

• 科学技术与社会 •

科学数据共享问题：一个博弈论视角的讨论

Data Policy: A Discussion from the Perspective of Game Theory

李思宇 / LI Siyu

(中国政法大学政治与公共管理学院, 北京, 100088)

(School of Political Science and Public Administration, China University of Political Science and Law, Beijing, 100088)

摘要: 本文试图用博弈论讨论科学数据共享问题, 通过一个简单的“鹰-鸽”型进化博弈模型, 从成本和收益的角度, 给出了数据共享实现的一个充分条件, 进而讨论了推进科学数据共享的可行的政策措施。

关键词: 科学数据共享 博弈论 数据政策

Abstract: Problems of scientific data sharing are discussed from the perspective of game theory. With the related cost and benefit considered, a sufficient condition for the implementation of data sharing policy is proposed through a simple “dove-eagle” evolutionary model. Discussion on the expected utility of individual agents in the game and the evolutionary fitness of the community has clear implications for the public policy of science and technology.

Key Words: Scientific data sharing; Game theory; Data policy

中图分类号: N0 文献标识码: A DOI: 10.15994/j.1000-0763.2020.12.010

引言

科学数据共享问题是一个被持续关注的复杂问题, 国内外的比较表明我国在相关方面仍有较大差距, 很多政策措施的效果并不尽如人意,^{[1]-[4]} 对此, 甚至有“数据共享困境”的说法。^[5] 提出破解这一难题的切实的具体措施, 是推动我国科技发展的一项重要的基础性工作。

一般认为, 科学数据共享问题更多地属于伦理问题和政策问题。然而从现代博弈论的角度说, 无论是公共政策选择还是规范伦理, 都是某种策略博弈的结果。^{[6], [7]} 因此, 博弈论是讨论科学数据共享问题的一个有效的工具。^{[8], [9]} 在这方面,

高峰和王剑曾基于合作博弈讨论农业科学数据平台的共享管理机制问题;^[10] 庄倩和何琳把数据共享表述为一种演化博弈, 讨论了有利于数据共享的条件。^[11] 而事实上, 在关于推进数据共享的政策和对策的讨论中, “对策”的概念本身也已经蕴含了博弈论的思路(博弈论的另一种中文翻译就是“对策论”)。^{[12], [13]}

博弈论的基本概念是, 每一个参与者(agent)都是(部分)理性的个体。个体之间无法充分沟通。因而博弈的产出, 是一个大家都觉得自己“最不吃亏”的、但未必是全局最优的结果。一定意义上, 一个社会现象的最终产出(outcome), 取决于每一个参与者根据效益-代价的平衡所做出的理性选择。科研人员作为(部分)理性的个体, 这一基

收稿日期: 2019年2月27日

作者简介: 李思宇(1992-)女, 辽宁铁岭人, 中国政法大学政治与公共管理学院博士研究生, 研究方向为公共行政、发展管理。Email: 1838495008@qq.com

本假定应该是没有问题的。基于此，曾有工作用行为理论模型研究科研人员的数据共享意愿问题。^[14]除去科研人员个体外，团体的因素也不能忽略。张丽丽等对国内外发展现状的调研指出，“国内科学研究的整体环境有利于开放科学数据文化的孵化，但仍需多利益相关群体的综合协调”。^[4]该项调研认为“未来，积极温和的共享趋势仍占据主流，科学数据的公、私权力争论愈演愈烈”。^[15]科学数据共享中的利益平衡机制问题也被提上日程。

为讨论这一“多利益相关群体的综合协调”和“公、私权力争论”的解决办法，本文通过一个高度简化的博弈论模型，讨论数据共享的几个关键问题。关注的重点是，数据共享如何能不只是一种自上而下的“命令”或“约束”，而是通过某种“演化”，成为一种自下而上的要求。这是“治理”与“管理”在思路上的最重要的差别。

一、数据共享问题的简单模型

我们考虑一个极为简单的进化博弈模型。考虑一个群体（community）。其中有两类成员（agent，既可以是个体，也可以是单位或小的团体）。一类成员积极进行数据共享，另一类成员对数据共享不积极。其组合有4种情况：（一）如果一个积极的成员（幸好）碰到另一个积极的成员，那么，结果是双方出力、同时双方获益；（二）如果一个积极的成员（不幸）碰到一个不积极的成员；（三）或者一个不积极的成员（有幸）碰到一个积极的成员，那么，结果是积极的成员单方出力、但最后结果仍是双方获益；（四）如果一个不积极的成员碰到另一个不积极的成员，那么结果自然是谁也不出力、同时谁也不获益。将收益（benefit）和成本（cost）分别记为E和C，则上述4种情况可以表示为下面的博弈矩阵（game matrix）：

表1 四种情况下的博弈矩阵

	积极	不积极
积极	(E-C), (E-C)	(E-2C), E
不积极	E, (E-2C)	0, 0

考虑一个群体，开始时，都是积极的。但因为某种原因，出现了一个不积极的成员。借用生态学的语言，出现了一个“入侵异种”（invasion mutant），或者出现了一个“变异”的个体。更一

般性的表示，是以一定的概率p出现这样一个“变态”。考查这个群体如何“进化”，会为这一群体所能达到的“最终状态”提供线索。

这个模型显然是十分简单的。实际上，这类“鹰-鸽进化博弈”模型的意义，^{[8], [9]}类似于物理学中的元胞自动机（CA）模型^[16]或地震学中的弹簧-滑块模型（BK模型），^[17]以非常简单的概念，揭示复杂现象的本质。基于同样的精神，庄倩和何琳将科学数据共享体系中的科研人员的共享行为划分为“分享信息”与“不分享信息”两种，并将其放在演化博弈的理论框架中，分析科学数据共享的动态演化过程。^[11]本文对模型的处理，其基本精神与庄倩和何琳等学者相似，但数学更为简洁，同时更为强调所得结果的政策涵义的讨论。

二、期望效用值与“进化适应性”

考虑一个积极的成员。她/他/它（有幸）碰见另一个同样积极的成员的概率是(1-p)，这种情况下她/他/它的期望效用值（expected utility）是(E-C)。她/他/它（不幸）碰见一个不积极的成员的概率是p，这种情况下她/他/它的期望效用值是(E-2C)。这样，她/他/它的进化适应性（evolutionary fitness），是两项的加权和（以概率为加权系数）：

$$(1-p) \cdot (E-C) + p \cdot (E-2C) = E - (1+p) \cdot C \quad (1)$$

同样，一个不积极的成员，她/他/它的进化适应性是：

$$(1-p) \cdot E + p \cdot 0 = (1-p) \cdot E \quad (2)$$

其中，(1-p)是她/他/它（有幸）碰上一个积极成员从而“净占便宜”的概率，p是她/他/它碰上另一个不积极的成员从而无便宜可占的概率。

式（1）所表示的是积极策略的进化适应性；式（2）所表示的是不积极策略的进化适应性。因此积极策略的进化稳定性（evolutionary stability）要求：

$$E - (1+p) \cdot C > (1-p) \cdot E \quad (3)$$

或者

$$p \cdot (E-C) > C \quad (3')$$

或者说，满足上式后，不积极的“变异”个体，会随着群体的“进化”而逐渐消失。因此上式一定意义上说是最终形成积极策略的一个充分条件。由于p在不等号的“大于”一侧，权取其最小值p=1/N，其中N是群体中总的成员数，则此时上式还可以表

示为:

$$E > C(N+1) \quad (4)$$

当这个条件满足时,那个“入侵的异型”会随着进化过程逐步消失,从而群体的最终状态是积极状态。

三、模型结果对于数据共享 相关公共政策的意义

上述模型,显然是一个过于简化的模型。但从这一极为简单的模型中,仍可得到很多启发性的结论。而这正是应用博弈论的目的。

1. “成本侧”的相关政策

式(4)中的右面是由 $C(N+1)$ 所表示的成本。降低这一项,就会促进数据共享的实现。值得强调的是,对于越大的个体集合或者共同体(community),降低C的要求就越高。因此可以理解的是,小范围的数据共享,通常还比较容易做到,但全社会的数据共享就难一些。我国在数据共享方面的差距,在多大程度上与我国自身的体量有关,值得探讨。尽管没有实证的证据,但一个间接的证据是另一个大国印度,其数据共享程度也很不理想。

从国家研究机构、国家主管部门的角度,一个关键,是要通过基础设施投入,使个体成员所承担的数据共享的成本C尽可能地小,即,将数据共享变得容易。N越大,这项工作就越重要。这一点,在其他方面(例如经济领域)已有很好的经验:理论上,个人信息的数据是最没有“共享”的条件,但是网购的方便、快捷使很多人不惜“让渡”这些信息的保密的权利。而从经济领域广为引用的熊彼特创新理论的角度看,如同企业的创新形态按层次的由低到高表现为产品创新、工艺创新、原材料创新、市场创新、平台创新,数据共享基础设施的建设,属于“平台创新”,在数据共享的推进中属于最高层次,具有决定性意义。^[18]一定程度上说,我国数据共享问题的突破,也许就在于类似于“支付宝”那样使全社会的数据共享变得十分容易的基础设施。在这方面,我们尚有较长的一段路要走。五年前的一项实证研究指出,我国科学数据共享平台“存在政策法规尚不完善、缺少数据描述与组织标准或规范、数据可获取性低、平台提供的服务形式较少、各个平台的科研效能差距较大等问题”。^[19]目前,这种情况虽有改善,但仍存在较大差距。

这里的C的理解,应该不仅包含做数据共享(这件事本身)的成本,而且还包括把数据共享(出去)的成本。一个重要的事情是,把数据提交共享的人,也能“对称地”和方便地使用其他人共享的数据才行。如果大家觉得提交数据容易、获得数据难,从而把自己的数据提交共享是“吃了亏”,那么数据共享的动力也就没有了。庄倩和何琳的研究结果表明,从演化博弈的角度看,建立健全的保障制度,阻止机会主义和“搭便车”行为的出现,的确可以促进科学数据的共享。^[11]然而,正如尽管在网购、电商中也有很多让消费者吃亏的情况需要监管(与网购等相联系的对个人信息的掠夺性获取也绝不是应该鼓励的),网购、电商本身却以极大的方便、快捷促进了一个新的生产-消费形态乃至一种新的生活方式的形成。这是从成本的角度看需要解决的主要问题。就是说,数据共享问题的解决,很大程度上取决于获得数据的便捷性;数据共享要求的实现,很大程度上取决于为数据共享本身所做的基础设施建设和服务。

2. “收益侧”的相关政策

尽可能地提高由E所表示的收益,是促进数据共享的一个重要的政策选择。其中很重要的一点,是在科研活动的绩效考评指标中对引用的强调。

SCI的一个最值得借鉴的思维方式就是,科学出版物,不仅数量是重要的,而且质量也是重要的;不仅发表是重要的,而且引用也是重要的。值得注意的是,以往对引用的强调,即强调引用那些已经形成了较为成熟的成果的工作(通常以科研论文、报告的形式发表),已越来越与分工精细、技术含量高、多方协作的现代科技不相适应。也因此,目前国际上共同推进的一个做法,是赋予观测数据源以“数字对象识别码”(DOI),以使数据集本身也能像论文那样被直接引用。^[20]

另一个值得注意的问题是,共享平台建设项目很多,但好像都跟最初期望的目标有不同程度的差距。有些甚至是“有平台,没数据”,或者“精品平台,垃圾数据”。解决这个问题,需要强调的是通过“准入”政策,以“共享准入”的荣誉,提升数据共享的“付出性”收益。

这里的政策建议(例如强调“准入”)在一定程度上反直觉。但其所考虑的一个关键问题是,如何能使数据共享不再是一种义务,而是一种像专利和“影响因子(IF)”那样的收获或肯定。从而个

人或单位努力要做的,不再是“保护”自己的数据尽可能“不被共享”(这是目前在很多地方实际发生的事情),而是尽可能“争取”自己的数据作为“好数据”通过共享而被引用。另一方面,强调“准入”,一个直接的收益就是,共享的数据都是“好”数据,国家级共享的数据是最好的数据。

四、结论和讨论

本文通过一个简单的“鹰-鸽”型进化博弈模型,从成本和收益的角度,给出了数据共享实现的一个充分条件,并进而讨论了推进科学数据共享的可行的政策措施。在“成本侧”,主要是建设数据共享基础设施;在“收益侧”,主要是强调引用、强调对共享数据实行“准入”政策。思路的重点是,仅仅强调科学数据“必须共享”,无论如何是过于简单化的。把数据共享政策的重心放在“防止不共享”上,也已被证明是事倍功半。因此需要探讨,数据共享如何能不只是一种自上而下的“命令”或“约束”,而是通过某种“演化”,成为一种自下而上的需求。

博弈论的基本思想,不在结论,而在“换位思考”思考问题的角度。这里并不刻意强调什么是“对”的、什么是“不对”的。实际上,“人民内部矛盾”的特点就是,大家都是对的,大家都有道理,或者说,你有你的道理、我有我的道理。博弈论要处理的问题是,不同的“道理”,怎样通过博弈,达成一个“好的”均衡。——不是所有的均衡都是“好的”均衡,比如博弈论中的经典案例“公地悲剧”,就不是一个“好的”均衡;但另一个经典案例“核威慑”,也不能说是一个“坏的”均衡。^{[8], [9]}

值得指出的是,在数据共享博弈中,国家研究机构、国家主管部门,是以一个重要的参与者(player,有时可以处理成dummy player)的角色出现的,其作用,是以某种“调控”的方式,通过合适的公共政策,促进达成一个“好”的“均衡”。

[参考文献]

- [1] 黄鼎成、郭增艳. 科学数据共享管理研究[M]. 北京: 中国科学技术出版社, 2002.
- [2] 黄如花、邱春艳. 国外科学数据共享研究综述[J]. 情报资料工作, 2013, (4): 24-30.
- [3] 司莉、庄小喆、王思敏、吴方枝、钱绮琪. 2005年以来国外科学数据管理与共享研究进展与启示[J]. 国家图书馆学刊, 2013, 22(3): 40-49.
- [4] 张丽丽、温亮明、石蕾、郑晓欢、黎建辉. 国内外科学数据管理与开放共享的最新进展[J]. 中国科学院院刊, 2018, 33(8): 774-782.
- [5] 吴忠良. 科技投入的“困境”或“悖论”及其“游戏规则”解[J]. 自然辩证法通讯, 2017, 39(3): 153-159.
- [6] Harsanyi, J. C. *Handbook of Game Theory With Economic Applications*[M]. Netherlands: North Holland, 1992, 669-707.
- [7] Verbeek, B., Morris, C. 'Game Theory and Ethics'[J]. *Stanford Encyclopedia of Philosophy*, 2010, (6): 137-142.
- [8] Dixit, A., Skeath, S., Reiley, D. *Games of Strategy*[M]. New York: W. W. Norton & Company, 1999, 245-267.
- [9] Harrington, J. E. *Games, Strategies and Decision Making*[M]. Washington: Worth Publishers, 2009, 399-410.
- [10] 高峰、王剑. 基于合作博弈的农业科学数据平台共享管理机制研究[J]. 技术与创新管理, 2013, 34(5): 457-460.
- [11] 庄倩、何琳. 科学数据共享中科研人员共享行为的演化博弈分析[J]. 情报杂志, 2015, 34(8): 152-157.
- [12] 黄如花、王斌、周志峰. 促进我国科学数据共享的对策[J]. 图书馆, 2014, (3): 7-13.
- [13] 李秋月、何玮雯. 我国科学数据权益保护问题及对策——基于共享政策的文本分析[J]. 图书馆, 2018, (1): 74-80.
- [14] 何琳、常颖聪. 科研人员数据共享意愿研究[J]. 图书与情报, 2014, (5): 125-131.
- [15] 司莉、辛娟娟. 科学数据共享中的利益平衡机制研究[J]. 图书馆学研究, 2015, (1): 13-16.
- [16] Turcotte, D. L. *Fractals and Chaos in Geology and Geophysics*[M]. Cambridge: Cambridge University Press, 1992.
- [17] Burridge, R., Knopoff, L. 'Model and Theoretical Seismicity'[J]. *Bulletin of the Seismological Society of America*, 1967, 57(3): 341-371.
- [18] 朱红恒. 熊彼特的创新理论及启示[J]. 社会科学家, 2005, (1): 59-61.
- [19] 司莉、李月婷、邢文明、华小琴、李鑫、辛娟娟. 我国科学数据共享平台绩效评估实证研究[J]. 图书馆理论与实践, 2014, (9): 3035.
- [20] Neumann, J., Brase, J. 'DataCite and DOI Names for Research Data'[J]. *Journal of Computer-Aided Molecular Design*, 2014, 28(10): 1035-1041.

[责任编辑 李斌 赵超]