

科学数据共享治理：模式选择与情景分析^{*}

张丽丽

摘 要 本文综合运用文献述评、理论移植、情景分析等方法,将科学数据共享的模式抽象与案例情景再现相结合。基于研究进展与局限性,切入新制度经济学视角,利用数据资产专用性程度与交易频次,从占优的组织形态角度,将科学数据共享归纳为组织内部的数据共享模式、有管制的数据共享模式、中间体形态的数据共享模式、个体间的数据分享模式和市场形态的数据共享模式。结合情境分析,重点选取资源池模式、数据出版模式与数据交易模式三种面向开放服务的主流共享实践进行演绎。围绕共享模式的供给与需求主线,比较三种模式的优劣势、机会和挑战。研究认为,资源池模式是当前及今后相当长时间内最为主要的科学数据共享模式,同时数据出版和数据交易作为重要的模式补充,发展机遇与挑战并存。科学数据共享治理应根据情景与数据属性进行灵活多样的模式选择。表 5。参考文献 34。

关键词 科学数据共享 共享模式 情景分析 新制度经济学

分类号 G203

Scientific Data Sharing Governance: Model Selection and Scenario Analysis

ZHANG Lili

ABSTRACT

The rough road of data sharing well needs pattern-matching sharing models and operating mechanisms. Based on the current state of scientific data sharing in big data era, this paper applies compound methods of literature review, theoretical transplantation, and scenario analysis. The progress of scientific data sharing model study has been summarized and research limitations have been pointed out. Furthermore, reanalysis from the view of new institutional economics has been suggested and the design of model abstraction combining with real case scenario analysis has been proposed.

For literature review, summarization is carried out by a micro-level perspective regarding research data themselves, a meso-level view focusing on operating organizations together with macro-level insights into institution. Then research data sharing activities and sharing models have been described and models in some specific disciplines as well as in other information resources fields have also been mentioned.

Moreover, theories of new institutional economics have been introduced. Judging from the data asset specificity and data trading frequency, different research data sharing activities tend to select different

^{*} 本文系国家自然科学基金项目“面向管理和决策的大数据资源共享与治理机制”(编号:91546125)的研究成果之一。(This article is an outcome of the project “Decision Making Oriented Massive Data Resources Sharing and Governance” (No.91546125) supported by National Natural Science Foundation of China.)

通信作者:张丽丽,Email:zhll@cnic.cn,ORCID:0000-0003-1847-6683 (Correspondence should be addressed to ZHANG Lili,Email:zhll@cnic.cn,ORCID:0000-0003-1847-6683)

organizational forms. Concerning those preferred organizational arrangements, there are mainly five kinds of data sharing models, including intra-organization level data sharing model, controlled data sharing model, intermediate-organization level data sharing model, individual data exchange model, and free market data sharing model.

Upon the abstraction of five logical models, scenario analysis represents with mainstream data sharing cases including data resources pool model, data publishing model and data market model. Following the guidance of supply-demand chain within data sharing models, stakeholders and their interactions tracing the data flow are reviewed. Meanwhile, comparisons about advantages, disadvantages, opportunities and challenges within the three mainstream sharing models for open data are fully analyzed and the prospect for the future development is also discussed.

Finally, we find out that data resources pool model is the most important way for scientific data sharing and its predominance will prevail for quite a long time in the future. However, due to the externality source of incentives and top-down pattern of data sharing path, data resources pool model can't help the data suppliers exert their subjective initiative to the utmost. Therefore, the cost for supervision and evaluation of such model is high. Besides, the data publishing model successfully facilitates the work of data sharing and helps data suppliers gain their scholarly reputations as well. However, data publishing model still needs further development in many aspects, such as the establishment of scholarly reputation and acknowledgement, sustainability of business model, balance of open data and copyrights protection and etc. In a way, data publishing model may become the real mainstream in the future scientific research community. What's more, the introduction of free market exchange makes the data market model a plus for data sharing model selection. While the hinders of such model include the high-tech challenges from the high specificity of data assets and the difficult negotiation between public funding data production and added value data trading.

Above all, throughout the overall development of scientific data sharing, model study shall continue forward. The mainstream data sharing models are bound to complement one another and jointly promote the scientific research by sharing data within but not limited to those existing models. Meanwhile, the inherent data sharing path dependence reshapes the real practices of ideal data sharing model to some extent. Besides, data sharing is inevitably subject to the existing framework of local culture and customs of research community. The reforms of data sharing models are more likely to take a gentle way step by step. Facing the current data sharing model arrangement, if the dominant data sharing model namely the data resources pool model could absorb more from others according to particular data attributes, open data can make more progress. 5 tabs. 34 refs.

KEY WORDS

Scientific data sharing. Sharing model. Scenario analysis. New institutional economics.

0 引言

“如果数字科技是现代科学变革的引擎,那么数据就是燃料。然而事实上,很多科研领域

燃料短缺”^[1]。数据匮乏现象与如火如荼的大数据时代相悖,究其原因,数据共享短板是主要障碍之一。捉襟见肘的数据管理能力和预算、强烈的数据产权意识与多元化的数据需求相互制衡,充满荆棘的数据共享之路迫切需要与之

匹配的模式选择与运行发展机制。立足大数据时代的科学数据共享现状,本文在总结现有共享模式实践的基础上,引入新制度经济学视角,将有效率的共享模式选择与情景分析相结合,期望再现研究活动中科学数据共享诸多模式的运行情况,并总结提出未来的治理思路。

1 科学数据共享的发展与模式梳理

所谓共享模式是科学数据共享活动中逐步形成并适应当前共享工作需要的普遍范式。不同的共享模式意味着不同的驱动机制与权责关系,以及不同的管理方式与质量绩效等。作为复杂的系统性问题,识别科学数据的特征,并选择与之匹配的数据共享模式有助于推进数据共享复杂问题的分层化解。

1.1 科学数据共享一般模式

不同的分类方法带给我们纷繁多样的科学数据共享模式。微观层面上围绕科学数据对象,中观层面关注共享运行组织,宏观层面探讨共享制度,我们可将现有科学数据共享的模式实践梳理如下。

首先,从科学数据来源看,共享可分为“专业化生产模式、联合共建共享模式和数据交换模式”^[2]。从技术应用角度出发,有学者提出基于完全开放数据库的共享、基于查询接口的共享、基于元数据的共享和基于开放网格服务架构(OGSA-DAI)数据集成的共享^[3]。从科学数据的权益性质出发,共享模式可分为“公益性和商业化”^[4]两种。从数据服务场景出发的模式实践更加丰富:①大科学装置为核心的高通量数据共享模式,如来自欧洲粒子中心(CERN)的多国科学家利用强大的共享协作网络,年平均数据交互量超过30PB^[5];②围绕大科学计划的联合研究项目共享模式,例如超级国际计划之一“地球环境领域的国际地球生物圈计划(IGBP)”拥有70个国家委员会代表的多领域研究项目^[6];③数据中心(存储库)为代表的服务机

构共享模式,如DRYAD、FIGSHARE等开放平台所提供的数据管理与开放服务;④Nature Scientific Data、PANGAEA等通过出版形式提供的数据共享模式。

其次,从组织形态划分,科学数据共享模式包括“国家政策驱动、部门之间交换、企业发展带动以及国际组织参与”四类^[7],或者“国际组织协作共建、政策驱动共建、主体(项目)合作共建和地域协作共建”四大模式^[8],或者事业性运行模式和商业化运行模式两种类型^[9]。

宏观层面,共享模式的形成还得益于共享制度如政策运动、标准规范、协议工具等的发展壮大。在政策推动的共享中,开放获取倡议提出了金色道路和绿色道路两种共享模式^[10];从国际组织到国家政府机构,一系列开放共享政策倡导全面开放的公益性共享模式,如经济合作与发展组织(OECD)^[11]强调开放数据与可持续性原则,国际科学理事会——世界数据系统(ICSU-WDS)^[12]以及全球地学观测系统(GEO)^[13]等,倡导不同层面(如数据、元数据、数据产品等)“全面开放”,《G8开放数据宪章和技术附件》^[14]提出为提升政府服务能力而面向公众开放数据的要求。2005年以后陆续兴起的国家政府部门开放数据门户建设(如data.gov)也成为国家层面助推科研数据开放的中坚力量。在以标准规范为主导的数据共享活动中,共享模式则多依附于行业领域实践而存在,例如我国的地震科学数据^[15]、国家食品药品数据(CFDAB/T-0303-2013)、民航科学数据(MH/T 0029-2009)等实践。以协议工具为驱动的数据共享模式则在国际范围内达成更广泛的共识,例如从保护原创作品权益扩展而来的“知识共享”协议(Creative Common),以保障重用为根本目标的定位标识工具“数字对象标识符(DOI)”等,皆成为推动数据共享的模式范例。

当然,所谓的模式分类并非一成不变,各种分类之间也存在着交叉重叠,共享实践常常是多种模式运行自洽的综合体系。例如,典型的服务场景中,大科学计划与大科学装置总是联

手推动数据全生命周期的共享活动;大型跨国数据组织也并不会局限于单一模式,交叉多元的组织形态更为常见。制度导向的宏观共享模式就更加需要一系列不同层次的法律、政策、标准规范与技术工具协调一致,同时也针对特定科学数据呈现出特定的组织形态,等等。

1.2 特定领域的共享模式

在不同的科学领域,数据的规模、算法模型与处理方式等存在显著差异,共享模式的选择也有很多特色化表现。表 1 收集了一些面向特定领域的科学数据共享模式的具体案例,这些案例为数据共享模式赋予了多样化的实践内涵。

表 1 科学数据共享实例

共享标的	模式选择	模式评价
气象科学数据	后台管理模式 ^[16]	便利数据发布,具有时效性、稳定性、安全性
公共卫生科学数据	权限管理模式 ^[17]	基于需求分析与角色分类,有效地解决相关数据对权限管理与最大限度数据分享的需要
基因和临床数据	GA4GH 模式 ^[18,19]	保障全球基因数据安全、尊重隐私与互惠分享
欧盟区个人隐私数据 ^[18]	充裕模式 安全港模式 有约束力的企业规则模式 标准合同模式 问责制模式 第三方认证模式	不同政策法规框架所支持的运行模式不能普遍适用,但多边制度安排、规模化的多角色参与、敏感数据的先期(安全)保证、违反要求的配套罚则与源自数据用户自身的责任意识等值得关注
地球系统科学数据	地理网络共享模式 ^[20]	重点关注面向服务的政府地理信息统一组织和管理模式,为松耦合集成化应用提供可供参考的积极案例
	设施即服务(IaaS)、数据资源即服务(DaaS)、数据功能即服务(SaaS)共享模式 ^[21]	倡导创建“人人都是数据的提供者,人人都是数据的使用者”的数据共享服务环境
野生动植物数据	英国国家生物多样性网络 ^[22]	该模式的规范框架包括引言、目标、数据资源、数据使用、数据分享、条款和条件、其他信息、建议和解释说明八部分,具有规范性、完整性和可操作性

1.3 其他领域的共享模式

如果我们将视角扩展到信息共享的大环境,不难发现,有关政府、企业等多种组织形式下的信息共享模式选择会给我们带来启发。例如,针对政府领域,杨兴凯^[23]从系统架构方式、

数据组织方式与政府部门关系等角度,总结提炼出不同的信息共享模式分类(见表 2)。这种分类外在直观,不失为厘清政府信息共享运行问题的一种模式解决方案。

表 2 政府跨部门信息共享模式^[23]

角度	模式分类
基于系统架构方式	点对点信息共享模式、星形信息共享模式、总线形信息共享模式
基于数据组织方式	集中式信息共享模式、分布式信息共享模式
基于政府部门关系	垂直信息共享模式、水平信息共享模式、交叉信息共享模式

姚玲杰^[24]等面向知识共享提出单向式和互动式两类,认为单向式共享通过知识源对知识接受者知识发送或接受者主动获取而最终实现知识转移;互动式共享则以及及时性、互动性见称,通

过供求双方互动而成为一种有效的交流实践方式。周永红^[25]等提出,围绕共享意愿、共享层次、共享范围、共享机制等不同角度,企业的知识共享模式具有多样化的选择分类(见表3)。

表3 企业知识共享模式^[25]

角度	模式分类
基于共享意愿	自愿型的知识共享模式、强制型的知识共享模式
基于共享层次	同一价值链层次上的知识共享模式、上下游价值链层次上的知识共享模式
基于共享范围	联盟伙伴对称型知识共享模式、盟主与其他成员企业非对称型知识共享模式
基于共享机制	市场机制主导的知识共享模式(市场机制主导的联盟内部知识共享、市场机制主导的联盟外部知识共享)、非市场机制主导的知识共享模式(企业联盟内部的知识免费开放共享、政府力量引导知识共享、社会力量推动知识共享)

1.4 模式研究小结

通过梳理不难发现,依形成路径划分,科学数据共享模式的归纳存在两种不同层次的操作。对于具体化模式,如表1所提供的多数例子,共享模式与应用情景联系更为紧密,这些模式生发于一定的基础实践,并在一段时间内逐步演化成熟,呈现出模式化的特征,进而成为值得推广的共享模式。而模式形成的另一种路径则是众多实践案例的高度抽象化,如表2、表3所进行的阐释。两种层次的共享模式各具意义:前一种与共享情景结合更紧密,生动灵活而稳定性不足,另一种则更像是开发语言中的“类”,具有更高的概括性指导意义,但具体应用常常与前一类实践相结合。

然而,仅仅归纳出模式研究发展现状与表征仍显不足。表1示例众多,却不能帮助我们在全局上把握科学数据共享模式的运行机理及相互关系。而表2、表3所示的分类方法与1.1节的阐释类似,它们都面临着数据共享交叉融合的现实态势。当然,理想的模式分类并非划清界限就束之高阁,共享模式的分类应该能够充分考虑共享对象特质,形成完整的概括性分类,并能对当前共享实践进行合理解释,对未来发展有所启发。

有鉴于此,再次回到我们的研究对象——科学数据共享。这是一个有意思的话题,一个

数据集是否会共享,共享边界在哪里,共享质量又如何?对于同一数据生产者而言,面对不同场景与治理形态,答案也总存在微妙变化。由此,以下对于科学数据共享模式的分析将模式抽象与模式情景相结合,以诠释科学数据共享的运行机理。

2 有效率的模式选择与治理

关于新制度经济学分析方法的引入对于研究信息共享制度问题的适用性前提,笔者在相关论文中已做过详细阐释^[26]。在此基础上,该理论体系对于一类特定形态的信息——科学数据同样具有一定的指导意义。共享活动中,贯穿始末的数据流所进行的融通过程可以抽象成为一种数据资产的交换过程,而模式的选择即作为一种相对条件下“有效率”的制度选择。以下即利用制度经济学的相关理论思想,探讨其对于模式选择与效率权衡的启发。

2.1 新制度经济学思想的引入

新制度经济学代表人物之一威廉姆森^[27]等人认为,利用资产专用性程度与交易频次可以判断资产所适用的组织形态。其中,资产专用性特指资产本身的专用性、资产选址的专用性以及人力资本的专用性等。

对于科学数据共享活动而言,在其他条件不变的前提下,数据资产的专用性程度越高或者交易频次越低(或规模不经济),成本就越高(如需要更新的技术、更高水平的人才、更苛刻的管理条件等),组织形态也更倾向利用“有形之手”的强大力量来实现共享工作,例如国家大科学计划中高度专用性的数据从经费支持、管理运维到公开共享无疑都印刻着“自上而下”的标签。反之,数据资产的专用性程度越低,交易频次越高,意味着来自市场多元力量的介入可能性就越大,例如利用公开数据源提供增值服务的市场单元。此外,介于两者之间的大量数据资产则倾向于采取混合形式的中间组织来

达成共享目标,例如众多国际数据联盟中,各成员是拥有各自利益取向的独立单元,联盟整体则具有相对稳定、统一的共享规则、一定的准入门槛和清晰的边界。

然而,这样的分析尚显单薄,有效率的科学数据共享模式选择,不仅仅意味着恰如其分的科学数据对象表征,也包括完整的模式运行治理。进一步借鉴“比较制度经济学”代表人物青木昌彦有关组织形态选择与治理机制方面的探讨^[28],整理得到科学数据共享治理形态的分析框架:从科学数据的专用性出发,不同性质的数据具有不同的组织形态选择倾向,组织环境又赋予其特定的激励与治理规则(见表4)。

表4 科学数据共享治理形态

数据专用性	组织形态	激励措施 (来源)	实施者	治理依据与规则		共享路径	数据流向	共享情景 (举例)
高专用性 /高交易频率	组织内部	共享网络的 规定(外部)	组织	组织共享沟通 网络的规范(终 止关系,惩罚不 诚实的交易者)		自上而下	环路为主	企业内部 数据部门
高专用性 /低交易频率	政府部门	行政命令 (外部)	司法机关	法制系统(惩治 欺骗行为)		自上而下	单向为主	资源池
混合专用性	中间组织	联盟规范 学术声望 (内外结合)	俱乐部	第三 方治 理	驱逐不 诚实交 易者	纵横联合	单向或环路	数据联盟
		出版规范 学术声望 (内外结合)	出版系统		传播欺 骗行为而失 信	纵横联合	环路为主	数据出版
弱专用性/ 规模不经济	个体间	自我实现 学术声望 (内部)	自我	个人信任,道德 准则(对欺骗 行为的摒弃)		自下而上	双向	个人主页 的数据 分享
弱专用性/ 规模经济	市场	经济收益 预期(内部)	市场交易者	合同(惩罚违 反合同)		自下而上	双向	数据市场

2.2 共享模式及其治理

结合数据专用性特征与占优的组织形态,可将数据共享区分为以下五种情况。

(1) 组织内部的数据共享模式

自上而下的数据共享活动中,微观层面的典型模式即是企业内部的共享网络。这类数据

兼具高专用性和高交易频次特征。企业边界的存在,使高成本投入的数据能同时满足高频次的访问需求,很多企业的战略研发部门即提供这样的内部数据共享服务。此类数据共享活动围绕组织目标而定,激励措施由数据生产者以外的共享网络力量主导。

(2) 有管制的数据共享模式

同样在自上而下的数据共享活动中,宏观层面的代表性模式即政府主导的、有管制的数据共享。作为高专用性、低交易频次数据共享的宏观模式代表,这类数据共享由政府部门通过法律手段与法制系统实现,对共享不当行为采取强有力的惩治手段。这是当前国际上科学数据共享的主流模式之一,由自上而下的力量来推动数据共享的全流程管理工作,管理方具有强势的管理能力与手段。然而,共享活动的激励措施仍然是外部性的,监督与控制成本总是出奇的高。

(3) 中间体形态的数据共享模式

科学研究中的数据并不能仅仅靠自下而上的渠道得以满足,数据的生产者或拥有者仍需要谋求数据共享的机会,甚至需要更多外力推动来促成共享活动。数据联盟以“俱乐部”的小团体形态,打破原有组织边界,使数据在有限人员与组织范围内更高效地交流互换。这在很多领域颇为常见,如基因组研究三大数据库(NCBI、Ensembl和UCSC)所进行的实践。然而这也带来了有关数据垄断的担忧。强有力的组织领导力使“俱乐部”汇集了大量数据资源,组织的声望与影响力、清晰的组织边界与确定的组织规则在一定程度上排斥了组织外的数据集结。数据资产的寡头垄断态势在中间体组织颇为常见,然而一旦其变相成为“有管制的数据共享”,由于其并不同时具备“政府”的角色定位,组织以外的群体势必成为共享弱势方。

在此基础上,数据出版提供了另外一种方式,通过出版社第三方可信平台,为数据作者的数据质量背书,以促进数据的合理、高效利用,同时又通过数据引用评价反哺数据生产者的劳动付出。这些数据共享活动的治理依靠第三方出版机构的规则进行,通过强有力的舆论传播机制,将伪造、抄袭等欺骗行为纳入科研共同体的失信黑名单。在这样的共享活动中,来自第三方的计量评价为数据供给者提供了适当的激励机制。

总之,中间体形态的数据共享中,来自外部的激励措施与内化激励并存,因而这类数据共

享模式具有不容忽视的重要地位。

(4) 个体间的数据分享模式

在介入门槛不高、预期收益不足的数据共享方面,来自科学家个体的自愿分享占据了主体地位。这一类共享活动中,数据规模与服务能力有限,分布零散,数据共享活动完全依靠科学家自律来保障共享数据的质量,也通过科学家群体的有限规模的交流评价来实现共享数据的治理。个体间的数据分享模式可提供灵活的共享交互与丰富的数据种类,然而“私人间的数据分享并不能随着时间的推移而形成规模化效应,数据及其暗含的知识若不经深度培育将很快衰减殆尽”^[29]。为更好地满足规模化的数据共享需要,数据市场开始接管小众化的数据共享活动,使之能够胜任批量处理工作的需要,实现更大规模的数据分享工作。

(5) 市场形态的数据共享模式

市场形态的数据共享代表着更低的共享门槛与可观的规模经济收益预期。这一类数据共享采取合同形式规范双方权责。作为一类自下而上的数据共享模式代表,“无形之手”的强大力量使之成为主流共享模式的补充。

以上分类仍显抽象,以下笔者将选取其中具有大众化服务意义的主流共享模式实践展开分析,以再现数据共享情景,主要包括资源池模式、数据出版模式和数据交易模式等。

3 科学数据共享主要模式情景分析

结合上述分析,下一步的情景分析将以供求关系为主线,再现相关利益群体的科学数据共享交互情况。

3.1 资源池模式

作为有管制的数据共享模式代表,这一类共享活动中数据的收集存储和分享利用并重,日益增长的数据共享需求主要靠数据生产不断积累来满足,笔者将其形象地称为“资源池模式”。

这一模式经常用于诠释公共财政经费支持

的数据共享。共享活动的延续性多取决于上游经费支持力度与制度规范的要求。基于公益属性,资源池模式采取自上而下强有力的推行手段,其数据共享路径多为持续的单向传播。数据供给者在一定的制度要求下,将面向公众提供数据共享服务视为一种义务和责任。资源池模式的数据共享以不断积累的数据供给为基础,通过扩大数据规模进而拓展共享规模,凭借数据生产与获取能力的不断提高,扩展数据共享服务的能力和水平。数据的动态更新迭代在促进数据积累的同时,推进共享工作前进。这是科学研究提供给公众社会开放服务最为有效的、主流的数据分享方式。我国科研信息化建设中涌现出一大批行业领域的数据共享工程都是这类模式的实践代表,他们通过公共财政经费倾斜来保障相关数据的生产与收割,同时又强调数据共享的深度与广度。

然而资源池模式的数据共享仍显粗放。尽管IT工具的应用从一定程度上提高了共享数据的完整性、系统性,如统计用户日志来衡量数据价值等,但资源池模式的数据堆砌式生长仍然存在缺陷。共享供给者受制于资助机构的强制政策要求,虽然管理形式上也通过包括同行专家评议、IT工具监测管控等手段来保障数据量和数据质量,但这些措施并没有考虑到供给者的自身诉求,数据的共享仍呈现被动状态。同样,由于既定政策,数据供给活动也很少考虑个体需求者的数据诉求。资源池模式从形式和规模上为推动科研数据开放共享提供了坚实基础,但要想在更大范围内挖掘“休眠数据”的深层价值,还需要调动数据供给者的积极性,倾听实际的数据需求,探索实现多模式并举。

3.2 数据出版模式

数据出版^[30]是为满足公众需要而对特定数据或数据集发布利用的一系列过程,该模式主要通过第三方治理使数据拥有者获得学术声望。结合Lawrence等^[31]的观点,笔者认为,从数据主体地位与发展过程方面可将数据出版分

为三大类:①历史悠久的学术论文辅助数据出版、附录数据出版,即认为数据是科研附属产物;②独立的数据出版,包括直接出版数据集、同时出版数据集及其描述文档(如数据论文)两种形式,由于数据作为科学研究的重要成果独立出现并更易于获取和重用,独立的数据出版模式成为当前颇受瞩目的数据出版形式;③广义的数据中心的数据出版。

数据出版模式中的数据流动开始逐步将数据供给与需求联系到一起。数据的供给者通过数据出版实现自己的学术价值,出版社则通过交换获得支持运行的基础资源;作为服务过程,数据用户与出版商凭借一定的货币或服务评价实现交换。共享双方交换的过程皆以第三方出版社为枢纽,这种交换不是简单的以物易物,而是以相对成熟的标准化协议或合同的形式出现,这使数据面向大规模开放服务成为可能。为了更好地推动数据应用,更多的数据出版平台将数据用户的反馈评价正式纳入其中(如大众评议,Crowd Rating)。数据共享不再是自上而下倾倒,来自个体用户的针对性建议和意见使这一模式运转有了更多交互的可能。

数据出版在出版社和国际组织中都有所实践。《地学系统科学数据》(*Earth System Science Data*, ESSD)于2008年起出版地球系统科学数据,以维护科学数据集资源的可信度,《自然》自2014年起不定期出版数据论文,并提供与数据论文相关的数据机构库存储与访问服务支持。F1000Research、PloS、eLife、GBIF等机构分别以科研出版物附属物的形式出版各自领域的科学数据。《中国科学数据》作为国内较早的多学科领域的数据论文出版机构,自2015年成立以来也进行了数据出版实践。

作为推动科学数据共享的有效途径之一,数据出版模式既提升了科研数据的可信度,又提高了学术造假成本,同时为更好地拥有、使用和推广数据成果带来可能。不过,相较于传统学术论文出版,数据出版的学术地位和行业认可度、运营模式的可持续性、数据开放与版权保

护等方面仍有待发展。数据引用及其计量评价源于传统学术体系,又有所不同,亟待建立一套既适应科研活动规律又能满足自身发展需要的评价体系^[32],从而调动数据供给方的积极性,切实推动数据出版共享。

3.3 数据交易模式

数据市场的交易模式利用“看不见的手”——市场力量,为权属明晰的数据买卖提供第三种选择。数据市场的数据交互为更好地发现数据价值提供了途径,同时也通过数据分析工具的使用与数据质量、范围的标识实现了数据的比较和价值发现。数据交易可提供相对可靠的数据质量控制与便捷的使用方法和工具,使数据价值更加容易被发现和获取。可以说,商业化的运作模式为持续挖掘科学数据的价值提供了广阔空间。

数据的巨大市场潜力吸引着传统 IT 公司、新兴数据公司纷纷打造数据交易平台,以期通过提供数据交易及其配套增值服务来获取更多商业价值的回报。例如,美国数据结构公司 (FACTUAL, 2007)、DATAMARKET (2008)、INFOCHIMPS (2009)、微软蓝云数据市场 (Microsoft Azure Data Market, 2010)、数据堂 (2011) 等典型数据交易平台的实践。

数据交易模式为面向开放服务的数据共享交流提供了新的视角和平台。该模式立足数据集特征,在数据发布方式、发布渠道、技术工具的应用等方面沿用传统数据集的共享管理方法,其亮点在于共享模式的内在运行机制不同于以往公益性的开放共享,它将数据共享活动视为一种经济交易活动推向市场。数据供给与需求的角色定位更为清晰,治理规则主要依靠经济市场“看不见的手”来调节,这为共享激励措施注入了新的元素,为面向开放服务的数据模式选择提供了新的契机和思路。

该模式仍处于发展上升期。仍在尝试数据交易的技术平台 FACTUAL、已被收购的 DATA-MARKET,以及 INFOCHIMPS 的战略转型、即将停止服务的 WINDOWS AZURE DATA MARKET 等,从另一方面说明,以市场交易形式实现数据共享仍存在一定困难:较高的数据专用性带来市场准入高门槛。与此同时,数据共享市场化需求规模仍显不足,商业模式仍待成熟完善。

3.4 模式比较评价

三种数据共享模式及其案例代表着当前面向开放服务环境的主要共享操作实践。通过比较分析,可以看出三种模式的利弊、机会与威胁 (见表 5)。

表 5 共享模式横向比较

模式类型	优势	劣势	机会	威胁
资源池模式	具有共享活动规模化示范化效应,量化共享效果明显	共享动机来源于义务和责任,忽视数据供给者主观能动性	为公共资源社会分享建立有力的平台和资源积累宝库	治理机制来源于法律规定,监督成本高昂或难于执行
数据出版模式	借鉴较成熟的出版运营模式,提升了共享数据实现科研价值的可能性	计量评价方法有待成熟;数据发布与数据长期保存职能割裂,利益链条复杂	具有实现数据共享义务与数据学术价值双赢机制,是面向开放服务的重要共享形式	棘手的数据出版难题 (如数据流、难于切割的海量数据集、复杂结构数据),数据的知识产权保护与利用
数据交易模式	经济利益刺激数据共享规模,共享质量保障	数据交易市场的成熟性,科研模式带来所有权结构的复杂性	广阔的市场空间,独辟新境的、面向开放服务不可或缺的共享模式补充	法律纠纷等

其中,资源池模式在数据交流方面采取持续单向性策略,其模式运转具有高度的政策和经费依赖性。开放服务的持续性需要“有形的手”来维持,数据共享多基于工作职责和义务,共享文化建设薄弱,供给者的主观能动性未得到很好的重视。数据的监督治理大多从规模、数量和时间等明确的量化指标考察,而囿于高昂的监督控制成本,共享数据的质量问题很难深入探究。这一模式的优势在于,共享活动效果立竿见影,充足的经费保障不断提升数据平台服务能力与数据资源积累规模。

数据出版依托文献出版的成熟运营模式,借鉴传统科研成果价值实现的方法来实现数据生产者的科研活动价值。但数据出版仍在发展上升期,与之相关的计量分析工具和方法等尚未成熟,数据的学术价值能否得到认可,数据生产者能否获得与传统科研成果出版一样的学术地位,这些都尚待探讨。此外,数据出版仅实现数据发布功能,数据使用需配套专业的存储资源库,利益链条的延伸与良好的合作模式是保证共享活动顺利实施的关键。与此同时,不同于以往的文献出版,数据对象独有的特征,如难于定义的数据版本(数据流的出版)、难于切割的数据集(海量数据的出版)以及复杂的数据结构等,带来了新的出版共享难题。

数据交易模式的优势则在于其通过“看不见的手”来调节数据的供给与需求,通过经济利益杠杆刺激数据共享规模的扩大,并通过“用脚投票”的方式有效控制共享数据的质量问题,有效地规避了资源池模式的监督难题,使数据供给者能够更好地发挥主观能动性来参与特定的数据共享活动。不过目前交易模式尚在发展之中,数据供给与数据需求仍不够成熟,科研活动的规模化、协作式、多经费来源现状,为数据产权界定带来了复杂的、难于切割的棘手问题。任何未获得完整、明确授权的擅自交易行为都可能带来法律纠纷,而这关乎科研共同体中科学家最关心的声誉问题。

4 共享模式展望与结语

基于情景再现不难看到,资源池模式在开放服务大环境中虽显粗糙,却是当前数据共享不可或缺的方式之一,并将在相当长一段时间内持续有力地支持公益性数据的原始积累。数据出版模式则较好地调动了数据生产者的积极性,既重视共享意愿的培养,也促进共享文化的形成,是面向开放服务日趋主流的数据共享模式之一。与此同时,由于市场力量的介入,数据交易模式将是未来面向开放服务不可或缺的模式选择,但由于数据资产专用性的高门槛一直存在,这类模式在相当长时间内仍将以科学研究共享活动中的补充形式存在。数据交易模式应用的前提是预先理顺公益性经费支持的数据生产与有偿共享服务的界限——只有具有一定附加值并能够独立核算成本投入的数据,方可根据行业领域情况尝试该模式的运用。

纵观科学数据共享的整体发展,有关模式的完善仍将持续,包括但不限于以上主流模式的发展势必相互补充,共同推动科研活动中的数据开放共享工作。与此同时,固有的共享路径依赖^[33]仍在一定程度上塑造着理想共享模式的现实形态。依附于科研活动,科学数据共享不可避免地受制于当时当地的文化与科研共同体的既有框架,模式的变革更倾向于采取一种温和的、循序渐进的路线。如果在以资源池模式为主导的共享模式中,能够根据不同的数据属性及其具体的组织形态选择更多模式予以补充,势将对现有的研究数据开放共享起到重要的推进作用。

综上,笔者梳理了现有的科学数据共享模式实践,引入新制度经济学视角,从科学数据专用属性、占优的组织形态选择与治理机制方面展开,将科学数据共享总结为组织内部的数据共享模式、有管制的数据共享模式、中间体形态的数据共享模式、个体间的数据分享模式以及市场形态的数据共享模式等五种。与之对应,

情景再现囊括了面向开放服务的主流共享模式,如资源池模式、数据出版模式、数据交易模式等。依托供求关系主线并梳理数据流动的利益链条,笔者重点探讨了三种共享情景的优劣势、机会与挑战。在与时俱进的科研共同体中,

“如果我们想继续将科学及其事业的真实可靠性视为一种获取知识的可靠途径,那么无论大数据与否,确保数据开放都将是重中之重”^[34]。为此,希望上述分析能够有助于不同情景下科学数据共享工作的组织与推进。

参考文献

- [1] Smith V. Data publication:towards a database of everything [J]. BMC Research Notes,2009(2):111-113.
- [2] 刘润达,赵辉,李大玲.科学数据共享平台之数据联盟模式初探[J]. 中国基础科学,2010(6):27-32.(Liu Runda,Zhao Hui,Li Daling. Preliminary study on data alliance model for scientific data sharing platform[J]. China Basic Science,2010(6):27-32.)
- [3] 王培正,张志强,吴一民.科学数据共享方式研究[J]. 科技管理研究,2010(17):201-204.(Wang Peizheng,Zhang Zhiqiang,Wu Yimin. Research on sharing mode of scientific data [J]. Science and Technology Management Research,2010(17):201-204.)
- [4] 王晴.论科学数据开放共享的运行模式、保障机制及优化策略[J]. 国家图书馆学刊,2014(1):3-9.(Wang Qing. Research on the operating modes,guarantee mechanism and optimization strategies for opening and sharing scientific data [J]. Journal of the National Library of China,2014(1):3-9.)
- [5] CERN.Computing[EB/OL]. [2017-02-09]. <http://home.cern/about/computing>.
- [6] IGBP. What we do[EB/OL]. [2017-02-09]. <http://www.igbp.net/about/whatwedo.4.2709bddd12c08a79de780001018.html>.
- [7] 左建安,陈雅.基于大数据环境的科学数据共享模式研究[J]. 情报杂志,2013(12):151-154.(Zuo Jian'an,Chen Ya. The analysis on sharing mode of scientific data in the environment of big data[J]. Journal of Intelligence,2013(12):151-154.)
- [8] 曾小红,王强,方佳.科学数据共建共享模式探讨[J]. 安徽农业科学,2009(29):14496-14498,14501.(Zeng Xiaohong,Wang Qiang,Fang Jia. Research on the joint construction sharing mode of scientific data[J]. Journal of Anhui Agricultural Sciences,2009(29):14496-14498,14501.)
- [9] 李娟,刘德洪,江洪.国际科学数据共享现状研究[J]. 图书馆建设,2009(2):19-21,25.(Li Juan,Liu Dehong,Jiang Hong. Research on international scientific data sharing [J]. Library Development,2009(2):19-21,25.)
- [10] Jeffery K.Open access:an introduction[N/OL]. ERCIM news,2006(1)[2017-02-09]. http://www.ercim.eu/publication/Ercim_News/enw64/EN64.pdf.
- [11] OECD. OECD principles and guidelines for access to research data from public funding[R/OL]. [2017-02-12]. www.oecd.org/science/sci-tech/38500813.pdf.
- [12] ICSU WORLD DATA SYSTEM. Data sharing principles[EB/OL]. [2017-02-09]. <https://www.icsu-wds.org/services/data-sharing-principles>.
- [13] GEO. The GEOSS data sharing principles post-2015[EB/OL]. [2017-02-09]. <http://www.earthobservations.org/dswg.php>.
- [14] G8 open data charter and technical annex[EB/OL]. [2017-02-09]. <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>.
- [15] 国家地震科学数据共享中心.地震科学数据共享管理办法[S/OL]. [2017-02-04]. <http://data.earthquake.cn/sjgsgz/index.html>.(China Earthquake Datacenter.Measures of earthquake scientific data sharing management [S/OL]. [2017-02-04]. <http://data.earthquake.cn/sjgsgz/index.html>.)
- [16] 高峰,王国复,孙超,等.后台管理模式在数据共享平台中的应用[J]. 应用气象学报,2011(3):367-374.(Gao Feng,Wang Guofu,Sun Chao,et al.Application of the background management model to data sharing service

- platform[J]. Journal of Applied Meteorological Science, 2011(3):367-374.)
- [17] 韩宏,樊秀娥,王京萍,等.公共卫生科学数据共享平台权限管理研究[J]. 中国公共卫生管理,2007(2):140-142. (Han hong, Fan Xiue, Wang Jingping, et al. Research on rights management in public health science data sharing platform[J]. Chinese Journal of Public Health Management, 2007(2):140-142.)
- [18] Kosseim P, Dove E, Baggaley C, et al. Building a data sharing model for global genomic research[J]. Genome Biology, 2014, 15(8):430.
- [19] Data Working Group in Global Alliance for Genomics and Health. Creating global data standards for Genomics [EB/OL]. [2016-11-04]. <http://ga4gh.org/#>.
- [20] 毛海峰.政府空间数据共享模式探讨[J]. 测绘信息与工程, 2006(1):43-44. (Mao Haifeng. Research on sharing model for government spatial data[J]. Journal of Geomatics, 2006(1):43-44.)
- [21] 苗茹, 诸云强, 宋佳, 等.基于云计算的地球系统科学数据共享研究与实践[J]. 地球信息科学学报, 2014(2):264-272. (Miao Ru, Zhu Yunqiang, Song Jia, et al. Services of earth system science data sharing based on cloud computing[J]. Journal of Geo-Information Science, 2014(2):264-272.)
- [22] NBN. NBN model data sharing and use policy [EB/OL]. [2016-08-03]. <http://nbn.org.uk/national-biodiversity-network/archive-information/nbn-model-data-sharing-and-use-policy>.
- [23] 杨兴凯.政府部门间信息共享模式与决策方法[M]. 北京:科学出版社, 2014:108-125. (Yang Xingkai. Governmental information sharing models and decision making methods [M]. Beijing: Science Press, 2014:108-125.)
- [24] 姚玲杰.高等学校知识共享的新模式[J]. 大学图书馆学报, 2015, 33(2):86-88, 93. (Yao Lingjie. Exploration of knowledge sharing mode in universities [J]. Journal of Academic Library and Information Science, 2015, 33(2):86-88, 93.)
- [25] 周永红, 吴银燕, 宫春梅.基于企业联盟的知识共享模式分析[J]. 情报理论与实践, 2014, 37(12):57-60, 46. (Zhou Yonghong, Wu Yinyan, Gong Chunmei. Analysis of the knowledge sharing model based on enterprise alliance [J]. Information Studies: Theory & Application, 2014, 37(12):57-60, 46.)
- [26] 张丽丽.信息共享的制度经济学浅析[J]. 图书情报工作, 2013, 57(19):57-61. (Zhang Lili. Analysis of information sharing based on institutional economics [J]. Library and Information Service, 2013, 57(19):57-61.)
- [27] 威廉姆森.资本主义经济制度——论企业签约与市场签约[M]. 段译才, 王伟.译.北京:商务印书馆, 2002:78-90, 104-113. (Williamson O E. The economic institution of capitalism: firm, markets, relational contracting [M]. Duan Yicai, Wang Wei, Trans. Beijing: The Commercial Press, 2002:67-78, 104-113.)
- [28] Aoki M. Toward a comparative institutional analysis [M]. Cambridge: The MIT Press, 2001:79.
- [29] Borgman C. Big data, little data, no data-scholarship in the networked world[M]. Massachusetts: The MIT Press, 2015:239.
- [30] Assante M, Candela L, Castelli D, et al. Are scientific data repositories coping with research data publishing[J]. Data Science Journal, 2016, 15:1-24.
- [31] Lawrence B, Jones C, Matthews B, et al. Citation and peer review of data: moving towards formal data publication [J]. The International Journal of Digital Curation, 2011, 2(6):4-37.
- [32] Li J, Wu C, Zhang L, et al. Survey and analysis of scientific data publishing[J]. China Scientific Data, 2016(1). DOI: 10.11922/csdata.120.2015.0009.
- [33] Rodriguez J P, Loomis S R, Weeres J G. The cost of institution: information and freedom in expanding economics [M]. New York: Palgrave Macmillan, 2007:74-78.
- [34] Science International. Open data in a big data world[R]. International Council for Science (ICSU), International Social Science Council (ISSC), World Academy of Sciences (TWAS), InterAcademy Partnership (IAP). Paris, 2015.

张丽丽 中国科学院计算机网络信息中心助理研究员。北京 100190。

(收稿日期:2016-11-17;修回日期:2017-02-20)