“数据化”的社会与“大数据”的未来

□ 韩 晗

社会的“数据化”（datazation）是数字时代人类社会必然的发展趋势，而“大数据”则是观察、认识与分析现在及未来社会的重要窗口。《现在的大数据》从“大数据”的概念出发，阐释了“大数据”对不同领域的作用与价值。但该书对作为社会变革动力与人类社会结构组织形式的“大数据”仍研究不足，它忽视了“大数据”的人文本质以及带来的社会问题。因此，在未来如何实现全社会共享作为公共资源的“大数据”，并使其打破不同学科与知识结构之间的壁垒，成为促进社会管理、决策的重要工具，才是“大数据”今后研究的重中之重。

一

“大数据”（big data）是近两年学界、业界普遍关注、常谈常新的一个热门话题，尤其在2013年，对这一问题的关注热度几乎到达峰值，据不完全统计，在该年度内，国内外关于大数据的论文、调研报告总量已经超过了5万篇，相关图书亦有近百种。

由数据统筹、智能开发与信息管理领域国际权威机构奥莱利（O’Relily）传媒机构主编并出版的《现在的大数据》（*Big Data Now*）出版于2012年年底，是笔者所见迄今为止关于“大数据”问题相对最为全面（涉及面广）、客观（并非一位作者完成，而是收集了各类不同作者的稿件或对话录）的著述之一。尽管《现在的大数据》中的作者多半为奥莱利传媒机构的高层人员，但也从一个独特的视角反映了美国学界对于“大数据”的代表性看法。

在书中，编者从大数据的定义出发，阐述了大数据的采集、应用、利弊以及与公共卫生的关系，认为在当下的语境下，大数据作为趋势早已不可避免。研究者与决策者务必要打破既定的学科知识壁垒，使得对大数据的开发、使用成为社会资源配置与发展的动力。

与其他关于“大数据”的研究成果一样，《现在的大数据》关于“大数据”的定义并未有特立独行之处，而是选用了艾德·丹彼尔（Edd Dumbill）[1]的一

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.ne

篇文章，认为大数据有大量（volume）、高速（velocity）与多样（variety）的特征，与迈尔·舍恩伯格（V. Mayer-schonberger）与肯尼斯·库克耶（Kenneth Cukier）相比，丹彼尔并未谈及大数据所具备的“精确”（veracity）特征。而且通观全篇甚至全书，“精确”这个单词并未出现过一次，这恰恰反映了《现在的大数据》对于“大数据”的一个核心观点：不确定性。

作为一个处于探索期、建设期的概念，大数据的精确性只是相对的，即作为工具的大数据可以有效地帮助决策者迅速、准确地寻找到既定目标，但这并不意味着“大数据”自身存在着精确性。在丹彼尔看来，大数据的基础是“云”（cloud），即一种涵盖整个互联网体系的庞大技术平台。

与此同时，杰瑞米·霍瓦德（Jeremy Howard）[2]在《大数据的工具、技术与策略》（该书第三章第一节）中，认为大数据的意义核心在于可以导致决策者最优选择，并认为这种最优选择可以给企业带来最为直接的利益，而这种优选的前提是制造出一个相对客观有效的模型。阿瑞亚·哈格海格（Aria Haghighi）在该章最后一段引用杰瑞米·霍瓦德的一句话便是对这一章最好的总结：“建立一种用于优化的预测性模型（Predictive Modelling），将会是下一个前沿话题。”[3]

该书第四章“大数据的应用”其实并无出彩之处。这一章主要由马克·斯洛卡姆（Mac Slocum）的《超越电子表格的传奇》[4]、阿拉斯达尔·阿兰（Alasdair Allan）[5]的《挖掘天文学的文献》以及迈克·鲁克德斯（Mike Loukides）[6]的《数据的阴暗面》三篇文章组成，中间穿插有罗伯特·辛普森（Robert simpson）的访谈。

该章用翔实的图表与数据来证明大数据的具体作用。一是协助企业在面对海量的、即将爆炸的数据库该做出何样的处理决策，二是大数据在天文学中的应用。实际上这两个观点在“大数据”领域早已是老生常谈，因为“大数据”的概念最早源于计算机、天文学与基因工程这三个理工科领域，而企业管理、客户选择则是近两年“大数据”最为普遍的应用。而这一章与该书的最后一章《大数据与医疗卫生》亦有异曲同工之处，后者旨在强调公共卫生管理中大数据应该扮演一个信息及时收集的角色，进而避免疫情的传播。实际上，早在2009年“甲流”在全球泛滥时，国际卫生机构就已经联合谷歌（Google）公司的用户搜索在全球定位“甲流”患者的可能存在处。在此之后的三年里，关于大数据在公共卫生领域（包括制药、医疗不同领域）里的应用层出不穷，尤其对于流行病、传染病、地方病的监控中使用广泛，今日再反观两者之间的关系，难以将其认同为“前沿问题。”

通观全书，相对最为出彩之处在第五章《在大数据中该关注什么?》，该章由阿里斯塔尔·克罗尔（Alistair Croll）[7]

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.ne

的《我们不知道“大数据”与每个人的公民权息息相关》与《三种大数据》、布拉德雷·沃塔克（Bradley Voytek）[8]的《自动化科学、深数据与信息的矛盾》、吉姆·斯塔克迪尔（Jim Stogdill）的《大数据方案里的鸡和蛋》与安迪·克里克（Andy Kirk）的《行走在形象化批评的钢索之上》五篇文章组成，这五篇文章应是该书的精华所在。

阿里斯塔尔·克罗尔认为，“大数据”成为趋势之后，其三大特性快速（fast）、庞大（big）与多样化（varied）将导致的最大一个问题或将是其受伦理、法律的制约，即对公民隐私的公开。尤其是“棱镜”事件之后，互联网语境下每一个公民的隐私都是不一定能受到保障的。在《三种大数据》中，克罗尔澄清了“大数据”的定义，他认为，以前学界对于“大数据”的定义是模糊的——“所有的东西都在互联网上，而互联网拥有海量的数据，因此所有的东西都是大数据”。[9]

沃塔克提出了“深数据”的概念，认为在应用上，“大数据”是与“深数据”息息相关的，所谓深，就是对于信息的细致性开发利用；斯塔克迪尔从自身的科研经验入手，谈及大数据对于客户平台资源整合的意义；克里克则独辟蹊径，从“可视化的生态系统”出发，认为“大数据”会将人类带向“从快餐到美食”的不合理需求，他意图构建一种合理性的分析，促使对“大数据”的研究与挖掘走向合理化。

二

毋庸置疑，《现在的大数据》虽立足前沿、内容全面，但笔者认为，《现在的大数据》仍存在着一定的不足，显示出了研究者视野的局限性。

首先，《现在的大数据》并未认识到“大数据”并非只是一种信息传播、聚集的方式，更不是理工、管理类学科的专有名词，而已经成为时下乃至未来人类社会结构的组成形态，未来的社会是“数据化”的社会，“大数据”的未来势必要走向更为开阔的范畴。因此，对于“大数据”的研究、分析与使用决不能停留在某个或是某种学科内部里，而是应拓展到更为广阔的世界当中。

“数据化”（datazation）是全球化背景下，人类社会在数字时代的发展必然趋势。数据化，就是人类在信息传播、人际交往乃至日常生活的过程中，为了便于沟通、传播与保存，将一切客观存在均处理为数据，进而使得整个人类社会成为了一个庞大的数据库。历史地看，自计算机发明至今的 40 年里，人类所处的时代只是“数据化”的第一阶段，即“数据化 1.0”；但自 21 世纪的第一个 10 年之后，以 3G 网络技术、云平台与 Android 系统的发明则将“数据化”迅速推进了第二阶段，即“数据化 2.0”的“大数据”时代。

与先前相比，“大数据”时代的社会“数据化”要更深、更广、更富有影响力。数据已经从人类知识的保存形式

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.ne

变成了人类社会的组织形式，人与人、人与社群以及社群与社群的联系已经完全由数据所取代并控制。大数据时代的数据，不再是简单的符码信息的堆砌，而变成了人类社会的数码符号。在这样的语境下，审理“数据化”的社会与大数据的未来趋势，显然有着前瞻性的现实价值。

因此，对“大数据”的研究，必须认识、关注人类社会“数据化”的现状。“数据化”不但导致社会结构呈现出了以互联网为框架的数据化形态，而且使得传统的人际关系、信息交流变成了即时、迅捷的数据交换。而“大数据”已经成为了人类社会在“数据化”之后的一个虚拟世界里的镜像。

笔者认为，研究“大数据”的本质就是对人类社会的研究。《现在的大数据》对这点有一定的认识，但并不深刻。造成这一问题的原因在于，目前“大数据”问题的研究主力军（包括《现在的大数据》的主力撰稿人）并非人文（如历史、哲学与思想史）与社科（如经济、传播、社会与伦理）学科的研究人员，而是计算机、统计学、数学与信息管理等理工学科的工程师。他们将研究焦点过多地停留在技术上，强调“如何获得大数据”或“如何使得大数据更加精确”。这种对于技术中心主义的推崇，使得其研究目的逐渐偏离了社会实践的需要。

其次，“数据化”的社会实际上已经产生了种种的社会问题，这些问题以文化问题、心理问题、伦理问题、法律问题、宗教问题、性别问题与道德问题等不同的形式表现出来，构成了“大数据”时代的社会疾患。如果对“大数据”的研究、讨论还停留在技术层面或学科内部，并忽视这些问题的存在，那么，这些社会疾患将会发展为人类的全球性共同灾难。

《现在的大数据》也在一定程度上承认：“大数据”时代最大的特征就是技术中心。事实上，凭借技术，人类可以将全世界不同国籍种族、文化背景的人连成一张巨大的网络，使得每个人都以“数据化”的形态存在。这种“数据化”尽管便捷、迅速，却忽视了人与人之间的差异性，使得原本立体、多元的社会结构趋向于扁平、单一化。“大数据”下如何使得原本不同的文化、宗教继续保持其多元化特质？

随着个人通信终端及其网络的发达，未来“数据化”的社会必然会促使人类在相互交往中更加依赖通信技术。社会性网络服务（Social Networking Services）、交互式信息平台（Interactive Information platform）、搜索引擎（Search engine）与“云数据框架”（Cloud data framework）的进一步普及与应用，会促使人类的交往从以往的“点对点”（point-to-point）的人际交流变为“点对多点”（point-to-multipoint）甚至“多点对多点”（multipoint-to-multipoint）的

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.ne

分众、大众传播。“大数据”下新的人类交往方式势必会带来隐私危机、信息风险、财产安全与网络暴力等新问题，这些当是“大数据”研究者着力考虑的范畴。

最近20年里，随着互联网技术的发达，人类社会开始呈现出了前所未有的新问题、新挑战。知识开始作为信息化的产品在赛博空间中解域（deterritorialization）流动。“大数据”促使全球化、解域化进一步加剧、加深。任意一个信息都可以作为“数据”在“大数据”的天空里翱翔、壮大，尤其是近5年以推特（Twitter）、微博、脸谱（Facebook）以及微信为代表的社会性网络服务网站的日趋普及，诸多信息（包括大量的虚假、垃圾信息）在经历过刻意的包装、修饰与几次点对多点的传播之后，进而很容易形成近似于“蝴蝶效应”的“网络迷因”（Internet meme）。[10]当被装扮为数据的语言、信息成为一种暴力形式的时候，我们又该如何从“大数据”着手进行应对？

遗憾的是，《现在的大数据》均未针对上述这些问题做出有效的阐释。当然，我们可以认为这是由于这本书的写作者多为理工技术学科人员，并不谙熟人文社科领域之故，这未免是该著的微瑕。但实际上就当下“大数据”的整体研究状况而言，无论是学界还是业界，对于上述问题仍显得关注不足，构成了“大数据”研究的盲区与缺陷。

三

《现在的大数据》一书目前虽未译成中文，但在短短的一年多时间里，却在欧美学界爆发出了不可忽视的影响力。据笔者不完全统计，截至2014年1月1日，该书在英语学界已经被引用近1.5万次。从影响力上看，该书可与纳特·希尔佛（Nate Silver）的《信号与噪声》、维克托·迈尔·舍恩伯格及肯尼斯·库克耶（Kenneth Cukier）合著的《大数据时代》合称“大数据”研究的“三典”之一。

但实际上，这三部书都共同反映了西方学界“大数据”研究的盲区与缺陷。笔者认为，《现在的大数据》所谈到的问题，实际上也基本上为其他研究著述所涉及，而它所忽视的问题，亦基本上反映了目前研究界的不足。它们共同反映了时下学界对于“大数据”研究的缺位之处，这将使得“大数据”研究在今后不但难以继续推进，而且甚至会被混淆概念、以讹传讹，使其可被探讨的空间逐渐萎缩，最终失去应有的社会意义。以《现在的大数据》一书来管窥目前国际学界对“大数据”的研究现状，无疑有着独特的现实意义。

因此本论认为，若是在“数据化”的社会中讨论“大数据”的未来，其立场、姿态与方法都必须与“大数据”的应用息息相关，即认同“大数据”为全社会共享的一种公共资源。一方面，在

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.ne

应用的过程中，使用者该如何规避可能出现的风险与问题；另一方面，研究者该以何种方式打破目前“大数据”既定的学科壁垒，进而使其成为促进社会管理、决策的重要工具。

首先，作为人类社会的投影与镜像，“大数据”所带来的问题，也是人类社会在当下所遇到的问题或其反映。尽管“大数据”规模庞大、瞬息万变且传播迅速，但它无法僭越人类社会的共有本质与逻辑基础。因而，规避“大数据”所带来的风险与问题，必须与人类现实社会相结合。

笔者主张，“大数据”改变最深的并不是资源配置、人际交往与信息传播的方式，而是人类的意识形态与生活习惯。自第一次工业革命至今，人类不断反省技术中心主义所带来的“异化”问题，从当年马尔库塞、本雅明提出的人类的“机械化”到阿甘本、德勒兹主张的“电子化”以及苏珊·桑塔格所认为的“信息化”及至今日的“数据化”，反映了人类在“单向度”的钢索上越走越远，技术越发达，人文精神所遭遇的挑战越大，对于道德、伦理、信仰等意识形态层面的反思则越显得必要。

实际上，早在数据化1.0时代就已经暴露出人类社会因为早期“数据化”所带来的一系列问题。从被斥为“网络暴力”到斯诺登曝光的“棱镜风波”，无一不证明了互联网所带来的各种负面因素，以至于至今中国大陆的网吧都标注“未成年人不得入内”，世界上许多国家都实行了违背互联网开放性原则的网络管制政策。而“数据化”几乎已经将互联网普及每个人身上。在机场、校园、商场与餐厅都可以使用的免费WiFi、运营商提供的4G网络服务与便捷、廉价的Android4.0系统硬件一旦紧密结合，整个人类的“大数据”必然会呈几何级数倍增。人类社会的分层将会从资本鸿沟、技术鸿沟、信息鸿沟迅速地过渡到数据鸿沟，在未来的人类社会中，无论谁真正地占有了数据，谁就占有了未来世界的金字塔顶。

“数据化”社会所带来的各种问题是一个庞大的学术课题，在本论中完全无法一一详述，甚至连列举都不可能，之所以提出这一问题，乃是为了抛砖引玉，提醒学界、业界对此问题应给予足够的关注。因为，在全球化的语境下，与100多年前的“机械化”相比，“数据化”对于人类历史变革的影响必然要深远、巨大得多，我们要做好足够的准备。

其次，与此同时，我们必须认识到“大数据”反映了人类社会的整体性、全面性的变化，而且这一现象在未来的十年甚至几十年里还会随着通信技术的迅速发展而进一步强化。“数据化”构成了全球化的现代化动力这一点已经成为了不可逆转的事实。如何挖掘“大数据”在促进社会发展、变革进程的不同应用范式，当是另一个值得关注的问题。

诚如前文所示，“大数据”的本质是时下乃至未来人类社会结构的组成形态，社会管理者所面对的是一个复杂的

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.ne

社会（无论是街区、城市、机构还是国家），不同民族、信仰、收入、文化背景的人共同聚居在一起，无论是个体心理还是群体心理，都是一个相对复杂的集合。传统意义上的社会管理实际上是“人对多人”的引导性管理，这样的管理方式在“大数据”时代已经显得力不从心。而且，在管理的过程中对于个人隐私、信息安全的保护又应该采取何种方式？因此，以“人对数据”这一形式进行有针对性的服务性管理，使得整个管理过程更为有效与人性化，当是今后社会管理的必然趋势。

“大数据”是互联网时代的进一步发展，因此，“数据化”的社会对社会管理提出了新的诉求。这不只是社会管理的业界与技术领域的责任，更是社会科学界的共同责任。在“大数据”语境下，促进传播、法律、伦理、心理乃至哲学等人文社科领域的研究转向，应是未来这些学科的重心之一。

目前学界、业界对于大数据的应用主要有两个方面。一个是纯粹理工学科的技术应用，譬如，生命科学、天文学等涉及庞大数据的学科等；另一个则是涉及对于广义人力资源的信息管理，譬如，人口普查、传染病人群定位、客户群锁定或嫌疑犯排查，等等。而真正意义上对于社会管理、决策，“大数据”所应扮演的角色及其意义并未被完全认识、挖掘，因此，更谈不上规避“大数据”可能带来的风险与问题。如何真正地将“大数据”为我所用，在大时代下发挥其应有的价值，带动人类社会的共同发展与进步，而不是象牙塔、实验室或交易所里的成人游戏，这是摆在我们所有人面前的历史责任。

注释

[1] 艾德·丹彼尔，知名作家与分析技术工程师，供职于奥莱利传媒机构，系奥莱利开源大会及开放资源委员会（O'Reilly Strata Conference and the O'Reilly Open Source Convention）主席。

[2] 杰瑞米·霍瓦德，卡格（Kaggle）公司总裁兼首席科学家。

[3] *O'relily: Big Data Now 2012* [M]. O'relily. 2012, p. 35.

[4] 马克·斯洛卡姆，奥莱利传媒机构网站主编。

[5] 阿拉斯达尔·阿兰，国际知名 IOS 程序专家，奥莱利传媒机构实验室工程师。

[6] 迈克·鲁克德斯，奥莱利内容策略机构副总裁。

[7] 阿里斯塔尔·克罗尔，著名作家，曾任奥莱利开源大会主席。

[8] 布拉德雷·沃塔克，神经学家，加州大学圣地亚哥分校助理教授。

[9] *O'relily: Big Data Now 2012* [M]. O'relily, 2012, p. 60.

[10] 对于这一问题，理查德·费斯（Richard Face）在 *The Book of F * cking Hilarious Internet Memes*（*Oculus Publishers, 2012*）一书中有较为详细的介绍，此处不再详述。

作者单位：中国科学院自然科学史研究所

（责任编辑 张娟）

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.ne