

科学数据共享的 知识产权保护机制研究

李 琰◎著



KEXUE SHUJU GONGXIANG DE
ZHISHI CHANQUAN BAOHU JIZHI YANJIU

 人 民 出 版 社

科学数据共享的 知识产权保护机制研究

KEXUE SHUJU GONGXIANG DE
ZHISHI CHANQUAN BAOHU JIZHI YANJIU

李 琰◎著



责任编辑:茅友生

封面设计:胡欣欣

图书在版编目(CIP)数据

科学数据共享的知识产权保护机制研究/李琰 著. —北京:人民出版社,2019.6
ISBN 978-7-01-020417-8

I. ①科… II. ①李… III. ①知识产权保护-研究-中国 IV. ①D923.404
中国版本图书馆 CIP 数据核字(2019)第 029927 号

科学数据共享的知识产权保护机制研究

KEXUE SHUJU GONGXIANG DE ZHISHI CHANQUAN BAOHU JIZHI YANJIU

李 琰 著

人民出版社 出版发行
(100706 北京市东城区隆福寺街 99 号)

天津文林印务有限公司印刷 新华书店经销

2019 年 6 月第 1 版 2019 年 6 月北京第 1 次印刷

开本:710 毫米×1000 毫米 1/16 印张:18.75

字数:298 千字

ISBN 978-7-01-020417-8 定价:69.00 元

邮购地址 100706 北京市东城区隆福寺街 99 号

人民东方图书销售中心 电话 (010)65250042 65289539

版权所有·侵权必究

凡购买本社图书,如有印制质量问题,我社负责调换。

服务电话:(010)65250042

前 言

就科学研究的一般过程而言,科学数据在人类科学发展过程中扮演了关键角色。科学数据既是科学研究活动开展的基础和源头,又是科学研究活动中最有价值的产物和输出。现代人类科学研究活动越来越依赖于科学数据的创造和利用,科学研究的范式发生了转向数据驱动(Data Driven)的革命,以科学数据、数据挖掘和知识发现为基础的研究方法正在改变着人类的科学进程。近期,直接以数据为研究对象,以利用统计和机器学习方法来发现和提取各类交叉学科领域知识为目标的数据科学取得显著进展,其学科思想以及方法论已经广泛渗入各个主流科学研究领域。

目前,对于数据的认识、挖掘与渴求早已超出纯粹学术研究的范畴,从商业经济、公共管理到文化教育等社会生活的方方面面,通过发现数据、收集数据和分析数据,在数据资产上实现更强决策力、洞察力和业务优化能力的做法已经形成一股席卷全球的热潮。全世界进入以数据多样化(Variety)、海量性(Volume)、高速性(Velocity)和价值性(Value)为特征的大数据时代。

我国于2015年发布《促进大数据发展行动纲要》,提出要大力推动政府信息系统和公共数据互联开放共享,加快政府信息平台整合,消除信息孤岛,推进数据资源向社会开放,增强政府公信力,引导社会发展,服务公众企业;以企业为主体,营造宽松公平环境,加大大数据关键技术研发、产业发展和人才培养力度,着力推进数据汇集和发掘,深化大数据在各行业创新应用,促进大数据产业健康发展;完善法规制度和标准体系,科学规范利用大数据,切实保障数据安全。2016年《中华人民共和国国民经济和社会发展第十三个五年规划纲要》出台,我国政府明确提出实施国家大数据战略、加快政府数据开放共享等内容。

由此可见,我国对于数据资源价值与应用的关注已经上升到国家战略的

层面,而公共数据的开放与共享是国家行动和计划的重中之重。如何在国家层面上实现数据共享以及怎样提高数据共享水平一直是研究者和决策者所关注的热点问题,西方发达国家以及我国就此问题投入巨大的资源并进行了长期的数据共享探索、实践和相关制度建设。

在传统学术研究中,科学数据共享(Science Data Sharing)是指科学数据在数据使用者之间以各种形式进行的科学数据交流、分享和再利用活动。科学数据共享行为始于人类科学研究活动之初,既是公开透明的科学研究精神之自然体现,科学研究成果正确性检验的客观需要,也是人类从科学研究活动中获取巨大回报的必然要求。

如果我们放宽视野,从社会化大生产的角度仔细审视,就会意识到,数据共享究其本意是尽可能提高全社会对于科学数据再利用的规模和水平,不断从各方面挖掘其潜在价值,以求实现数据内在价值最大化。既然涉及数据的价值及其实现,从经济学视角考察,数据作为国民经济中各类主体科研创新活动的智慧结晶,作为数字经济时代微观经济活动的重要生产资料,数据资源共享问题的实质是资源有效配置问题。

科学数据共享发生和完成于多个具有独立权利要求的主体,这个过程必然伴随着数据跨越多个权利领域边界的流动,意味着数据这一信息财产的财产权及附属所有权利束的必然改变,意味着共享可视为一种特殊交易行为,同样受制于相关资源配置制度本身。基于此,数据共享活动中所形成的各种财产权利的归属形式及其各种权利的具体组合——与数据相关的产权制度,成为理解和改善数据资源共享效率的关键所在。

在本书中,笔者创新性地采用法经济学的分析思路,通过梳理各类数据资源的经济学属性及分布,分析数据所属各类主体及其特征,厘清附着于数据资源之上的各类产权制度特征,最终得到围绕数据共享的各种形式知识产权制度安排结构。而此结构正是进一步设计数据共享相关知识产权制度所需的基础。

前述研究内容主要立足于数据及其所依附的各类“制度空间”分析,而对于数据从产生、保护、读取、更改、迁移、存档到获取整个数据生命周期的分析则构成了本书展开的另一条线索——“时间演化”维度。

数据在各个时期内涉及的参与主体不尽相同,为数据生产的分工和社会

化创造了条件。数据作为数字经济时代最重要的生产要素,其生产过程的社会化、独立化、专业化是数字经济发展的基础。随之而来的结果是,数字经济时代围绕科学数据逐渐形成一个轮廓比较清晰的生态系统:由各类公共部门、私人部门以及社会公众组成,分化出数据生产者、使用和消费者、发布者(市场中介)、增值者等角色,发挥不同的功能和作用,拥有不同的利益目标。它们既互相依赖又存在竞争,彼此形成复杂交错的网络关系结构,数据资源通过关系网络在成员之间流动、共享、循环并增值。它们互利共存,共同维持系统的延续和发展,形成共生的生态系统。

基于此,本书认为,围绕科学数据社会化生产的所有部门与成员逐渐演化成宏大、复杂并且多样的科学数据生态系统。而维持一个健康的、生机勃勃的生态系统,需要满足以下要求:生产者在生物群落中起基础性作用;多样性是系统稳定发展的保障;系统的健康取决于各个子系统的协调运行、有序合作和各个成员的平衡发展。这一认识为本书设计良好的数据共享制度提供了极大的启发:对于系统中不同部门的作用,应当在整个生态系统的大背景下全面考量,追求利益与开放共享并不存在天然矛盾;制度安排应当以系统具有蓬勃的生命力为目标。

形成以上两个维度的理论研究框架之后,本书将研究的焦点转移到现实性问题之上。在参与国家科学数据共享工程的过程中,经过与国家科学数据共享工程各领域专家们多次深入访谈,在分析研究西方发达国家在科学数据共享实践中成败得失经验的基础上,本书系统整理消化了多种来源多个领域内的文献资料,对于各类共享制度安排形成明晰而透彻的认识,而这正是以往实践与研究中经常出现混乱冲突的地方。随着研究的深入,笔者发现:数据相关知识产权制度安排是决定数据共享能否顺利实现的最为关键部分,是解决数据共享问题的最重要着力点。合理的制度安排能够极大降低数据资源配置的成本,从而实现数据资源在全社会的高效合理配置。在科学数据生态系统中,各部分所扮演角色及社会属性,决定了其所适用的数据知识产权安排不同。可以灵活运用以信息公开法、传统和新型知识产权法律、合同法以及以荣誉奖励为核心的知识产权制度系统,通过产权制度安排“形式”上的保护,以保护促进数据的生产、发布、使用、流通和再利用,从而实现“实质”上高效率与低成本的数据共享,最终缔造一个富有生机与活力,不断创造社会与经济价

值的数据生态系统。

本书系笔者所主持的国家社会科学基金西部项目“国家科学数据共享战略的知识产权保护机制研究”后续成果。国家科学数据共享工程是我国国家科学数据共享战略的具体实现和重要基础,是我国科技发展基础条件大平台的重要组成部分,也是国家科技创新体系建设的重要内容。国家科学数据共享目标是在国家统一规划下通过整合离散的科学数据资源,构建面向全社会的管理与共享服务体系,以期实现对科学数据资源的高效利用。它是国家创新体系中急需发展的重要基础设施。本书的出发点,一方面试图为国家科学数据共享战略提供思路和机制设计;另一方面也是希望通过系统研究对我国未来大数据产业以及公共应用领域的制度设计进行深入探索。笔者希望能够起到抛砖引玉的作用,为相关领域的决策者、研究者和从业者提供若干有益的启发。

目 录

前 言	(1)
第一章 引 论	(1)
第一节 研究背景	(1)
一、科学数据与科学数据库	(1)
二、科学数据共享及其重大意义	(5)
三、国家科学数据共享战略的内容	(8)
第二节 科学数据共享亟待解决的问题	(12)
一、科学数据的生产与共享问题	(12)
二、多种因素共同作用下的系统性风险	(19)
第三节 国内外研究现状	(31)
一、国外研究现状	(31)
二、国内研究现状	(37)
第四节 研究目标和思路	(55)
一、科学数据的多重属性与科学数据共享	(56)
二、知识产权保护机制是科学数据共享目标实现的关键	(57)
三、法经济学视角下的科学数据共享知识产权保护机制	(60)
四、研究思路总结	(64)
第五节 研究方法	(66)
第六节 创新之处	(68)
第二章 科学数据共享的法律制度环境	(72)
第一节 政府信息公开法律制度	(72)
第二节 传统知识产权保护法律制度	(77)
一、著作权法	(82)

二、专利法	(90)
第三节 反不正当竞争法	(96)
第四节 数据库特殊权利保护法律	(103)
第五节 基于合同法的科学数据知识产权保护	(108)
第三章 科学数据共享系统分析	(119)
第一节 科学数据的产权结构	(119)
一、知识产品的公共领域、知识产权保护领域和准公共领域	(119)
二、各领域内的科学数据资源	(135)
第二节 科学数据的生命周期	(144)
第三节 科学数据共享的参与主体	(146)
一、公共部门	(147)
二、私人部门	(154)
第四节 科学数据共享系统	(156)
一、科学数据的生产社会化	(156)
二、科学数据生态系统与科学数据共享	(159)
第四章 科学数据共享的法经济学分析	(165)
第一节 科学数据的经济物品属性与科学数据共享	(166)
第二节 科学数据共享的经济学实质	(170)
一、科学数据的产权问题	(170)
二、产权理论	(172)
三、资源配置	(175)
四、科学数据共享问题的经济学实质——优化科学数据 资源配置	(180)
第三节 知识产权保护机制与科学数据共享	(183)
一、制度及其作用	(183)
二、科学数据相关知识产权制度	(185)
三、合同法、许可协议与知识产权保护机制	(192)
第五章 知识产权保护机制研究	(203)
第一节 科学数据生态系统中的成本分析	(203)
第二节 国家科学数据共享战略知识产权保护机制的总体目标	(212)

第三节 国家科学数据共享战略知识产权保护机制的构成 (217)

 一、政府信息公开相关法律体系 (217)

 二、政府科学数据共享管理政策 (222)

 三、科学数据共享许可协议体系 (228)

第四节 构建我国科学数据共享战略知识产权保护机制 (258)

 一、健全我国政府信息公开相关法律 (258)

 二、强化政府科学数据共享管理政策 (261)

 三、构建科学数据共享许可协议体系 (267)

参考文献 (276)

后 记 (288)

第一章 引 论

第一节 研究背景

一、科学数据与科学数据库

数据(Data)是对客观事物的符号表示,是用于表示客观事物的未经加工的原始素材,如图形符号、数字、字母等。或者说,数据是通过物理观察得来的事实和概念,是关于现实世界中地方、事件、其他对象或概念的描述,是构成信息和知识的原始材料。数据是反映客观事物属性的数值,是进行各种统计、计算、科学研究或技术设计的根本依据。具体而言,科学数据^①可以是人类收集、观察及创造的,具体表现为表格、数字、图像、多媒体或其他格式的各种实验数据、观测数据、统计数据、仿真数据;它也可以是论文之中附带的实验数据,还可以是独立的原始数据,如对数据进行描述的元数据、数据集以及与数据相关的出版物。

科学研究过程通常始于对某些问题的一个或一系列的假设,而假设的提出又往往以已有观察和数据为基础,之后便是通过设计实验来证实假设,最终实验结果以一系列数据的形式来证实或反驳假设。通过这样的科学研究方法,人类的知识不断增长,对于自然和社会的理解亦得以不断加深。从科学研究的一般过程来看,科学数据在人类科学发展过程中扮演了关键角色。科学数据既是科学研究活动开展的基础和源头,又是科学研究活动中最有价值的产物和输出。目前,人类科学研究活动越来越依赖于科学数据的创造和利用,科学研究的范式发生了转向数据驱动的革命,以科学数据、数据挖掘和知识发现为基础的研究方法正在改变着人类的科学进程。

① 黄永文等:《国外开放科学数据研究综述》,载《现代图书情报技术》2013年第5期。

科学数据是国家创新体系中的基石。从社会化大生产的角度观察,科学数据是数字经济时代中微观经济活动的重要生产资料,是国民经济各类主体科研创新活动的智慧结晶。知识经济时代,各类经济活动主体无论其背景如何,都逐渐主动或被动地参与到围绕科学数据产品开展的生产和共享活动中,科学数据的价值与内涵也就不断地被挖掘拓展。以经济领域中的证券业、保险业和银行业为例,数据是它们业务活动中的核心内容,创造、组织、积累和利用关于人类经济活动的海量科学数据,是其业务活动的永续主题。对这些数据的量化研究以及在此基础上进行增值服务的创新活动,在主观上有助于它们探索和发现新的商业模式,从客观上也极大地丰富和拓展了人类关于自身社会经济活动的知识。

另外,从科学探索的宏观视角来看,数字经济时代日益多元化的科学研究活动主体大大开拓了科学活动的疆域,人类科学活动不再局限于公共部门资助的大学和研究机构,追逐着超额利润的私营部门凭借其敏锐的商业触角,正置身于科学活动的前沿,并书写着人类探索未知领域的精彩画卷。

与之相伴的现象是,科学数据已经成为数字经济时代最基本的生产要素,科学数据的共享与再利用活动早已超越了公共部门间的内部协作,成为公私部门以及社会公众的日常行为。因此,现代科学数据的内涵得到了极大丰富,数据的质量、数量、种类都快速增长,科学数据共享已经成为全社会各个层面参与社会政治经济生活的客观要求,参与科学数据共享活动的主体前所未有地丰富。

在这种情形下,本书将各类产权归属复杂的科学数据纳入视野,以科学数据生态系统的系统论视角去观察和分析共享活动。反之,如果仅将科学数据视为狭义的公立研究主体之所得,势必将研究视角限制得过于狭窄,从而将大量重要数据共享活动和研究对象拒之门外,导致窥一斑而无法知全豹的错误,更遑论这与国外相关研究的通行做法不符。因此在本书中,科学数据这一概念不仅仅描述了公共资助科学研究活动所产生的数据,还包含由各类社会组织机构以及社会公众所创造的形式多样的诸多科学数据。

一般而言,科学数据具备如下关键特征:

1. 正外部性。科学数据是一种特殊的信息。而信息本身具有无形性、易于扩散传播,并具有非排他性和非竞争性的经济特性,其公共物品的属性极为

明显,存在很强的正外部性效应。因此就科学数据而言,无论其由谁生产,也无论其产权制度安排如何,其对于全人类福祉的溢出效应都是绝对存在的。但是,要保证科学数据正外部性作用的良好实现,就必须借助设计合理的制度安排。唯有通过恰当可行的制度安排,才能在将科学数据所蕴含的巨大价值转变为现实中实际价值的同时,保护科学数据创作者的合法权益,并促进科学数据的社会化生产,最终实现各方主体利益平衡以及全社会福利的提升。任何过分强调单方面利益的做法都会损害科学数据共享活动的长期利益。

2. 价值的潜在性。与有形的物质财富明显不同的是,科学数据的价值不会随着使用次数的增多而逐渐耗尽。恰恰相反,随着科学发展和人类认识的深入,科学数据所潜藏的、以前未引起充分注意的科学现象才会得到研究,其潜在的重要性和价值才能逐渐被认识。不仅如此,随着来自不同领域之多种科学数据的联合使用,人类未曾探索过的科学新领域的大门不断打开,未知的科学现象相继被人类发现并了解,而其前所未有的商业应用也随之得以开发。科学数据的使用价值,因为广泛地整合使用而得到指数式的增长。例如,蛋白质排序数据与 X 射线晶体学相结合,令其价值得到极大提升。又如地理信息系统,整合了多种来源的数据,为人们提供了深刻了解自然和建设环境的新契机,极大丰富了人们对所生活的地域以及周边环境的新认知。地理信息系统衍生出诸多的重要商业应用,包括环境管理、城市规划、路径规划和导航,紧急预案与应对,农产增产等。因此,各类科学数据再利用活动,对于科学、经济和社会意义重大。

此外,某些科学数据还具有唯一性。很多科学研究对象所伴随的物理现象或社会情景,因为时间的流逝而不可再现,无法在事后重新获取,因而具有极其重要的科学价值,比如关于自然界的大多数观测数据。此类科学数据往往只能从单一来源获取,因此,其生产者对这些数据具有天然的独一无二的垄断性,其他任何个人和组织都无法替代其垄断地位。另外,通常情况下,用于采集科学数据的基础设施建设所需成本巨大,以致很多科学数据的获取成本极其昂贵,例如,借助大型仪器设备产生的空间观测数据,这进一步强化了生产者的科学数据垄断地位。

随着人类信息技术的发展和认识世界水平的不断提高,科学数据的范围与来源飞速拓展,人类进入了前所未有的大数据时代。事实上,自计算机发明

以来,科学数据的收集、操作和发布过程几乎完全数字化,自然界和人类社会活动的方方面面都能够以数据的形式观察和记录。随着数字化时代的到来,以前无法认识或者被忽视的领域,自然而然被纳入科学研究的领域内,因而,对相关现象的观察记录也自然成为科学数据。伴随着人类科学的进步,海量科学数据亟待积累、整理和重新使用。这些数据资源方法,对于科学研究进程至关重要,为了对其高效管理和使用,数据库技术应运而生。

所谓科学数据库,是人们为了解决特定的科学问题,把原始分散的相关科学数据,按照一定的组织方式存储在一起,所形成的用途广泛、结构化的科学数据集合。科学数据库的建立通常会涉及以下内容:将连续的科学观测记录加以精确地数字化;将科学数据组织为有适当结构、格式、描述和编排的数据库;为科学数据库创造必要的支持和分析软件;采用适合的质量保证和质量控制措施;采用一定技术手段对数据库进行发布以供再利用;对科学数据库的备份与维护。随着科学数据库规模的增长,完成这些工作需要投入越来越多的各种资源,并且构成了科学数据库成本的主要部分。此外,还存在着一类非结构化的科学数据,以专利、科学文献以及某些实验记录为载体。与结构化的科学数据库相比,非结构化科学数据,零散分布于载体中、缺乏良好的组织与排列,对其进行共享再利用要耗费更多的资源与精力,也更难以对其进行增值开发,一般而言,往往更缺乏通用性。但是这类科学数据不仅覆盖的学科门类更为庞大,而且存在数量也颇为可观,同样蕴含着重要的科学价值及商业价值。

随着人类科学进程中各类科学数据库的建立和运行,对各类科学数据库所存储的海量信息进行数据挖掘和知识发现的研究方法,已经成为现代科学研究活动不可或缺的重要组成部分。在数据驱动的前沿领域,每天都有新的模式、规律和理论,被发现并被成功转化为商业应用,这类融合科研和商业的创新性研究活动日益成为数字经济时代的发动机。人们对于科学数据库不断增长的渴求,激发了对于科学数据库商业化投资开发的热潮。科学数据不仅是基础性科学研究的起点和新信息,也是各类新型商业应用的基础,是知识经济时代重要的无形财富。由此,世界各国关于科学数据产权化的论调一时甚嚣尘上,数据库财产权立法保护的呼声日益高涨,私人部门,甚至公共部门,对于其所建设科学数据库的保护手段也层出不穷、不断强化。在这种纷乱复杂的局面下,作为人类公共财富的科学数据应该何去何从,应该如何保护、发展

和再利用,已经成为科学数据建设和使用中亟待解决的头等大事。

二、科学数据共享及其重大意义

科学数据共享,是指在数据使用者之间,以各种形式进行的科学数据交流、分享和再利用活动,其目的是提高全社会对于科学数据再利用的规模 and 水平,实现科学数据内在价值最大化。科学数据共享行为始于人类科学研究活动之初,是公开透明的科学研究精神之自然体现^①,是科学研究成果正确性检验的客观需要,亦是人类从科学研究活动中获得巨大回报的必然要求。目前,世界各国政府相继投入大量资源,建立包括从硬件、软件、数据到制度规范在内的各类公共科学数据中心,以多种模式为科学研究、商业活动及其公民,提供开放的数据共享服务。各类公立、私立的研究组织、商业企业及研究者之间开展了灵活多样的数据共享合作计划。各国政府间也就关系人类前途命运的各类重大科学问题开展了长期持久的大科学数据共享计划。人类迈向数字经济时代的今天,科学数据共享已经成为国家信息化战略的重要组成部分,在国计民生中发挥着巨大的基础性作用。

伴随着技术进步和时代发展,人类对于科学数据共享的认识也呈动态的螺旋式上升,不断地拓宽和升华。科学数据共享运动的蓬勃发展,始于 20 世纪 60 年代西方国家一系列大科学研究项目。国外学术界对于数据共享活动的认识理解,从概念的提出到内涵的完善,也经历了较长的时间,以至数据共享这一概念并行存在着多种不同术语的描述,比如常见的 Data Open、Data Open Acces、Data Reuse 以及 Data Sharing,这些概念所描述的内涵及其侧重有着很多或细微或明显的差异,不仅体现了数据共享活动的不同层面,也带有时代的烙印,如果不加分辨必然会为研究者带来诸多误解。因此,在此处首先要特别指明,本书中的“数据共享”一词基本等同于 Data Sharing,指一切形式的数据共享与再利用行为,不仅包括自由开放、完全免费的公益性共享行为,也包括商业性的科学数据再利用行为。其次,科学数据共享有别于经常见诸于报端的技术性的数据共享,本书中科学数据共享是指各类主体的科学数据资

^① 科学的精神气质,它的特征是 CUDOS,即公有主义(Communalism)、普遍主义(Universalism)、无私利性(Disinterestedness)、独创性(Originality)、怀疑主义(Skepticism)。Ziman, John (2000), *Real Science: What It is, and What It Means*, Cambridge: Cambridge University Press.

源在全社会实现优化配置的过程;而后者主要是指在各类组织内部通过技术手段,促进生产、经营、管理性数据在信息系统间的流动和再利用,消除信息系统中数据的重复、冗余和不一致现象的行为。最后,还必须特别注意到,科学数据共享这一概念实际代表了错综复杂的科学数据生态系统的数据流动行为,它是在由法律、政策、经济、具体学科等多个维度构成的社会空间中的动态演进过程,对其了解、调整和优化,必须结合这些背景因素。

随着人类社会进入数字经济时代,科学数据共享的意义和价值显得尤为重大,我们结合国内外最近的研究资料,从科学、经济与社会这三个基本方面对其总结如下:

1. 科学数据共享是科学研究活动的重要基础。首先,通过对科学数据共享,可以有效避免重复投入和重复劳动,加速科学研究的进程,使科学家的研究工作可以快速建立在前人的研究成果之上。还能够增进科学界同行间的交流与合作,促进同行业具备各种研究条件的组织间合作、交流与互补,促进不同领域内科学家的合作,从而获取更具创新性的研究成果,提高整个科学界的研究水平。其次,开放共享的科学数据,是大型基础科学研究活动的必要条件。基础研究主要指对自然界如何运作以及自然界的诸多方面彼此如何连接运行的研究。人类社会使用这些研究成果来拓展人类的知识空间,并应用这些知识,以各种各样的方式创造社会财富并提高公众的福利。自 20 世纪 60 年代以来,随着基础科学领域全方位的拓展,以投资强度巨大、多学科交叉、依靠昂贵复杂的大型试验设备、研究目标宏大为特征的大科学(Big Science)项目大行其道,这些研究大大地丰富了人类对世界的认识。科学家个人间、研究机构间以及政府间的合作对于这些项目的成功至关重要,而合作的基础正是研究过程中资源的开放与共享。以国际人类基因组计划为例,中国完成了占全部工程 1% 的人类 3 号染色体短臂上约 3000 万个碱基对的测序任务,并与世界各国共享了整个人类基因组数据,获得了巨大的回报与成果。再次,对于科学研究者个人而言,通过数据共享可以促进对所在领域的理解,便于使用他人的研究数据来验证和改进自己的工作。最重要的是,数据共享对于提高自己的学术知名度极为有利——公开数据会吸引其他研究者来了解自己的工作、引用自己的研究成果、合作撰写研究论文并增加与本领域内的学术权威合作的机会。最后,科学数据共享是促进对科学研究成果的质量与价值进行检

验和评价的重要基础,是维护学术道德规范的重要保证,亦是公众参与社会政治生活和进行重大科学项目决策的重要依据。以“气候门”(Climate Gate)为例^①,在关于“全球气候是否变暖”这一与人类命运息息相关的重大科学问题上,2009年11月,多位世界顶级气候学家的邮件和文件被黑客公开,同时被指控涉嫌扭曲、隐藏关键气候数据,伪造气候变暖的假象,这一事件严重损害了政府间气候变化专门委员会(IPCC)的声誉。在经过官方调查之后,调查人员虽然肯定了研究成果的真实性,但也严厉批评了气候科学家的保密文化,认为这些科学家“未能展现适当程度的开放”,对外界索取数据的请求不愿提供帮助。

2. 科学数据共享是数字经济产业链的起点,科学数据的商业化开发和使用是未来经济增长的关键助推器。数字经济是信息和商务活动均已数字化的全新社会政治和经济系统。随着数据资源的开放与共享,越来越多的资本和技术会投入到挖掘和转化这些数据资源潜在价值的活动中,各种各样的商业应用逐渐构筑于其上;不仅如此,这些活动同时也将产生新的数据资源输出。这样一来,数据源源不断地从新的交叉领域生长出来,创新应用从错综复杂的数据资源联系中诞生和成长,一个能够自我更新与进化的数据生态系统也就随之不断演化。2010年欧盟委员会对2030年科学数据应用远景描述为^②:随着数字经济的飞速发展,物理和技术基础设施将变得不可见,数据本身将成为一种基础设施,对数据的无缝访问、利用、再利用和信任将变成最基础的行为。美国年度健康基础数据估值为3500亿美元这一事实,为以上看法提供了强有力的支持^③。事实上,超出欧盟预期的速度,随着传感器和云计算技术的飞速发展,以数据量巨大、数据类型繁多、价值密度低为特征的大数据(Big Data)时代已经来临,科技、金融、电信、工商、交通、卫生等行业正在实时生产,并积累着大量的科学数据资产,对在线问卷数据、交通气象数据、消费者超市购物

① 《英国“气候门”丑闻调查结论出炉》,资料来源:<http://www.ftchinese.com/story/001033464>。

② Riding the wave-how europe can gain from the rising tide of scientific data [EB/OL]. <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>.

③ Here Comes the Data Economy [EB/OL]. http://www.slate.com/articles/technology/future_tense/2012/09/brightscope_castlight_new_businesses_built_on_open_government_data_single.html.