

【产业发展】

数据交易之悖论与突破：不完全契约视角

龚 强 班铭媛 刘 冲

【摘 要】数据是数字经济的基础性和战略性资源,数据交易是推进数据要素市场化配置的必要前提。本文依据数据的“场景专用性”特点,基于不完全契约理论,揭示了数据交易存在“数据安全”与“数据流通”之悖论的理论机制。数据交易的“可用不可见”是突破数据交易悖论的关键,隐私计算技术能够提供可行的解决方案。“可用不可见”不仅是数据交易技术的变革,更是交易模式的创新,即从传统的“数据所有权交易”转变为“数据使用权交易”。相比数据所有权交易,数据使用权交易能给予更充分的数据生产与采集激励,促进数据资源市场化交易,实现数据要素的高水平积累与价值释放。然而,建立“可用不可见”的数据交易环境需要一定技术投资,市场主体私人投资激励可能不足。为此,有必要适度超前布局数据交易基础设施,从国家和区域层面加强共建共享统筹规划,解决共性基础技术难题,降低市场主体的交易成本,从而更广泛、更充分地释放数据动能。本文指出,数据安全不仅不与数据流通相悖,反而是数据流通价值释放的关键,数据安全将成为数字经济实现高质量发展的先决条件和重要保障。

【关 键 词】数据要素;数据交易;数据安全;不完全契约

【作者简介】龚强,中南财经政法大学文澜学院,北京大学数字金融研究中心,电子信箱:qianggong@zuel.edu.cn;班铭媛,中南财经政法大学文澜学院,电子信箱:banmingyuan2008@126.com(武汉 430073);刘冲(通讯作者),北京大学经济学院,电子信箱:pkuliuchong@pku.edu.cn(北京 100871)。

【原文出处】《经济研究》(京),2022.7.172~188

【基金项目】本研究得到国家自然科学基金重大项目(72192844)、国家自然科学基金面上项目(71773143, 72073146)、北京大学经济学院种子基金的资助。

一、引言

数据是数字经济的核心生产要素和创新动力源泉。习近平总书记指出,要“发挥数据的基础资源作用和创新引擎作用,加快形成以创新为主要引领和支撑的数字经济”。^①加快完善数据要素市场化配置,培育发展数据交易市场,是推进数字经济高质量发展的应有之义。然而,当前我国数据交易市场存在重重挑战,市场主体积极性不高,数据交易活跃度较低。尤其是数据交易中潜在的数据安全风险,成

为众多市场主体“不敢交易、不愿交易”的隐忧,形成“数据安全”与“数据流通”难以两全的悖论,极大制约了数据要素配置效率。如何在有效保障数据安全的前提下,最大程度促进数据交易与流通,以充分释放数据要素价值,是当前我国数据要素市场化配置亟待解决的问题。

数据要素是数字经济繁荣发展的基石,党中央高度重视数字经济发展,明确提出数字中国战略,围绕数据要素进行了一系列的顶层设计(相关政策

文件请见表1)。早在2015年,十八届五中全会就已经提出实施“国家大数据战略”。2019年,十九届四中全会首次将“数据”确认为生产要素之一,并提出数据要素“由市场评价贡献、按贡献决定报酬”。2020年4月,中共中央、国务院发布《关于构建更加完善的要素市场化配置体制机制的意见》,将“数据”与土地、劳动力、资本、技术并称为五种生产要素,提出“加快培育数据要素市场”,数据要素市场化配置上升为国家战略。2022年6月,习近平总书记主持召开中央全面深化改革委员会第二十六次会议,审议通过了《关于构建数据基础制度更好发挥数据要素作用的意见》,首次明确要建立数据产权制度、数据要素流通和交易制度,以及数据要素收益分配制度等三方面数据基础制度。从一系列顶层设计可以发现,数据要素相关政策不断具体化、制度化,从确立数据为一种新型生产要素,到提出培育数据要素市场,再到明晰数据要素市场建设所需要的基础制

度建设,尤其越来越强调数据安全、数据产权、数据要素市场三方面内容。同时,各省市也加快推进数据要素制度建设,截至2022年6月,已有18个省市公布了相关数据条例(含草案),其中深圳、上海、重庆、浙江等省市已进入条例实施阶段。^②

健全数据交易机制是推进数据要素市场化配置的先决保障。数据交易可以充分发挥市场的资源配置作用、激活数据价值,是数据资源转换为数据要素的重要一环。^③我国数据交易市场建设走在全球前列,具有先发优势。我国拥有丰富的数据资源,为数据交易市场发展奠定了禀赋优势。据国际数据公司IDC统计,2019年全球大数据存储量达41ZB,其中我国数据占全球数据总量比例高达23%。在国家大数据战略等一系列顶层设计和统筹布局下,我国数据交易市场建设持续推进。截至2019年末,我国已建成仅次于美国的世界第二大数据交易市场。^④目前已有四十余家大数据交易所(平台、中心),包括贵

表1 关于数据要素的国家层面相关政策文件

时间	政策文件名称	涉及数据要素流通的重点内容
2019年10月	《中共中央关于坚持和完善中国特色社会主义制度 推进国家治理体系和治理能力现代化若干重大问题的决定》	首次确认数据为生产要素之一。提出“健全劳动、资本、土地、知识、技术、管理、数据等生产要素由市场评价贡献、按贡献决定报酬的机制”。
2020年4月	《中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见》	将数据与土地、劳动力、资本、技术并列为五大生产要素。提出“加快培育数据要素市场”,包括“推进政府数据开放共享、提升社会数据资源价值、加强数据资源整合和安全保护”等。
2021年3月	《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》	提出“加快建立数据资源产权、交易流通、跨境传输和安全保护等基础制度和标准规范”。
2022年1月	《“十四五”数字经济发展规划》	提出“加快构建数据要素市场规则,培育市场主体、完善治理体系,促进数据要素市场流通”,提出“数据要素市场培育试点工程”,其中包括“开展数据确权及定价服务试验、推动数字技术在数据流通中的应用、培育发展数据交易平台”三方面任务。
2022年1月	《要素市场化配置综合改革试点总体方案》	提出“探索‘原始数据不出域、数据可用不可见’的交易范式,在保护个人隐私和确保数据安全的前提下,分级分类、分步有序推动部分领域数据流通应用。探索建立数据用途和用量控制制度,实现数据使用‘可控可计量’”。
2022年3月	《中共中央 国务院关于加快建设全国统一大市场的意见》	提出“加快培育数据要素市场,建立健全数据安全、权利保护、跨境传输管理、交易流通、开放共享、安全认证等基础制度和标准规范,深入开展数据资源调查,推动数据资源开发利用”。
2022年6月	《关于构建数据基础制度更好发挥数据要素作用的意见》(中央全面深化改革委员会第二十六次会议审议通过)	强调“数据基础制度建设事关国家发展和安全大局”,提出要“建立数据产权制度、数据要素流通和交易制度、数据要素收益分配制度”三方面数据基础制度。

注:资料来自中国政府网,作者整理。

阳大数据交易所、北京国际大数据交易所、上海数据交易所等。

然而,我国数据交易市场建设也面临着极大挑战,突出表现在数据交易过程中存在“数据安全”与“数据流通”的悖论。以全球首家建立的大数据交易所——贵阳大数据交易所为例。贵阳大数据交易所是在大数据国家战略背景下建立,然而,在几年来的发展过程中,数据交易的活跃度却不尽如人意。囿于技术能力的客观局限性,交易所陷入“数据安全”和“数据流通”难以两全的困境。为保护数据安全,交易所只能交易数据分析结果而非原始数据本身,这极大限制了数据交易对象范围;再者,为筛选客户以达到数据保护的目的,交易所设立了严格的会员准入制度,导致交易门槛较高、交易规模相对有限。^⑤在这种情况下,很多数据交易选择绕开交易所,进行场外数据交易。然而,游离于监管之外的黑市交易潜伏着更大的数据安全隐患,尤其是涉及到居民个人隐私的非法交易,引发了大数据杀熟、大数据电信诈骗等一系列危害国家和人民安全的重大社会问题。

本文研究发现,鼓励在“可用不可见”的安全交易环境下进行数据使用权交易,是打破数据交易市场发展瓶颈的关键。理论分析表明,如果忽视数据的非竞争性特征,效仿普通竞争性商品的所有权交易模式,将导致数据出售方不再是数据所有权的唯一拥有者,损害数据产业链上游的数据采集激励,制约数据要素积累水平。隐私计算技术(包括联邦学习、安全多方计算、差分隐私等)有望成为推动数据交易模式转型的契机,其经济本质在于,突破“数据所有权交易”这一传统交易模式的桎梏,探索“数据使用权交易”的新型交易模式。隐私计算技术能够为数据的“可用不可见”提供安全环境,确保数据所有权在交易过程中不发生转移,这有助于提高数据采集激励,促进数据交易流通,实现更高水平的数据要素积累与价值释放。从实践来看,本文的理论发现也较好解释并顺应了国内数据交易所的发展趋势。仍以贵阳大数据交易所为例,在经历低迷困境

后,贵阳大数据交易所重组优化后再度启动运营,新的交易所摒弃了既有交易模式,将充分利用隐私计算、区块链、云计算等数字技术,力求实现原始数据“可用不可见”,数据产品“可控可计量”,流通行为“可信可追溯”。^⑥

本文也从理论上揭示了加快推动数据交易基础设施建设的必要性和迫切性。从技术功能角度,隐私计算技术能够实现“数不出库、数尽其用”,从而平衡数据安全与数据流通的需求;但从技术成本角度,当前隐私计算技术在复杂场景应用还有一定局限性,需要一定的技术攻关投入,而这些成本并非所有市场主体都能负担得起,在短期内企业未必有激励选择“可用不可见”交易。本文指出,从社会最优角度,应支持数据交易所加强数据交易基础设施建设,尤其是解决共性基础技术难题,以平台服务形式将技术开放给市场主体使用,降低中小数据企业的技术门槛,从而更充分地激活数据要素价值。这一理论发现也从技术赋能角度解释了我国数据交易所发挥基建作用的趋势。例如,北京国际大数据交易所明确提出要“打造国内领先的数据交易基础设施”,这一战略定位是非常可取的,这既是打破数据交易悖论、推动数据交易模式转型的当务之急,也是广泛激励市场主体参与、做大做强做强我国数据要素市场的必要之举。

本文系统探讨了数据交易的不完全契约特征。数据要素具有虚拟性、非竞争性、价值不确定性等独特的技术经济特征(蔡继明等,2022;蔡跃洲和马文君,2021),也正因如此,数据交易具有典型的不完全契约特征。具体而言,由于数据具有非竞争性,同一数据可以在同一时点下被不同主体同时消费(Jones & Tonetti, 2020)。这意味着,即使数据卖方在交易事前制定了限制竞争的相关条款,如禁止买方重复使用或转售数据,也是不可置信的。因为数据买方一旦购得数据,即可自由支配数据用途,从此不再依赖于数据卖方,相当于变相切断与数据卖方的契约关系。因此,交易双方很难制定出能够预见各种情况并有足够约束力的完全契约。

进一步,我们首创性地提出数据的“场景专用性”概念,即数据只有与具体应用场景相结合,数据价值才得以体现。这一概念与谢康等(2020b)、林志杰和孟政炫(2021)的发现也具有内在一致性,即数据要素必须要与数据分析性劳动、技术、资本等要素结合后,才能实现其经济价值。正因为数据具有“场景专用性”,所以数据价值具有高度不确定性,交易双方在交易前无法准确预见数据未来能够产生多少经济价值(Dosis & Sand-Zantman, 2022),故难以在事前设计周密的价格条款。^⑦“场景专用性”不仅精炼地概括了数据要素的独特技术经济特征,同时为我们运用不完全契约理论分析数据交易问题提供了一个很好的逻辑起点。

本文的主要贡献可以概括为以下三个方面:

第一,为数据交易市场的交易机制设计与优化提供了理论基础。自国家大数据战略实施以来,各省市加速推进数据交易市场建设,然而,据实地调研,这些数据交易所的实际运营表现大多未及预期,目前实际运营的数据交易所只有三分之一左右。为什么数据交易所没有像其他交易所那样充分地激发交易活力?数据交易所应当如何扭转局面并寻求转型突破口?这些现实问题亟需相关理论进行解释和指导。本文立足于我国数据交易市场建设的实践经验,从理论上揭示了当前大数据交易市场面临的发展挑战以及突围关键。研究发现,数据“可用不可见”是促进数据交易的关键,其本质上是“数据所有权交易”模式转化为“数据使用权交易”模式,后者能更好地激励数据采集,更有效地促进数据价值释放,推动数据交易市场形成可持续的商业模式,为数字经济高质量发展打下坚实的数据要素基础。

第二,本文指出,数据安全不仅不与数据流通相悖,反而是数据流通价值释放的关键,这为加强数据安全治理提供了理论依据。本文研究表明,数据安全不仅因为关乎国家安全而具有极端重要性,同时也是实现数据价值的前提保障。以牺牲数据安全为代价的数据流通,必定会与数据价值实现的初衷背道而驰。若在没有安全技术保障下直接进行数据交

易,不仅严重威胁数据安全,同时极大损害了数据采集的激励,不利于数据交易的可持续性,从长期来看,制约数据要素的积累与价值实现。因此,数据安全和数据流通并非是“非此即彼”的矛盾关系,正因为要通过数据流通充分释放数据价值,才要严管数据安全。守住数据安全底线,就是促进数据价值实现,也是数据要素市场化改革的先决保障。

第三,从经济角度揭示了隐私计算技术的经济本质和可观潜力,并揭示了推进数据交易基础设施建设的必要性和迫切性。习近平总书记在中共中央第三十四次集体学习时强调,“要加快新型基础设施建设,打通经济社会发展的信息‘大动脉’。”隐私计算技术为促进数据安全流通、培育壮大数据产业链提供了关键保障。不过,此类共性基础技术的系统性、通用性强、前期投资成本高,市场主体投资激励不强,需要充分发挥新型举国体制的资源配置优势,在国家层面加强统筹共建共享规划。这既是对市场主体投资激励不足的有力补充,也能有效避免重复建设、运营低效等问题。

本文余下内容安排为:第二节回顾数据要素与数据交易相关文献;第三节介绍理论框架;第四、五节对数据所有权交易和使用权交易两种交易方式进行理论分析;第六节从数据要素积累及其带来的社会财富角度对两种数据交易方式进行对比;第七节总结并提出相关政策建议。

二、文献回顾

数据要素驱动的数字产业化和产业数字化正在掀起一场数字经济革命,成为中国经济增长的重要支撑(蔡跃洲和牛新星, 2021; 徐翔和赵墨非, 2020)。数字技术的应用有效提升了企业生产的自动化水平和生产效率(Aghion et al., 2018),提升企业研发创新能力和效率(肖静华等, 2020; 刘意等, 2020; 张明超等, 2021),推动供应链协作变革(陈剑和刘辉辉, 2021; 孙新波等, 2019),再调整劳动要素需求(王永钦和董雯, 2020; 周广肃等, 2021)。随着数字经济持续蓬勃发展,越来越多的研究将视角“下沉”至驱动数字经济的底层生产要素——数据要素。数据作为新

的生产要素,有着区别于传统生产要素的独特技术经济特征(熊巧琴和汤珂,2021;徐翔等,2021;蔡跃洲和马文君,2021)。概括而言,数据要素具有虚拟性、非竞争性、正外部性、价值不确定性、极强专用性、边际产出递增、边际成本几近于零等独特特征。此外,数据资源并不能独自地发挥经济价值,数据资源若要成为现实生产要素,必须要与劳动、资本、技术等传统生产要素进行协调和配合,因此,数据要素与传统要素共同构成互补性资产(谢康等,2020b;林志杰和孟政炫,2021)。

在数字经济时代,数据伴随生产、分配、流通、销售等各环节商业活动而产生,是经济产出的副产物(Jones & Tonetti, 2020; Farboodi & Veldkamp, 2021)。越来越多的企业开始采用“数据驱动决策”模式(Brynjolfsson & McElheran, 2016),即企业根据数据资源调整优化生产计划和营销方式,提高生产和销售效率,从而进一步提高数据资源积累效率,形成正向反馈循环的闭环机制(data feedback loop)。数据要素驱动生产效率提升有多种实现机制:一是,从技术角度,数据要素有助于驱动生产技术的改进(Jones & Tonetti, 2020; Cong et al., 2021);二是,从信息角度,数据具有信息属性,有助于减少不确定性,提高预测的精确性,进而提高生产效率(Farboodi & Veldkamp, 2021; Farboodi et al., 2019; Bajari et al., 2019);三是,从与劳动要素结合角度,数据要素可以通过数据的初始存量、前期收集处理数据所投入的劳动以及当期在收集处理数据所投入的劳动等途径提高绝对生产力(蔡继明等,2022);四是,从组织结构角度,数据要素有助于驱动企业组织制度从传统工业经济的科层制组织,过渡至更适应数字经济特征的网格制组织(谢康等,2020a)。

鉴于数据要素的重要意义,促进数据交易流通,是健全完善数据要素市场、激活数据要素潜能的重要一环。很多研究探讨了数据共享的价值。数据共享有助于发挥数据的非竞争性特征,让更多企业可以使用开放数据来进行技术研发创新(Beraja et al., 2022)。例如,Akcigit et al.(2016)研究发现,专利交易

市场有助于创新思想的流通、提高企业生产率、促进经济增长。同时,数据开放共享也有助于增强市场竞争,防止拥有市场垄断地位的企业滥用其市场地位,进而从根本上解决企业的“大数据杀熟”倾向(李三希等,2021)。尽管数据流通有诸多益处,但拥有数据的企业出于维护自身优势的考虑,往往不愿意将数据分享给其他企业,这会形成数据孤岛,极大地阻碍了数据充分发挥其经济价值(Jones & Tonetti, 2020)。

为促进数据流通,数据交易中心应运而生,为数据交易流通提供了平台。Azcoitia & Laoutari(2022)对全球97家数据交易中介开展了详尽的调查研究。^⑧基于全球数据交易市场的蓬勃发展,国外很多学者探讨了数据交易(或共享)的相关理论问题。例如:数据定价策略(Bergemann et al., 2018; Bergemann & Bonatti, 2015)、数据交易的信息设计(Ichihashi, 2020; Bergemann et al., 2018)、数据交易市场竞争(Ichihashi, 2021; Ali et al., 2022)、数据产权归属的福利影响(Ali et al., 2022; Dosis & Sand-Zantman, 2022)、数据外部性特征引发“数据过度共享”的担忧(Choi et al., 2019; Acemoglu et al., 2022; Bergemann et al., 2019)等。

数据交易流通的重要性是不言而喻的,但从实践来看,当前我国数据交易市场建设仍面临多重挑战。我国数据交易市场仍在不断探索建设中,数据定价机制不完善、数据权属不明确、数据安全难保障等瓶颈难题亟待解决(蔡跃洲和马文君,2021;汤琪,2016)。这些瓶颈难题与数据要素所具有的独特技术经济特征有密不可分的关系。数据要素的生产函数、供给需求分析等方面均与传统生产要素有本质区别,为数据要素定价带来极大困难。^⑨产权界定清晰是数据要素市场健康发展的基础,数据权属及其分配规则不清,已成为当前数字经济发展的制度障碍(申卫星,2020),数据归谁所有、由谁使用、收益归谁等数据权利归属问题仍有待于解决(肖旭和戚聿东,2021)。

隐私计算技术为突破数据交易困境提供了可行思路,成为完善数据要素市场化配置的重要抓手。^⑩

目前关注隐私计算技术在数据交易中应用的经济学分析还相对较少。Abowd & Schmutte(2019)关注到这一问题,但其分析的侧重点与本文有所不同。Abowd & Schmutte(2019)更聚焦于差分隐私算法如何平衡统计数据中的用户隐私保护与统计精确性之间的关系。本文则以当前数据交易所面临的“数据流通”与“数据安全”的现实矛盾作为研究出发点,侧重于从不完全契约视角考察隐私计算技术平衡数据交易与数据安全的内在机制。

三、理论框架

我们考虑市场中的代表性数据产业链。产业链的上游为数据采集方(下文简称“采集方”),从事数据生产、搜集、标注、清洗、聚合等工作。产业链的下游为数据分析方(下文简称“分析方”),从事数据建模与分析工作,直接面向终端市场,为消费者提供数据分析服务。采集方完成数据采集工作后,将数据出售给分析方,并由分析方进行数据分析和数据服务销售。

(一)交易流程

采集方和分析方每一期都有机会进行数据交易,直至无穷期,市场贴现因子为 ρ 。考虑交易时间序列 $t=1,2,\dots,\infty$ 。以第 t 期为例,交易流程分为如下四步(见图1):

第一步,采集方进行数据采集投资。采集数据的规模为 $q(t)$,投资成本为 $C(q(t))=\alpha q(t)^2$,其中 $\alpha>0$ 为成本系数。第二步,采集方与分析方进行数据交易。与不完全契约理论一致(Grossman & Hart, 1986),双方在合同中约定,待分析方在终端消费者市场获得销售利润后,双方采用协商定价方式,分析方向采集方

支付数据购买费用。^①支付规则将在模型分析部分详细阐述。第三步,分析方在终端消费者市场获得销售利润,销售利润为 $R(t)$ 。第四步,分析方向采集方支付数据购买费用 $p(t)$ 。接着,时间进入 $t+1$ 期,重复与第 t 期类似的交易过程。

(二)数据的“场景专用性”与经济价值

数据的“场景专用性”体现于,数据资源本身并不具备经济价值,只有把数据流通至“能工巧匠”手中,在具体应用场景下把数据“活”用起来,才能转化为数据要素,产生经济价值。例如,企业的工商、财务、知识产权等信息本是零散分布的,价值密度很低,但如果将其应用于银行贷款审批,就可以绘制企业画像,训练大数据风控模型,节约贷款成本;或将其应用于政府产业规划,就可以绘制产业图谱,锁定产业链薄弱环节,定向实施补链强链计划,节约财政资金。从这个例子可以看出,同一数据可以应用于不同场景,在不同场景下数据产生的经济价值也不尽相同。因此,讨论数据价值,必须要置于具体应用场景之下,这也是建立数据要素交易市场的初衷与目标——只有让数据流通起来,才能借助市场的力量,有效地寻找应用场景,从而在最大程度上释放数据的经济价值。

为刻画数据的“场景专用性”,模型假设,只有分析方具备在具体应用场景下进行数据分析的能力,而采集方则不具备,采集方和分析方有激励进行数据交易,因为唯有此才能释放数据的经济价值。^②因此,只有采集方和分析方进行数据交易时,数据产生总收益 $R(t)$;否则,双方收益均为0。从另一角度,也可以理解为,只有分析方拥有具体应用场景下所需

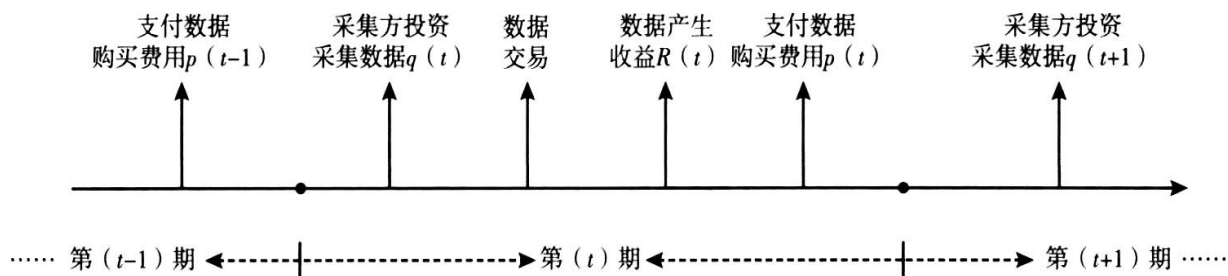


图1 交易流程(以第 t 期为例)

的各种其他生产要素,只有当数据和与之匹配的其他生产要素结合后,才能产生经济价值,这与谢康等(2020b)、林志杰和孟政炫(2021)、蔡继明等(2022)等研究观点是一致的。

在分析方的具体应用场景下,数据产生的经济价值 $R(t)$ 与数据规模、数据时效性两方面因素有关。实证分析表明,数据价值与其时效性正相关,历史时间较早的数据对预测当前情况的作用相对有限(Chiou & Tucker, 2017)。为刻画数据的时效性,定义有效数据存量 $Q(t) = \sum_{k=0}^t \lambda^{t-k} q(k)$, 表示第 t 期数据交易时采集方采集的全部有效数据。此处的“有效”指,考虑数据由于时效性下降而产生的价值折旧。其中, λ 为折旧因子, $0 < \lambda < 1$, λ 越小,表示数据折旧幅度越大。有效数据存量的迭代关系为 $Q(t) = q(t) + \lambda Q(t-1)$, 初始条件为 $Q(0) = 0$ 。在第 t 期双方交易时,可供交易的有效数据存量为 $Q(t)$, 若有效数据存量全部用于数据分析,将产生收益 $R(t) = \beta Q(t)$, 其中 $\beta > 0$ 。此处,为简化计算并不失一般性,将数据收益 $R(t)$ 表示为有效数据存量 $Q(t)$ 的线性函数,以刻画数据随数据规模增加而价值增加、随时效性下降而价值下降的两方面特性。

(三) 数据交易方式

我们考虑两种代表性数据交易方式。第一种交易方式是数据所有权交易,指没有人在安全技术环境下,采集方直接将数据出售给分析方。在这种交易方式下,分析方在一次性交易后,即永久性拥有该数据,并可以在未来任何时间点随意使用。例如,在实务操作中,数据以 excel 表、数据包等形式交易便类似于这种交易方式。

第二种交易方式是数据使用权交易,指采集方不直接出售数据,而是在“可用不可见”的安全环境下,将数据以固定期限服务的形式出售给分析方,分析方必须在采集方所提供的技术环境下进行数据分析工作,无法将数据输出到技术环境之外使用。例如,联邦学习、安全多方计算等隐私计算技术在数据融合中的应用,本质上都类似于这种交易方式。在模型中,我们假设在这种交易方式下,分析方仅能够

在交易当期使用数据,如果下一期想继续使用该数据,必须重新购买。同时,采集方为部署“可用不可见”的技术环境,需要每期投入固定的技术投资成本 FC , 例如投入云计算资源等。

从实践角度,上述两种数据交易模式的核心差异在于,数据是否在安全技术环境下进行交易。从理论角度,安全技术环境的从无到有,本质上是从交易数据的所有权,转变为交易数据的使用权。为了凸显两种数据交易模式的经济本质区别,同时兼顾表述的简洁性,下文用“数据所有权交易”作为“无安全技术保障的数据交易”的代称,以上角标 PR 标识;用“数据使用权交易”作为“有安全技术保障的数据交易”的代称,以上角标 UR 标识。

四、数据所有权交易分析

由于数据具有非竞争性,不同主体可以在同一时间使用同一数据。如果将数据直接出售,采集方和分析方将共同享有数据的所有权。本节将通过理论分析证明,当进行数据所有权交易时,数据的非竞争性将影响数据交易的可持续性,并制约数据要素的长期积累。

与 Grossman & Hart(1986)的理论框架一致,待分析方获得数据分析收益后,交易双方进行纳什讨价还价,假设交易双方议价能力相同且均为 $\frac{1}{2}$, 分析方向采集方支付金额 $p^{pr}(t)$, 如(1)式。分析方的收益由两部分组成:一是自身保留收益,即若本期无数据交易发生,分析方所能获得的收益;二是与采集方分享由数据交易产生的价值。

$$R(t) - p^{pr}(t) = \underbrace{\lambda R(t-1)}_{\text{分析方保留收益}} + \underbrace{\frac{1}{2}(R(t) - \lambda R(t-1))}_{\text{双方分配的价值}} \quad (1)$$

(1)式从交易契约角度诠释了数据的非竞争性。首先,关于分析方的保留收益,分析方在一次性购买数据后,从此之后都可以利用相同数据,独立完成数据分析工作,而无需再依赖于采集方。因此,分析方在第 t 期的保留收益为 $\lambda R(t-1)$ 。需注意,经过一期后,数据价值会相应折旧。其次,关于双方分配的价值,由于分析方可以重复使用过去已购买的数据,分析方仅会购买相较于已有数据的新增数据,

因此双方需要分配的价值仅为新增数据产生的收益,即 $R(t)-\lambda R(t-1)$ 。即使交易双方在事前禁止分析方重复多期使用数据,但一旦数据完成交易,分析方便掌握了数据所有权,以此摆脱对采集方的依赖,因此采集方的事前约定实际上是没有事后约束力的。在极端情况下,如果采集方不再新增或扩充数据,数据交易将会变成“一锤子买卖”而难以持续发展下去。

采集方确定每一期数据采集规模,以最大化长期总收益。对于任意 $\tau \in \mathbb{N}$, 给定有效数据存量 $Q(\tau)$, 将采集方从第 τ 期至无穷期的未来最优总收益记作 $V^{PR}(Q(\tau))$, 即(2)式:

$$V^{PR}(Q(\tau)) = \max_{\{Q(t+1)\}_{t=\tau}^{\infty}} \sum_{t=\tau}^{\infty} \rho^{t-\tau} [p^{PR}(t) - C(t)] \quad (2)$$

(2)式决策方程需要满足(1)、(3)、(4)式约束条件:

$$Q(t) = q(t) + \lambda Q(t-1) \quad (3)$$

$$R(t) = \beta Q(t) \quad (4)$$

其中(3)式为数据积累的动态迭代约束;(4)式为有效数据存量在当期产生的收益。根据采集方的目标函数及其对应的约束条件,可以得到对应的贝尔曼方程(5)式:

$$V^{PR}(Q(t-1)) = \max_{Q(t)} \left\{ \frac{1}{2} \beta (Q(t) - \lambda Q(t-1)) - \alpha (Q(t) - \lambda Q(t-1))^2 + \rho V^{PR}(Q(t)) \right\} \quad (5)$$

根据(5)式,可以得到跨期替代关系,即欧拉方程(6)式:

$$\frac{\partial C(Q(t); Q(t-1))}{\partial Q(t)} - \frac{1}{2} \frac{\partial R(Q(t); Q(t-1))}{\partial Q(t)} = \rho \frac{\partial V(Q(t))}{\partial Q(t)} \quad (6)$$

(6)式的经济涵义是,对采集方而言,当多采集边际单位的数据所带来的当期边际收益(或损失)恰好等于下一期的边际损失(或收益)时,当期数据采集规模的增减均不会实现跨期净收益提升,从而达成最优决策。在数据所有权交易的情况下,新采集的边际单位数据在下一期产生的边际收益为0。因此,当期数据采集决策不会影响未来决策,采集方面临的决策问题与时间无关,这是由数据所有权交易是基于“数据增量”而非“数据存量”进行交易的特点所

决定的。求解(5)式,得到定理1。

定理1:在数据所有权交易的情况下,采集方的最优决策是每期采集 $q^{PR*} = \frac{\beta}{4\alpha}$ 单位的数据。采集方未来最优总收益为 $V^{PR*} = \frac{\beta^2}{16\alpha(1-\rho)}$ 。

随着数据价值折旧,有效数据存量的增速将逐渐趋缓。当有效数据存量满足 $Q(t)=Q(t-1)$ 时,有效数据存量将达到稳态值。结合定理1和约束条件(2)式,得到定理2。

定理2:在数据所有权交易的情况下,有效数据存量将达到稳态值 $Q^{PR*} = \frac{\beta}{4\alpha(1-\lambda)}$ 。

由定理2,易知: $\frac{\partial Q^{PR*}}{\partial \lambda} > 0$, 即当数据折旧速度越慢时,采集方的数据采集激励越强,有效数据存量的积累水平越高。

当有效数据存量达到稳态时,社会财富增加值 (social wealth increment) SWI^{PR*} 为当期数据产生的经济价值,扣除当期数据采集成本。

$$SWI^{PR*} = R(Q^{PR*}) - C(q^{PR*}) = \frac{4\beta^2 - \beta^2(1-\lambda)}{16\alpha(1-\lambda)}$$

易知: $\frac{\partial SWI^{PR*}}{\partial \lambda} > 0$, 当数据折旧速度越慢时,数据要素创造的社会财富水平越高。

五、数据使用权交易分析

隐私计算技术为数据的“可用不可见”交易提供了技术可行性,能够保障在数据交易过程中数据所有权不发生转移。本节将证明,数据使用权交易相比于数据所有权交易,在数据采集激励、数据要素积累、社会财富增长等方面均有本质差异。

交易双方在“可用不可见”技术环境下进行数据使用权交易时,仍按照纳什讨价还价规则,分析方支付给采集方的数据购买费用为 $p^{UR}(t)$, 由(7)式决定:

$$R(t) - p^{UR}(t) = \underbrace{0}_{\text{分析方的保留收益}} + \underbrace{\frac{1}{2}(R(t) - 0)}_{\text{双方分配的数据价值}} \quad (7)$$

对比(1)和(7)式,两种数据交易方式的定价规则存在两方面差异。第一,分析方的保留收益不同。当在“可用不可见”的技术环境下进行数据使用权交

易时,分析方只能在当期使用数据,使用期结束后,如果不继续购买数据使用服务,将无法继续进行数据分析,因此分析方在数据交易前的保留收益为0。第二,双方分配的价值基础不同。在数据所有权交易情况下,分析方可以选择自留历史数据,双方仅就新增数据所产生的价值进行讨价还价;而在数据使用权交易的情形下,分析方的分析工作依赖于采集方提供的数据和技术环境支持,因此,数据产生的全部收益必须由双方共享。

与数据所有权交易问题类似,在数据使用权交易下,采集方确定每一期数据采集规模,以最大化长期总收益。对于任意 $t \in \mathbb{N}$, 给定有效数据存量 $Q(t)$, 将采集方从第 t 期至无穷期的未来最优总收益记作 $V^{UR}(Q(t))$, 如(8)式:

$$V^{UR}(Q(t)) = \max_{\{Q(t+1)\}_{t=\tau}^{\infty}} \sum_{i=\tau}^{\infty} \rho^{i-\tau} [p^{UR}(t) - C(t) - FC] \quad (8)$$

根据采集方的目标函数及其对应的约束条件,得到对应的贝尔曼方程(9)式:

$$V^{UR}(Q(t-1)) = \max_{\{Q(t)\}} \left\{ \frac{1}{2} \beta Q(t) - \alpha(Q(t) - \lambda Q(t-1))^2 - FC + \rho V^{UR}(Q(t)) \right\} \quad (9)$$

需注意,与数据所有权交易不同,在数据使用权交易下,采集方新采集的数据会累积到数据存量中,并在下一期为采集方带来相应收益。因此,采集方面临的动态决策是,短暂性牺牲部分当期净收益,用于数据采集投资,以换取未来更多数据存量积累和收益变现。求解(9)式,得到定理3。

定理3:在数据使用权交易的情况下,采集方的最优决策是每期采集 $q^{UR*} = \frac{\beta}{4\alpha(1-\lambda\rho)}$ 单位的数据。采集方未来最优总收益为 $V^{UR}(Q(t)) = \varphi + \phi Q(t)$ 。其中, $\varphi = \frac{\beta^2}{16\alpha(1-\rho\lambda)^2(1-\rho)} - \frac{FC}{1-\rho}$, $\phi = \frac{\beta\lambda}{2(1-\rho\lambda)}$ 。

对比定理3与定理1可知,在数据使用权交易方式下,采集方未来总收益与有效数据存量 $Q(t)$ 存在正向关系。其原因是,在数据使用权交易方式下,数据所有权始终掌握在采集方手中,采集方拥有的有效数据存量越多,未来总收益越大。易知: $\frac{\partial q^{PR*}}{\partial \rho} > 0$,

$\frac{\partial q^{PR*}}{\partial \lambda} > 0$, 即采集方越有耐心,数据折旧速度越慢,采集方越有强的数据采集激励。

类似地,考虑有效数据存储量的稳态水平。当有效数据存量满足 $Q(t) = Q(t-1)$ 时,有效数据存量将达到稳态值。结合定理3和约束条件(3)式,得到定理4。

定理4:在数据使用权交易的情况下,有效数据存量将达到稳态值 $Q^{UR*} = \frac{\beta}{4\alpha(1-\lambda\rho)(1-\lambda)}$ 。

基于定理4,采集方长期总收益为 $V^{UR*} = \frac{\beta^2(\lambda+1-2\lambda\rho)}{16\alpha(1-\lambda\rho)^2(1-\lambda)(1-\rho)} - \frac{FC}{1-\rho}$ 当有效数据存量达到稳态时,用社会财富增加值 SWI^{UR*} 表示数据要素产生的社会净收益。

$$SWI^{UR*} = R(Q^{UR*}) - C(q^{UR*}) - FC = \frac{\beta^2}{4\alpha(1-\rho\lambda)(1-\lambda)} - \frac{\beta^2}{16\alpha(1-\rho\lambda)^2} - FC$$

六、两种数据交易方式的对比

本节将对两种数据交易方式:首先,比较两种交易方式的稳态水平;其次,分析两种交易方式达到稳态前动态路径的异同;最后,考虑数据发生部分泄露情形,将理论结果进行一般化推广。

(一)两种数据交易方式的稳定状态对比

对比定理1和定理3,即两种数据交易方式下有效数据存储量的稳态,可以得到命题1和命题2。

命题1:相比于数据所有权交易,在数据使用权交易的情况下,每期数据采集规模更大,有效数据存储量的稳态水平更高,即 $q^{UR*} > q^{PR*}$, $Q^{UR*} > Q^{PR*}$ 。

命题1表明,在“可用不可见”的安全交易环境下进行数据使用权交易,有助于更加充分地激发数据生产和采集的潜力。易知, $q^{UR*} = \frac{1}{1-\lambda\rho} q^{PR*}$, $Q^{UR*} = \frac{1}{1-\lambda\rho} Q^{PR*}$, 其中, $\frac{1}{1-\lambda\rho} > 1$, 表示在数据使用权交易情形下积累单位数据的长期收益。

命题2:当“可用不可见”的技术投资成本 $FC < \overline{FC}$, 在两种数据交易模式中,采集方倾向于选择数据使用权交易, $V^{UR*} > V^{PR*}$ 。③

命题2表明,当“可用不可见”的技术使用成本较低时,采集方在数据使用权交易情况下的长期收益水平更高。特别地,当“可用不可见”的安全交易环境无需消耗任何成本($FC=0$)时,采集方在数据使用权交易情形下的收益恒大于数据所有权交易的情形下的收益。

进一步,对比两种交易方式下社会财富增加值,得到命题3。

命题3:当“可用不可见”的技术投资成本 $FC < \hat{FC}$ 时,在两种数据交易方式中,数据使用权交易更有益于社会财富增长,即 $SWI^{UR*} > SWI^{PR*}$ 。^④

图2展示了命题2和命题3的经济涵义。当 $FC > \hat{FC}$ 时,“可用不可见”安全交易的技术投资成本

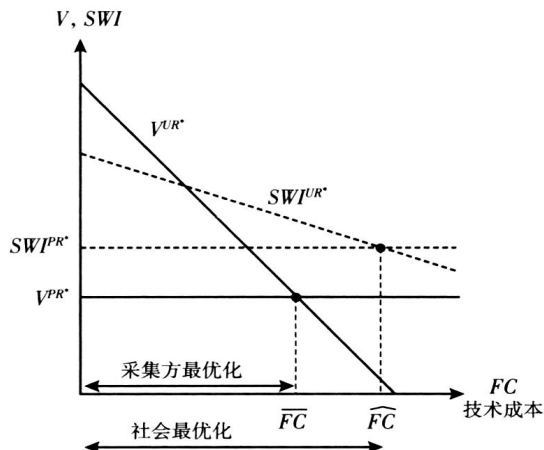


图2 技术进步与数据交易方式选择

过高,无论从社会角度,还是采集方角度,数据所有权交易均为最优选择;当 $\overline{FC} < FC < \hat{FC}$ 时,对社会而言,数据使用权交易更优,但采集方从自身利益出发,“可用不可见”安全交易的技术投资成本负担仍然过高,数据所有权交易是最优选择;只有当 $FC < \overline{FC}$,即随着“可用不可见”安全交易技术不断进步成熟,使得技术投资成本足够低时,采集方才有足够激励采用数据使用权交易方式。

综上所述,当“可用不可见”安全交易环境的技术投资成本较高,且由私人部门承担时,市场主体可能会因激励不相容,而放弃对社会整体而言更优的数据交易方式。为此,从提高数据要素配置效率角度,应鼓励数据交易中心加强数据交易基础设施的支撑技术创新,发挥数字新基建的引领和赋能作用,由此降低中小数据采集企业的技术使用的成本门槛,更加充分地释放市场主体的数据采集激励,促进数据要素实现更高水平、更高效率的市场化配置。

(二)两种数据交易方式的动态路径对比

图3展示了两种数据交易方式下数据要素积累及其所带来的社会财富增长的动态路径。^⑤从数据要素积累动态看,相比于数据所有权交易,数据使用权交易的数据积累速度更快,且达到有效数据存量稳态水平更高,这与命题1的结论是一致的。从社会财富增长看,尽管在短期内数据所有权交易的社会

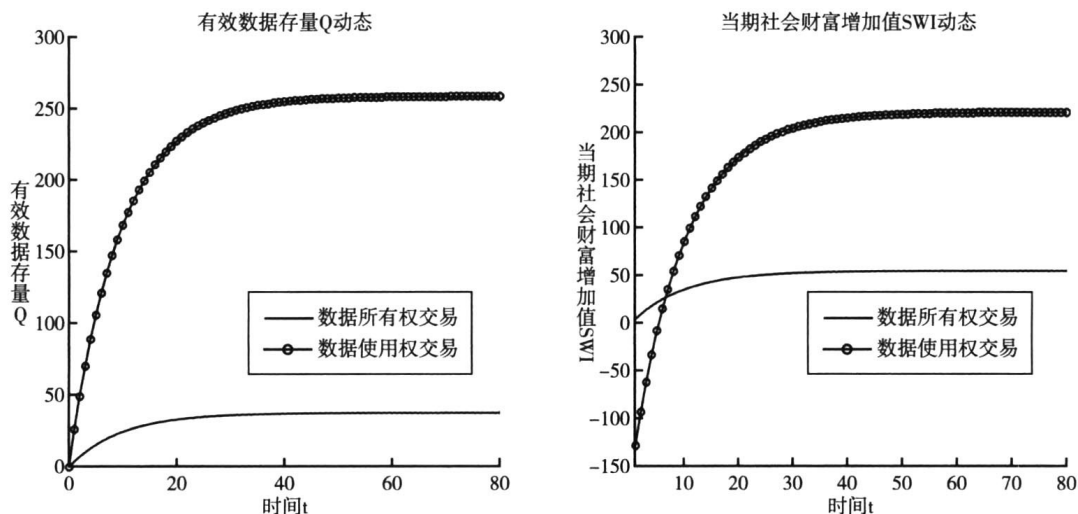


图3 数据积累与社会财富增长路径对比

财富增长水平略高,但从长远来看,数据使用权交易对社会财富增长更为有益。在“可用不可见”的安全交易环境下进行数据所有权交易时,数据交易不会转移数据所有权,数据所有权仍然归属于采集方,这意味着历史积累的数据资源能够持续性为采集方带来收益。因此,采集方将更加关注数据要素的长期积累,而非短期的成本和收益。采集方出于长远利益考虑,会更倾向于扩大数据采集规模(尽管这可能会加重短期的数据采集成本负担),快速积累较多数据,从而在未来实现更加可观的收益。从社会整体财富角度,数据要素积累水平越高,越有助于发挥数据要素的经济价值,因此,数据使用权交易是一种更符合长远社会利益的交易制度安排。

进一步,对两种数据交易方式下社会财富增长路径进行比较静态分析。我们重点关注两个外生参数,即数据“可用不可见”的安全交易技术投资成本FC,以及数据价值系数 β (见图4)。数值分析结果表明:一方面,随着数据“可用不可见”的安全交易技术投资成本FC不断降低,对比数据所有权交易和数据使用权交易下的社会财富增长,前者追赶后者的速度越来越快,且前者最终达到的社会福利稳态水平不断提高。另一方面,当数据价值系数 β 逐渐提高时,两种数据交易方式的社会财富增长均有一定提升,但数据使用权交易情形下的社会财富增长幅度

更大。根据上述分析可知,伴随数据要素在经济发展中的价值不断凸显,隐私计算等安全交易技术日趋成熟而成本下降,数据使用权交易模式更有助于充分释放数据价值,促进社会财富增长。

(三)进一步分析:考虑数据泄露冲击的情形

在基准模型,我们假设“可用不可见”的安全交易环境能够完好地保护数据所有权归属于采集方,这一假设是比较理想化的。一种更现实的情形是,尽管采集方投入了相关技术成本,但仍无法完全杜绝数据泄露的可能性。为刻画数据泄露的情形,我们考虑采集方和分析方在 $t=\{1,2,\dots,\tau-1\}$ 期内进行数据使用权交易,但从第 τ 期开始出现 θ 比例的数据泄露冲击,且冲击将一直延续至无穷期。与基准模型分析的思路类似,我们得到定理5。

定理5:当数据发生部分泄露的情况下,采集方的最优决策是每期 $q^{DL*}=\frac{\beta(1-\theta\lambda\rho)}{4\alpha(1-\rho\lambda)}$ 单位的数据。采集方

未来总收益为 $V^{DL}(Q(t))=\mu+vQ(t)$ 。其中, $\mu=-\frac{\alpha}{1-\rho}q^{DL*2}+\frac{\beta(1-\theta\lambda\rho)}{2(1-\lambda\rho)(1-\rho)}q^{DL*}-\frac{FC}{1-\rho}$, $v=\frac{\beta\lambda(1-\theta)}{2(1-\rho\lambda)}$ 。

易知: $q^{DL*}=(1-\theta)q^{UR*}+\theta q^{PR*}$, $Q^{DL*}=(1-\theta)Q^{UR*}+\theta Q^{PR*}$ 。这意味着,数据泄露将对数据要素积累带来不利影响。当采集方预期未来将持续存在数据泄露时,数据采集激励会受损,导致新采集的增量数据难抵数

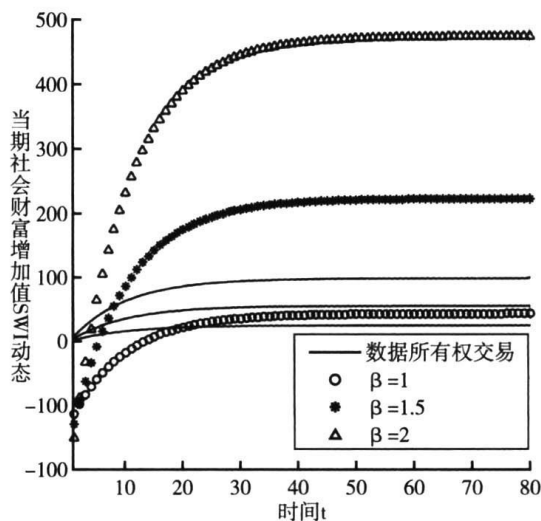
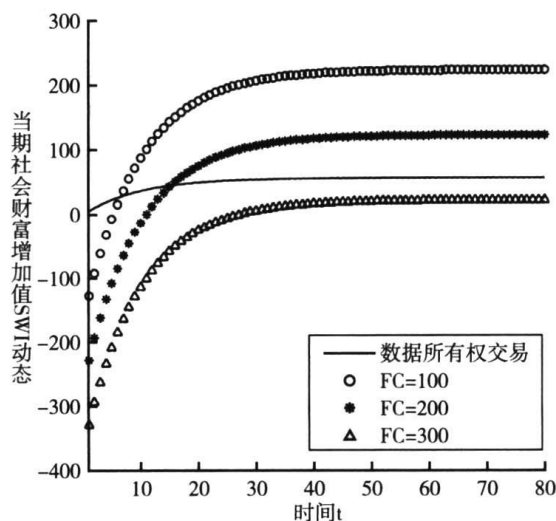


图4 社会财富增长路径的比较静态分析

据存量的折旧消耗,有效数据存量逐渐降低,并收敛至低稳态水平。

关于数据泄露情形的探讨,再次突出了数据安全对于数字经济发展的意义。数字技术迭代速度快,数据安全风险变化频,这些因素都使得保障数据安全是一个需要长期投入且动态调整的目标。为此,数据交易基础设施建设也应秉持“常抓不懈、小步快走”原则,以动态维持相关技术的前沿性和有效性,切实为数据交易安全保驾护航。同时,要强化法律法规的指引和规制作用,对危害数据安全的非法分子严格依法惩治,为数据交易市场的安全、可控、高质量发展提供制度性保障。

七、总结与政策启示

本文基于不完全契约理论,对数据交易的机制设计问题进行系统理论分析。研究发现,数据的“可用不可见”是突破“数据流通”与“数据安全”悖论的关键。由于数据具有非竞争性,如果进行数据所有权交易,将导致数据所有权同时归属于交易双方,这会降低数据采集方的激励,不利于数据交易市场的可持续发展。隐私计算技术为实现数据交易“可用不可见”提供了技术方案。“可用不可见”不仅是数据交易技术的变革,更是交易模式的创新,即从传统的“数据所有权交易”转变为“数据使用权交易”。相比于数据所有权交易,数据使用权交易能够更有效地提升数据采集激励,实现数据要素的高水平积累,充分发挥数据价值,推动数字经济的高质量发展。^⑥

基于本文理论分析,我们可以得到以下三方面的政策启示:

(一)数据安全不仅不与数据流通相悖,反而是数据流通价值释放的关键

数据安全事关国家安全大局。守住数据安全底线,是在数字时代牢牢把握总体国家安全观的应有之义。纵观党中央出台的与数据要素相关的政策文件,无一不将“数据安全”摆在首要位置,并贯穿于数据要素市场化的全过程和各领域。自大数据战略实施以来,国家已相继出台《国家安全法》《网

络安全法》《数据安全法》《个人信息保护法》等多部法律文件,为数据安全保护构筑起坚实的制度保障。从数据基础制度体系建设看,数据安全治理制度已经走到了数据产权制度、数据要素交易制度、数据要素收益分配制度等其他制度建设的前面。实际上,本文从经济学视角阐释了,为什么要将数据安全法律制度作为数据基础制度体系的“先行制度”。研究表明,只有在保障数据安全的前提下,才能推进数据要素交易流通的行稳致远,这也符合数据要素市场的长期建设目标。反之,如果数据安全得不到有效保障,数据流通既无法“行稳”,即潜在威慑国家和社会安全,也难以“致远”,即数据要素积累与价值实现往往也是短暂的、不可持续的。同时,从技术与法律的匹配性角度,《数据安全法》等法律文件的出台,也有助于进一步促进隐私计算等安全技术落地,通过制度先行、带动技术应用。因此,必须要将数据安全制度置于数据要素制度设计的第一要义,为数字经济实现高质量发展提供先决条件和重要保障。

从市场主体角度,本文的启示意义在于,指出数据安全不仅不与数据流通相悖,反而是提升数据流通价值的关键,从而增强企业保障数据安全合规的主观激励。在国家 and 地方相继出台关于数据安全的一系列法律法规后,尤其是《数据安全法》开始实施以来,数据产业发展进一步规范。但也有一些声音担忧,这是否会给行业带上“紧箍咒”,甚至抑制数据交易市场的活力。本文分析则表明,若市场主体脱离数据安全而“侥幸”进行数据交易,不仅于国家和人民利益“不义”,也于自身长远利益“不利”。本文建议,数据产业相关市场主体应筑牢数据安全合规意识,积极探索应用数据安全技术,做好数据安全评估工作,以数据安全合规能力作为核心竞争力、以提升自身市场价值。此外,对数据密集型行业而言,相关主管部门或行业协会应积极推动数据安全监管与自律指引建设。例如,中国人民银行发布《金融业数据能力建设指引》,明确指出“可用不可见”为金融业数据能力建设五大基本原则之一,为提高金融机构

的数据合规应用能力指明了方向。

(二)加快推进数据交易基础设施建设

隐私计算技术尽管能为数据安全制度落地提供有力的技术支撑,但目前在大规模、高效率的商用化方面仍面临着一系列局限和挑战。具体而言:从安全性角度,当前的隐私计算存在算法协议安全、开发应用安全等亟待解决的问题;从性能角度,隐私计算在理论上要比明文计算付出更大的计算和存储代价,比如同态计算的密文扩张规模可达1到4个数量级。而考虑到隐私计算是一种多方同步计算,性能的瓶颈会出现在最薄弱的环节,即计算或通信资源最受限的参与方将限制整个计算平台的性能(王思源和闫树,2022;周午凡等,2022)。^⑩基于这些技术瓶颈因素,目前只有头部企业能够负担起隐私计算技术的投资成本,而大部分中小企业只能望而却步。此外,企业为了满足数据安全合规评估要求,也需要负担极高的成本。有鉴于此,本文指出,囿于隐私计算技术的产业化成熟度与投资成本,数据采集企业在短期内未必有激励选择“可用不可见”的数据交易方式。

为实现数据交易市场高质量建设和发展,建议政府部门在国家和区域层面适度超前布局数据交易基础设施,统筹规划、集约建设、开放共享一批数据安全技术公共服务平台。一方面,从技术攻关角度,政府可以充分发挥“集中力量办大事”的优势,聚集财力和智力资源,加快突破隐私计算技术的“卡脖子”问题,提升技术的产业化成熟度,降成本、提效率。另一方面,从市场培育角度,政府可以依托现有数据交易所,以公共服务平台的形式,将数据安全技术开放给中小数据企业使用,让更多中小企业在合规的条件下“活下来”,发挥数据交易的创新引领作用,培育更多数字经济领域专精特新“小巨人”企业。目前,包括北京国际大数据交易所、上海数据交易所在内的多家数据交易所,都提出要建设以隐私计算、区块链等技术为底层支撑的数据交易基础设施,这对于推动数据要素市场高质量发展是非常好的开端。

(三)数据交易需要考虑“场景专用性”

本文基于数据的“场景专用性”概念,阐释了为何很多数据交易所会采用协商定价模式。理论分析表明,数据具有“场景专用性”,数据在不同应用场景下的价值不尽相同,数据价格的确定离不开具体应用场景。数据买方相比于数据卖方,通常能够掌握更多关于数据应用场景的信息,因而更加清楚数据价值。在数据卖方事前不确定数据价值的情况下,如果在交易前确定数据价格,那么可能会出现的情形是:分析方为了自己利益最大化,就会利用其信息优势,采取“钓鱼策略”,也即故意诱导采集方相信数据价值很低,以此来压低数据价格,分析方在获得数据后,再将数据的高价值变现,这会对采集方非常不利。因此,对采集方更有利的做法是,待分析方产生收益后进行分成,如此能够确保分析方为实现自己利益最大化而将数据高价值变现,达成最有效率且公允的数据定价。^⑪

从现实观察看,很多数据交易所已经认识到应用场景对于数据交易的重要性,并探索基于应用场景的数据交易模式。例如,上海数据交易所明确表示“不合规不挂牌,无场景不交易”,要求使用方在申请数据产品时,必须明确使用场景并签订相关合约,一旦出现滥用数据的情况将按照合约接受处罚。北京国际大数据交易所提出“数据可用不可见,用途可控可计量”的交易范式,其中强调“用途可控”也有明晰数据应用场景的类似思想。基于本文理论,这些实践不仅是为了保障数据在可控安全范围内使用,实际上也是探索数据要素价格形成机制的有益尝试。

进一步,基于数据的“场景专用性”,有效市场和有为政府在发挥数据要素价值方面显得尤为重要。正因为数据只有找到合适的场景才能充分释放其价值,才需要发挥有效市场的资源配置优势,尽可能多而广地探索数据应用场景。不过,资本的逐利性可能也会使数据被用于不当用途,例如大数据诈骗,危害国计民生。因此,同样需要“有为政府”发挥好作用,在数据要素的应用场景方面加强引导和规范,防

范数据的无序滥用,让数据要素在“国家所需、地方所能、群众所盼、未来所向”的应用领域充分释放其价值。

作者感谢匿名审稿专家的宝贵建议。当然,文责自负。

注释:

①习近平总书记在主持中共中央政治局第二次集体学习“实施国家大数据战略加快建设数字中国”时强调。

②其中,贵州、天津、海南、山西、吉林、安徽、山东、福建、黑龙江和辽宁出台了大数据条例,深圳、上海、重庆和浙江出台了数据条例,四川、广西、江西、河南等地公布了相关数据条例草案。资料来源:《18省市公布〈数据条例〉》,澎湃新闻。

③根据信通院《数据价值化与数据要素市场发展报告(2021年)》的概念界定,“数据资源”并不直接等同于“数据要素”。数据资源与数据要素的区别在于数据资源是否产生了经济效益。

④资料来源:中国信通院《数据价值化与数据要素市场发展报告(2021年)》。

⑤资料来源:《理想很丰满 现实很骨感 贵阳大数据交易所这六年》,证券日报网。

⑥资料来源:《“东数西算”,寻找算力的西部支点》,《科技日报》;《贵州首发全新数据交易规则体系,解决数据交易痛点难点问题》,北青网。

⑦由数据价值的不确定性引发的数据交易契约的不完全性,与Arrow(1962)曾提出的“信息悖论”有一定类似之处,熊巧琴和汤珂(2021)对这一问题进行了探讨。熊巧琴和汤珂(2021)指出,“买方如果交易前不了解该数据资产的详细信息,会较难明确该数据能带来的效用价值;但如果卖方了解数据的全部信息,购买该数据对卖方的价值降低。”实际上,本文讨论的“数据安全”与“数据流通”之间的悖论,与上述所说的“信息悖论”从经济学本质上有一定相似性。本文是从现实问题出发,将“信息悖论”进一步具像化凝练为“数据安全”与“数据流通”之间的悖论,并从不完全契约视角对这一问题进行了理论分析。

⑧ Azcoitia & Laoutari(2022)将数据交易中中介分为三类。第一类是数据供应商。数据供应商主要出售来自互联网

或独家采集的数据,大多数供应商通过API或批量下载的方式将数据提供给买方客户。第二类数据交易模式是数据市场。数据市场与数据供应商最大不同之处在于,它是一个双边市场,主要负责搭建平台,以此连接数据供应商和买方客户。数据主要通过API或批量下载的方式进行交付,买卖双方均可在平台上对数据交易进行管理。第三类是个人数据管理系统,定位于服务个人用户的数据管理与交易。个人数据管理系统能够搜集、管理用户数据,并允许用户将个人数据授权交易。关于各类数据交易中中介模式的代表性平台或机构,感兴趣的读者可以参考Azcoitia & Laoutari(2022)的表4,本文受篇幅限制,不再一一列举。

⑨吴绪亮:《制度设计是数据要素定价的关键》,《经济日报》2020年10月23日。

⑩李芃达:《隐私保护计算让数据资产增值》,《经济日报》2021年9月7日。

⑪目前国内数据交易所普遍采用的是协商定价模式,这是本文关于数据定价方式假设的现实基础。例如,根据中国信通院发布《数据价值化与数据要素市场发展报告(2021年)》的表1显示,国内绝大多数的数据交易所均采用协商定价模式。据作者了解,协商定价模式被国内数据交易所广泛采用的一个重要原因是,数据能够产生多少价值与其应用的具体场景是密切相关的,同一数据在不同场景下数据的价值往往不尽相同,交易所很难做到统一定价。为此,数据交易所通常会针对不同的应用场景与个性化需求,与客户协商定价。不过,也需指出,国外部分数据交易所也会采用其他数据定价模式,包括固定价格、按量定价、按使用定价、拍卖等(Azcoitia & Laoutaris, 2022)。为了突出中国数据交易市场的问题,本文不再一一讨论其他数据定价模式。

⑫这是一个略严格但放松后不影响主要结论的假设。若假设采集方也具备一定的场景应用分析能力,并不影响双方数据交易的动机。因为采集方既可以自己使用数据,也可以把数据卖给分析方,这是数据的非竞争性使然。不过,若做此假设,还要继续讨论采集方与分析方在产品市场中的竞争问题,这是后续很有价值的研究问题,但本文为了突出核心问题,不再展开讨论。

$$\textcircled{13} \text{其中, } \overline{FC} = \frac{\beta^2(\lambda\rho(1-\lambda)(2-\lambda\rho) + 2\lambda(1-\rho))}{16\alpha(1-\lambda\rho)^2(1-\lambda)}.$$

$$\textcircled{14} \text{其中, } \hat{FC} = \frac{\beta^2\lambda\rho(4(1-\rho\lambda) - (1-\lambda)(2-\lambda\rho))}{16\alpha(1-\rho\lambda)^2(1-\lambda)}.$$

$$\textcircled{15} \text{参数设定: } \rho=0.95, \lambda=0.9, \alpha=0.1, \beta=1.5, FC=100.$$

⑩本文也为后续研究探讨数据交易市场竞争、数据要素收益分配制度等重要问题提供了一般性的分析框架。在数据交易市场竞争方面,后续研究可以进一步考虑数据产业链下游存在多个数据分析方。一家下游企业购得数据后可以选择转售收据,变成第二家数据卖方,与既有的上游企业形成竞争关系,并基于此探讨不同的数据交易模式对数据交易市场竞争的影响。在数据要素收益分配方面,本文目前暂未考虑数据采集和分析所需要的劳动投入,后续研究可以进一步将人力资本纳入数据要素收益分配框架,考虑家庭部门劳动力在不同数据产业部门的分配,并进一步探讨数据要素报酬分配的制度设计。

⑪实际上,除了投资成本外,隐私计算技术目前仍面临着一些非成本方面的瓶颈。例如,在生态瓶颈方面,隐私计算涉及的隐私保护技术和算法非常多,且算法的复杂度、性能、优势场景等各不相同,难以形成统一的算法安全基础,有待建立涵盖主流隐私计算产品的系统性安全分级标准;在合规瓶颈方面,隐私计算技术的使用仍需面对数据来源的合法性问题,如果在隐私计算技术启用的起点,数据提供方的数据来源存在瑕疵甚至不法,那么后续计算的过程、计算的结果都将面临侵权的指责和责任的承担。因此,尽管从理论上隐私计算可以为平衡数据流通和安全之间的矛盾提供技术支撑,但也要客观认识到,隐私计算落地过程中还需要更好的工具,更好的解释安全性,以及生态能力更加结合业务场景。当下的隐私计算相关技术尚处于萌芽与上升阶段,最终真正成熟实现可用的技术状态还需时日,但未来可期。

⑫现实中也有一些数据交易所采用了事前定价模式,但这和本文并不冲突。因为在那些事前定价的情形下,买卖双方关于数据价值的信息不对称程度较低,数据买方很难利用信息不对称对卖方实施“钓鱼策略”。

参考文献:

- [1]蔡继明、刘媛、高宏、陈臣,2022:《数据要素参与价值创造的途径——基于广义价值论的一般均衡分析》,《管理世界》第7期。
- [2]蔡跃洲、马文君,2021:《数据要素对高质量发展影响与数据流动制约》,《数量经济技术经济研究》第3期。
- [3]蔡跃洲、牛新星,2021:《中国数字经济增加值规模测算及结构分析》,《中国社会科学》第11期。
- [4]陈剑、刘文辉,2021:《数智化使能运营管理变革:从供

应链到供应链生态系统》,《管理世界》第11期。

- [5]李三希、武珂璠、鲍仁杰,2021:《大数据、个人信息保护和价格歧视——基于垂直差异化双寡头模型的分析》,《经济研究》第1期。
- [6]林志杰、孟政炫,2021:《数据生产要素的结合机制——互补性资产视角》,《北京交通大学学报(社会科学版)》第2期。
- [7]刘意、谢康、邓弘林,2020:《数据驱动的产品研发转型:组织惯例适应性变革视角的案例研究》,《管理世界》第3期。
- [8]申卫星,2020:《论数据用益权》,《中国社会科学》第11期。
- [9]孙新波、钱雨、张明超、李金柱,2019:《大数据驱动企业供应链敏捷性的实现机理研究》,《管理世界》第9期。
- [10]汤琪,2016:《大数据交易中的产权问题研究》,《图书与情报》第4期。
- [11]王思源、闫树,2022:《隐私计算面临的挑战与发展趋势浅析》,《通信世界》第2期。
- [12]王永钦、董雯,2020:《机器人的兴起如何影响中国劳动力市场?——来自制造业上市公司的证据》,《经济研究》第10期。
- [13]肖静华、胡杨颂、吴瑶,2020:《成长品:数据驱动的企业与用户互动创新案例研究》,《管理世界》第3期。
- [14]肖旭、戚聿东,2021:《数据要素的价值属性》,《经济与管理研究》第7期。
- [15]谢康、吴瑶、肖静华,2020a:《数据驱动的组织结构适应性创新——数字经济的创新逻辑(三)》,《北京交通大学学报(社会科学版)》第3期。
- [16]谢康、夏正豪、肖静华,2020b:《大数据成为现实生产要素的企业实现机制:产品创新视角》,《中国工业经济》第5期。
- [17]熊巧琴、汤珂,2021:《数据要素的界权、交易和定价研究进展》,《经济学动态》第2期。
- [18]徐翔、厉克奥博、田晓轩,2021:《数据生产要素研究进展》,《经济学动态》第4期。
- [19]徐翔、赵墨非,2020:《数据资本与经济增长路径》,《经济研究》第10期。
- [20]张明超、孙新波、钱雨,2021:《数据赋能驱动智能制造企业C2M反向定制模式创新实现机理》,《管理学报》第8期。
- [21]周广肃、李力行、孟岭生,2021:《智能化对中国劳动力市场的影响——基于就业广度和强度的分析》,《金融研究》第6期。
- [22]周午凡、董宏伟、张丽霞,2022:《国外隐私计算最新进

展及对我国的启示》，《通信世界》第5期。

[23]Abowd, J. M., and I. M. Schmutte, 2019, "An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices", *American Economic Review*, 109(1), 171–202.

[24]Acemoglu, D., M. Ali, M. Azarakhsh, and A. Ozdaglar, 2022, "Too Much Data: Prices and Inefficiencies in Data Markets", *American Economic Journal: Microeconomics*, forthcoming.

[25]Aghion, P., F. J. Benjamin, and I. J. Charles, 2018, "Artificial Intelligence and Economic Growth", NBER Working Paper w. 23928.

[26]Akcigit, U., M. A. Celik, and J. Greenwood, 2016, "Buy, Keep or Sell: Economic Growth and the Market for Ideas", *Econometrica*, 84(3), 943–984.

[27]Ali, S. N., G. Lewis, and S. Vasserman, 2012, "Voluntary Disclosure and Personalized Pricing", *Review of Economic Studies*, forthcoming.

[28]Arrow, K., 1962, "Economic Welfare and the Allocation of Resources for Invention", NBER Chapters in "The Rate and Direction of Inventive Activity: Economic and Social Factors", 609–626.

[29]Azcoitia, S. A., and N. Laoutaris, 2022, "A Survey of Data Marketplaces and Their Business Models", Working Paper.

[30]Bajari, P., V. Chernozhukov, A. Hortacsu, and J. Suzuki, 2019, "The Impact of Big Data on Firm Performance: An Empirical Investigation", *American Economic Association Papers and Proceedings*, 109(5), 33–77.

[31]Beraja, M., D. Y. Yang, and N. Yuchtman, 2022, "Data-intensive Innovation and the State: Evidence from AI Firms in China", *Review of Economic Studies*, forthcoming.

[32]Bergemann, D., and A. Bonatti, 2015, "Selling Cookies", *American Economic Journal: Microeconomics*, 7(3), 259–294.

[33]Bergemann, D., A. Bonatti, and A. Smolin, 2018, "The Design and Price of Information", *American Economic Review*, 108(1), 1–48.

[34]Bergemann, D., A. Bonatti, and T. Gan, 2019, "The Economics of Social Data", *RAND Journal of Economics*, 53(2), 263–296.

[35]Brynjolfsson, E., and K. McElheran, 2016, "The Rapid Adoption of Data-driven Decision-making", *American Economic Review*, 106(5), 133–139.

[36]Chiou, L., and C. Tucker, 2017, "Search Engines and Data Retention: Implications for Privacy and Antitrust", NBER Working Paper, w. 23815.

[37]Choi, J., D. Jeon, and B. Kim, 2019, "Privacy and Personal Data Collection with Information Externalities", *Journal of Public Economics*, 173, 113–124.

[38]Cong, L. W., D. Xie, and L. Zhang, 2021, "Knowledge Accumulation, Privacy, and Growth in a Data Economy", *Management Science*, 67(10), 6480–6492.

[39]Dosis, A., and W. Sand-Zantman, 2022, "The Ownership of Data", *Journal of Law, Economics and Organization*, forthcoming.

[40]Farboodi, M., and L. Veldkamp, 2021, "A Growth Model of the Data Economy", NBER Working Paper, w. 28427.

[41]Farboodi, M., R. Mihet, T. Philippon, and L. Veldkamp, 2019, "Big Data and Firm Dynamics", *American Economic Association Papers and Proceedings*, 109, 38–42.

[42]Grossman, S. J., and O. D. Hart, 1986, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration", *Journal of Political Economy*, 94, 691–719.

[43]Ichihashi, S., 2020, "Online Privacy and Information Disclosure by Consumers", *American Economic Review*, 110(2), 569–595.

[44]Ichihashi, S., 2021, "Competing Data Intermediaries", *RAND Journal of Economics*, 53(3), 515–537.

[45]Jones, C. I., and C. Tonetti, 2020, "Nonrivalry and the Economics of Data", *American Economic Review*, 110(9), 2819–2858.