

国际健康大数据研究计划发展及启示^{*}

崔春舜 徐 畅 高东平

(中国医学科学院/北京协和医学院医学信息研究所 北京 100020)

〔摘要〕 对美国、欧盟等发达国家和地区的大数据研究计划中健康相关部分进行调研,分析其特征,最后提出我国医疗大数据建设发展建议,包括建立数据统一监管体系、统一数据标准与结构、开放共享数据资源以及开展多学科数据科学人才培养。

〔关键词〕 医疗健康;大数据;研究计划

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2019.12.002

Development and Enlightenment of International Medical Big Data Research Plans CUI Chunshun, XU Chang, GAO Dongping, Institute of Medical Information, CAMS&PUMC, Beijing 100020, China

〔Abstract〕 The study on the health-related part of the big data research plans implemented in the United States, the European Union and other developed countries and regions has been carried out. The paper analyzes their specific characteristics, and finally puts forward some suggestions on the building and development of medical big data in China, including building unified data supervising and regulating system, unified standards and structures of data, opening to share data resources, and talent training for multidisciplinary data.

〔Keywords〕 medical health; big data; research plan

1 引言

随着社会生活数字化进程的加快,海量数据源源不断产生并呈现爆炸性的增长,预计 2035 年全球数据总量达到 19 267ZB,见图 1。大数据分析最初是在计算生物学、生物医学工程以及医学等领域发展起来的,通常是将大量的数据通过模型训练转换为“关于数据的数据”,用以预测趋势、发现关系以及检测异常。2012 年联合国全球脉动倡议

(UN Global Pulse) 组织发布“大数据促进发展:挑战与机遇”(Big Data for Development: Challenges & Opportunities)^[1],用来推动各国政府机构以及各行各业认识到数据科学是消除贫困、饥饿与疾病强有力的工具。其后许多国家纷纷发布一系列大数据研究计划,大力推进大数据技术研究和应用。本研究主要对国外大数据研究计划中医学部分进行综述,以期对我国大数据研究计划的制定提供借鉴参考。

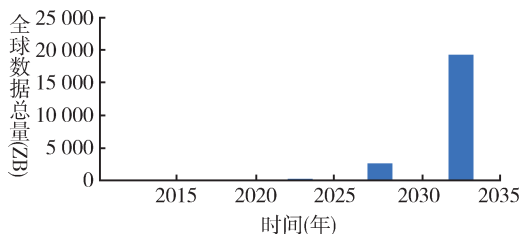


图 1 新摩尔定律主导下 2015-2035 年全球数据总量增长态势^[2]

〔收稿日期〕 2019-07-02

〔作者简介〕 崔春舜,硕士研究生,发表论文 6 篇;通讯作者:高东平,博士。

〔基金项目〕 中国医学科学院医学与健康科技创新工程(项目编号:2016-I2M-2-004)。

2 国外大数据研究计划

2.1 美国

2.1.1 早期研究计划概述 美国作为较早的发布国家层面大数据计划的国家,对我国大数据未来的发展规划和监管具有较好的借鉴意义。早在 2009 年,美国发布《利用数字数据的力量》(Harnessing The Power of Digital Data)^[3]提出发展大数据,之后发布《大数据研究和发展计划》确立大数据发展的国家战略,接着陆续发布《数据—知识—行动》^[4]、《大数据:把握机遇,维护价值》、《大数据和隐私:技术视角》^[5]、《精准医疗计划》、《大数据报告:算法系统、机会与公民权利》以及《联邦大数据研发战略计划》等文件。

2012 年 3 月美国政府发布《大数据研究和发展计划》(Big Data Research and Development Initiative)^[6]并成立大数据高级指导小组,计划中 6 个联邦政府部门发布其重点项目,其中卫生与公众服务部(United States Department of Health and Human Services, HHS)的内容涉及 3 个部门:疾病控制中心(Centers for Disease Control and Prevention, CDC)建立疫情与细菌快速检测反应网络;医疗保险和医疗补助服务中心(Centers for Medicare and Medicaid Services, CMS)开发基于 Hadoop 的医疗保险和医疗补助数据库以期 CMS 决定优先项目提供参考;食品和药品管理局(Food and Drug Administration, FDA)开发虚拟实验室,使任何地点都能在 1 小时内实现虚拟实验室功能。美国国立卫生研究院(National Institutes of Health, NIH)建立提供共享服务的癌症影像档案(The Cancer Imaging Archive, TCIA)并建成心血管研究网络(The Cardiovascular Research Grid, TCRG)用于分析、匿名化和共享数据,另外还包括神经成像信息工具和资源交流中心(The Neuroimaging Informatics Tools and Resource Clearinghouse, NITRC)、可扩展神经影像档案工具包(The Extensible Neuroimaging Archive Toolkit, XNAT)以及计算机解剖学和多维建模资源洛杉矶神经成像实验室(The Los Angeles Laboratory of Neuro Imaging, LONI)等项目^[7]。

2014 年美国政府发布白皮书《大数据:把握机

遇,维护价值》(Big Data: Seizing Opportunities Preserving Values)^[8],其中指出电子病历的普及以及医疗保险偿付机制的转变正在促进学习型医疗系统的形成,推动按患者康复程度而非治疗次数收费模式的发展。白皮书中同时指出将基因数据与医疗保健数据链接的生物医学数据库成为隐私保护的重点,因此建议构建国家层面统一的数据标准和结构使外部在受控的方式下访问数据。作为大数据白皮书的延续,2016 年美国发布《大数据报告:算法系统、机会与公民权利》(Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Right)^[9]。在 2014 年的白皮书中提到数据隐私暴露之后的基因歧视问题,而在 2016 年的白皮书中重点阐述关于歧视的问题,在贷款、就业、教育以及刑事司法 4 个方面进行剖析。

2015 年美国宣布《精准医疗计划》(Precision Medicine Initiative, PMI),致力于治愈癌症和糖尿病等疾病,目的是让研究人员、医疗保健提供者和患者共同合作发展个性化护理。PMI 于 2016 年启动,1.3 亿美元被分配给 NIH 用于建立一个全国性的大型研究参与者群体,称为队列,7 000 万美元被分配给国家癌症研究所,作为肿瘤 PMI 的一部分,领导癌症基因组学的工作。

2.1.2 未来研究战略 2016 年 5 月美国政府发布《联邦大数据研发战略计划》(The Federal Big Data Research and Development Strategic Plan)^[10],该计划指出美国下一代大数据研究的 7 大战略包括通过促进数据共享和管理的策略提高数据价值、安全的大数据收集共享和使用、改善国家层面大数据教育和培训的格局以及建立和加强国家大数据创新生态系统的联系。计划中提到 NIH 于 2015 年开始提供“大数据到知识(BD2K)”^[11]机构培训拨款,为本科生和研究生提供计算机科学、信息学、统计学、数学以及生物医学科学的综合培训^[12]。

2.2 欧盟

2.2.1 早期研究计划概述 欧盟作为世界最大的经济体之一,其大数据研究计划的导向性对欧洲整体的未来发展以及我国大数据相关领域的发展具有重要指导性。2011 年欧盟正式发布《开放数据:创新、增长和透明治理的引擎》(Open data: an en-

gine for innovation, growth and transparent governance)^[13] 报告,其围绕开放数据这一主题内容,致力于调整法律框架以适用数据、动员融资工具支持开放数据以及促进成员国之间的协作和经验分享。此后先后发布《数据价值链战略计划》、《释放欧洲云计算服务潜力》、《迈向繁荣的数据驱动经济》、《大数据价值战略研究和创新议程》以及《欧盟资助的健康、福利和老龄信息与通信技术领域的研究与创新》等计划。

2012 年欧盟发布《数据价值链战略计划》(A Strategy on The Data Value Chain)^[14],围绕 3 个价值主张,首先是培育一个协调一致的欧洲数据生态系统,其次是促进跨部门、跨语言和跨国界的数据服务和产品开发,最后是改善从数据中提取价值的框架条件,包括基础设施、标准以及有利的政策和法律环境。同年 9 月《释放欧洲云计算服务潜力》(Unleashing the Potential of Cloud Computing in Europe)^[15] 报告发布,称若政策干预成功,到 2020 年欧盟的云计算收入会增加到近 800 亿欧元,欧盟国内生产总值将实现 1.6 亿欧元的净增长,产生 380 万个就业岗位,同时提到整合各国的云计划如法国的 Andromède、英国的 G-Cloud 以及德国的 Trusted Cloud 等。2014 年欧盟发布《迈向繁荣的数据驱动经济》(Towards a Thriving Data-driven Economy)^[16],概述未来数据驱动经济发展的特点,其中提到“灯塔数据倡议”(Lighthouse Data Initiatives)个性化医疗^[17],其目的是建立欧洲范围的个性化医疗数据生态系统,致力于实现提供癌症基因组,建立激励数据贡献者的机制;提供先进的生物信息学工具,同时优化基因组管道的性能和准确性,为基因组学的临床部署提供环境;为公共卫生基因组学基础设施提供平台,同时满足系统、数据格式和互操作性的多样性;通过使用数据相关技术在家中“治疗”和监测患者而非住院等。此外还提出基于个性化医疗的数据生态系统结构需要实现数据收集、管理集成、共享、分析、可视化以及决策支持的功能,以包括软件、网络、存储、高性能计算以及云计算在内的技术设施建设为基础,收集 5 大类数据包括患者/公民数据(如移动医疗、营养数据等)、科学数据(如生物库、OMIC 数据库、临床试验、文献等)、医学数据(如基因组学、影像学

等)、公共健康数据以及医保数据^[18]。同年 4 月发布的《大数据价值战略研究和创新议程》(European Big Data Value Strategic Research & Innovation Agenda)^[20] 总结实施数据驱动经济的 4 项战略包括 I-Spaces、灯塔项目、技术项目以及合作与协调项目。在 2017 年 10 月更新的议程中提到未来发展的 5 个优先技术领域:数据分析以提高对数据的理解;优化用于分析静态和动态数据的架构;确保数据保护和匿名化的机制;高级可视化和用户体验;数据管理工程。

2018 年 6 月欧盟更新《欧盟资助的健康、福利和老龄信息与通信技术领域的研究与创新》(EU-funded Research & Innovation in the field of Information and Communications Technology for Health, Wellbeing & Ageing)^[20] 报告,提出资助 4 类项目:帮助患者和医疗保健专业人员预防性地帮助人们保持健康;创新医疗保健系统及其工作方式,提高组织协同工作能力;健康老龄化的 ICT 解决方案;由中小企业资助的项目,以加速健康、福利和老龄化 ICT 解决方案的市场引进。

2.2.2 未来研究战略 “地平线 2020”(Horizon 2020)^[21] 计划在 7 年(2014-2020 年)期间将提供近 800 亿欧元用于欧洲保持一流的科学水平。2014-2017 年间该项目在健康、人口变化和福祉领域投入 20 亿欧元,在其 2018-2020 年计划文件中列出 6 项目标:将个性化医疗方法有效整合到医疗保健服务和系统中;对抗传染病和抗微生物耐药性日益增长的威胁;满足最弱势群体的需求和控制慢性病;探索环境(包括气候变化和空气质量)对健康的作用,制定缓解措施;探索健康创新和医疗保健的数字化潜力,包括建立欧洲健康研究和创新云;通过探索先进技术的应用,促进欧洲医疗领域和行业的创新,改善劳动力健康,促进监管科学发展。

2.3 其他发达国家

瑞士国家重点科研计划(NFP)大数据专项(Big Data, NFP75)^[22] 于 2017 年正式启动,计划投入资金 2.5 亿瑞士法郎,2017-2020 年为期 4 年。该计划在健康领域目前公布的项目主要涉及为卫生部门大数据使用的道德评估。新加坡智慧国家(Smart Nation)^[23] 倡议于 2014 年正式启动。在医疗

领域主要有 4 部分投入,包括医疗保健中的辅助技术如机器人技术、HealthHub 一站式在线健康信息和服务门户网站、国家步行挑战健康 365 应用程序以及远程医疗。日本于 2014 年对《创建最尖端 IT 国家宣言》^[24]进行更新,其中提到通过提供适当的地域医疗、看护等以实现健康长寿社会。其关于健康的内容主要有两点,首先是提供高效的医疗护理服务,构建远程医疗、家庭看护以及生活支援服务为主体共享合作的医疗护理结构,其次是推进医疗健康信息数据的利用,基于个体生活方式数据进行适当且持续性的健康促进和重症化预防的方案建议。综上所述,许多发达国家在发布或前瞻性布局未来医疗大数据研究计划时会优先考虑建立独立机构,此机构具有监管、引导以及协调等功能,作为医疗大数据研究的引导者,辅助延展传统医疗部门功能,实现医疗部门的功能升级。其次结合本国目前研究现状,优选重点领域进行重点投资,如欧盟着重于跨国医疗大数据的研究发展建设,而日本老龄化较严重,则比较注重健康管理数据的研究计划。最后是配套设施的建设,医院、诊所等医疗场所的硬件设施往往不能满足医疗大数据收集与处理需求,因此硬件设备的配套更新是医疗大数据发展的重要支撑。

3 对现阶段我国医疗大数据发展的启示

3.1 我国现阶段医疗大数据建设存在的问题

大数据标准化方面我国与国外差距较大。美国医疗机构的电子病历已经是强制执行,我国包括三甲医院在内的所有医院尚未完全普及。美国在推广结构化的电子病历时,对首先使用结构化电子病历的机构给予奖励,待使用率达到一定程度,政府开始对未使用机构进行惩罚。我国则可通过试点的方式来普及。医疗数据中心建设方面,国内现存大大小小的数据中心,中心之间数据互通工作的开展不尽如人意,数据孤岛不利于医疗大数据的应用发展。并且数据中心的专业性并不能满足所有专业团队的需求,因此有很多专业团队与生物信息技术公司合作建立各自的数据库,这就导致更多数据孤岛的产生,形成恶性循环。伦理方面涉及数据安全、个人隐私保护问题。保护信息安全有几个层面,信息是国家的资源,对于国外研究机构的数据要协议

使用不能无偿使用,这方面的政策需要及时跟进。人才培养方面,在医疗大数据研究发展过程中,独立培养的计算机技术人才与医学人才之间的知识壁垒较高,造成研究结果往往达不到预期效果,因此交叉学科人才的培养体系亟需建立。

3.2 建设国家数据统一监管体系

2017 年全国信息安全标准化技术委员会发布《信息安全技术个人信息安全规范》,但其只是推荐性质的标准,国家层面并未出台具有法律效力的数据监管文件。建议设立独立且专业的监管机构促进相关法律法规的出台并对数据在全社会的流动进行有序控制,同时要平衡好开放与隐私间的关系。

3.3 构建国家层面统一的数据标准和结构

统一的数据标准有利于形成统一的访问管控,对于数据隐私保护以及存储较为有利。需要构建的标准包括医学术语、临床路径、电子病历、健康档案标准等。统一的数据标准有利于消除信息壁垒,形成统一的数据理解,更好地促进机构之间的业务协同,消除信息孤岛。统一的数据标准有利于快速整合不同数据源数据,进行高效、高质量分析,为数据共享开放奠定基础。

3.4 开放共享数据资源

多个国外政府都在推动数据的开放共享,借助更多的非本机构专业人员实现数据的最大化利用。较为成熟的如英国 NHS 网站提供大量开放数据并为访问者提供免费科普性质的资料,形成统一的健康共享门户。我国亟待建设权威的健康资源门户,在其中开放共享健康数据。

3.5 开展多学科数据科学人才培养

数据科学人才是医学大数据发展的重要支撑。政策制定者需要考虑在教育层面加大对统计学、计算机科学以及生物医学等多学科人才的培养,如对生物医学人才加大数据学科的培训力度,使其成长为生物医学领域的的数据科学人才。

4 结语

随着医疗行业的快速发展,通过对各层次的健

康大数据进行挖掘和分析, 提高行业运营效率与质量, 为人民健康提供保障。

参考文献

- 1 United Nations Global Pulse. Big Data for Development: challenges & opportunities [EB/OL]. [2019-02-24]. <https://www.unglobalpulse.org/projects/BigDataforDevelopment>.
- 2 IDC Corporation. International Data Corporation [EB/OL]. [2019-02-24]. <https://www.idc.com/>.
- 3 The Networking and Information Technology Research and Development. Harnessing The Power of Digital Data for Science and Society [EB/OL]. [2019-02-24]. https://www.nitrd.gov/About/Harnessing_Power_Web.pdf.
- 4 The Networking and Information Technology Research and Development. Big Data Research and Development Initiative [EB/OL]. [2019-02-25]. https://www.nitrd.gov/nitrdgroups/index.php?title=Data_to_Knowledge_to_Action.
- 5 The White House. Big Data and Privacy: a technical perspective [EB/OL]. [2019-02-26]. <https://obamawhitehouse.archives.gov/the-press-office/2015/11/16/fact-sheet-pcast-report-big-data-and-privacy-technological-perspective>.
- 6 The White House. Data - Knowledge - Action [EB/OL]. [2019-02-27]. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final.pdf.
- 7 李德升, 付伟. 美国推动大数据发展的做法及启示 [J]. 金融电子化, 2013 (6): 37-40.
- 8 The White House. Big Data: seizing opportunities preserving values [EB/OL]. [2019-02-27]. https://obamawhitehouse.archives.gov/sites/default/files/docs/20150204_Big_Data_Seizing_Opportunities_Preserving_Values_Memo.pdf.
- 9 The White House. Big Data: a report on algorithmic systems, opportunity and civil right [EB/OL]. [2019-02-28]. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.
- 10 The Networking and Information Technology Research and Development. The Federal Big Data Research and Development Strategic Plan [EB/OL]. [2019-03-01]. <https://www.nitrd.gov/PUBS/bigdatardstrategicplan.pdf>.
- 11 National Institutes of Health. Big Data to Knowledge [EB/OL]. [2019-03-02]. <https://commonfund.nih.gov/bd2k>.
- 12 Bourne P E, Bonazzi V, Dunn M, et al. The NIH Big Data to Knowledge (BD2K) Initiative [J]. Journal of the American Medical Informatics Association, 2015, 22 (6): 1114.
- 13 European Commission. Open Data: an engine for innovation, growth and transparent governance [EB/OL]. [2019-03-02]. http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=1190.
- 14 European Commission. A Strategy on The Data Value Chain [EB/OL]. [2019-03-03]. http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=3488.
- 15 European Commission. Unleashing the Potential of Cloud Computing in Europe [EB/OL]. [2019-03-04]. <https://www.pdpjournals.com/docs/88053.pdf>.
- 16 European Commission. Towards a Thriving Data-driven Economy [EB/OL]. [2019-03-04]. http://ec.europa.eu/newsroom/dae/document.cfm?action=display&doc_id=6210.
- 17 European Commission Health and Food Safety. Lighthouse Initiative on Personalised Medicine [EB/OL]. [2019-03-04]. <https://euapm.eu/big-data-activities-lighthouse-initiative.html>.
- 18 European Commission Health and Food Safety. Big Data Roadmap for Personalised Medicine [EB/OL]. [2019-03-05]. https://euapm.eu/pdf/EAPM_Big_Data_Roadmap_for_Personalised_Medicine.pdf.
- 19 European Commission. European Big Data Value Strategic Research & Innovation Agenda [EB/OL]. [2019-03-05]. http://bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf.
- 20 European Commission. EU-funded Research & Innovation in the field of ICT for Health, Wellbeing & Ageing [EB/OL]. [2019-03-06]. http://ec.europa.eu/newsroom/document.cfm?doc_id=2852.
- 21 European Commission. Horizon 2020 [EB/OL]. [2019-03-06]. <https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020#Article>.
- 22 Schweizerischer Nationalfonds. NFP75 [EB/OL]. [2019-03-06]. <http://www.nfp75.ch/de/projekte/modul-2-gesellschaftliche-und-regulatorische-herausforderungen/projekt-vayena>.
- 23 Smart Nation and Digital Government Office. Smart Nation [EB/OL]. [2019-03-06]. <https://www.smartnation.sg/what-is-smart-nation/useful-apps>.
- 24 Prime Minister's Office of Japan. 创建最尖端 IT 国家宣言 [EB/OL]. [2019-03-06]. <https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20140624/siryoul.pdf>.