

数据集市及其在交易中的价值*

■ 赵栋祥¹ 陈烨¹ 张斌²

¹ 武汉大学信息资源研究中心 武汉 430072 ² 武汉大学中国传统文化研究中心 武汉 430072

摘要: [目的/意义]数据集市是一种新兴的数据服务形态,通过分析其产生发展的环境、理论基础和探索实践,为大数据环境下的数据服务创新提供一些思考和启示。[方法/过程]从大数据环境对数据管理和服务的影响出发,分析数据集市产生环境;基于对大数据管理流程和大数据价值链的分析,论述数据集市的理论基础;选取数据集市在实践应用中的典型案例,总结其较为成熟的服务模式;通过梳理从传统信息环境到大数据环境对数据价值认知的演变,分析数据集市的价值发现。[结果/结论]大数据环境从多方面影响着数据集市的产生和发展;数据生命周期和大数据价值链是数据集市的重要理论基础;数据银行和众包是数据集市的重要服务模式;科研、商业和政务是数据集市价值发现的典型领域。

关键词: 数据集市 数据交易 数据价值 大数据环境 服务创新

分类号: G203

DOI: 10.13266/j.issn.0252-3116.2017.13.001

随着“互联网+”时代的到来,大数据成为与云计算、移动互联网、物联网和人工智能并驾齐驱,推动经济发展和行业融合的强大引擎^[1]。信息技术的进步推动着数据环境的转变,数据的采集、处理、交易和应用等都产生了显著变化,持续创新商业模式和服务业态^[2]。大数据交易作为大数据环境下一种新兴数据服务形态,具有广阔的发展空间。国外数据交易市场始于 2008 年,数据服务提供商包括美国的位置数据及服务开放平台 Factual、实时数据交易市场 BDEX、大数据平台提供商 Infoclimps 以及日本的富士通数据市场 Data Plaza 等^[3]。国内数据交易虽然起步较晚,但发展迅猛,据《2016 年中国大数据交易白皮书》预计,2020 年国内大数据交易市场规模将达到 545 亿元^[4]。

“数据交易服务”引起了学者们的广泛关注。美国是数据交易服务领域的领头羊,其数据交易产业是在数据经纪产业的基础上发展起来的,数据集市(data marketplace)也延续了互联网发展早期的数据经纪商的角色^[5]。该术语来源于 A. A. Armstrong 和 E. H. Durfee 的研究成果,他们认为信息集市是由许多从事信息内容和服务买卖的代理人组成,并模拟了数字图

书馆联盟的信息交易活动^[6]。德国明斯特大学欧洲信息系统研究中心的研究团队围绕数据集市这一研究主题发表了一系列文章。如 A. Muschalle 等^[7]基于访谈资料分析了数据集市的含义、定价策略和面临的挑战。F. Schomm、F. Stahl 和 G. Vossen^[8]阐述了数据集市和数据供应商的含义、特征和功能,并选取 46 个产品或服务为研究样本,从核心产品、时间特性、领域、数据来源、定价模式、访问方法、输出格式、语种、目标客户、可信性、规模和成熟度进行分类和描述性统计,从而全面调查和比较数据交易市场的发展现状以及预测发展趋势。该团队利用 2013 年 7-8 月进行的第二次市场调查数据,从退出者、新进入者和坚守者的角度比较两次调查结果,并利用 Jaccard 相似系数分析了数据集市不同维度之间的共现关系^[9-10]。基于前两次市场调查的研究结果,该研究团队继续追踪数据集市这一新兴数据服务模式的发展现状和未来的发展趋势^[11]。F. Stahl 等^[12]从新古典经济学视角解释了数据集市的内涵,认为信息通讯技术促进了虚拟化的数据交易市场的产生和发展,并基于数据交易活动的参与方、数据交易平台的所有权,建立了包含 6 种商业模式的数据集

* 本文系国家自然科学基金重点国际合作项目“大数据环境下的知识组织与服务创新研究”(项目编号:71420107026)和武汉大学自主科研项目“知识组织视角下的大数据环境解构”研究成果之一。

作者简介:赵栋祥(ORCID:0000-0001-5285-951X),博士研究生,E-mail:zhao_dongxiang@163.com;陈烨(ORCID:0000-0002-7619-3246),博士研究生;张斌(ORCID:0000-0002-5591-7874),博士后。

收稿日期:2017-03-30 修回日期:2017-05-27 本文起止页码:5-12 本文责任编辑:王传清

市模型。此外,还有对特定数据交易平台的研究。如 K. Möller 和 L. Dodds^[13] 概述了英国在线数据平台 Kasabi 的功能,它基于关联数据原理,为数据发布者提供了一种简单的方式来发布、链接和交易数据,帮助应用开发人员以 API 的形式访问和接入所需数据。T. Cao 等^[14] 介绍了动态的数据市场 MARSA,它建立了数据供应商和数据消费者之间开放和可扩展的数据通信机制。国内关于数据交易服务的研究主要从两个方面展开:

(1) 数据交易中涉及的法律问题。王融^[15] 探讨了大数据交易的数据所有权归属这一核心法律问题,认为原始底层的个人数据,所有权归用户本人所有;基于原始数据,经过充分匿名化获得的数据集,企业享有限制性的所有权。汤琪^[16] 在调查国内外大数据交易的法律政策与实践现状基础上,对交易过程中涉及的授权合法性、交易安全、交易成本、交易公平以及隐私保护等产权问题进行了分析,对我国的大数据交易从法律政策的制定、行业法规的建立、产权环境的改善等方面提出建议。史宇航^[17] 认为应该加强对个人数据交易的法律规制,打击非法数据交易,引导个人数据交易的有序进行。张敏^[18] 在分析我国大数据交易的实践情况和大数据交易平台交易规则的基础上,认为大数据交易应以交易安全和数据自由流通为原则,并通过行政法规确立自律监管与行政监管并行的监管模式,对于交易主体、交易范围、交易价格、交易质量等不同的监管事项,确定法律与行业规范分别监管的范围。

(2) 数据交易服务的发展现状、问题和对策。杨琪和龚南宁^[19] 认为数据源头、数据中介和最终用户构成了数据流通和交易的主体,分析了国内外大数据流通和交易市场的现状及特点,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条的不完整以及对数据资源流通过程中商业秘密和个人隐私泄露的恐惧,并建议从数据商品化、社会认知建立和市场主体权益保护 3 个方面推动数据资源流通。李骥宇^[20] 认为大数据交易主要是指将大数据作为一种资源,通过各种交易方式转移到需求者手中获得收益的过程;并将大数据交易分为基础大数据源、大数据技术服务、大数据应用交易 3 种模式。穆会军^[21] 以华中大数据交易所为例,从出售数据、采购数据及交易成交额等方面对其数据交易情况进行了初步分析,认为企业机构等数据交易意愿强烈、政府来源的数据需求最为旺盛、数据交易难以有效满足买方需求、数据交易管理过程存在漏洞。唐斯斯和刘叶婷^[22] 总结了我国大数据交易主

要类型:基于大数据交易所(中心)的大数据交易、基于行业数据的大数据交易、基于数据资源企业推动的大数据交易和互联网企业派生出的大数据交易;分析了交易环境、数据质量等问题,并建议从标准法规、数据开放、模式创新等方面寻求突破。宋梅青^[23] 针对现有的大数据交易平台存在数据供需错配、大数据资源定价困难、数据的时效性不强三大不足,设计了一种全新的融合数据分析服务的大数据交易平台。

综上所述,目前关于数据交易服务的研究大都集中在发展现状、交易模式、存在问题和对策建议等方面,而对于数据交易服务的产生环境、理论基础以及价值发现等问题则很少涉及。考虑到数据交易平台种类繁多,称谓不一,如数据交易所、数据交易中心、数据市场、数据交易公司等,因此,本文用“数据集市”来统一各类数据交易平台。笔者首先分析数据集市的产生环境;从信息管理的角度论述数据集市的两大理论基础,即数据生命周期和大数据价值链;总结数据集市的两种重要服务模式,即数据银行和众包模式;并选择科研、商业和政务 3 个典型领域,分析数据集市与价值发现的关系。

1 数据集市的产生环境

大数据的形成是一个渐进的过程,其来源分为 3 个方面:被动产生的数据(如传统的数据库管理系统)、主动产生的数据(Web2.0 环境下的用户生成内容,如微博、微信等社交媒体)和自动产生的数据(物联网、RFID 技术)。大数据环境下,数据管理和服务的内涵、外延、特点和类型都发生了显著变化。大数据环境对数据管理和服务的影响是全面而深刻的^[24-26],主要表现在以下两个方面:

(1) 传统数据管理和服务模式难以适应大数据环境。海量异构、多源实时的数据特征给数据管理和服务的整个过程都带来了严峻挑战,如获取满足应用需求的高质量数据资源、实现低成本高可靠性的数据存储、从海量数据中挖掘出对用户有价值的信息,并以直观形象、易于理解的方式呈现出来变得更加困难。传统技术手段无法有效实现高质量大数据资源的采集和获取,常规的数据管理工具和软件难以在一定时间内对其完成处理和分析,数据应用的领域和范围开始扩大和转移,传统数据管理和服务模式亟待变革。

(2) 数据管理和服务的需求增长。大数据环境下,由于传统数据管理和服务模式的不足,围绕数据生命周期的全过程,相应的数据管理和服务的需求逐渐

增长。政府积累了海量的交通、气象、税收、征信、人口、国土资源等各类数据,但却受限于数据分析处理能力,难以有效转化为科学决策、公共服务体系建设和社会治理创新的坚实基础。随着经济转型升级和“互联网+”的深入,企业面对组织内部和外部的海量数据,迫切需求从中挖掘出有价值的知识,并实现大数据的业务支撑和扩展应用。此外,数据泛滥和知识贫乏的矛盾越来越突出,个人所处的情境、待解决的问题以及科学知识的结构日益复杂,传统的数据管理和服务已经难以满足当前的需求增长。

麦肯锡在其研究报告^[27]中指出,大数据为数据信息服务提供了丰富的资源、技术和应用场景,在催生新兴创业公司和服务形态的同时,也倒逼传统行业企业的变革与创新,一定程度上促进了数据管理和服务新模式的涌现。一方面,传统数据管理和服务模式难以适应大数据环境;另一方面,大数据环境下数据管理和服务的需求快速增长。正是在这样的背景下,数据集市得以产生和发展。

《辞海》对“集市”的解释是“农村或城市中定期买卖货物的市场”^[28]。那么,“数据集市”就是买卖数据的市场。即数据集市是指以数据资源(包括数据集、数据 API 等)为交易对象,为数据拥有方和数据需求方建立数据供需关系、提供数据交易机制的服务形态。笔者用图的形式直观清晰地呈现出大数据环境、数据集市、数据集、数据管理和数据科学之间的关系,见图 1。如前所述,大数据环境是数据集市产生和发展的重要外部环境;数据集市以数据集、数据 API 等数据资源为交易对象。数据科学是一门关于数据的科学或者研究数据的科学,包括数据获取、数据存储和管理、数据安全、数据分析、数据可视化等相关技术方法和软件工具^[29]。数据管理是利用计算机技术对数据进行有效的收集、存储、处理和应用的过程,数据管理技术的发展有助于应对大数据环境的挑战^[24]。由此可见,数据管理、数据科学可以为数据集市的运行和发展提供理论和技术支撑。

2 数据集市的理论基础

作为一种新的数字信息资源服务方式,数据集市的本质是数据交易服务平台,如贵阳大数据交易所、中关村数海大数据交易平台、聚合数据等。它与传统买卖市场、商品集市的显著差异是交易行为和交易空间的虚拟化。数据集市有效地促进了数据要素的交易和流通,为数据价值发现提供了更多可能性,是大数据商

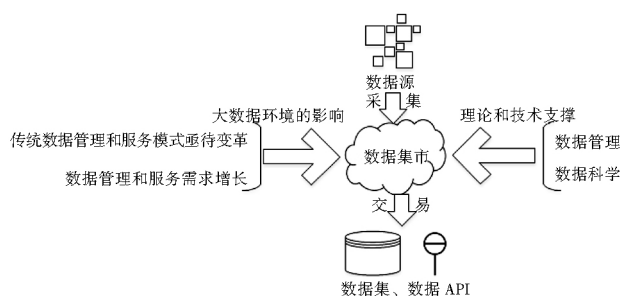


图1 大数据环境、数据集市、数据集、数据管理和数据科学之间的关系

业模式创新 3.0 时代需要关注的热点^[30]。数据集市的产生和发展除了受到大数据环境的影响外,还具有一定的理论基础。

2.1 数据生命周期

数据生命周期是数据集市的信息管理学逻辑和重要理论基础之一。英国最大的人文社会科学数据保存机构 UK Data Archive 认为,数据生命周期是一个包括数据创建、数据处理、数据分析、数据保存、数据访问和数据复用 6 个阶段的循环过程^[31]。从数据交易和数据集市的视角来看,数据产生后,通过多种方式被采集和获取,然后经过清洗、脱敏、关联、融合等数据处理过程,使之成为可供交易的数据产品,租售给用户以满足其各类数据挖掘和应用需求,获得数据源支持的用户产生的业务数据又成为新的数据源,由此进入新的数据生命周期。

由此可见,数据生命周期是一个完整的循环过程,各环节相辅相成。然而,在大数据环境下,数据生命周期的运转并不十分顺畅,主要体现为数据资源的供需矛盾。一方面,数据拥有者虽占有大量数据资源,但数据的使用场景却十分有限,寻找潜在需求方的成本很高;另一方面,大量用户需要数据资源支持和拓展自身业务,却苦于找不到相匹配的数据拥有者。

数据集市的主要目标就是缓解数据资源的供需矛盾,便捷和优化数据交易,因此其核心活动都是围绕数据交易展开(见图 2)。数据拥有方、数据需求方和数据集市是数据交易服务的主要参与方,数据集市是连接数据供给方和数据需求方的桥梁和纽带,数据供给方和数据需求方在一定条件下可以转换。其中,数据交易是必不可少的活动和功能,数据获取和数据处理则因数据集市的类型而有所差异。调查发现,数据集市可以分为两类:第一类只进行数据交易活动,如中关村数海大数据交易平台;第二类除了数据交易之外,还承担一定的数据获取和数据处理职能,如数据堂。这

两种类型的数据集市都在数据生命周期的不同环节中发挥着重要作用。



图 2 数据生命周期及各参与方的关系

注：虚线圆角方框内是数据集市的活动与功能；其中，虚线直角方框表示非必须的活动与功能；实线直角方框表示必备的活动与功能

2.2 大数据价值链

大数据的获取、处理、存储、交易、分析、解释、常规应用以及转移应用等过程使大数据价值逐渐得以实现，这一系列过程共同组成了“大数据价值链”^[32]。“大数据价值链”是伴随数据生命周期和数据服务流程的大数据价值发现、价值创造和价值实现乃至增值过程，是数据集市的又一重要理论基础。

数据集市以数据价值为导向，基于自身的技术优势，向上溯及大数据资源链条，向下延伸至用户服务、大数据应用环节，逐步建立起基于“大数据价值链”的数据服务流程。以数据堂为代表的第二类数据集市在数据生命周期中的参与度更高，因此下面的数据服务流程主要参考了第二类数据集市，并以第一类作为补充，如图 3 所示：

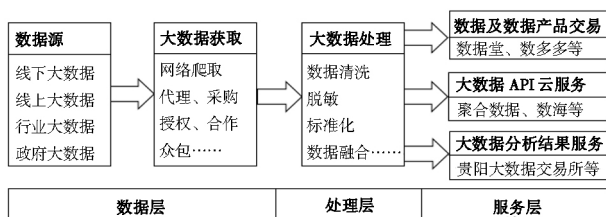


图 3 数据集市的数据交易服务流程

2.2.1 数据层 数据层是数据集市提供数据交易服务的基础，实现大数据资源的多源采集和获取。线下数据、互联网数据、政府数据以及企业在主营业务中产生的行业数据是数据集市的主要数据来源，众包、采购、网络爬取、授权代理和外部合作是其主要的数据获取方式。如金融和经济数据交易平台 Quandl 通过网络爬虫、众包和社区的贡献积累数据；数据堂基于众包平台采集了海量的线下数据资源；贵阳大数据交易所与超过 500 家企业签约，使他们成为数据交易会员；数多多依靠八爪鱼网络数据爬取平台上每天成千上万用户产生的亿级数据作为支持。

2.2.2 处理层 处理层是数据集市提供数据交易服务的关键，实现数据清洗、脱敏、标准化、结构化和数据

融合等。受限于数据质量等原因，原始数据不能被直接交易，必须经过特定处理才能转化为数据资产，因此，数据处理是数据服务流程的关键。如美国的位置数据提供商 Factual 对全球地点数据进行采集，包括数千万个地点的商业设施的名称、地址、电话、分类、经纬度、工作时间、网站、电子邮件等各类数据，并基于对地理信息的深刻理解，在广告投放和市场营销方面为用户提供帮助^[33]；九次方大数据为政府不同部门提供底层数据的抽取、清洗、建模和分析服务；数据堂依托大数据处理平台和众包标注平台，使原始数据成为标准化、可应用、可交易的数据产品；贵阳大数据交易所根据需求方要求，对数据进行清洗、分析、建模及可视化等操作之后再出售。

2.2.3 服务层 服务层是数据集市提供数据交易服务的最终实现，为用户提供各种数据源支持。用户的数据需求得以满足后，可以通过大数据分析、深度挖掘、转移应用等方式发现新的数据价值。目前，按交易对象和形式划分，数据集市的服务方式主要有 3 种^[10]：基础数据及数据产品交易，如数据堂、数多多等；满足在线数据调用需求的 API 云服务，如千数堂、聚合数据等；大数据分析结果的交易，如贵阳大数据交易所等。

3 数据集市的服务模式

3.1 数据银行

数据可以在交易中实现价值。但是，由于原始底层数据的质量不高，有时不能被直接交易，因此必须经过一定的数据处理环节才能转化为可以变现的数据资产。数据银行在这个过程中发挥着至关重要的作用，是数据集市的重要服务模式之一。数据银行以数据资产为核心，致力于实现数据资源的标准化、商品化和资产化，承载着数据资产形成、管理、交易和流通的职责。

数据银行主要有 3 个功能：①整合、融合各数据提供方的原始数据；②对大数据进行初步分析、过滤和分类；③向数据需求方提供统一的服务平台和接口。对照前文的数据生命周期分析内容，可以发现数据银行的 3 个功能与数据生命周期的数据获取、数据处理和数据交易大致相似。数据银行模式在以数据堂为代表的第二类数据集市中的应用更为全面和深入。

由图 4 可见，从整个大数据生态圈来看，数据银行模式的角色和作用显而易见。云计算服务商主要负责提供计算、存储和带宽等基础能力；数据提供方提供各种形式的海量数据资源，包括政府大数据、行业大

数据、互联网大数据以及线下大数据等;数据集市基于数据银行模式,经过数据获取、聚合、处理等过程,提供数据交易服务;数据应用商通过多源数据融合、创新转移等方式拓展新的数据应用场景,进一步发现数据潜在价值。由此可见,基于“数据银行”模式的数据集市是大数据生态圈的关键,是连接数据提供方和数据应用商(即数据需求方)的纽带。

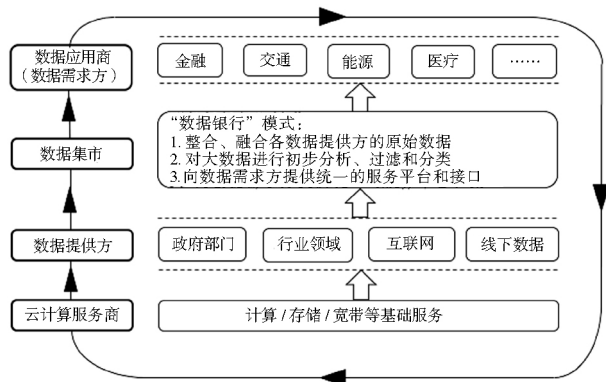


图4 数据银行模式在大数据生态圈中的角色和作用

3.2 众包模式

数据采集是数据交易服务流程和大数据价值链上最基础的一个环节。目前常用的数据采集手段主要有社会调查、实验等人工方法以及传感器、射频识别(RFID)、芯片等计算机辅助方法^[34]。在大数据时代,需要采集海量的底层原始数据,考虑到成本和效率的要求,很多时候已无法基于传统的采集手段来实现。而且,根据目前的技术发展现状,数据的采集、标注和清洗等重复性工作还很难完全实现自动化。大数据的特点和大数据应用的扩展和深化,使得传统的数据采集手段和获取渠道难以满足科学研究和产业发展的需要。因此,建立适应不同环境、满足不同需求的多元互补的大数据采集体系显得格外重要。

在实践应用中,数据集市逐渐探索出基于众包的服务模式。如Inrix聚合和众包来自1 000万个来源的实时交通信息,包括汽车、出租车、货车、卡车等,随后将这些数据出售给移动应用开发商和汽车厂商,还与交通运输服务机构合作,帮助管理道路网络、向公众提供旅游信息等^[35]。数据堂、数多多等数据集市类企业将众包模式引入大数据采集和标注,充分借助群体智慧以弥补自身资源能力上的不足,以较低的成本快速有效地积累了实时多源的大数据资源,逐步建立了立足于大数据行业的先发优势。在众包采集方面,数据堂建立统一集成的基于众包模式的大数据采集平台,以任务的形式将需要采集的数据发布在众客堂上,众

客完成数据采集任务以获得相应的物质奖励。众客作为移动数据采集器,实时、高效地采集和提供各类线下数据,如通话录音、各地方言、照片图片等。类似的还有数多多的八爪鱼数据采集平台、遥感集市的“遥感众包”“遥感威客”平台等。在众包标注方面,数据堂借助在线众包标注平台发布任务,激励众客对图片、语音、文本、视频等初级数据进行加工处理,如标识发音人性别、判断噪音类型等。

综上所述,数据采集是众包模式在数据集市中最典型的应用场景。同时,在海量数据的加工和标注等需要耗费较高人力和时间成本的任务中,众包模式提供了一种成本可控、规模易伸缩的实现途径,也具有较大的应用空间。这与分享经济充分利用互联网将闲散资源与需求对接以解决供需失衡的理念是契合的。

4 数据集市与价值发现

从数据中发现价值一直是信息管理领域中的重要课题。传统数据环境下,为了实现从数据到价值的跨越,需要经历从信息到知识,再到智能的层层升华与淬炼,信息链理论^[36]和数据、信息和知识的金字塔模型^[37]是这一思维方式的典型体现。在大数据环境下,除了传统的数据价值实现路径外,数据本身的价值逐渐突显,即数据或数据产品可以像商品一样,通过租售、交易等方式直接获得价值。舍恩伯格^[38]认为,大数据时代,思维和技术是最有价值的,但是最终大部分的价值还是必须从数据本身中挖掘,数据才是最核心的部分。

数据集市与价值发现具有密切关系,即在交易中发现数据价值。这里的“价值”至少包含3层含义:首先,指数据自身的价值,数据经过获取、处理、交易等人类劳动实现价值发现和增值;其次,指对数据集中各行动主体的价值,数据购买方获得数据资产,数据供给方因租售数据获利,同时数据集市也从交易活动赚取酬劳;最后,数据集市及其数据交易活动促进了数据要素的流通,提高了社会经济系统的总价值。

大数据作为一种重要的稀缺资源和战略资产,已经不同程度地渗透到各个行业领域和部门。数据集市促进了数据资源的交易和流通,其深度应用不仅有助于科学家的科学研究、企业的商业活动,还有利于提升政府的公共管理和社会服务水平。典型案例如表1所示:

表1 数据集市与价值发现的典型案例

领域	典型案例
科研	王永等 ^[39] 提出了一种从中文网络评论中提取产品特征的新方法,并利用手机评论语料验证了该方法的有效性 刘梦超等 ^[40] 对从数据堂获取的近7个月的HTTP请求日志进行分析挖掘,从而发现用户的上网行为模式和潜在上网规律 张四海和王芷阳 ^[41] 利用北京出租车行业2012年间两个月的GPS跟踪数据,分析了出租车司机拒载乘客的影响因素和行为决策模式
商业	DataMarket是微软公司Windows Azure Marketplace的一部分,主要为买卖双方建立一个进行数据和应用程序交易的统一市场 ^[42] 。内容合作伙伴可以发布收集到的数据到DataMarket上,以提高它的可被发现性;同时,用户可以轻松地发现、探索、订阅和使用可信、优质的数据 贵阳大数据交易所是国内第一家大数据交易所,截至2016年9月1日,交易额累积突破1亿元,交易框架协议接近3亿元,可交易数据产品接近4000个,可交易的数据总量超过60PB 聚合数据加入阿里云星火计划、与腾讯众创空间达成战略合作,为其服务开发者和项目团队提供基础数据等服务,提高开发效率,降低开发成本,实现数据管理轻量化、效率最大化
政务	2015年4月,九次方大数据与贵阳市人民政府合作成立的贵阳大数据交易所正式挂牌运营,参与工信部《大数据产业“十三五”发展规划》、全国信标委《大数据交易标准》《大数据技术标准》《大数据安全标准》《大数据应用标准》等标准的起草和制定工作 贵阳、银川等经济欠发达地区把握大数据发展机遇,通过政府大数据资产运营、大数据交易中心项目等举措,推动了智慧城市建设、政府数据开放共享和社会公共服务优化 九次方大数据与贵阳、甘肃、广东等40多个省市在政府大数据资产运营等方面展开合作,并设计研发了4000多个政府大数据应用场景,有力地促进政府数据公开,使数据共享成为可能并服务社会民生

(1) 科研领域。随着 e-Research 和数据密集型科研的发展,科学家对数据的需求日益增长,然而科研数据的获取并不容易。数据集市的出现,则缩小了科学家数据需求与数据获取之间的鸿沟,为科研数据获取与利用提供了方便。数据集市上有大量可用于科学研究的数据,如微博用户关系数据、豆瓣图书评分数据、电商评论、中文情感极性词典、中文分词测评语料以及上网日志等。这些数据不仅数量庞大,而且种类繁多,可以满足多种科研需求。调查发现,面向科研领域的数据集市不在少数,如数多多、数据堂、类聚、数聚工场等。

(2) 商业领域。虽然商业领域的数据利用由来已久,但是对于数据的深度挖掘,以及海量多源数据的关联分析仍有待提高。大数据引发了商业变革,让人们开始重新审视数据、技术、思维对效率、价值和利润的影响。基于大数据资源的用户行为模式挖掘、用户需求识别、用户行为偏好分析,对于潜在合作伙伴发现、精细用户群体划分、定向广告投放和市场营销、多维度科学评估、智能监测预警等具有重要意义,尤其是在当前的传统行业转型升级和“大众创业、万众创新”背景

下,商业是大数据价值发现最具有生命力和吸引力的领域之一。

(3) 政务领域。政府拥有的数据比任何公司、企业都多,而且数据质量相对较高,但是在信息技术的应用、数据开放共享和数字信息资源深度利用上,与私营领域相比,还是明显落后的。但是近年来,政府不断推动大数据在社会治理、宏观调控、智慧城市、电子政务、民生服务、公共安全等领域的应用^[43]。总体来说,政务领域的价值发现主要通过“政产学研”合作建立大数据联盟的形式完成,以九次方大数据为代表的数据集市中在其中发挥着重要作用。政府提供数据资源,数据集市提供数据处理、数据分析、应用场景设计等增值服务,最大限度地发现政府大数据的潜在价值。

5 结语

作为一种新兴的数字信息资源服务方式,数据集市在提供数据交易服务的过程中,不仅完成了数据本身的价值发现和价值增值,而且使数据交易活动各参与方都得以分享价值。在大数据环境下,随着数字经济的发展、数据科学的兴起和数据需求的增长,数据集市的意义更加明显。展望未来,数据交易和数据集市将呈现以下发展趋势:①实时数据的交易需求和市场规模将持续增长,如医疗健康监测数据、金融证券行业数据、物联网传感器数据等。②直接提供原始数据的数据交易模式将不再那么流行,而越来越普遍的是提供经过处理的数据及增值服务。③对数据质量管理和控制的重视程度将日益提升。④大数据技术在数据集市和大数据产业发展中的作用将更加突出,而纯粹的集市类的形态将逐渐弱化,如 InfoChimps 逐渐从数据集市转型为大数据平台提供商。

可以预见,在政策的持续推动下,数据集市将具有更大的成长空间和发展前景。《促进大数据发展行动纲要》中明确提出,要“引导培育大数据交易市场,开展面向应用的数据交易市场试点,探索开展大数据衍生产品交易,鼓励产业链各环节市场主体进行数据交换和交易,促进数据资源流通,建立健全数据资源交易机制和定价机制,规范交易行为”。《国民经济和社会发展规划第十三个五年规划》中提出“实施国家大数据战略,推进数据资源开放共享”。此外,数据交易中的行业规范、定价模式、权益归属、数据安全隐私保护以及数据质量控制等也在一定程度上影响数据集市的发

展, 这些问题同样值得管理者、从业者和研究者的关注。

本文旨在通过对数据集市的概述, 增进对数据集市产生环境、理论基础、服务模式和价值发现的认识, 从而更好地应对大数据环境的影响与挑战, 推动国家大数据战略的实施。当然, 本文只是对数据集市的初步概述, 而没有对国内数据集市的发展现状进行全面市场调查, 因此缺少足够的数据支持。这将是下一步的研究重点。

致谢: 感谢马费成教授在本文成文过程中给予的启迪和帮助!

参考文献:

- [1] 中国政府网. 国务院关于印发促进大数据发展行动纲要的通知[EB/OL]. [2016-09-11]. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.
- [2] 张斌, 马费成. 大数据环境下数字信息资源服务创新[J]. 情报理论与实践, 2014, 37(6): 28-33.
- [3] 上海市经济和信息化委员会. 全球五大顶级大数据交易平台[EB/OL]. [2016-12-26]. <http://www.sheitc.gov.cn/jjyw/669713.htm>.
- [4] 贵阳大数据交易所. 2016年中国大数据交易白皮书[R]. 贵阳: 贵阳大数据交易所, 2016.
- [5] RODERICK L. Discipline and power in the digital age: the case of the US consumer data broker industry [J]. Critical sociology, 2014, 40(5): 729-746.
- [6] ARMSTRONG A A, DURFEE E H. Mixing and memory: emergent cooperation in an information marketplace[C]// International conference on multi agent systems. Paris: IEEE, 1998: 34-41.
- [7] MUSCHALLE A, STAHL F, LÖSER A, et al. Pricing approaches for data markets[C]// Workshop business intelligence for the real time enterprise. Heidelberg: Springer, 2013: 129-144.
- [8] SCHOMM F, STAHL F, VOSSEN G. Marketplaces for data: an initial survey [J]. SIGMOD record, 2013, 42(1): 15-26.
- [9] STAHL F, SCHOMM F, VOSSEN G. The data marketplace survey revisited [R]. Münster: ERCIS - European Research Center for Information Systems, 2014.
- [10] STAHL F, SCHOMM F, VOSSEN G. Data marketplaces: an emerging species [J]. Frontiers in artificial intelligence & applications, 2014: 145-158.
- [11] STAHL F, SCHOMM F, VOMFELL L, et al. Marketplaces for digital data: Quo Vadi? [R]. Münster: ERCIS - European Research Center for Information Systems, 2015.
- [12] STAHL F, SCHOMM F, VOSSEN G, et al. A classification framework for data marketplaces [J]. Vietnam journal of computer science, 2016, 3(3): 137-143.
- [13] MÖLLER K, DODDS L. The Kasabi information marketplace[EB/OL]. [2016-12-26]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.309.9599&rep=rep1&type=pdf>.
- [14] CAO T, PHAM T, VU Q, et al. MARSA: a marketplace for real-time human sensing data [J]. ACM transactions on Internet technology, 2016, 16(3): 1-21.
- [15] 王融. 关于大数据交易核心法律问题——数据所有权的探讨[J]. 大数据, 2015, 1(2): 49-55.
- [16] 汤琪. 大数据交易中的产权问题研究[J]. 图书与情报, 2016(4): 38-45.
- [17] 史宇航. 个人数据交易的法律规制[J]. 情报理论与实践, 2016, 39(5): 34-39.
- [18] 张敏. 交易安全视域下我国大数据交易的法律监管[J]. 情报杂志, 2017(2): 127-133.
- [19] 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, 1(2): 38-48.
- [20] 李骥宇. 大数据交易模式的探讨[J]. 移动通信, 2016, 40(5): 41-44.
- [21] 穆会军. 国内大数据交易平台建设及交易情况的相关分析——以华中大数据交易所为例[J]. 信息系统工程, 2016(9): 123-124.
- [22] 唐斯斯, 刘叶婷. 我国大数据交易亟待突破[J]. 中国发展观察, 2016(13): 19-21.
- [23] 宋梅青. 融合数据分析服务的大数据交易平台研究[J]. 图书情报知识, 2017(2): 13-19.
- [24] 孟小峰, 慈祥. 大数据管理: 概念、技术与挑战[J]. 计算机研究与发展, 2013, 50(11): 146-169.
- [25] 郭贺铨. 大数据时代的机遇与挑战[J]. 求是, 2013(4): 47-49.
- [26] 马建光, 姜巍. 大数据的概念、特征及其应用[J]. 国防科技, 2013, 34(2): 10-17.
- [27] McKinsey Global Institute. Big data: the next frontier for innovation, competition and productivity [EB/OL]. [2016-12-26]. <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>.
- [28] 夏征农, 陈至立. 辞海[M]. 上海: 上海辞书出版社, 2010.
- [29] 叶鹰, 马费成. 数据科学兴起及其与信息科学的关联[J]. 情报学报, 2015(6): 575-580.
- [30] 周涛. 大数据商业模式创新三部曲[J]. 北大商业评论, 2015(3): 60-67.
- [31] UK DATA ARCHIVE. Research data lifecycle[EB/OL]. [2016-12-26]. <http://www.data-archive.ac.uk/create-manage/life-cycle>.
- [32] 李天柱, 马佳, 吕健露, 等. 大数据价值孵化机制研究[J]. 科学学研究, 2016, 34(3): 321-329.
- [33] BELL T. Big data, big local[EB/OL]. [2017-05-25]. <https://www.factual.com/blog/big-data-big-local>.

- [34] 刘智慧,张泉灵. 大数据技术研究综述[J]. 浙江大学学报(工学版), 2014, 48(6): 961-962.
- [35] INRIX traffic data services: tapping into real-time traffic flow data [EB/OL]. [2017-05-25]. http://cta.ornl.gov/TRBenergy/trb_documents/2012_presentations/192_%20T%20Trepanier%20-%20INRIX%20LeeSchipper%20Data%20Workshop%20TRB2012.pdf.
- [36] 梁战平. 情报学若干问题辨析[J]. 情报理论与实践, 2003, 26(3): 193-198.
- [37] 涂子沛. 大数据[M]. 桂林: 广西师范大学出版社, 2012.
- [38] 舍恩伯格, 库克耶. 大数据时代: 生活、工作与思维的大变革[M]. 杨燕, 周涛, 译. 杭州: 浙江人民出版社, 2013.
- [39] 王永, 张勤, 杨晓洁. 中文网络评论中产品特征提取方法研究[J]. 现代图书情报技术, 2013(12): 70-73.
- [40] 刘梦超, 肖基毅, 陈荣, 等. 数据挖掘在用户上网行为分析中的应用研究[J]. 电脑知识与技术, 2012(31): 7409-7412.
- [41] ZHANG S, WANG Z. Inferring passenger denial behavior of taxi drivers from large-scale taxi traces [J]. PLOS ONE, 2016, 11(11): 1-21.
- [42] Microsoft Azure. Windows Azure Marketplace DataMarket 概述 [EB/OL]. [2017-05-25]. <https://blogs.msdn.microsoft.com/azchina/2011/04/07/windows-azure-marketplace-datamarket/>.
- [43] 中国计算机学会大数据专家委员会. 中国大数据技术与产业发展报告[M]. 北京: 机械工业出版社, 2016.

作者贡献说明:

赵栋祥: 论文撰写与修改;

陈焯: 论文修改;

张斌: 论文设计与修改。

Data Marketplace and Its Values in Data Trading

Zhao Dongxiang¹ Chen Ye¹ Zhang Bin²

¹ Center for Studies of Information Resources of Wuhan University, Wuhan 430072

² Center of Traditional Chinese Cultural Studies of Wuhan University, Wuhan 430072

Abstract: [Purpose/significance] Big data environment accelerates development and innovation in the field of data and information service. The purpose of this article is to provide reference and insight for relative practice and research work, with analyzing the development environment, theoretical foundation and exploratory practice of data marketplaces. [Method/process] Firstly, the development environment of data marketplaces was discussed from the perspective of the effects of big data on data management and service. Then we analyzed the theoretical foundation of data marketplaces, namely information lifecycle and value chain of big data. And then we analyzed the theoretical foundation of data marketplaces on the basis of the management processes and value chains for big data. Thirdly, the typically successful cases were discussed in order to discover the general patterns of data marketplaces. Finally we analyzed how data marketplaces help to discovery the value data trading. [Result/conclusion] Big data environment stimulates and promotes the emergence and development of data marketplaces to some extent. Information lifecycle and value chains of big data are important theoretical foundation of data marketplaces. The data bank and crowdsourcing are important service patterns of data marketplaces. Scientific research, business and government are the typically potential fields for data value discovery of data marketplaces.

Keywords: data marketplace data trading data value big data environment service innovation