

大数据时代的中国地理学

廖小罕

中国科学院地理科学与资源研究所, 北京 100101

摘要: 本文阐述了地理大数据时代特点和地理时空大数据对地理学研究的重要性, 指出遥感卫星数据是数据量最大的地理大数据。2012年开始国家遥感中心组织力量利用中外遥感卫星数据对全球的生态环境进行分析, 发布了系列年度报告, 为全球环境问题研究提出了解决方案, 为中国地理研究从区域研究拓展到全球研究发挥了重要作用。2018年国务院提出《科学数据管理办法》, 对有效管理地理大数据和推动数据共享具有重要意义。同年中国地理学会地理大数据工作委员会成立, 在解决国家需求、地理学发展、地理大数据人才队伍建设等方面发挥了积极作用。当代地理学研究过程通过获得大数据“燃料”的注入, 解决问题的能力将更强, 得出的分析结果能够进一步适应实际生产实践的需要和多元化的需求。在指数性爆发的地理大数据支撑下, 地理学理论、方法等必将不断得到检验和完善, 必将能够进一步满足“全域、全时、全要素”等新时代的要求, 为社会进步和经济发展中出现的资源环境问题提供解决方案并发挥重要的作用。

关键词: 地理大数据; 中国地理学; 遥感; 全球观测; 生态环境

DOI: 10.3974/geodp.2019.04.01

1 前言

当代信息化、数字化、自动化的进程产生了全球范围不同尺度的巨量、丰富、杂乱的数据资源。我们既面临着大数据的挑战, 也迎来大数据带来的机遇^[1]。有文献统计, 从人类文明开始到2003年, 人类共创造了5 EB数据(1 EB=10⁹ GB), 最近10年, 数据量呈爆炸式增长, 到2020年全球产生的数据量预计近40 ZB(1 ZB=10³ EB)^[2]。这样的发展趋势, 超出所有人的预测, 数据的爆炸式增长正深刻地影响着我们的科研。

《Nature》、《Science》相继出版专刊, 讨论大数据研究的机遇与挑战^[3-4]。2012年3月, 美国政府宣布启动“大数据研究和发展计划”^[5]。2012年5月, 联合国发布题为“大数据发展, 机遇与挑战”的白皮书^[1]。2011年12月, 中国资本市场发布第一篇大数据主题报告“大数据时代即将到来”^[6]。2015年8月, 国务院印发《促进大数据发展行动纲要》的通知^[7]。2017年12月8日, 习近平在中共中央政治局第二次集体学习时强调实施国家大数据战略, 加快建设数字中国^[8]。2018年3月, 国务院提出《科学数据管理办法》(国办发〔2018〕17号)^[9]。2018年5月, 习近平在2018中国国际大数据产业博览会贺信中提出全面实施国家大数据战略^[10]。

收稿日期: 2019-09-23; 修订日期: 2019-12-10; 出版日期: 2019-12-24

作者信息: 廖小罕 S-3620-2018, 中国科学院地理科学与资源研究所, liaoxh@igsrr.ac.cn

引用格式: 廖小罕. 大数据时代的中国地理学[J]. 全球变化数据学报, 2019, 3(4): 311-316. DOI: 10.3974/geodp.2019.04.01.

大数据正从各行各业源源不断地加速产生,大数据时代已经到来。拥有数据的规模和运用数据的能力也是一个国家综合国力的重要标志。对数据的占有和控制也将成为国家间和企业间纷纷抢占的制高点。大数据时代的大浪潮,波及到社会各领域,也渗透到地理学及其各个领域和研究方向。

2 地理大数据时代特点

地理学主要研究自然要素和人文要素的空间结构和时间过程,以及两者间交互作用的内部机制与外部表征。地理科学离不开对空间结构、时间过程、表面特征、极端环境的研究。从数据角度来说,这些科学研究对数据的空间分辨率、时间分辨率、光谱分辨率以及数据的可获得性的要求至关重要。这些数据获取越详细越准确,对事物的认知程度就越高,对真理的认识就越前进。因此,从一开始,地理学就张开双臂欢迎大数据时代。

地理数据是典型的大数据,具有如下特点^[11-13]:

- (1) 体量大,遥感、无线传感网、自发地理信息(Volunteered Geographic Information, VGI)等数据体量从TB级到PB级甚至EB级;
- (2) 更新快,空天地一体化传感网络数据快速生成与传输VGI数据随时随地生成;
- (3) 模式多,包括结构化的遥感、地面观测网络数据以及半结构或无结构的VGI数据;
- (4) 精确性不确定,大量被污染数据存在,使用前需要开展繁重的数据清洗工作;
- (5) 价值大,可广泛用于农业服务、城市管理、资源调查、环境保护和防灾减灾等领域。

世界上80%的信息与地理位置有关^[14],例如,国土资源(包括土地、矿产、地质环境等)管理、城市规划、交通出行、水利、农业、林业、环境保护和应急决策等信息都可以通过空间可视化、空间查询、专题制图和空间分析等手段开展研究获得。

地理时空数据是大数据的重要组成部分,目前可获取的方式很多,如遥感卫星(包括陆地、大气、海洋、海冰系列等卫星)产生的对地观测大数据。全球仅遥感卫星数量就超过1,000颗^[15],卫星遥感数据存档量达到EB级,是地理大数据中数据量最大的数据。一些常态化的观测数据也很多,如气象、水文数据等,现在全球有上千个陆基气象、水文,海基浮标、深海探测等观测网络。自发地理信息大数据,包括专业平台的移动GPS、标注、影像等结构化数据,社交网络所产生包含时空信息的非结构化数据。现在全球互联网用户数突破40亿^[16],所产生的自发地理信息大数据在急剧增长,其他还有各种模式所生产的大数据,如地球系统、大气、陆面、海冰等模式,CMIP6第六次评估这些模式所生产的模拟、预测地球科学大数据约为20-40PB^[17]。

大数据时代的地理学研究体现出新的特点,包括:

- (1) 从传统的表观研究和量化到数据深度挖掘和解决复杂关联;
- (2) 地理大数据具有可扩展性,植入时空信息的大数据都可以作为地理大数据依靠空间分析开展数据挖掘分析;
- (3) 新技术、新方法、新手段在地理研究中大量使用;
- (4) 传输网络化、数据云端化、计算云端化和存储分布式等无处不在;

(5) 第四科学研究范式的广泛应用, 即数据驱动科学发现。

例如, 目前正在进行的国土资源勘探调查工作(国土三调: 耕地、永久基本农田、林地调查), 动用了许多第一次调查时没有采用过的数据和技术方法^[18-20]: 土地利用、基础地理、管理业务数据库等专业大数据以及社会大数据, 卫星遥感和无人机获取的大数据和超级计算、深度学习、云计算等方法。通过这些新的技术方法、数据得到图斑身份编码技术、地理搜索引擎、地理智能感知引擎下的变化发现做进一步的质量评价、决策支持等。又如, 基于手机信号开展的城市人口空间密度分布研究, 分析一天内人口的迁徙等, 这些工作在过去的研究中不借助大数据技术是很难获取的。

要驾驭这一切, 离不开地理信息系统的运用。地理信息系统是一种采集、存储、管理、分析、显示与应用地理信息的计算机系统, 是分析和处理现实世界(资源与环境)海量地理数据的通用技术^[21]。地理信息系统经历了以下几个发展阶段: 20世纪70年代是GIS起步发展阶段, 主要是地图数据输入、地图数据管理、空间数据统计、地图交互制作等技术研发阶段; 20世纪80年代是GIS巩固发展阶段, 经历了混合数据模型、空间综合分析、专业软件模块、资源环境应用等发展过程, 这两个时代是地理信息系统技术发展阶段^[22-23], 又称大型机时代; 20世纪90年代, 是地理信息科学发展阶段, 也称PC机时代, 经历了从系列化组件、专业应用系统、大型数据库、网络化服务的专业应用发展到时空认知、地理本体、系统模拟、Web GIS的科学体系的构建; 21世纪进入互联网时代, 是地理信息的大众服务阶段, 主要是位置服务、计算服务、网格GIS、虚拟环境的运用; 最近10年进入大数据时代, 是开展新一代GIS的社会服务阶段, 也称为地理信息世界发展阶段, 是进行自动化处理、规模化存储、高效化计算、知识化服务的阶段^[24]。

值得一提的是, 地理大数据当中, 卫星遥感数据占据的比例最大, 遥感数据让中国地理研究从中国拓展到全球。以前的科学研究, 发达国家在研究全球问题时更多的还是采用一手的原始观测数据。近代以来中国科学家对世界其他地区的地理、环境和生态的研究主要是利用国际上已有的数据, 缺乏一手数据, 在一些关键时刻中国科学家可能还没有话语权。现在依靠国内外遥感卫星, 中国地理科学家已经有条件发表对于跨国和全球问题的看法了。截止目前, 我国已经成为接近美国拥有全球遥感卫星序列最全、数量最大的国家^[25]。现在中国科学家利用外国遥感卫星基础性数据和基于自己的算法产生了许多在国际上有影响的数据产品和分析结果, 同时利用中国遥感卫星获取世界各地数据发出中国的声音。例如, 从2012年开始, 国家遥感中心组织力量利用中外的遥感卫星数据对全球的生态环境进行分析, 并提供了一系列年度报告^[26], 至今已经开展了8年, 涉及生态、植被、土地利用、农业、湿地和城镇化, 在地球观测组织等组织实施的国际科学计划中产生很大影响。

3 地理学大数据“燃料”的注入促进学科解决实际问题能力的提高

地理学在认识自然的过程中积累了一系列理论、方法和各种研究成果, 这些成果在改造自然过程中发挥了重要作用。然而, 随着社会的发展进步和经济总量的增加, 资源环境领域问题日益突出并错综复杂: 一方面是传统简单化的模型已经不能用于解决实际复杂

问题,需要在大样本基础上发展和完善;另一方面,许多理论、方法需要足够准确的边界约束和初始场的数据。地理学发展史表明,可靠、多因子分析的结果往往与精度和输入的变量数呈正相关。学科综合性、区域性和交叉性很强的地理学理论和方法是否可以发挥好相应的作用,结论是否可靠,预测是否准确,有足够保障的、可信赖的数据至关重要。当代地理学通过大数据“燃料”的注入^[27],解决问题的能力将会更高,得出的分析结果更加与实际吻合(图1)。

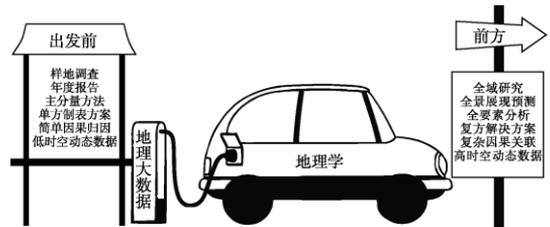


图1 大数据“注入”地理学促进学科理论和方法研究内涵进一步发展比喻图

依靠地理大数据带来的进步主要是因为但不限于如下几个方面:

第一,借助大数据,地理学研究范围可以从抽样到全域覆盖,比如依靠卫星遥感数据全球植被生物量、作物产量等估算可从传统样地调查到快速、大范围积分。这种从样地到全域的研究已经广泛用于旅游、自然资源、农业等研究领域;

第二,依靠大数据,地理学研究从所见所闻的当下场景拓展到高精度历史环境重建和高可靠未来情景预测。这种具备全景视野的研究在依靠数值模型开展区域发展、全球变化、地球模拟等研究等中已经得到广泛应用;

第三,通过大数据挖掘、高分辨率时空数据拟合与时空关联机理分析,将从过去简单主分量因果分析发展到考虑全要素自变量和因变量,基于多元因素,提供复方解决方案。

第四,大数据带来的大样本可以对过去基于相对小的样本形成的理论和方法进行检验并不断完善,提高科学研究的可靠性和帮助科研人员解决实际存在问题的能力。总之,在指数性爆发的大数据支撑下,围绕前述地理学研究出现的“全域、全时、全要素”等方面的突破,地理科学必将迎来新的发展,在为社会经济这一“巨系统”出现的资源环境问题提供解决方案中发挥重要的作用。

4 做好地理大数据应用管理的若干考虑

应用好地理大数据离不开有序的数据管理、知识产权制度、数据安全和质量控制等。2018年8月中国地理学会地理大数据工作委员会成立^[28],是我国地理学界做出的一份努力。该工作委员会提出来的问题需求,对促进解决地理学发展对地理大数据最关切的瓶颈问题具有一定代表性,主要包括:

(1) 数据成果管理与数据共享问题

以国务院《科学数据管理办法》为指导,建立数据出版与共享平台、参与国家科学数据中心建设、促进学会会员将地理科学数据通过数据出版实现知识产权保护、数据安全可靠、数据有效共享,力争用5年左右时间配合国家大数据战略和实施部署,基本解决地理数据共享的关键瓶颈问题。

(2) 数据知识产权保护和科学评价问题

建立数据知识产权认证的标准、规范、方法和途径,推动数据成果纳入科学成果评价体系,力争在1-2年时间内做好试点、示范工作,争取用5年的时间在全国推广。

（3）全球数据薄弱问题

中国是一个大国，中国地理学家不仅要研究中国区域性问题的同时，也要研究全球性问题。中国地理学会将通过召开全球数据研发和共享研讨会，与全球生态环境遥感监测等国家相关科技计划紧密沟通，力争在5年时间内配合国家发展战略和联合国可持续发展目标，发布有代表性、基础性的全球数据集。

（4）学术论文与科学数据脱节问题

中国地理学会地理大数据工作委员会与学术编辑委员会将密切配合，采取共同行动，通过出版带动学术论文与原创科学数据关联出版。力争用5年的时间将中国地理学会主办（联合主办）的学术刊物发表的学术论文与数据关联出版的比例从目前的不足1%提升至30%以上。

（5）数据计算环境建设问题

数据计算是数据挖掘、科学发现的关键环节，中国地理学会通过评选和弘扬地理大数据优秀计算环境实用案例，促进从数据到科学发现、从数据到社会可持续发展的转化，从而实现地理大数据的真正价值。

（6）地理大数据为可持续发展服务的科学普及问题

推动地理科学大数据在社会可持续发展起到更加积极的作用是地理大数据工作委员会的重要任务之一。中国地理学会拟采取发挥地理大数据的优势，以地理特色、地理标志、地理传统产品（“三地”产品）生态环境源地理大数据为试点，探讨地理大数据服务于国家和地方可持续发展、保障国家和地方生态安全的机制和途径，力争用5年时间做出可以推广的试点经验。

（7）地理大数据人才队伍建设问题

中国地理学会地理大数据工作委员会将继续推动地理大数据百校传播活动，推动教材编写、大学开课的进程，力争到2025年完成100所院校的大数据普及工作，10所以上大学增设地理大数据课程。

（8）数据安全和科学道德、伦理问题

数据安全和科学道德、伦理问题永远是中国地理学会必须强调和优先考虑的重点问题。中国地理学会地理大数据工作委员会将进一步加强地理大数据安全和科学道德、科学伦理的宣传和普及工作，制定出行业规范和行动指南。

5 结束语

随着地理大数据时代的来临，中国地理学迎来了新的发展机遇。地理学理论、方法和区域及全球范围内的应用实践必将通过大样本数据不断得到检验和完善；地理大数据的“全域、全时、全要素”发展态势，特别是中国地球观测卫星数据的迅猛发展，中国地理学研究和发展必将借助大数据“燃料”的注入走向世界研究前列，为全球包括中国的社会进步和经济发展，特别为资源环境新挑战问题的解决发挥重要作用。

致谢：本文基于2019年9月21日中国地理学会地理大数据工作委员会年会报告整理完成，马军花、邓晓明为本文稿的修改和完善提供帮助，作者谨此深表谢忱。

参考文献

- [1] UN Global Pulse. Big data for development: challenges & opportunities [R/OL]. [2012-10-02]. <http://www.unglobalpulse.org/projects/BigDataforDevelopment>.
- [2] <https://www.idc.com/>.
- [3] The big challenges of big data [OL]. <https://www.nature.com/>.
- [4] Special online collection: dealing with data [OL]. <https://www.sciencemag.org/site/special/data>.
- [5] The federal big data research and development strategic plan [OL]. <https://www.nitrd.gov>.
- [6] <https://blog.sina.com.cn/zhaogd>.
- [7] 国务院. 促进大数据发展行动纲要(国发[2015]50号). 2015-8-31.
- [8] http://www.xinhuanet.com/politics/leaders/2017-12/09/c_1122084706.htm.
- [9] 国务院. 科学数据管理办法(国办发[2018]17号). 2018-03-17.
- [10] 习近平. 2018 中国国际大数据产业博览会致贺信[R]. 人民日报, 2018 年 05 月 27 日 01 版.
- [11] 郭华东. 大数据大科学大发现——大数据与科学发现国际研讨会综述[J]. 中国科学院院刊, 2014, 299(4): 500–506.
- [12] 郭华东. 地球大数据科学工程[J]. 中国科学院院刊, 2018, 33(8): 818–824.
- [13] 韩平, 张建武, 王佑武. 浅析地理信息与位置大数据在地图编制中的应用[J]. 甘肃科技, 2016, 32(1): 334–336.
- [14] 曹世博. 2017 年遥感卫星市场综述(下) [J]. 中国航天, 2018(6): 73–78.
- [15] <https://new.qq.com/omn/20190128/20190128B00CCS.html>.
- [16] <http://www.xinhuanet.com/expo/zt/sjdlxxdh/index.htm>. 2018.
- [17] <https://esgf.llnl.gov/search/cmip6>.
- [18] 郑冀英. 基于第三次全国国土调查的技术探究分析[J]. 中国资源综合利用, 2019, 37(9): 64–66.
- [19] 朱伟, 王军仓, 袁荣才. 浅谈智能移动终端在国土三调中的应用[J]. 科技与信息, 2017(9): 155.
- [20] 广西自然资源厅自然资源调查监测处. 三调新在哪?——一张图读懂第三次全国国土调查新变化[J]. 南方国土资源, 2019(6): 18–19.
- [21] 陈述彭. 地理信息系统导论[M]. 北京: 科学出版社, 2003.
- [22] 周成虎. 地理信息系统的新时代: 网格地理信息系统[J]. 地理信息世界, 2007(4): 17.
- [23] 张洪岩, 王钦敏, 周成虎等. “数字地球”与地理信息科学[J]. 地球信息科学, 2001(4): 1–4.
- [24] 张胜利. 地理信息服务能力的提升有力地促进社会进步[J]. 测绘技术装备, 2015, 17(1): 53–54.
- [25] <http://www.broadcast.hc360.com>.
- [26] <http://www.nrscc.gov.cn/>.
- [27] 王鹏. 大数据插上机器学习翅膀为 5G 提供新燃料[J]. 通信世界, 2019(11): 47–48.
- [28] 中国地理学会. 中国地理学会地理大数据工作委员会正式成立[R]. 全球变化数据学报, 2018, 2(3): 354–356. DOI: 10.3974/geodp.2018.03.18.