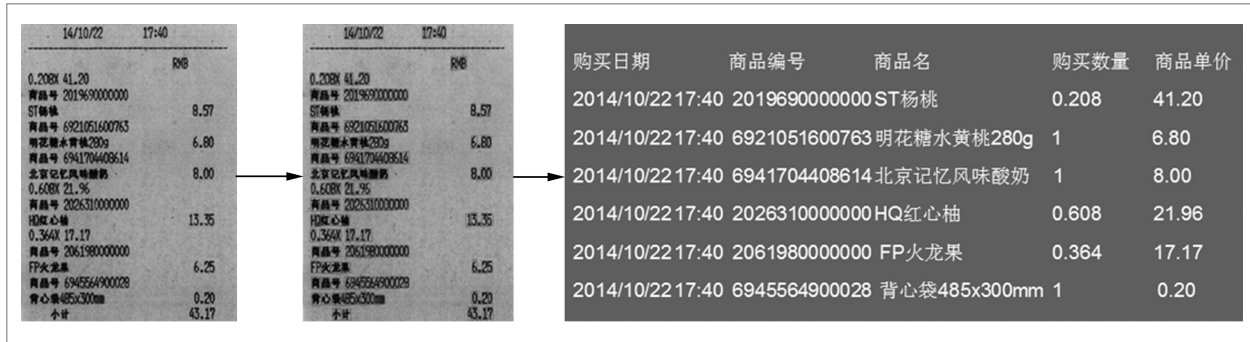


图5 纸质票据的处理过程

巨大的差距，严重抑制了数据流通和交易的活跃度。而数据类型越单一、应用场景越聚焦，技术和业务模式相对就越容易成型，数据产品或服务才越有可能实现规模化的生产和交付。因此，数据集市型的平台由于很难对所有数据都进行深度的价值提升，正逐渐淡出或转型。

3.2 对于数据流通的恐惧和回避心理

价值认知的鸿沟和避险心理。数据交易的目的是促进数据的流动和价值体现，但是不同数据拥有者对数据资产的价值和风险认知存在较大差异。因为对数据中蕴含的信息缺乏足够的洞察，很多数据拥有者不放心让自身的数据进入流通环节，担心用户隐私或企业机密泄露。

对数据外部性认知不足。数据拥有者无法意识到自身数据的资产属性，缺乏足够的动力将自己的数据公开。比如，搜索引擎出于服务的目的，记录了用户搜索所输入的关键字，而这些数据可以被卫生部门用来进行疾病的监控与防治。数据能够对数据生产业务相关方之外的第三方产生影响。从另一方面而言，政府、企业或组织都还未充分认识到引入外部数据可以对自身工作或业务起到巨大的提升作用。

数据资源的垄断意识较强。大多数数据源企业仍然较为关注自己的小生态圈，尤其是一些大型企业往往不愿意把自己的

数据资源向自己业务圈外的市场提供，从而形成了多个规模和性质各异的数据封闭生态。尤其在我国，全社会普遍还未形成数据分析的传统。在大数据氛围的激发下，对于拥有海量数据资源的企业或机构而言，撬动自身数据资源支撑业务的优先级必然会高过通过数据交易获得收益。

4 对于推动数据资源流通的建议

针对上面提到的两个主要问题，可以通过商品化改造、社会认知提升、安全顾虑消解等几个方面着手解决。

4.1 数据的商品化改造环节

对比美国的数据市场形态，对于价值链条缺位的问题，最关键的是培育数据的商品化环节，包括纠错、补漏、标准化、整合等方面。数据商品化改造的核心目的在于降低数据利用的门槛，为用户端进行数据分析和挖掘提供便利。数据商品化环节是数据价值链条上的共性需求，比如美国三大信用机构之一的Equifax，专门引入外部的技术力量进行重复数据的删除和清洗工作。

对于商品化改造效果的评估，不同数据必然有不同的标准。不过总体而言，有一些共性的抽象指标可以作为参考，比如准确性（是否真实反映了客观现实）、完整

性(数据是否有缺漏)、一致性(同类数据表述形式是否一致、多次出现的同一数据是否存在差异)、及时性(数据是否根据客观现实的变化而及时更新)和可用性(是否容易被用户理解、是否便于通过用户现有的技术手段进行分析和挖掘)等。除了上述共性指向之外,针对每类具体的数据都有必要根据用户共有的需求,提炼出针对具体行业的数据商品化标准。比如,语音数据所需的元数据包括是否包含有效语音、语音对应的文字、语言本身的描述、噪音情况、时长、录音设备描述以及发音人的性别、年龄、地域、口音等。

除根据数据市场的客观规律打造专业的商品化改造环节之外,根据数据价值链条还处于发展初期、广大用户端企业缺乏专业数据处理和分析手段的现状,我国还具有孕育新型产业形态的可能性。就我国具体情况而言,由于大部分企业尤其是传统行业用户缺乏对数据进行改造的能力,很难直接对原始形态的数据进行挖掘和分析,加之文本、图像和音视频等非结构化数据的激增,使得市场对更为“强大”的数据中介层有着强烈的需求。此外,一个兼具技术、信息安全和法律责任的数据价值转化渠道,还可以在很大程度上消除数据拥有者的顾虑,使之能够对自身数据资源在市场上的流通具有一定的控制力,从而比多对多的交易集市类平台更具安全感。与银行所具有的金融资产供需对接的中介功能相似,这类数据中介的新形态可以称之为数据银行。数据银行和数据中介的对比如图6所示。

数据银行的特点在于承担了部分源头和用户端的职能,将资源的对接提升到资产改造、优化和组合的层面,更加强调对数据价值的深挖。比如,数据银行可能根据用户端的要求,定向采集当前并未实现电子化的信息(这与数据中介只是实现各现有数据源的物理汇集有着根本区别),

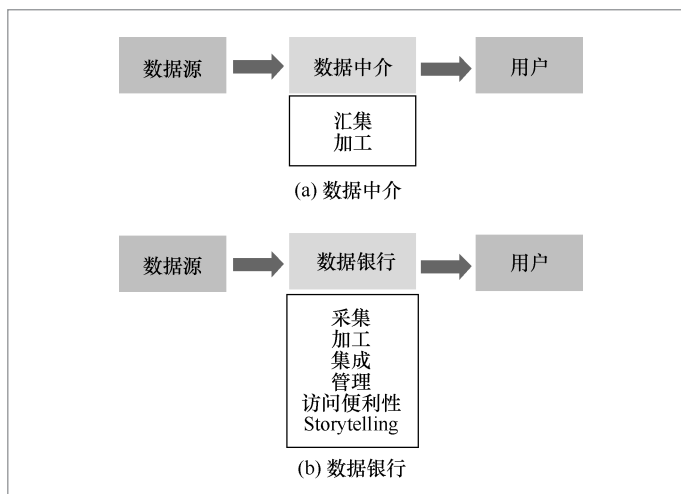


图6 数据银行和数据中介的对比

或在市场需求和技术都就位的前提下,开展一定的分析活动,包括对非结构化数据的转化和标注。

同时,数据银行还更多地承担起粘合作用,比如,通过API(application programming interface,应用程序编程接口)或应用开发平台的形式,保证用户可以方便地访问和利用所需数据;在开拓数据源和最终用户时,需要对数据应用场景和方案进行预先的储备,以此让市场供需两端确信数据流通所能带来的价值,即storytelling、布道或咨询的职能。从实际的情况来看,国内的综合性流通平台已经有朝数据银行形态转变的趋势。

数据银行与交易型平台都为数据资源的流通提供了规模化的通路,但是两者具有重要的区别。交易平台更关注数据的买卖交易,以规模化为主要追求目标,是一种商品集散的概念;而数据银行则在追求流通规模的同时,注重数据价值的深层挖掘,是资产改造、组合和融通的平台,核心能力不在于实现客户数量的最大化,而是让数据资源能深度嵌入到特定领域的价值链条中。直观地说,数据银行就是以价值实现链条为基础的横向聚合,可能更符合数据市场发展的客观规律和我国国情。

4.2 以政府数据开放带动社会认知的提升

站在数据拥有者的角度,对数据的外流心怀恐惧是很正常的:用户数据被外界挖掘(即便已经经过脱敏和处理),需要承担暴露隐私的风险;数据内容极有可能暴露企业的战略和具体方法;担心数据归属问题导致企业无形资产受损。而在数据用户一端,同样会担心隐私问题导致的连带风险以及质疑引入外部数据的投入产出比。这些因素使得数据交易在北美从未被完全晒在阳光下,诸如Axicom公司和Experian公司等数据中介巨头都拒绝向调查委员会透露自身的数据源头以及最终的用户^[2];2012年,当美国政府要求数据中介保障公民获知自身信息被采集情况的权利时,Axciom公司就以没有通过姓名查询数据库的功能为由拒绝。

尽管如此,上述种种顾虑在我国多是由于舆论影响而形成的,不能成为行动迟缓的正当理由。而且,通过正确认知的树立,疑惑和顾虑是可以被淡化乃至消除的。

消除市场参与者的顾虑,达成正确认知,最有效的方法就是以政府数据开放为切入口,为全社会起到正面的示范作用,原因在于以下几点。

(1)从企业角度来说,在数据意识刚刚觉醒的阶段,如果自身数据还未充分利用于本身业务,则自然会将数据首先视为一种差异化竞争的资本。要想打破数据藩篱,必须用现实案例说明数据外部性的巨大价值,才能让企业意识到数据的流通也可以带来可观的收益。

(2)政府数据的开放最可能快速催生巨大的经济和社会价值,从而形成巨大的示范作用。政府以及相关机构拥有最高价值的数据(如统计、税收、治安、土地、就业、环境、交通等各类数据),可以为各行

各业所充分利用。当前,美国、英国、法国、日本、德国、加拿大、澳大利亚、新西兰和新加坡等都纷纷提出了政府或公共数据开放的战略,并付诸实施。

(3)横向比较,我国政府数据的规模和价值总量更为可观。由于国土、人口、经济规模以及政治体制的原因,我国已成为仅次于美国的数据大国。其中,据麦肯锡全球研究院分析,我国1/3的数据属于政府和公共事业单位,在政府决策、公共管理和经济发展方面具有极大的挖潜可能。

在具体的行动上,北京、上海等地都建设了政府数据开放网站,贵阳和武汉等地也上线了大数据交易平台,但是仍然可能受机密数据和个人隐私泄露风险的羁绊。尽管国情不同,在此仍以美国政府数据开放为例,希望能起到一定的借鉴作用。美国政府首任首席信息官昆德拉在负责Data.gov建设的过程中,对于数据原始性、数据安全、数据质量、数据价值和是否物理汇集等主要争议,仅就原始性和是否物理汇集做出了明确的应对,而对于其他问题则采取迂回的策略,从没有争议的数据入手,快速上线、不断迭代。

从上面美国政府数据开放推进的例子还可以看出,对于大数据交易乃至大数据产业的发展,快速起步、不断迭代的模式可能比传统的自顶向下、集中化、大一统的模式更为适用,尤其是在发展初期更是如此。同时,应该正确识别市场发展的真正推动力量,在具体推进过程中多从最终用户入手,让各行各业的业务和技术力量参与、发声。

最后,大数据本就不是一个严格的、有明确界限定的技术或商业词汇,理解内涵远比呆板的词句更为重要,而且各方面的探索才刚刚起步。因此,如果伴随着政府数据开放的动作有相关的标准或规范出台,需要考虑留出足够的弹性和想象空间,

谨防成为数据市场和产业本身发展的限制因素。

4.3 信息安全顾虑的消解

数据商品中可能蕴含涉及国家机密、企业核心信息和个人隐私的内容，当前数据源最担心是数据不受控制地流动。同时，在单份数据内容安全的情况下，对多个数据整合在一起进行关联分析，也有可能造成隐私的泄露^[3]。对于消除信息安全方面的顾虑，可分解到制度建设和数据中介机构业务两个层面进行考虑。

(1) 制度层面的举措

作为数据市场的关键主体，数据中介是相关制度的主要针对对象。具体的举措可以包括资质认定、信息披露和日常评估等。其中，需要特别注意的是信息披露环节，要在披露的形式和范围上精心设计，否则可能导致数据交易市场停滞不前。

相对于事前预防，更重要的是事后追责。风险永远存在，正如关联、跨域分析可能会造成的隐私泄露，事前预防的难度极大。因此，除上述措施之外，更为重要的是事后的惩戒制度。只有足够高的违规成本，才能比较彻底地消除数据中介和最终用户泄密的可能性，并促使数据中介切实承担起确保数据资源安全的责任。可以作为侧面印证的是，有观点认为，Facebook、LinkedIn等企业因为财力雄厚才能承受各类法律官司和公关成本^[4]。

此外，在立法和规范之间，应该首先以行业自律规范类的形式来确保数据的信息安全。可以作为旁证的是，在认为数据服务商总体上是为社会带来更多正面影响的前提下，美国政府一直以行业组织的形式来规范相关企业和机构的行为，直至数据规模增长到能够导致较大负面效应时，才开始制定相应的法律法规。

(2) 数据中介的举措

站在数据中介的角度，需要在两个界面上对数据安全作出应对。首先，要在经济利益、资产归属、技术实现、业务限定等各个方面打消数据源的顾虑。根据数据源性质、数据所包含的信息、未来数据服务面向的受众等因素不同，在具体协议制定上必然会有所不同，但有些注意事项是具有普适性的。

- 数据产品和服务的受众控制。对数据中介可以售卖的受众加以限制，或者对数据服务接口的访问进行限制（如果以API形式提供）。

- 对可流通的数据范围进行规定。对数据中哪些信息可以进入流通市场加以限制，防止商业机密或个人隐私外泄。

- 系统对接的安全性要求。与数据源一起就系统对接、数据传输和存储等环节进行协商。

- 就收益和版权事宜进行协商。这一点主要针对需要进行深度改造的数据商品，双方针对预期的清洗和加工环节进行评估，对数据增值幅度和市场定价进行沟通，达成收益双赢的局面。另外，对数据版权进行明确的规定，规避可能发生的纠纷。

- 对违约行为的追责。在数据源稳定性、更新频率、数据扩散范围等方面商定明确的责任界定和处罚协议，打消源头的顾虑，并为对外的数据服务提供保障。

对于与最终用户的条款，可以参照与数据源之间的协议，并且可以通过API等技术手段进行辅助，控制数据的流向、访问对象和扩散范围。

5 结束语

从对国内外数据流通市场的分析和对比可以看出，我国大数据流通和交易仍缺

乏基本的产业链支撑,同时大数据相关企业尤其是位于产业链最上游的数据源对数据资源的外流存在较大的疑虑。数据资源的流通和交易必须以相对成熟的数据价值链条为基础,否则数据价值的挖掘和体现只能停留在较低层面,远不能发挥数据在社会管理和经济发展中本应起到的作用。无论政府、行业还是科研机构,应聚焦重点应用领域,着力数据采集、处理和分析环节的培育,避免目标泛化、过分追求规模效应以及对隐私和安全的过分忧虑,这对国内各大数据交易平台以及相关规划工作尤为重要。

参考文献

- [1] 数据堂(北京)科技有限公司.大数据产业调研及分析报告. http://www.datamall.com/article-news_gongsidongtai-164, 2015
- [2] Datatang (Beijing) Technology Co, Ltd. Research and analysis report of big data industry. http://www.datamall.com/article-news_gongsidongtai-164, 2015
- [2] Committee on Commerce, Science, and Transportation. A review of the data broker industry: collection, use, and sale of consumer data for marketing purposes. http://www.commerce.senate.gov/public/?a=Files.Serve&File_id=0d2b3642-6221-4888-a631-08f2f255b577, 2013
- [3] Samarati P, Sweeney L. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Proceedings of the IEEE Symposium on Research in Security and Privacy, Oakland, USA, 1998
- [4] Ken G. Business Models for the Data Economy. Sebastopol: O' Reilly Media Inc, 2013

作者简介



杨琪,男,数据堂(北京)科技股份有限公司企业发展部副总监,主要研究方向为数据交易、大数据产业、基础架构。先后获得国家科技进步奖二等奖、信息产业重大技术发明奖、北京市科学技术奖一等奖、教育部科技进步奖二等奖。



龚南宁,男,数据堂(北京)科技股份有限公司对外合作部副总监,主要研究方向为数据商品化、数据银行。

收稿日期: 2015-07-03

论文引用格式: 杨琪,龚南宁.我国大数据交易的主要问题及建议.大数据,2015017

Yang Q, Gong N N. Reflections on big data exchange of China. Big Data Research, 2015017