

数据中介的多元模式探析

李振华, 王同益

蚂蚁集团研究院, 浙江 杭州 310013

摘要

数据中介有望成为促进数据流通的中坚力量, 通过多样化的创新实践, 助力数据要素市场的建设。介绍了多种类型的数据中介, 不同的数据中介着重解决不同的实践难题, 比如数据交易平台着重解决供需双方的信息不对称问题, Plaid等开放银行服务商着重解决数据标准和数据接口的统一转换问题, 数据信托可以优化“授权同意”这一个人信息分享路径。建议坚持问题导向原则, 根据场景需要, 积极探索、创新多元化的数据中介形式, 建设良好可信的数据生态, 充分释放数据要素的价值。

关键词

数据中介; 多元模式; 数据生态

中图分类号: F28; F49

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2022068

Analysis on various patterns of data intermediary

LI Zhenhua, WANG Tongyi

Ant Group Research Institute, Hangzhou 310013, China

Abstract

Data intermediaries are expected to become the backbone of promoting data circulation through diversified innovative practices. Various patterns of data intermediaries were introduced. Different data intermediaries focus on solving different practical problems. For example, the data transaction platform focused on solving the information asymmetry between the supply and demand sides, open banking service providers such as Plaid focused on the unified conversion of data standards and data interfaces, and data trust could optimize the personal information sharing the path of authorization and consent. It was suggested to adhere to the problem orientation, actively explore and innovate diversified data intermediary patterns according to the needs of the scenario, build a healthy and reliable data ecosystem, and fully release the value of data.

Key words

data intermediary, various pattern, data ecosystem

0 引言

数据具有非竞争性和范围经济两个重要特点,为了实现数据的社会价值最大化,促进数据分享是必要途径和手段。数据的非竞争性指一家公司收集的数据可用于多种目的,如果其他公司能够访问这些数据并确保用于其他目的(比如生产其他产品或提供其他服务)不会对原公司产生直接的竞争效应,则社会将会受益。数据的范围经济指与保存在“数据孤岛”中相比,数据集互补和汇聚后可以提取更多的有价值信息,实现更高的经济或社会效益^[1]。

然而,数据分享是有成本的,当交易费用较高时,数据分享的规模会非常有限甚至难以实现。比如,如果数据的分享不可控,那么“搭便车”和数据过度使用就无法避免,就不会有人愿意投资数据的生产行业,最终将导致“无米下锅”;如果数据的权利人过多,每个当事人都有权阻止其他人使用该资源或相互设置使用障碍,同时却没有人拥有有效的使用权,则很难对数据的分享达成一致,导致资源的闲置和使用不足,发生“反公地悲剧”。数据分享障碍还包括数据持有者与数据使用者之间的供需信息匹配问题,数据搜寻、汇聚的成本问题,数据价值发现的专业化能力问题,数据传输的标准和低成本的互操作问题,数据分享过程中安全和合规责任的合理分配和执行问题等。

由此可见,降低交易费用是促进数据分享和流通的关键,而数据分享中介在其中扮演着极其重要的角色。根据英国开放数据研究所(Open Data Institute, ODI)和经济合作与发展组织(Organization for Economic Co-operation and Development, OECD)发布的报告,数据

中介(data intermediary)指各类有助于更好地访问或分享数据的中介机构^[2-3]。数据中介的形式并不拘于一格,它们可能涵盖一系列不同的业务活动和治理模型,只要是通过提供各类新颖的、技术支持的解决方案,实现数据在提供方和使用方之间安全和无摩擦分享的,都可被视为数据中介。例如,出于隐私考虑,一些人会反对将他们的健康数据用于研究,尽管他们知道分享数据可能能够带来社会效益,如果中介机构能够搭建受控的数据实验室,保证在数据使用过程中患者身份和隐私信息不泄露,那些原先持反对意见的人就有可能改观。再比如,数据中介可能是行业龙头或者第三方机构通过搭建“共建共享”平台,实现数据在成员机构间的共享;科技企业可以通过应用程序接口(application programming interface, API)开放平台等技术搭建不同数据源的互联互通管道,这将有效降低数据在不同主体间的传输成本。

对于数据分享中存在的问题,可以从5个维度、6个方面来理解:动力(缺少分享数据的激励)、顾虑(包括商业、道德和声誉风险,以及法律和监管风险两方面)、渠道(缺乏关于数据可用性或价值的信息)、成本(数据访问/分享的成本过高)、公共性(数据公共价值发挥不够)。**表1**展示了数据分享中存在的问题以及数据中介可能扮演的角色。

1 数据中介的典型类型

本文主要介绍三大类数据中介:第一类数据中介与传统商品市场的中介相似,核心在于解决市场中的信息不对称问题,并发挥规模效应,从而降低平均成本,这类数据中介包括数据交易平台、数据经纪商(data broker),也包括商业征信机构

表1 数据中介的潜在角色和价值

问题	问题描述	数据中介潜在角色
动力: 缺少分享数据的激励	数据提供者无法从数据分享中获得回报或者回报低于成本, 因而没有足够的动力来分享或提供对其数据的访问	运行“共建共享”模式, 某个组织只有贡献了数据才能访问更大的数据集; 帮助个人更好地实现信息携带权; 凭借对需求端的渠道和洞察优势, 帮助数据提供方提高数据交换价值
顾虑: 商业、道德和声誉风险	数据提供者担心数据使用者对己方商业竞争优势产生威胁, 或者数据使用者滥用数据, 对己方社会形象产生负面影响, 从而不愿意对外分享数据	向数据主体和数据提供者保证, 数据只能被经过认可的个人/组织访问并用于特定目的
顾虑: 法律和监管风险	可能违反个人信息保护、知识产权保护或监管要求的风险和担忧也会阻碍数据的分享	提供法律专业知识、技术解决方案(比如可信隐私计算)、通用数据标准(比如个人信息去标识化和匿名化技术标准), 以促进合法合规的数据分享
渠道: 缺乏关于数据可用性或价值的信息	数据提供者可能对其数据的潜在用途缺乏足够的了解, 而数据使用者则对有哪些数据可用以及如何获取缺乏足够的了解	通过相关信息的汇聚, 数据使用者能够更轻松地发现特定数据集和更容易定价, 数据提供商能够让他们的数据产生价值
成本: 数据访问/分享的成本过高	由于利益基础、技术设施和标准缺乏或者规模效应不足, 成本(谈判成本、对接成本、技术成本等)可能过高	通过API开放平台等技术搭建不同数据源的互联互通网络, 有效降低数据传输成本; 通过专业化分工, 形成规模效应
公共性: 数据公共价值发挥不够	数据访问/分享可能带来经济和社会等公共效益	为了公共利益研究, 搭建受信任的研究环境, 提供对敏感数据的安全访问

① 微软的 Azure 数据市场于 2018 年关闭。Kasabi 于 2010 年启动, 2012 年关闭。BuzzData 于 2010 年启动, 2013 年 8 月 1 日关闭。Infochimps 于 2009 年进入大数据市场, 2013 年被 CSC 收购, 不再提供数据服务。Timetric 专注于统计和时间序列数据, 成立于 2008 年, 于 2018 年被 GlobalData 收购, 后者是一家咨询公司而非交易市场。社交数据提供商 Gnip 成立于 2008 年, 2014 年被推特(Twitter)收购。DataStreamX 市场于 2019 年 3 月停止运营。Freebase 由 Meta Web Technologies 于 2007 年推出, 被谷歌收购后于 2016 年关闭。xDayta 于 2013 年推出, 2015 年关闭。

和消费者个人征信机构等; 第二类数据中介与增强个人对个人信息的控制权和携带权的政策背景有关, 目的是在信息分享过程中更多地体现个人的意志, 更好地保护个人的权益, 个人信息管理系统(personal information management system, PIMS)和数据信托(data trust)是典型代表, 另外也包括为了更高效、更低成本地实现个人信息携带移转的技术服务商, 如开放银行服务商; 第三类数据中介与数据易被复制、易被滥用的特性有关, 这些特性使得数据提供方面临较强的商业、道德和声誉风险, 第三类中介旨在通过可信技术环境或可信组织, 消除数据提供方对这方面的担忧, 这类中介包括数据沙箱(data sandbox)、数据共享平台等。

1.1 数据交易平台

交易平台有助于解决信息不对称问题, 从而更好地匹配数据供需双方, 基于这样的目的, 数据交易平台(中心、所)开

始兴起。2015年4月, 贵州省人民政府批准成立了全国第一家大数据交易所——贵阳大数据交易所, 其基本情况见表2。在之后的几年中, 武汉、哈尔滨、西安、广州、青岛、上海、沈阳、成都等地纷纷建立大数据交易所或交易中心。

尽管数据交易平台被寄予厚望, 但具体方案仍有待探索试验。2010年前后海外兴起的Infochimps、Azure、BuzzData、Timetric、DataStreamX、xDayta等交易机构纷纷关闭或转业^①。2014—2019年, 我国挂牌成立的多家数据交易中心(所)的数据交易业务也非常有限。数据交易平台未能成功发展起来有多方面的原因: 一是供给方面, 公共数据开放不足, 没有发挥示范和杠杆效应, 私营数据由于确权和合规等问题不敢交易; 二是合规与安全方面, 缺乏法律法规和标准, 缺乏技术手段与可预测监管; 三是商业竞争层面, 未实现规模效应, 产品特色与交易成本缺乏竞争力; 四是互联互通与有效治理层面, 未形成产品、流通、治理的统一市场标准。不过诸多条件正在逐步改

表2 贵阳大数据交易所基本情况

基本情况	具体信息
成立时间	2015年4月
所有制	政府指导、社会参与、市场化运作（成立时的股东包括贵州阳光产权交易所、九次方财富资讯（北京）、郑州市迅捷贸易、北京亚信数据等，2021年重组后由国有企业云上贵州大数据集团100%控股）
交易时间	24 h/天×365天
数据类型	30大类（金融、医疗、海关、能源、社交、法院、交通、通信、物流、气象、地理……），初期交易粗加工的原始数据，后改为交易定制化数据产品
定价模式	针对每一个数据品种设计自动计价公式，交易系统根据数据品种、时间跨度、数据深度、数据完整性、数据样本覆盖和数据时效性实时定价
成交机制	①买方根据挂牌价应约；②卖方选择买方应约价与其成交；③分拆成交
收费模式	①会员会费，5万元/年；②交易提成，10%
对数据提供方的要求	“宽进严管法”：进入门槛低（宽进），但对于违规或违法、数据造假、数据欺诈、数据来源不合法的供应商，有撤销其资格、将其加入交易所黑名单等处罚方式（严管）
对数据买方的要求	不允许个人参与购买，外资需要进行额外的资格审核

善，新的探索仍在进行中。改善中的条件包括：一是增加新的动力，北京、广州、深圳拟集中通过数据交易所开放公共数据，有效改善数据交易平台的供给不足问题；二是制定新的规则，深圳、上海等地发布数据条例，明确数据权益保护机制，这将在一定程度上消除数据提供者的顾虑；三是使用新的技术，隐私计算、区块链等技术的应用将为数据交易模式的创新提供更多可能性；四是运用新的模式，建设数据生态，引入征信机构和大数据机构进行数据二次开发利用，这将有效提高数据产品和服务的竞争力^[4]。

1.2 数据经纪商

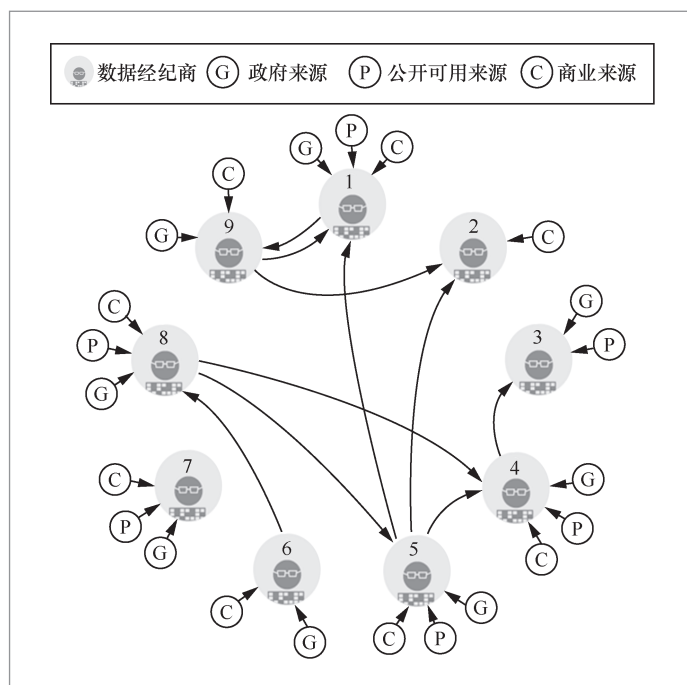
数据经纪商指从事数据收集、处理并对外提供数据服务的机构。相对于数据交易平台，数据经纪商往往积极参与额外数据的收集与聚合，这是两者的重要区别。数据交易平台往往是被动中介，数据经纪商也可以通过数据交易平台购买或者提供数据集。这些数据包括个别公司和公民披露或提供的数据、来自安装传感器的公司的数据、从互联网“爬取”的数据、来自非营利组织和公共部门机构的数据（例如

地球观测数据，人口、健康数据，以及其他统计数据）。数据经纪商也通过广泛的商业渠道购买所需用户个人数据，数据经纪商之间也会互为数据源。最早的数据经纪商就是商业征信机构和消费者个人征信机构，不过随着数据范围的拓展和数据量的增加，征信机构聚焦于利用真实可靠的数据，出具固定模板的征信报告，与一般的从事大数据交易的数据经纪商有一定的区别。

数据经纪商是美国数据交易服务的主要提供者。美国的个人信息保护模式呈现出“大隐私权保护模式”特征，即通过不断丰富隐私权的内容来应对网络时代的个人信息问题，但因为个人信息的控制与美国宪法所保障的言论及出版自由在现实中存在一定的冲突，而且美国政府高度重视数字经济产业的快速发展，因此在不严重侵犯个人隐私权的前提下，美国更倾向于个人信息的自由流动。美国对个人信息的保护也没有形成统一的立法，而是分散在不同的法律中。在此背景下，美国形成了规模庞大的数据交易市场，也产生了专门提供数据交易服务的数据经纪商。表3列举了美国九大数据经纪商的基本情况，图1展示了美国九大数据经纪商的数据来源^[5]。

表3 美国九大数据经纪商基本情况

机构名称	基本情况
Corelogic	包含超过10亿条历史交易数据、超过3亿条抵押贷款申请信息，面向不动产抵押贷款机构、保险机构、房产税务等机构提供服务，客户数超过100万个，涵盖99%以上美国住宅和商业地产。2020年营业收入为16.4亿美元，同比增长14%，盈利3.3亿美元
LiveRamp (2018年前名为Acxiom)	数据库涵盖全球7亿个人，包括涉及大部分美国用户的3 000个数据段，为市场营销提供用户数据和分析服务。2020年营业收入为4.43亿美元，2019年和2020年平均增长24.5%，净亏损9 030万美元
Datalogix	涵盖大部分美国家庭、涉及金额1万亿美元以上的用户交易信息，为营销者提供有意义的见解，提高广告的有效性和可衡量性，也提供预测性分析和集成媒体解决方案
eBureau	最早只是分析某人是否可能成为潜在的优质客户或者某笔交易是否存在商业欺诈，后来发展为向其客户提供数以亿计的用户消费记录，向营销商、金融公司、在线零售商以及其他商业主体提供预测评级和数据分析服务（于2017年被TransUnion收购）
ID Analytics	主要提供以身份认证、交易欺诈检测和认证为目的的数据分析服务（于2020年被LexisNexis Risk Solutions以3.75亿美元收购）
Intelius	向商业机构和客户提供背景调查和公开记录信息，其数据库包含200亿条以上的公开记录信息（会员制，每月会费为19.95美元）
PeekYou	能够分析60家社交媒体网站、新闻来源、网站主页、博客平台内容，帮助查找朋友、亲戚和同事信息
Rapleaf	提供80%的美国电子邮件地址的数据，使营销人员能非常容易地了解他们的客户及其联系方式；帮助品牌细分用户并做个性化推荐，提高客户服务体验（于2013年被TowerData收购）
Recorded Future	通过互联网捕捉用户和企业的各类历史数据，利用该类历史数据分析用户和企业的未来行为轨迹，为企业提供情报，进行品牌维护、风险管理和安全保障

图1 美国九大数据经纪商的数据来源示意图^[5]

从监管趋势来看，提高数据经纪商的透明度和提供便捷的退出机制是核心要求。美国数据经纪商的发展的确在很大

程度上促进了数据的流通，提供的产品有助于预防诈骗、提高产品供应、实现广告精准投放。不过数据经纪商到处收集用户信息的行为也在一定程度上侵犯了用户的隐私权，数据经纪商掌握的大量数据一旦遭到泄露，将造成难以弥补的伤害和损失。美国联邦贸易委员会（Federal Trade Commission, FTC）建议国会提高数据中介透明度和增强用户的控制能力：要求数据经纪商设置一个互联网门户网站，更清晰地说明数据信息的收集来源和使用方法，并为用户提供选择退出机制。美国多地出台了数据经纪商管理办法，2018年5月，佛蒙特州通过数据经纪商专项监管法案，加大对数据经纪商的监管，要求建立年度注册制度，强化信息安全责任，禁止为违法目的获取数据，同时增强消费者权益保护，包括明确退出方法、退出权适用范围以及是否允许消费者委托授权第三方行使退出权，增加消费者信用报告知情权和控制权。2019年10月，加利福尼亚州也颁布了关于数据经纪商监管的法案。2019年7月30日，两

名议员向FTC建议,提高数据经纪商的监管层级,成立受FTC监督的全国性注册机构,并要求数据经纪商每年向该机构进行注册。

1.3 个人信息管理系统

个人信息管理系统根植于“我的数据我做主”(MyData)的理念,旨在让数据主体对他们的个人数据拥有更多的控制权。数据主体可以从银行、互联网公司、医院和政府等提供商那里将个人数据副本导入PIMS,通过PIMS实现个人数据访问的一站式管理,并有权随时撤回访问权限。

这方面较早的尝试是英国的Midata计划(后改名为MyData计划)。英国商业、创新和技能部(Department for Business, Innovation and Skills, BIS)于2011年试验推出一项消费者赋权计划——Midata计划,让消费者可以获取和携带自己的消费数据。后又将试验范围进一步扩大至金融、能源、医疗、交通等领域。基本思路是允许消费者下载消费记录副本,并在需要

时自行提供给第三方。但试验发现这不利于个人对资料的管理,由于可能出现资料被篡改、格式不统一等问题,第三方验证和重用的成本也很高,于是出现了专门的管理机构PIMS。PIMS提供的服务和功能可能包括:一是通过PIMS实现个人信息的一站式查询;二是为用户提供个人数据存储和数据整理;三是作为交互界面,实现用户与数据提供方和使用方的授权同意管理。两种MyData实践模式如图2所示。

某些PIMS(如Nesta DECODE)是非营利性企业,但也有不少PIMS(如digi.me和Solid)是商业企业,这表明可能会出现一个满足个人更好地存储并控制自己数据需求的市場。例如,digi.me为用户提供个人数据存储,用户可以在其中整理自己的数据,并与digi.me平台集成的应用程序和公司分享。通过digi.me,用户可以将社交媒体平台等在线提供商中关于自己的数据导出,可以自由选择数据存储的位置(比如存储到Dropbox、Google Drive等个人云中),可以控制与其他机构分享哪些个人数据以及以何种条件分享,个人也可以撤销这

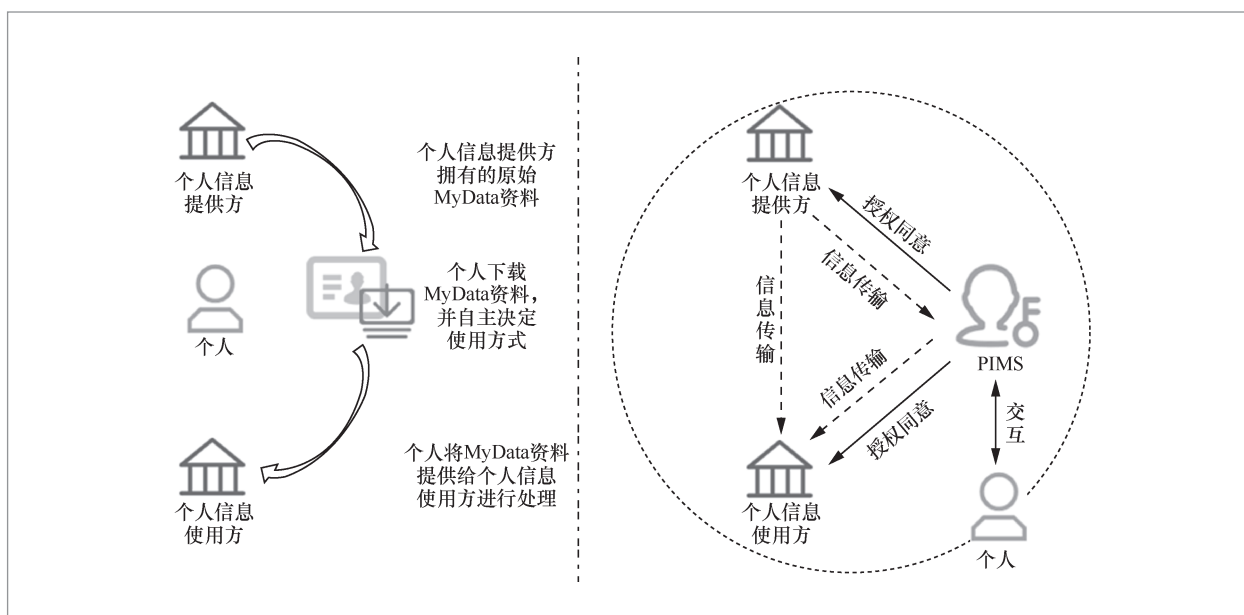


图2 两种 MyData 实践模式

些机构对其个人数据的访问权限^[6]。PIMS的其他示例还有，Mine帮助个人向相关机构提出“删除个人数据”的请求，促进个人信息/数据删除权的实现；韩国的MyData运营商基于信用传送要求权，收集分散的个人信用信息，为用户提供一站式查询、金融产品咨询、资产管理等服务。除了上述示例，PIMS还可以帮助个人用户自动退出可能被认为不受认可或者违规用途的数据分享，如赌博。

1.4 数据信托

数据信托指将数据纳入信托法律关系，在受托人、委托人、受益人三方之间建立相互信赖和责任的法律机制。2016年，美国耶鲁大学教授杰克·巴金 (Jack Balkin) 在隐私数据保护领域首次提出采用信托工具解释数据主体与数据控制人之间的关系的主张。ODI将数据信托定义为提供独立的第三方数据管理的法律结构。数据信托作为一种新型信托制度受到了很多关注，被 *MIT Technology Review* 评为2021年“十大突破性技术”之一。

设立数据信托的意义，既包括提高治理框架的透明度和信任度，也包括降低管理和分享数据所需的成本和技能要求。公共数据治理在这方面已有不少试验。比如，ODI发起了3个试验项目，一是与大伦敦政府和格林威治区一起探索创建关于城市空间的数据信托，重点关注电动汽车停车位的数据和住宅加热传感器收集的数据；二是通过WILDLABS技术中心探索建立数据信托以应对国际非法野生动物贸易问题，重点关注图像和声学数据以及边境部门获得的数据；三是与食品和饮料制造商和零售商一起探索建立数据信托以解决全球食物浪费问题，重点关注食物浪费和销售数据。2018年10月，Alphabet子公司

的Sidewalk Labs提议使用公民数据信托来管理其在多伦多地区为智能城市项目收集的数据。

不过，数据信托最大的价值可能在于优化了“授权同意”这一个人信息分享路径。在“授权同意”的路径下，个人信息的分享需要取得数据主体的“同意”甚至“单独同意”，但这一路径与现代数据管理需要与数据的流通利用需要常常无法很好地匹配：从数据主体角度来看，由于识别潜在风险的能力弱、权益遭受损害后维权成本高，同意个人信息向第三方分享的概率大大降低；从数据提供方和数据使用方角度来看，数据的分享往往是规模化的，每一次数据（集）的分享都需要分别向每一个个体征求同意，成本过于高昂。数据信托的本质在于由受托机构按照委托人的意愿，以自己的名义对委托人的数据进行管理和处理，第三方获取数据时只需要获得数据信托机构的同意，无须向各个委托人征求同意，这将大大提高数据授权的效率和管理的专业性。

1.5 开放银行服务商

开放银行是个人信息携带权落地最早也相对成熟的领域。而在开放银行建设过程中，开放银行服务商是推动开放银行生态系统建设的中坚力量。目前，全球已经有30多个国家和地区采纳或正在考虑采纳开放银行模式。银行开放自身服务和数据，通过数据聚合、产品创新等方式与合作伙伴一起获得新客户、增加客户触点。同时，客户能够通过第三方平台在非金融场景下获得金融服务，从而提升客户体验。例如，英国在开放银行方面的发展上已经取得一定成果。截至2021年第二季度，英国共有319个受监管的服务商（包括金融科技公司）处于开放银行生态系统中；超过250万个英

国消费者和企业正在使用开放银行产品来管理财务、获得信贷并进行支付；API调用量从2018年的平均557万次/月增加到2022年2月的8.65亿次/月。

开放银行服务商的一个重要价值在于通过技术创新解决数据市场碎片化和互操作性问题。开放银行的生态圈包含银行、科技公司、第三方商户、用户等多种参与方，参与方的多样性、独立性及复杂性可能会衍生出多样的技术形态，并导致各方路径、结构等技术方面与业务方面存在不兼容等现象。提高标准化程度固然是一个方向，可减少各类参与者之间的对接所产生的额外成本，但全部依赖统一的标准是不现实的，在提高特定行业特定数据标准化程度的同时，也需要鼓励市场创新，通过提供私人技术解决方案来解决数据市场碎片化和互操作性的问题。

美国的Plaid公司就是在这浪潮中发展起来的“领头羊”。Plaid通过开发的API产品及基于API的技术解决方案搭建起连接银行的“数据管道”，金融科技企业在开发App时将Plaid的API嵌入其中，就相当于实现了“一点接入”，无须再与各银行逐个进行对接，这大大简化了金融科技企业获取银行业务及消费者财务数据的过程。截至2022年4月，Plaid的客户包括12 000余个金融机构以及5 500余个金融科技企业，链接超过2亿个消费者账户，估

值已经高达134亿美元。Plaid的基本情况见表4。类似的代表性服务商还有美国的Yodlee、Finicity，英国的TrueLayer，瑞典的Tink。

1.6 数据沙箱

数据沙箱提供强级别的控制，有望提供对非常敏感数据、个人数据和专有数据的安全访问。数据沙箱由数据保管人提供，用来描述任何安全的相对孤立的环境，通过该环境，可以访问和分析数据，但无法从环境中复制或提取任何原始数据，并且仅在不敏感时才会导出分析结果。这些沙箱可以通过物理隔离手段或者技术手段（如可信执行环境）实现。数据沙箱通常要求分析代码在数据存储位置执行。

牛津大学的电子健康记录安全分析平台OpenSAFELY是一个非常典型的例子。该平台存放着英国95%人口的医疗保健记录，数据极其敏感。该平台允许独立研究人员将他们的分析代码上传到数据沙箱中进行分析，但研究人员既无法直接查看数据，也无法将数据转移出其所在的安全环境，实现了数据的可用不可见。OpenSAFELY使得使用敏感数据进行公共利益研究成为可能，同时维护了该数据的隐私性和机密性。

表4 Plaid的基本情况

基本情况	具体内容
公司定位	利用API提供金融数据共享服务的B2B (business to business) 金融科技企业，构建了连接消费者、传统金融机构和开发人员的技术基础架构API
覆盖地区	2012年成立于美国旧金山，目前已覆盖美国、加拿大、英国、法国、西班牙、爱尔兰、荷兰、德国
员工数	1 000余个
用户数	12 000余个金融机构，5 500余个金融科技企业，链接超过2亿个消费者账户
市场估值	134亿美元
总融资	7.34亿美元（2020年年底VISA欲以53亿美元收购，但迫于美国反垄断诉讼压力放弃）

1.7 数据共享平台

数据共享平台可在严重缺乏信任关系的竞争方之间搭建桥梁,促成数据共享解决方案的实现。在很多商业环境中,数据共享能为所有相关方带来好处,如减少检测欺诈或洗钱等金融犯罪、改善供应链管理、实现协作产品开发和设计,甚至解决全行业的问题,但可能由于竞争利益(如出于保护敏感知识产权或市场地位),难以开展数据共享。在这种情形下,数据共享平台作为一个受信任的第三方,可以发挥重要的作用。

根据需要,数据共享平台既可以由公共部门设立,也可以由私营机构设立。在某些情况下,政府可以作为或创建受信任的第三方。例如,我国的金融信用信息基础数据库由中国人民银行征信中心负责建设、运行和维护,实现了借贷信息在金融机构间的共享,有效助力了整个行业的风险防范和业务发展。私营机构设立数据共享平台的情形可能更加普遍:主要数据提供者形成一个联盟,指定一个现有的可信组织或创建一个新的可信组织或平台来负责平台运维。例如,HiLo海事风险管理是一项由航运业主要参与者发起的非营利性计划,旨在将航运公司的健康和绩效数据与竞争对手进行比较。HiLo充当受信任的第三方,它汇总来自航运公司的数据,以提高整个行业的健康性和安全性。作为提供数据的回报,航运公司会获得HiLo的分析建议,以改善自己的安全实践。ODI公布的数据显示,HiLo使救生艇事故减少了72%,机舱火灾减少了65%,燃料泄漏减少了25%,大大发挥了其作为数据共享平台的作用。

1.8 各种数据中介的比较优势

对应表1概括的数据分享存在的6类问

题,表5对不同数据中介能缓解甚至解决的问题进行了总结。

总体来看,不同的数据中介在解决不同问题时具有不同优势。数据交易平台着重解决供需双方的信息不对称问题,但在解决数据供给方的激励问题、供需双方的信任问题以及降低数据传输成本方面却不具有优势;数据经纪商能大幅降低数据搜寻、汇聚的成本,但在保护个人信息和满足合规要求方面存在不足;个人信息管理系统重新定义了个人与数据持有者之间的权益关系,有助于解决数据分享的激励和透明度问题,但不具备成本优势;数据信托可以优化“授权同意”这一个人信息分享路径,提高数据供给效率,但在解决信息不对称等问题方面不占优势;Plaid等开放银行服务商着重解决数据标准和数据接口的统一转换问题,可以有效降低数据传输成本,但难以解决数据分享后的安全可控问题;数据沙箱能够解决数据的可控安全问题,特别是敏感数据分享,但可能不利于细颗粒度数据的大规模融合和多场景使用;数据共享平台在一定程度上可以拆除竞争篱笆,实现数据的共建共享,但第三方自身仍然存在不完全可信的问题,数据供给方可能仍不愿意将核心数据分享出来。

数据流通中的障碍是多维的,主要矛盾往往因场景、数据类型、机构间信任关系等因素而不同,因此寄希望于单一类型的数据中介解决所有问题,可能是不现实的。这就有必要坚持问题导向,根据不同的场景、不同类型的数据、不同机构碰到的主要问题,以及技术可行性等因素综合考虑,寻求最合适的数据中介方案,或者是其中几类方案的创新组合形式。

2 总结与建议

在数字经济时代,促进数据的访问和

表5 不同数据中介能缓解甚至解决的问题

数据中介	动力：缺少分享数据的激励	顾虑：商业、道德和声誉风险	顾虑：法律和监管风险	渠道：缺乏关于数据可用性或价值的信息	成本：数据访问/分享的成本过高	公共性：数据公共价值发挥不够
数据交易平台				√		√
数据经纪商				√	√	√
个人信息管理系统	√		√	√		√
数据信托	√		√		√	√
开放银行服务商	√	√			√	√
数据沙箱		√	√			√
数据共享平台	√	√			√	√

注：√表示该类数据中介能够缓解甚至解决这方面的问题。

分享比以往任何时候都更加重要。但这无法一蹴而就，良好的生态需要逐步有序地建立。一方面，需要通过制度创新疏通政策堵点，降低法律法规的不确定性；另一方面，需要通过市场创新切实降低交易费用，寻找、构建多方共赢的激励机制和可持续的商业模式。

本文认为，发展数据中介是降低阻碍数据分享和流通的各类交易费用的关键。阻碍数据分享流通的问题可以分为5个维度、6个方面，分别是动力（缺少分享数据的激励）、顾虑（包括商业、道德和声誉风险以及法律和监管风险两方面）、渠道（缺乏关于数据可用性或价值的信息）、成本（数据访问/分享的成本过高）、公共性（数据公共价值发挥不够）。而数据中介可以分为三大类：第一类数据中介的核心目标在于解决市场中的信息不对称问题，并发挥规模效应，从而降低平均成本；第二类数据中介旨在增强个人对个人信息的控制权和携带权；第三类数据中介旨在通过可信技术环境或可信组织，消除数据易被复制、易被滥用的风险。

本文详细介绍了数据交易平台、数据经纪商、个人信息管理系统、数据信托、

开放银行服务商、数据沙箱和数据共享平台7种典型的数据中介及相应案例。不同的数据中介着重解决不同的实践难题：比如，数据交易平台属于第一类，着重解决供需双方的信息不对称问题；开放银行服务商属于第二类，着重解决数据标准和数据接口的统一转换问题，降低个人信息携带权的落地成本；数据沙箱属于第三类，着重解决数据的可控安全问题，特别是敏感数据分享。正因为不同的数据中介对于解决不同的问题有着各自的优势，所以不应寄希望于单一类型的数据中介解决所有问题，而应遵循适配性原则。

结合本文分析，有两点需重点注意：第一，交易费用的降低是数据分享和流通市场发展的前提，而要降低交易费用，需要坚持市场在资源配置中的决定性作用，大力发展数据中介；第二，数据流通中的障碍是多维的，主要矛盾往往因场景、数据类型、机构间的信任关系等因素而不同，而不同的数据中介在解决不同问题中有着各自的优势，因此有必要鼓励百花齐放，百家争鸣，鼓励发展多元化的数据中介形式。在具体实践中，需要具体问题具体分析，坚持问题导向原则，寻求最合适的数据中介方案，或者多种方案的创新组合形式。

参考文献:

- [1] MARTENS B, STREEL A D, GRAEF I, et al. Business-to-business data sharing: an economic and legal analysis[R]. 2020.
- [2] O'DONNELL F, GHANI R. Matchmakers of the data world: how “data intermediaries” can bring decision makers and data together to help combat COVID-19[R]. 2020.
- [3] Organisation for Economic Cooperation and Development. Enhancing access to and sharing of data: reconciling risks and benefits for data re-use across societies[R]. 2019.
- [4] 黄丽华, 窦一凡, 郭梦珂, 等. 数据流通市场中数据产品的特性及其交易模式[J]. 大数据, 2022, 8(3): 3-14.
- [5] HUANG L H, DOU Y F, GUO M K, et al. Features and transaction modes of data products in data markets[J]. Big Data Research, 2022, 8(3): 3-14.
- [5] RAMIREZ E, BRILL J, OHLHAUSEN M K, et al. Data brokers: a call for transparency and accountability[R]. 2014.
- [6] The Centre for Data Ethics and Innovation. Unlocking the value of data: exploring the role of data intermediaries[R]. 2021.

作者简介



李振华 (1976-), 男, 蚂蚁集团研究院院长, 中国计算机学会中国数字经济50人论坛首批委员, 中国金融学会理事, 中国证券业协会互联网证券专业委员会委员, 清华大学五道口金融学院民生财富管理研究中心研究员, 浙江大学-蚂蚁集团金融科技研究中心联合主任。曾作为二十国集团(G20)领导人杭州峰会数字普惠金融中方专家组成员之一, 参与《二十国集团数字普惠金融高级原则》制定和咨询工作。牵头蚂蚁集团在数字经济、数字信贷、数字信用、移动支付、数字货币、金融科技监管、金融风险与金融稳定等领域的专题研究。



王同益 (1988-), 男, 博士, 蚂蚁集团研究院高级专家, 主要研究方向为数字货币、数字支付、区块链、数据治理等金融科技创新及与之相关的经济、法律、技术等交叉性问题。

收稿日期: 2022-04-27

通信作者: 王同益, tongyi.wty@antgroup.com