

开放数据：大数据时代的发展趋势研究

邢超¹，石玲²

(1. 中国国际核聚变能源计划执行中心，北京 100862；

2. 中国科学技术部，北京 100862)

摘要：在“数据革命”时代，数据的自由流通最为关键。本文以创新视角分析了大数据背景下开放数据的概念、演变、原则等，特别指出科学数据开放对支持创新有重要意义。认为开放数据是世界发展趋势，没有开放数据，就没有创新，也就失去了发展的机遇。最后针对中国开放数据运动提出政策建议。

关键词：开放数据；大数据；创新与创业；科学数据开放

中图分类号：G203 **文献标识码：**A **DOI：**10.3772/j.issn.1009-8623.2016.10.012

依托海量的数据状态和云的数据处理能力，大数据以其大量（Volume）、高通量（Velocity）、多样性（Variety）、真实性（Veracity）的4V特点，带来了一场全方位的变革。在“数据革命”引发的开源世界里，数据的自由流通最为关键，必须打破数据孤岛。为此，开放数据运动应运而生。

1 开放数据的基本概念和发展

根据英国开放知识基金会（Open Knowledge Foundation）的定义，开放数据的概念需具备以下三个基本元素^[1]：

（1）非歧视性：数据若开放，则其对任何人都开放；

（2）机器可读性：数据若开放，则应是机器可读格式，例如：对于表格数据，应采用.csv格式，而非.pdf格式。

（3）开放授权性：数据若开放，则其对应授权条款应确保使用者自由免费访问、获取、使用、增值、演绎、复制、传播的权利。

基于上述元素不难看出，相较于数据共享而言，开放数据更秉承着开源世界所倡导的平等、自

由的价值观。开放数据强调的非歧视性和开放授权性，打破了传统数据共享中设定的“共享条件”和“特定共享方”的限制。

相对于政府一直推行的信息公开而言，开放数据所强调的机器可读性及其明确赋予数据使用者的自由增值利用（包括商业使用和非商业使用）和分享传播权利，更好地刺激了公众对政府数据资源的需求，并鼓励公众对政府数据增值利用。

美国联邦通信委员会（FCC）消费者和政府事务局局长乔尔·古林在《开放数据》一书中对开放数据进行了描述：开放数据是指公众、公司和机构可以接触到的，能用于确立新投资、寻找新的合作伙伴、发现新趋势，作出基于数据处理的决策，并能解决复杂问题的数据。

也就是说，开放数据不同于大数据，也有别于信息公开和数据共享，虽然它们之间确实有所重叠。开放数据的宗旨是提供免费、公开和透明的数据信息，这些数据能适用于任何领域，如，政府运作、商业经营等。开放数据本身并没有明显的商业价值，但经过公众和公司加工之后，可能产生巨大的商业价值。

第一作者简介：邢超（1975—），男，经济学博士，主要研究方向为经济管理，科技管理。

收稿日期：2016-10-09

数据开放的法律依据来源于政府资助行为必须公开、透明。以美国为例,开放数据运动最初完全是政府透明化在推动,政府透明化也是其原动力。这使开放数据运动从动力源泉上就超越了为提高数据使用效率、减少重复而提出的数据共享运动。美国最早有关开放数据运动的法律是2009年奥巴马总统签署的《开放政府指令》(Open Government Directive)^[2],旨在使民众更好地监督政府运作的方向和运转的方式,推进开放数据的发展。2014年美国又进一步推动了《数据法令》(Data Act)的颁布,全面推进了数据的开放。之后,美国政府,包括美国政府参与的国际组织,凡属美国纳税收入支持的机构与活动都必须公开、透明^[3]。

与此同时,美国重视有关数据开放生态的建立,积极探索数据开放的商业潜能,培育各类机构参与数据增值链条^[4]。2013年,国际知名咨询机构麦肯锡在其发布的研究报告《开放数据利用流动信息解锁创新与性能》(Open Data Unlocking Innovation and Performance with Liquid Information)^[5]中指出,全球每年开放数据产生的经济价值高达3万亿美元,其中不仅包括创造的新财富,也包含了节省的成本,显示出开放数据对经济社会的重要积极作用。

2 大数据时代的开放数据

人类即将进入数据时代。大数据时代是海量数据同完美计算能力结合的结果^[6]。移动互联网、物联网产生了海量的数据,大数据计算技术完美地解决了海量数据的收集、存储、计算和分析的问题。大数据时代开启了人类社会利用数据价值的新时代:

(1) 大数据提高了数据处理效率,增加了人类认知盈余,大幅度提高了人类改造世界的效率,加快了人类的进步。

(2) 大数据通过全局的数据让人类了解事物背后的真相。大数据使用全局的数据,不再拘泥于以往样本代替全体的统计方法和思路,更接近事物真相。

(3) 大数据有助于了解事物发展的客观规律,助力科学决策。利用大数据提供的分析结果有助于发现事物的发展规律,帮助人们进行科学决策。

(4) 大数据提供了相同事物的连接,有助于

客观了解人类行为。拥有了大数据技术,各类传感器传输反映人类行为,人类行为也更体现为数据表现,便于人类全面、客观、数据化地认识自我。

(5) 大数据改变了过去的经验思维,帮助人们建立数据思维。摒弃过去的经验思维和惯性思维,掌握客观规律,跳出了历史预测未来的困境。

与传统相比,大数据时代从思维、方法到技术手段都有明显不同,其中最明显的特点是大数据生态圈及其拓展的开放性。在大数据全链条中,无论是数据的收集、存储还是处理、分析,都是以开放数据作为大数据时代的源头之水。没有真正意义上的开放数据,大数据就只是大,缺乏多样性、持续性、整体性和普遍联系性,无法对原始数据进行加值使用,不具备大数据时代所赋予人类的真正意义上的数据革命,无法产生变革性的社会发展。

3 大数据时代的科学数据开放

21世纪以来,科学研究的复杂性、开放性、交叉性愈加显著,使得科学研究进入了“大科学”时代。大数据让“大科学”面临更多的挑战与机遇。

大量科学研究越来越多地依托海量科学仪器、望远镜、卫星、传感器网络等各种科研仪器和装置,以及模拟方法、智能终端的应用,这些都使科研领域推动了一个前所未有的“大数据”时代。海量数据的剧增,量变引发质变,也直接带来了科研模式的改变。继实验科学、理论科学、计算科学之后的第四种研究范式,即“数据密集科学”,成为大数据时代的新模式。

2015年12月科技国际会议在南非召开,会议主题为数据革命的机遇与挑战,提出数据开放将是科学创造未来的重要前提。大会发布了《大数据时代的开放数据》(Open Data in a Big Data World)^[7],倡议全球在科学研究和政策制定方面推进数据开放,以应对大数据时代的变革与挑战。科技国际会议由国际科学理事会(International Council of Science, ICSU)、国际科学院联盟(InterAcademy Partnership, IAP)、世界社会科学理事会(International Social Science Council, ISSC)和世界科学院(原名发展中国家科学院、第三世界科学院, World Academy of Sciences for the Advancement of Science in

Developing Countries, TWAS) 四个国际组织发起, 汇集了全球和地区 250 多个学术团体和组织, 共商大数据时代的科学数据开放。

《大数据时代的开放数据》这项倡议包括一系列开放数据的导则、创新链条中数据开放的推进与保护, 以及发展中国家参与开放数据的全球创新。

3.1 大数据时代的机遇

近一二十年的数字革命对现代社会产生的深刻变革远胜于古代的印刷术。无论是数据的获取、存储、制造还是广泛传播等方面, 在数量和速度上都有极大的提高。如 2003 年, 科学家耗费近 10 年时间、耗资 1 亿美元才完成了人类基因图谱, 而现在仅需要几天时间和 1 000 美元左右就可以完成^[8]。这样的变革刚刚开始, 就已经给人类经济社会生活带来了翻天覆地的变化。科学工作者作为这个变革的最前沿者, 已经体验到并将主导这种变革的发展^[9]。

互联网作为这一革命的平台, 不仅创造出了巨大的数据, 而且加快了数据的传输, 形成了“大数据”的一些基本特性, 如大量 (Volume)、高通量 (Velocity)、多样性 (Variety)、真实性 (Veracity) 的 4V 特点。

数字革命的核心特性是链接性。近几个世纪的科学进步主要来源于单元体、非耦合, 或弱耦合的系统。但是, 随着社会和经济的纵深发展, 全球机遇与挑战更多地需要用高耦合系统、整体性的科学研究方法来予以解决。只有通过成本更低、效率更高的大数据能力, 才能给人类带来更大的变化。

3.2 积极应对挑战, 数据开放势在必行

(1) 保持自我纠偏能力, 实现“智能开放”

数据仅是简单开放是远远不够的, 还需要保持自我纠偏的能力, 实现“智能开放”, 需要在以下五个特性方面进行完善: 可发现性, 网络搜索可轻易发现其存在; 可获取性, 数据能够被电子化输入或被电脑获取; 智能性, 有足够的背景信息; 可评价性, 用户可评价数据的适用性; 可用性, 有足够的元数据, 能够被用于创建派生数据和相关的代码, 并可与计算机的特性结合起来, 可被访问。

(2) 数据开放要适应科研规律

对于那些依赖高通量科学仪器等获取的科研数据, 渐进实现科研认知的领域要有不同的数据开

放的要求。数据开放要遵从和完善科研道德约束, 尊重科研智力创新, 它与数据公开推动创新同等重要。

(3) 数据开放要实现全球广泛参与

大数据和数据开放具有惠及发展中国家特别是最不发达国家的巨大潜力。但最不发达国家的国家研究体系通常存在资源不足的情况, 无法参与基于大数据和数据开放的全球科研合作, 这种鸿沟在近年来愈发明显。新兴国家和发达国家要积极帮助发展中国家特别是最不发达国家实现数据开放, 推动数据驱动型发展 (data-driven development)。

(4) 抢抓机遇, 实现有效的数据开放

抢抓机遇, 只有在国际层面、区域层面、国家层面乃至个人共同努力, 才能实现有效的数据开放。而致力于数据开放的国际组织, 如: 国际科学理事会、科学与技术数据理事会、研究数据联盟等, 它们的首要任务是推动和支持国际上数据开放获取、互通和可持续发展。

(5) 开放数据推动开放科学, 开放科学创造公共知识

认识到要解决当今世界的许多问题, 亟需加强科学界与更广泛社会各界的对话与合作, “开放科学”的理念应运而生。科学研究的真正内涵在于其作为公共产品应造福社会和人类。无论是公共部门还是私营部门, 都将从数据开放中获益。

3.3 推进开放数据的主要方面

数据出版是一个新概念, 是通常意义上科研学术文章出版的更高阶段。数据出版是数据开放的一种主要方式。无论是科学家自身, 还是出版商、学术资助机构、科学家所在科研机构 and 学术团体都应在数据开放和数据出版中承担各自的责任^[7]。

(1) 参与主体职责

· 科学家有责任、有义务认为科研数据是科研进展的有机部分, 在科研进展得到发表的同时, 推进科研数据的出版, 并确保低门槛获取和有效归置。

· 科研机构应营造支持开放数据的良好环境, 包括完善科研数据的管理服务, 提供数据管理培训, 建立数据管理的标准与措施激励等。

· 出版机构在审查过程中需注重数据审查, 并积极推进元数据出版, 为下阶段科学研究分析提供原始文本和数据。

· 资助机构在研究评价过程中,要把数据引用、数据创造贡献等纳入评价标准。资助内容要包括弥补数据开放出版过程中的成本。

· 专业学术机构应制定准则和政策,推进开放数据,并广泛地对开放数据准则和政策进行宣传推介。

· 图书馆要建立数据储存的标准,确保其长期可获取,并提供相应的开放数据服务。

(2) 数据开放的边界

除涉及隐私、安全、国家机密以及商业秘密的数据外,其他科学数据都应尽可能地开放。

(3) 数据开放的行为

· 数据引用。在学术出版中,需要对引用部分注明资料来源出处。

· 互通性。研究数据和元数据都可以实现互通验证。

· 无受限多次使用。数据出版后,没有限制性措施限制多次使用。

· 可连接性。开放数据,应可最大限度地与内容进行连接使用。

4 启示与建议

数据是政府掌握的核心资产,也是长期以来政府治理国家的重要工具。在大数据时代,开放数据运动对中国借鉴意义极强。我国政府积极推进政府改革,其核心是简政放权,增加政府透明度,实现公众对政府的监督,释放民间创新活力。无论是政策还是技术手段都需要开放数据的支持。一方面可因数据开放透明而减少成本开支,另一方面可激活新兴创新市场。当下“互联网+”的背景下,开放数据所激活的新兴创新市场实际就是在做“开放数据+”的工作,通过结合开放数据,一些传统的行业得到改变,而另一些创新的应用服务也成为可能。

“大众创新、万众创业”都可以基于数据开放实现更多、更新的业态,更全面地推进创新与创业。与此同时,开放政府数据(国家安全数据、保密和隐私数据除外)对催生创新和经济增值,推动数字经济发展,促进我国经济增长,提升我国的全球竞争力,创建一个更加富有责任感、更高效、更透明的政府具有重要的意义。

(1) 数据开放,势不可挡

在大数据时代,移动互联网、物联网将每个人

都变成了数据的提供者 and 应用者。开放的数据才是有生命力的数据,才可以推动和维持数据的可持续性。不孤立、普遍联系的数据,才可以实现大数据的互联互通,才使数据具有了灵魂。是数据开放才给予大数据时代生命和灵魂。数据开放是大数据时代到来的前提。

(2) 数据开放,是未来大数据时代竞争力的基点

开放、共享是大数据时代的主旋律。爱尔兰剧作家萧伯纳说过:“你有一个苹果,我有一个苹果,我们彼此交换,每人还是一个苹果;你有一种思想,我有一种思想,我们彼此交换,每人可拥有两种思想。”在大数据时代的竞争中,数据开放不仅是数据价值链条,是数据产生、获取、分析、应用和存储的基点,也是经济社会与科技创新理念的基础。

(3) 数据开放,法令先行

无论是科研数据中对于知识产权的保护和尊重,还是经济社会公开数据中涉及隐私和公共安全的内容,都需要通过制定法令来给予约束。推进政府信息首先实现数据公开更需要国家法律予以确立。研究和借鉴发达国家在数据开放中的法律法规体系建设经验,结合我国实际,加快制定出台符合数据开发和大数据发展规律的数据开放制度法规是当前的重点任务。

(4) 数据开放,政府先导

公开、透明是数据公开的原动力。追溯信息公开、数据开放的发展历程,政府接受来自公众的监督,始终是数据开放的核心。政府是一国最大的数据生产者和拥有者。在政府数据爆炸式增长的同时,公众的数据意识和使用能力在迅速提升。政府开放数据已是大势所趋。在这样的发展背景下,通过商业开发积极推动和科技手段的催化,大数据时代下的数据开放才具有最真实的意义。

(5) 中国开放数据必将充分发展

尽管欧美早于中国完成了数据门户(Government Data Portal)建设、政策和组织框架构建、商业开放和标准制定等工作,但我国大力推进的政务公开、“互联网+”“大众创业、万众创新”都将给我国数据开放提供更高的需求和不懈的发展

动力。目前,我国政府正从经验决策向数据决策转变,从行政主导向公众参与和服务政府转变。开放数据有助于促进政府改革,提升政府治理能力,对我国经济社会发展具有重要意义。■

参考文献:

- [1] Open Knowledge Foundation UK. How to Open Data[EB/OL]. [2016-08-26]. <https://okfn.org/opendata/how-to-open-data/>.
- [2] 高丰. 开放数据: 概念、现状与机遇 [J]. 大数据, 2015 (2): 9-18.
- [3] 陆健英, 郑磊, Sharon S. Dawes. 美国的政府数据开放: 历史、进展与启示 [J]. 电子政务, 2013 (6): 26~32.
- [4] The World Bank. Open data for economic growth[EB/OL]. (2014-06-25)[2016-08-27] <https://worldbank.org/content/dam/worldbank/document/Open-Data-for-Economic-Growth.pdf>.
- [5] McKinsey Global Institute. Open data: unlocking innovation and performance with liquid information[EB/OL]. (2013-10-2) [2016-08-27]. http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with+liquid+information.
- [6] 曹凌. 大数据创新: 欧盟开放数据战略研究 [J]. 情报理论与实践, 2013 (4): 118-122.
- [7] Science International. Open data in a Big data world[EB/OL]. (2015-12-07) [2016-08-28]. http://www.science-international.org/sites/default/files/reports/open-data-in-big-data-world_short_en.pdf.
- [8] CODATA. Current Best Practice for Research Data management Policies[EB/OL]. (2014-05-21) [2016-08-29]. <http://dx.doi.org/10.5281/zenodo.27872>.
- [9] Research Councils UK. RCUK Common Principles on Data Policy[EB/OL]. (2014-03-21) [2016-08-29]. <http://www.rcuk.ac.uk/research/datapolicy>.

Open Data: Research on the Development Trend of Big Data

XING Chao¹, SHI Ling²

(1. China International Nuclear Program Execution Center, Beijing 100862;

2. The Ministry of Science and Technology of China, Beijing 100862)

Abstract: In the Data Revolution era, the free flow of data is critical. The paper analyzes the concepts, revolution and principles of Open Data in the Big Data environment from the perspective of innovation, points out that Open Science Data is extremely important to support innovation. It believes Open Data is the world's development trend. Without Open Data, there will be neither innovation nor opportunity for development. Finally it gives suggestions for China on how to practice on the Open Data.

Key words: open data; big data; innovation and entrepreneurship; open science data