



第S69次学术讨论会·科学数据治理与利用的前沿和热点

加强生物医学大数据建设应用, 推动健康中国战略实施

王健伟¹, 尹岭², 刘德培^{3*}, 周伟⁴, 陈杰⁵, 杜冠华⁶, 唐其柱⁷, 马旭⁸, 陈景元²,
苏雪梅⁹, 吴静¹⁰, 李海燕¹¹, 曹晨光¹², 钱庆¹³, 翟晓辉¹⁴, 张胜发¹

1. 中国医学科学院北京协和医学院, 北京 100730;
2. 中国人民解放军总医院, 北京 100080;
3. 中国医学科学院基础医学研究所, 北京 100730;
4. 国家人口健康科学数据中心, 北京 100730;
5. 中国医学科学院北京协和医院, 北京 100005;
6. 中国医学科学院药物研究所, 北京 100050;
7. 武汉大学第一临床学院, 武汉 430060;
8. 国家卫生健康委科学技术研究所, 北京 100081;
9. 中国疾病预防控制中心信息中心, 北京 102211;
10. 中国疾病预防控制中心慢性病控制中心, 北京 100500;
11. 中国中医科学院中医药信息研究所, 北京 100700;
12. 国家药品监督管理局信息中心, 北京 100044;
13. 中国医学科学院医学信息研究所, 北京 100020;
14. 国家卫生健康委医疗管理服务指导中心, 北京 100044

* 联系人, E-mail: liudp@pumc.edu.cn

我国政府始终将人民的权益放在优先发展的地位, 不断提高人民健康水平, 主要健康指标总体上优于中高收入国家平均水平^[1]. 2017年, “健康中国”被写入党的十九大报告, 健康中国上升为国家战略^[2]. 2019年印发的《关于实施健康中国行动的意见》(https://www.gov.cn/zhengce/content/2019-07/15/content_5409492.htm)和《健康中国行动(2019~2030年)》(https://www.gov.cn/xinwen/2019-07/15/content_5409694.htm)明确了健康中国战略的总体目标, 部署了重点方向和任务. 在大力推进健康中国战略的同时, 我国生物医学大数据快速发展, 逐步成为国家重要基础性战略资源. 生物医学大数据广泛涉及人类生命健康相关的各个领域, 类型复杂多样, 不仅包括基础医学、临床诊疗、公共卫生、药物研发、卫生监测、健康管理、环境卫生等医疗卫生与人口健康数据, 还包括基因组学、微生物学等生物学数据^[3-5]. 如何利用生物医学大数据为健康中国战略提供理论和实践支撑, 助力健康中国战略实施和居民健康水平持续提升成为我国政府的重要议题.



王健伟 中国医学科学院北京协和医学院副院长、教授、研究员, 中国医学科学院学部委员. 主要从事呼吸道病毒感染致病机制与防控研究.



刘德培 中国医学科学院基础医学研究所教授、研究员、博士生导师, 中国工程院院士, 美国医学科学院院士, 第三世界科学院院士, 欧洲科学院院士, 重大疾病共性机制研究全国重点实验室主任, 国家人口健康科学数据中心主任. 主要从事基因表达调控、基因治疗与心血管疾病发病机制研究.

1 国外发达国家健康战略与生物医学大数据发展应用现状

美国的“健康国民”战略规划始于1980年颁布的“健康国民1990: 促进健康与预防疾病”, 其后陆续发布了“健康国民2000”“健康国民2010”“健康国民2020”等一系列国家健康战略^[6]. 同时, 美国国立卫生研究院(National Institutes of Health, NIH)早在1988年就成立了美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI), 该中心整合共享世界范围内的生物医学数据资源, 极大促进了生物医学大数据的发展与应用^[7]. 2013年初, NIH启动了大数据到知识(Big Data to Knowledge, BD2K)计划, 旨在支持新方法、软件、工具及培训项目的发展以寻求加速数据的处理与再发掘, 推动生物医学大数据向知识转化^[8]. 2015年, 美国启动为期5年的“精准医学计划”, 形成了高质量的生物医学数据库, 可获得的数据包括健康调查、体格检查、生物样本、电子健康记录、数字化健康信息等^[9]. 2017年, 美国国家医学图书馆(The United States National Library of Medicine, NLM)发布了《2017~2027战略规划——生物医学发现和数据驱动健康平台》(https://www.nlm.nih.gov/pubs/plan/lrp17/NLM_StrategicReport2017_2027.pdf), 并于2018年6月发布《数据科学战略规划》, 旨在基于科学数据驱动推动生物医学创新(https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf). 2021年NIH发布《NIH拓展战略规划(2021~2025年)》(<https://www.nih.gov/sites/default/files/about-nih/strategic-plan-fy2021-2025-508.pdf>), 提出“全民研究项目”(All of Us Research Program), 以期利用数据科学为全体美国人民的健康服务^[10]. 目前, 美国在生物医学数据发展与应用方面取得了重大成效, 形成了核酸序列数据库GenBank、美国癌症基因组图谱计划数据库(The Cancer Genome Atlas, TCGA)等具有全球影响力的数据资源库. 截至2023年6月, Genbank已经收集超过26亿序列数据和近21.8万亿核苷酸数据(<https://www.ncbi.nlm.nih.gov/genbank/statistics>). TCGA对2万多例原发性癌症进行了分子特征分析, 产生了超过2.5 PB的基因组、表观基因组、转录组和蛋白质组数据(<https://www.cancer.gov/ccg/research/genome-sequencing/tcga>). 同时, 美国建立了国家级的健康医疗数据开放共享平台(<https://healthdata.gov>), 逐步实现医疗服务、公共卫生、医疗保障、药品等方面健康医疗大数据共享, 有力推动了美国健康战略的发展和医药卫生创新应用.

欧洲联盟(简称欧盟)的健康战略规划始于2003年, 欧盟相继发布了“欧盟健康规划(2003~2007年)”“欧盟健康规划(2008~2013年)”“欧盟健康计划(2014~2020年)”和“欧盟健康计划(2021~2027年)”等健康战略或计划, 以期实现改善和促进健康、群体保护、加强卫生系统等总体目标^[11,12]. 同时, 欧盟及欧洲部分国家开展了“数据价值链战略计划”和“地平

线2020”计划等生物医学大数据与生命健康相关战略规划^[13,14]. 目前, 欧洲部分国家在生物医学数据发展与应用方面取得了明显成效, 欧洲生物信息学研究所(European Bioinformatics Institute, EBI)已建立包含跨基因组学、蛋白质组学等在内的大型生物信息公共数据库, 数据资源量已超过390 PB (https://www.embl.org/documents/wp-content/uploads/2023/06/2022_EMBL-AR_Digital_PDF.pdf), 为欧洲地区健康战略实施与生物医药卫生技术创新提供了数据与知识服务, 促进了欧洲地区居民健康水平的提升.

2 健康中国战略实施进展

2.1 健康中国战略的发展

我国健康战略规划始于2008年原卫生部组织专家开展的“健康中国2020”战略研究, 并在党的十八届五中全会明确提出了推进健康中国建设的任务^[15]. 2016年, 中共中央、国务院印发的《“健康中国2030”规划纲要》(https://www.gov.cn/xinwen/2016-10/25/content_5124174.htm)首次对健康中国建设提出了长期的部署工作. 2017年10月18日, 健康中国被写入党的十九大报告, 健康中国上升为国家战略^[2]. 为贯彻落实健康中国战略, 2019年7月, 国务院印发《关于实施健康中国行动的意见》《健康中国行动(2019~2030年)》和《健康中国行动组织实施和考核方案》(https://www.gov.cn/xinwen/2016-10/25/content_5124174.htm)等文件, 在全方位干预健康影响因素、维护全生命周期健康和防控多种重大疾病3个方面部署了15项重大行动任务, 并明确了124项具体量化指标对健康中国建设情况进行实时评估.

2.2 健康中国战略阶段性成效

自健康中国战略实施以来, 我国的健康和卫生事业发展取得了令人瞩目的成就, 截至2021年, 我国卫生总费用占国民生产总值的6.5%, 医保覆盖率稳定在95%以上. 2015~2021年, 人均寿命由76.34岁增至78.2岁, 居民健康素养水平由10.25%增至25.4%, 婴幼儿死亡率由8.1‰下降至5.0‰, 5岁以下儿童死亡率由10.7‰下降至6.0‰, 孕产妇死亡率由20.1/10万下降至16.1/10万, 肺结核等传染病发病率由63.4/10万下降至45.37/10万, 主要健康指标总体上优于中高收入国家平均水平^[16~18]. 同时, 我国医疗卫生服务体系不断完善, 服务可及性不断提高. 2015~2021年, 每千人口医疗卫生机构床位数从5.11张增长到6.70张, 执业(助理)医师数从2.22人增长到3.04人, 注册护士数从2.37人增长到3.56人^[19]. 同时, 我国基本公共卫生服务均等化水平进一步提高, 2021年国家免费提供14大类国家基本公共卫生服务项目, 人均基本公共卫生服务经费补助标准从2009年的15元提高到2021年的79元^[20]. 目前, 我国已建成全球最大疾病和健康危险因素监测网络, 2021年重大慢性病过早死亡率下降至15.3%, 比2015年下降了3.2%^[21].

3 生物医学大数据发展及其对健康中国战略的推动作用

3.1 国内生物医学大数据建设与发展现状

20世纪80年代末,生物医学数据和卫生健康信息化成为国家战略性新兴产业、关键技术创新工程、国家级发展规划等重要创新应用和战略布局内容,支持了生物大数据标准化和集成融合技术、心血管疾病和肿瘤疾病大数据处理分析与应用研究等一系列重点课题^[22]。“十二五”期间,我国生物医学大数据迎来了重要发展机遇期,《“十二五”生物医学发展规划》(http://cn.chinagate.cn/economics/zgsew/2011-11/29/content_24032674.htm)明确提出支持医学、生物信息学等综合交叉学科的发展,加强科学数据的采集、保护和共享。“十三五”期间,生物医学大数据与生物医学技术融合发展逐步成为生物医药卫生科技创新的重点。进入“十四五”以来,科技部发布了“生物与信息融合”“科技创新2030——新一代人工智能”等重点专项,进一步加强新一代信息技术和生物医学大数据深度交叉融合。

在生物医学大数据资源建设与整合方面,我国不断加强生物医学科学数据资源整合与共享服务平台建设,2011年启动了生物医学数据相关的国家科技基础条件平台建设,并于2018年启动了多个国家级生命科学数据中心的组建工作。目前,国家微生物科学数据中心整合微生物资源、微生物组学等数据资源量达6 PB,数据记录超52亿条(<https://nmdc.cn>);国家基因组科学数据中心整合人、动物、植物、微生物基因组数据达39 PB(<https://ngdc.cncb.ac.cn/?lang=zh>);国家人口健康科学数据中心整合生物医学、基础医学、临床医学、公共卫生、药学等专业领域数据资源量约1.22 PB,数据记录超182.4亿条(<https://www.ncmi.cn>)。我国还建成了中国慢性病前瞻性研究数据(China Kadoorie Biobank, CKB)、泰州队列、精准医学队列等具有国际影响力的生物医学科学数据资源。

同时,自《2006~2020年国家信息化发展战略》(https://www.gov.cn/xinwen/2016-07/27/content_5095336.htm?cid=303&wd=&eqid=8a5430a40003bee10000000364759777)发布以来,医疗卫生信息化建设和医疗服务信息化已成为我国信息化发展战略的重点。此后,我国相继出台了《关于加快推进人口健康信息化建设的指导意见》(https://www.gov.cn/gongbao/content/2014/content_2600086.htm)、《关于印发促进大数据发展行动纲要的通知》(https://www.gov.cn/gongbao/content/2015/content_2929345.htm)、《国务院关于印发“十三五”卫生与健康规划的通知》(https://www.gov.cn/zhengce/content/2017-01/10/content_5158488.htm)等一系列政策文件,为生物医学大数据应用发展提供了信息化支撑。随着医疗卫生信息化事业持续快速发展,我国已建立了全员人口数据库、居民电子健康档案数据库、电子病历数据库等基础数据库,电子病历、健康档案、公共卫生数据等不同类

型的生物医学大数据持续快速积累^[23]。数据显示,2021年我国数据产量达到6.6 ZB,据此估算当前我国生物医学大数据估计可达数百EB,为健康中国战略迈向纵深提供坚实的数据基础^[3]。

3.2 生物医学大数据在实施健康中国战略中的作用

党的十八大以来,我国坚持面向世界科技前沿、面向经济主战场、面向国家重大需求和面向人民生命健康,将生物医学大数据作为国家重要的基础性战略资源和实施健康中国战略的重要前提,高度重视生物医学大数据在医药卫生科技创新、公共卫生事件应对、医疗卫生服务改革、居民健康服务等方面的关键作用。

3.2.1 生物医学大数据在医药卫生科技创新中的重要应用

随着未来医学模式将向精准医学、智能医学、生物医学工程等以生物医学数据和计算方法为主的新型医学范式发生转变,大规模、高质量和开放共享的生物医学大数据越来越成为生物医药科技创新的基础条件。“十三五”以来,我国发布精准医学、国家重大慢病研究专项等国家重点研发计划专项,形成了一批具有较大科研价值的生物医学数据资源库,为临床、医药、公共卫生实践提供基础^[24,25]。国家“十三五”重点研发计划“精准医学研究”专项是生物技术和信息技术在医学临床实践交汇融合应用的典型,形成重大疾病的风险评估、早期筛查、个体化治疗等精准防治方案和临床决策系统,反映了“十三五”期间医药卫生科技发展的前沿方向^[26]。国家重点研发计划中“重大慢性非传染性疾病防控研究”“主动健康和老龄化科技应对”等专项都提出构建队列数据、医学影像数据、临床数据、组学数据等数据资源库,结合5G、区块链、人工智能等最新信息技术,在智能诊断、辅助医疗、互联网医疗、精准医疗等方面涌现出一批高质量科技成果。如国内学者使用多个项目转录组和DNA甲基化的生物医学数据证明,45种结直肠癌变体中29种易感性可能是由基因调节的顺式作用介导的^[27]。同时,生物医学大数据与数字化技术已应用于新药研发各个环节的创新研究,缩短了新药研发周期,降低了研发成本,优化了试验设计,加快了新药获批速度,极大促进了我国新药研发由仿制向创仿结合、自主创新转变。

3.2.2 生物医学大数据在驱动公共卫生事件应对中的重要作用

我国公共卫生事业发展面临传染性疾病和慢性非传染性疾病并发、公共卫生服务不均衡、公共卫生技术创新不足等诸多挑战,生物医学大数据的发展应用为突发公共卫生事件风险防控提供了新思路。针对突然暴发的新型冠状病毒感染(简称新冠感染)疫情,我国学者在全球率先完成新型冠状病毒测序并将数据面向全球公开发布,为我国乃至世界范围内的新冠疫情防控提供了基础数据支撑^[28]。国家基因

组科学数据中心、国家微生物科学数据中心分别建立发布了新型冠状病毒基因测序数据库、新型冠状病毒数据库^[29]。国家人口健康科学数据中心组织国家卫生健康委医疗服务指导中心、中国疾病预防控制中心等建立了新冠肺炎临床和流行病学数据库、冠状病毒数据库等数据资源库(<https://www.ncmi.cn/covid-19>),为新型冠状病毒的病毒结构、发病机理、疫苗等方面的研究提供了坚实基础。基于新冠感染疫情大数据结合地理信息时空位置数据,我国开展了健康码、行程码、核酸检测、疫苗接种等疫情防控的有效措施,提升疫情防控现场流调和风险人员排查的及时性、精准性,支撑了通信管理、公安、卫生健康(疾控)等部门数据共享和业务协同,实现数据动态维护、实时更新、实时共享、跨区协查,解决防疫信息协同度、风险人员定位精准度、及时区域协查等防控短板漏洞^[30-32]。

3.2.3 生物医学大数据在完善医疗卫生服务体系方面的重要作用

我国医药卫生服务体系改革尚处在进程中,一些体制性、机制性和结构性的深层次矛盾依然存在,生物医学大数据的发展应用可在完善医疗卫生服务体系、优化医院信息化管理、强化公共卫生监测、完善医保制度及推动医药安全源等方面发挥重要作用。“十三五”期间,我国建立的中国脑血管病大数据平台,积累了大量具有价值的脑卒中专病数据资源,包括每年450万国家病案首页数据、160万人中国卒中中心联盟上报数据、47万社区长期随访的队列数据等。基于该数据资源开展的多组学、临床诊疗、药品研发、机器学习等方面的研究,突破了脑卒中“防、治、管、康”过程的一些关键技术,推动我国全面开展卒中“防、治、管、康”全周期工作模式^[33-36]。在公共卫生服务领域,生物医学大数据的应用改变了我国公共卫生服务的模式,疫苗质量管理和监测、传染病直报和预警系统等都不再只依靠人工记录和报送,使公共卫生职能实现了智能化和快速化,推动了被动防疫向主动预防转化。在医疗保障领域,通过对医保支付环节的大数据监测,可以大幅降低人力资本的投入,避免人工失误,在设定的审计标准基础上迅速识别出违规支付行为,使事后审计逐步走向事前提醒和事中控制。在医药安全方面,基于生物医学大数据,我国建立了药品追溯制度,通过数据编码对药品流通进行监控,实现了药品“来源可查、去向可追”。

3.2.4 生物医学大数据推动居民健康促进方面的重要作用

随着工业化、城镇化、人口老龄化进程加快,经济发展进入新常态,健康危险因素增多,慢性病成为主要疾病负担,心理健康问题也日益凸显,给“全民健康”目标的实现带来了严峻挑战。面对当前居民健康问题的复杂性和多元性,生物医学大数据的发展应用可以为居民健康提升带来新的思路和路径。在我国持续推进的家庭医生签约服务工作过程中,

一些地区搭建了“互联网+”社区卫生健康管理服务体系,基于基本公共卫生服务数据和诊疗数据,针对高血压、糖尿病均设立了规范化的管理目标,可以为需求方的问诊决策提供更加全面的参考,在一定程度上促进了分级诊疗的实现。通过生物医学大数据的应用,居民改变了以往完全依赖医疗机构数据记录的局面,个人可以通过各种可穿戴设备和移动应用随时观测诸如体温、体脂、血糖、心率等基础健康数据,并通过基于基础数据的健康应用APP实时监测自己健康指标。同时,通过对气象和环境污染等大数据的分析,居民个人可据此采取应对环境影响的健康措施,减少环境影响造成的健康威胁。

4 生物医学大数据发展应用所面临的挑战

党的十八大以来,国务院先后印发的《“健康中国2030”规划纲要》《科学数据管理办法》将科学数据和卫生与健康放在了优先发展的战略地位,为生物医学大数据应用发展提供有利支撑。然而,生物医学大数据在驱动健康中国战略过程中仍然面临一些挑战。

4.1 生物医学大数据资源丰富,整合利用不足

虽然国家各级各类科研项目沉淀了大量具有较高价值的生物医学科学数据,但由于科研数据主题类型多样,数据资源关联程度低,加上数据来源少、数据规模小、数据质量差别大等问题,难以形成规模大、价值高、范围广的生物医学大数据,对健康中国战略的支撑作用有限。来源于医疗卫生服务、健康管理等机构电子病历、健康体检等真实世界数据规模大、范围广、价值高,是生物医学大数据的主要来源。然而,由于数据大部分是业务数据,数据来源复杂、存储来源分散、标准化程度低、缺乏数据统筹管理与综合利用的体制机制,绝大部分生物医学大数据无法应用于决策支持和科研创新,可用于健康中国建设的数据占比严重不足,不能充分发挥其对健康中国战略的支撑作用。此外,生物医学数据产生过程可能涉及的主体包括患者、医疗机构、医生等,缺乏责权利相互促进和相互制约的数据权属确认、开放共享、安全保护及数据监管等制度,生物医学大数据共享应用与安全保护措施难以平衡,导致大多数生物医学大数据不能有效共享和充分利用。

4.2 生物医学大数据平台基础薄弱,体系功能仍需完善

虽然我国已挂牌成立国家基因组科学数据中心、国家微生物科学数据中心、国家人口健康科学数据中心等在内的多个生物医学相关的国家科学数据中心,为生物医学科学数据汇交整合与共享应用提供了基础平台和组织保障。然而,当前国家科学数据中心的投入严重不足,基础设施、人才等方面基础薄弱,目前仍然处于服务型数据中心的发展阶段,

对比国外已逐步转型为研究型数据中心的生物医学大数据中心,我国生物医学科学数据中心对医药卫生科技创新与前沿性研究支撑的作用有限,亟须提档升级。同时,虽然近年来国家卫生健康委员会(简称卫健委)按照“1+5+X”的总体规划布局了国家健康医疗大数据中心等数据基础设施建设,然而,由于数据孤岛、数据权属、数据安全等方面的制约,生物医学大数据难以形成有效汇聚、融合和共享应用,生物医学大数据的潜在价值不能有效发挥。

4.3 生物医学大数据流失海外,危及国家安全

由于生物医学大数据整合汇聚、安全管理与整合利用等方面的体制机制不健全,导致大量生物医学数据流失海外。《2020年中国互联网网络安全报告》(http://www.cac.gov.cn/2021-07/21/c_1628454189500041.htm?from=groupmessage)显示,我国2020年仅基因数据通过网络出境就高达717万余次,新型冠状病毒数据出境次数达99万余次,医学影像数据出境497万余次。尤其是在科研领域,由于国内学者申请NIH等境外资金支持的在国内开展的科研项目结题时要求提交项目数据,《Nature》等国外高水平的生物医学期刊在作者投稿时要求提供研究数据,导致多达数十PB的国内生物医学科学数据存储在境外生物医学数据中心,背后则是成百亿元的国家科技投入,不仅给我国社会经济带来重大损失,还严重危害国家安全。

4.4 生物医学大数据安全和隐私保护与数据共享应用难以平衡

生物医学大数据不仅涉及个人属性数据、个人健康状况数据、生物识别数据等个人敏感和隐私信息,还包括国家人口基础信息、人类遗传资源信息等反映群体健康生理状况、族群特征等的重要数据。世界各国在大力推动生物医学大数据共享应用的同时,也通过各项法律法规或制度措施来保护数据隐私和安全^[37]。然而,由于我国隐私保护、数据安全和数据权属等相关法律法规和制度不完善,生物医学大数据安全、隐私数据保护与数据共享应用难以平衡。目前主要存在两个突出问题:一是过度数据保护而不共享数据,导致数据不可发现、不可访问、不可重用,大量具有丰富价值的数据被闲置,数据资源价值的挖掘利用严重不足;二是已共享应用的数据由于数据安全和隐私保护不到位,导致未经同意或非法收集、窃取、处理、共享、使用和交易生物医学大数据的风险不断增大,对个人、社会和国家带来严重危害^[38]。

4.5 生物医学大数据标准欠缺,标准应用不足

近年来,我国已制定《卫生信息基本数据集编制规范》《电子病历基本数据集》等生物医学大数据相关标准与规范。然而,大部分生物医学大数据采自不同的机构、部门和

业务系统,不同类型机构间标准、接口规范不统一、术语代码类标准不健全、内容与格式不统一、数据质量良莠不齐,导致基础数据采集、处理和整合困难,数据在各个机构、部门和业务系统海量沉积,数据利用率极低,数据资源价值不能有效挖掘^[3]。同时,我国生物医学大数据标准应用管理体制机制不健全,国家对地方及行业的标准应用管理缺乏有效的衔接和管理机制,部分数据标准并未真正落到应用。

4.6 复合型数据人才不足

生物医学大数据发展过程中仍旧面临着众多制约因素,其中大数据人才的稀缺则是关键。根据现有人才市场上对大数据人才的需求状况,生物医学大数据人才应是掌握基本医疗知识、数学、统计学、数据分析、机器学习和自然语言处理等多方面知识的复合型人才。针对生物医学大数据未来数据体量大、数据类型多、价值高、更新速度快的特征,我国亟须建设和完善教学科研领域重点专业的创新教育体系和专业课程体系。此外,生物医学大数据复合型人才培养所需实验室、数据库以及大数据分析平台等硬件建设、产学研基地建设仍然缺乏,难以为复合型数据人才技术和实践能力提升提供培养环境。

5 推动生物医学大数据发展应用,助力健康中国战略实施的建议

5.1 推动业务数据向科学数据转化,完善生物医学数据资源建设

建议制定生物医学相关业务数据向科学数据转化的制度,制定相关业务数据向科学数据转化的标准、规范、技术等,通过科研立项、委托服务等方式,全面推动真实世界的生物医学大数据向生物医学科学数据转化,打造我国自主、可控、安全、共享的生物医学科学数据资源体系。建议明确国家财政支持的生物医学科学数据权属归国家所有,由科学数据中心负责保藏、管理和开放共享。建立数据出版、数据标识、引用规范和权益保护制度,形成标识清楚、产权清晰、可规范引用和共享的生物医学科学数据资源库。

5.2 加大数据中心建设投入,完善研究型数据中心建设

建议加大生物医学数据中心的投入,加强科学数据中心基础设施和服务体系建设,建立和完善数据中心运行体制机制,畅通汇交渠道,实现数据汇聚,为生物医学大数据汇聚提供平台和服务支撑。落实生物医学科学数据开放共享制度,分级分类开展生物医学大数据共享服务。建议进一步完善数据中心功能,推动现有生物医学大数据中心逐步转变为研究型数据中心,并进一步向平台型的研究数据中心转变。

5.3 创新生物医学大数据建设体制机制, 防止数据流失海外

建议发挥我国政治体制和制度优势, 加强顶层设计和统筹协调, 联合科技部、卫健委、医疗科研机构等项目数据或行业主管部门, 规划优先收集和重点保存利用的关键领域数据, 制定和建立项目数据、行业数据、社会数据等生物医学大数据汇交、收集和管理的相关政策和办法, 加强对医疗卫生服务机构、科研院所、高校等法人单位数据资源建设, 完善健康医疗科学数据的汇交机制和渠道, 要求财政资金支持的科研数据和学术期刊论文支撑数据汇交至国家生物医学大数据中心, 防止生物医学大数据外流, 维护国家安全。

5.4 加强数据安全制度体系和技术攻关, 推动生物医学数据安全共享

建议加快生物医学大数据安全制度和体系建设, 通过健全管理责任、评价考核制度、风险评估制度、操作规程、技术规范等手段, 规范生物医学大数据安全管理, 保护数据免受泄露、窃取、篡改、非法使用等。建议加快建立生物医学大数据出版和共享保护机制, 建立数据标识、引用规范和数据权益归属制度, 保护生物医学大数据各利益相关方合法权益。建议加强生物医学大数据相关数据安全与隐私保护技术攻关, 重点突破数据安全传输存储、分类分级、数据脱敏、隐私计算等关键技术, 推动数据安全多方计算、联邦学习、全同态加密等数据开发利用支撑技术的部署应用, 实现安全前提下的数据开放共享与创新应用。

5.5 完善生物医学数据标准规范, 加强标准规范应用

建议探索建立由基础标准、数据标准、管理标准、技术标准、服务与应用标准和专业标准等组成的生物医学数据标准体系框架, 建立数据质量控制体系, 提高生物医学大数据的数据质量和价值。建议进一步探索和完善生物医学大数据的采集、处理、存储和交换的标准规范, 促进各级各类数据生产机构、管理机构 and 交易平台之间的数据交换与共享。建议建立和完善生物医学大数据标准规范实施相关措施和可落地的实施细则, 引入数据质量和标准的考核机制, 提高数据标准规范的执行与应用。

5.6 加大生物医学数据复合型人才培养

建议加强生物医学大数据相关专业课程建设力度, 明确大数据学科为核心、医学知识为辅的知识体系构建方向, 从数据科学的基础知识和发展、常用工具和方法、应用场景和案例等各个方面构建起规范的课程体系。建议加强对数据中心等相关单位内部人员数据素养的培训, 建设完善多层次、多类型专业人才培养方案, 更新从业人员的知识结构, 加强专业知识学习, 提高员工的综合科研能力和科学数据管理创新能力, 逐步打造具备医学信息学、数据科学、计算机科学、知识组织、生物医学等学科理论和技术背景的高学术水平专业团队。建议加大生物医学大数据复合型人才培养所需的实验室、数据库以及大数据分析平台等硬件建设和产学研基地建设, 建立复合型数据人才技术和实践能力培训基地。

致谢 感谢中国工程院重大咨询研究项目(2019-ZD-23)资助。

推荐阅读文献

- 1 Yao J H. Persisting in prioritizing the people's health and building a healthy China (in Chinese). Red Flag Manuscr, 2023, (2): 44–48 [姚建红. 坚持健康优先 建设健康中国. 红旗文稿, 2023, (2): 44–48]
- 2 Li B. Implementing the Healthy China Initiative (in Chinese). Popul Fam Plan, 2018, (1): 7–11 [李斌. 实施健康中国战略. 人口与计划生育, 2018, (1): 7–11]
- 3 Zhang S F, Luo W, Ma Y H, et al. Considerations and suggestions on accelerating the transformation of health and medical big data into scientific data (in Chinese). J Med Inf, 2022, 43: 8–13 [张胜发, 罗葳, 马玉环, 等. 关于加快健康医疗大数据向科学数据转化的思考与建议. 医学信息学杂志, 2022, 43: 8–13]
- 4 Liu D P, Yin L. Medical science data sharing practice in China (in Chinese). Comput Appl Softw, 2006, 10: 1–3 [刘德培, 尹岭. 中国医药卫生科学数据共享实践. 计算机应用与软件, 2006, 10: 1–3]
- 5 Ning K, Chen T. Big data for biomedical research: Current status and prospective (in Chinese). Chin Sci Bull, 2015, 60: 534–546 [宁康, 陈挺. 生物医学大数据的现状与展望. 科学通报, 2015, 60: 534–546]
- 6 Yin C L, Wu J Y, Zou J T, et al. Comparative analysis of the national health programs between China and America (in Chinese). Chin J Health Policy, 2017, 10: 45–52 [尹纯礼, 吴静雅, 邹佳彤, 等. 中美国家健康战略比较分析及启示. 中国卫生政策研究, 2017, 10: 45–52]
- 7 Shu T, Liang M H. An analysis of the content of the US Federal Government's Medical Informationization Strategic Plan (2015–2020) (in Chinese). China Digit Med, 2015, 10: 2–4 [舒婷, 梁铭会. 美国联邦政府医疗信息化战略规划(2015–2020)内容简析. 中国数字医学, 2015, 10: 2–4]

- 8 Ohno-Machado L. NIH's Big Data to Knowledge initiative and the advancement of biomedical informatics. *J Am Med Inf Assoc*, 2014, 21: 193
- 9 Adams S A, Petersen C. Precision medicine: Opportunities, possibilities, and challenges for patients and providers. *J Am Med Inf Assoc*, 2016, 23: 787–790
- 10 Fair A, Watson K S, Cohn E G, et al. Innovation in large-scale research programs: Elevating research participants to governance roles through the all of us research program engagement core. *Acad Med*, 2022, 97: 1794–1798
- 11 Xu C, Yang Y, Liu Y R, et al. The enlightenment of European Union Healthy Strategy on healthy China construction (in Chinese). *Chin Health Econ*, 2020, 39: 65–68 [徐畅, 杨渊, 刘雅茹, 等. 欧盟健康规划对健康中国建设的启示. *中国卫生经济*, 2020, 39: 65–68]
- 12 Samarasekera U. New EU health programme comes into force. *Lancet*, 2021, 397: 1252–1253
- 13 Hong N, Qian Q, Fang A, et al. Research progresses and comparative analysis of biomedical linked data (in Chinese). *Libr Inf Serv*, 2012, 56: 123–129 [洪娜, 钱庆, 方安, 等. 生物医学关联数据研究进展与比较分析. *图书情报工作*, 2012, 56: 123–129]
- 14 Zhang Y E, Wang Y Z. The policy and enlightenment of EU's scientific data management and open access: Taking the EU horizon 2020 program as an example (in Chinese). *Libr Inf Serv*, 2017, 61: 70–76 [张玉娥, 王永珍. 欧盟科研数据管理与开放获取政策及其启示——以“欧盟地平线2020”计划为例. *图书情报工作*, 2017, 61: 70–76]
- 15 Dai T, Liu X X, Wei X. Characteristics and revelations of national and sectoral industry in the mid-term and long-term strategic plans (in Chinese). *Chin J Health Policy*, 2009, 2: 26–31 [代涛, 刘晓曦, 韦潇. 国家和部分行业制定中长期战略规划的特点与启示. *中国卫生政策研究*, 2009, 2: 26–31]
- 16 Department of Planning, Development and Informatization Technology, National Health Commission of the People's Republic of China. Statistical bulletin of China's health development in 2021 (in Chinese). *Chin J Viral Dis*, 2022, 12: 321–330 [国家卫生健康委员会规划发展与信息化司. 2021年我国卫生健康事业发展统计公报. *中国病毒病杂志*, 2022, 12: 321–330]
- 17 Bagcchi S. WHO's global tuberculosis report 2022. *Lancet Microbe*, 2023, 4: e20
- 18 Wei X Y, Du R R, Yang Y, et al. The implementation process and effectiveness of the Healthy China Initiative Since the 18th National Congress (in Chinese). *J Med Res*, 2022, 51: 1–5 [魏晓瑶, 杜然然, 杨渊, 等. 十八大以来健康中国战略的实施历程与成效. *医学研究杂志*, 2022, 51: 1–5]
- 19 National Health Commission of the People's Republic of China. 2021 China Health Statistics Yearbook (in Chinese). Beijing: China Union Medical University Press, 2022 [国家卫生健康委员会. 2021中国卫生健康统计年鉴. 北京: 中国协和医科大学出版社, 2022]
- 20 Cheng N, Song D P, Cui Y R. Analysis on the status of the implementation of national basic public health service program (in Chinese). *Chin Health Econ*, 2022, 41: 60–62 [程念, 宋大平, 崔雅茹. 国家基本公共卫生服务项目实施现状及问题分析. *中国卫生经济*, 2022, 41: 60–62]
- 21 Li H X. China has world largest monitoring network for disease, health risk factors (in Chinese). *Chin J Mod Drug Appl*, 2022, 16: 195–196 [李和新. 中国建成全球最大疾病和健康危险因素监测网络. *中国现代药物应用*, 2022, 16: 195–196]
- 22 Yao S H, Luo A J, Feng L. Research progress of foreign medical big data and its implications (in Chinese). *J Med Inf*, 2016, 37: 16–21 [姚山虎, 罗爱静, 冯麟. 国外医学大数据研究进展及其启示. *医学信息学杂志*, 2016, 37: 16–21]
- 23 Jin X T. Big Data in Healthcare (in Chinese). Beijing: People's Medical Publishing House, 2018 [金小桃. 健康医疗大数据. 北京: 人民卫生出版社, 2018]
- 24 Wang J, Chen X, Lu H, et al. Feature-shared adaptive-boost deep learning for invasiveness classification of pulmonary subsolid nodules in CT images. *Med Phys*, 2020, 47: 1738–1749
- 25 Huo J, Ouyang X, Si L, et al. Automatic grading assessments for knee MRI cartilage defects via self-ensembling semi-supervised learning with dual-consistency. *Med Image Anal*, 2022, 80: 102508
- 26 Jiang H L, Jia C Y, Li M S, et al. Strengthening independent innovation in medicine and promoting the construction of a healthy China (in Chinese). *Acta Acad Med Sin*, 2022, 44: 658–660 [蒋慧莉, 贾春岩, 李明穗, 等. 强化医药自主创新, 推进健康中国建设. *中国医学科学院学报*, 2022, 44: 658–660]
- 27 Yuan Y, Bao J, Chen Z, et al. Multi-omics analysis to identify susceptibility genes for colorectal cancer. *Hum Mol Genet*, 2021, 30: 321–330
- 28 Guo L, Ren L, Yang S, et al. Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). *Clin Infect Dis*, 2020, 71: 778–785
- 29 Gao Y J. Recommendation of hospital library information resources for prevention and control of novel coronavirus pneumonia (in Chinese). *Chin J Med Libr Inf Sci*, 2020, 29: 63–68 [高雅洁. 医院图书馆对新型冠状病毒肺炎的信息资源推介. *中华医学图书馆情报杂志*, 2020, 29: 63–68]
- 30 Sun J J, Yang B. Research on provincial epidemic prevention and control management platform based on spatio-temporal location big data (in Chinese). *Surv Mapp*, 2022, 45: 162–164 [孙敬杰, 杨斌. 基于时空位置大数据的省级疫情防控管理平台研究. *测绘*, 2022, 45: 162–164]
- 31 Liu J, Liang W N, Liu M, et al. Development of the COVID-19 cluster containment evaluation system using the emergency management theory (in Chinese). *Chin Gen Pract*, 2021, 24: 2122–2126 [刘珏, 梁万年, 刘民, 等. 基于应急管理理论的新型冠状病毒肺炎聚集性疫情防控效果评价指标体系的构建. *中国全科医学*, 2021, 24: 2122–2126]
- 32 Wu Q X. The platform-based governance: The reform of government governance in a “digitalized battle against COVID-19” (in Chinese). *Jiangsu Soc Sci*, 2022, 6: 90–99 [吴青熹. 平台型治理: “数字抗疫”中的政府治理变革. *江苏社会科学*, 2022, 6: 90–99]

- 33 Cheng S, Zeng Y L, Li H, et al. Multiomics research assists in the development of new stroke drugs (in Chinese). *Chin J Stroke*, 2022, 17: 213–215 [程丝, 曾亚龙, 李昊, 等. 多组学研究助力卒中新药研发. *中国卒中杂志*, 2022, 17: 213–215]
- 34 Chen S D, Yu W R, Huang X Y, et al. A study on 1-year functional prognosis prediction of newly diagnosed acute ischemic stroke patients based on machine learning (in Chinese). *Chin J Stroke*, 2022, 17: 265–271 [陈思玓, 俞蔚然, 黄馨莹, 等. 基于机器学习的新发急性缺血性卒中患者1年功能预后预测研究. *中国卒中杂志*, 2022, 17: 265–271]
- 35 Wang Y J, Li Z X, Gu H Q, et al. Chinese stroke report 2020 (Chinese Version) (Part 1) (in Chinese). *Chin J Stroke*, 2022, 17: 433–447 [王拥军, 李子孝, 谷鸿秋, 等. 中国卒中报告2020(中文版)(1). *中国卒中杂志*, 2022, 17: 433–447]
- 36 Wang Y J, Li Z X, Gu H Q, et al. Chinese stroke report 2020 (Chinese Version) (Part 2) (in Chinese). *Chin J Stroke*, 2022, 17: 553–567 [王拥军, 李子孝, 谷鸿秋, 等. 中国卒中报告2020(中文版)(2). *中国卒中杂志*, 2022, 17: 553–567]
- 37 Chen S, Duan B, Zhu C, et al. Privacy-preserving integration of multiple institutional data for single-cell type identification with scPrivacy. *Sci China Life Sci*, 2023, 66: 1183–1195
- 38 Chen Y. Research on healthcare data sharing and personal information protection (in Chinese). *J Intell*, 2023, 42: 192–199 [陈怡. 健康医疗数据共享与个人信息保护研究. *情报杂志*, 2023, 42: 192–199]

Summary for “加强生物医学大数据建设应用, 推动健康中国战略实施”

Strengthening the construction and application of biomedical big data to promote the implementation of the Healthy China Initiative

Jianwei Wang¹, Ling Yin², Depei Liu^{3*}, Wei Zhou⁴, Jie Chen⁵, Guanhua Du⁶, Qizhu Tang⁷, Xu Ma⁸, Jingyuan Chen², Xuemei Su⁹, Jing Wu¹⁰, Haiyan Li¹¹, Chenguang Cao¹², Qing Qian¹³, Xiaohui Zhai¹⁴ & Shengfa Zhang¹

¹ Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China;

² People's Liberation Army of China (PLA) General Hospital, Beijing 100080, China;

³ Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, Beijing 100730, China;

⁴ National Population Health Data Center, Beijing 100730, China;

⁵ Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing 100005, China;

⁶ Institute of Materia Medica, Chinese Academy of Medical Sciences, Beijing 100050, China;

⁷ The First Clinical School of Wuhan University, Wuhan 430060, China;

⁸ Institute of Science and Technology, National Health Commission of the People Republic of China, Beijing 100081, China;

⁹ Information Center, Chinese Center for Disease Control and Prevention, Beijing 102211, China;

¹⁰ Chronic Disease Control Center, Chinese Center for Disease Control and Prevention, Beijing 100500, China;

¹¹ Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China;

¹² Center for Information, National Medical Products Administration, Beijing 100044, China;

¹³ Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100020, China;

¹⁴ National Center for Medical Service Administration, National Health Commission of the People Republic of China, Beijing 100044, China

* Corresponding author, E-mail: liudp@pumc.edu.cn

The Chinese government has always prioritized the rights and interests of people, continuously striving to improve their level of health. In 2017, the “Healthy China Initiative” was clearly defined as a national strategy. In 2019, the “Opinions of the State Council on Implementing the Healthy China Action” and the “Healthy China Action (2019–2030)” clearly defined the overall goals of the Healthy China Initiative and established key directions and tasks. While vigorously promoting the Healthy China Initiative, China's biomedical big data shows an explosive growth trend and has become an essential fundamental strategic resource for driving the comprehensive and in-depth implementation of the Healthy China Initiative. However, China still needs to improve its construction of biomedical big data resources, data services, data sharing, innovative applications, and other aspects. There is an urgent need to strengthen the development and application of biomedical big data to give full play to its supporting role in the Healthy China Initiative.

By reviewing the development status of health strategy and biomedical big data in China and elsewhere, this study found that many countries have introduced national health strategies or actions to improve residents' health status. At the same time, domestic and international efforts are continuously strengthening the integration and sharing of biomedical scientific data resources and the construction of data platforms. Global biomedical big data is accumulating rapidly, providing a solid data foundation for countries' health strategies to move forward in depth.

With the implementation of the Healthy China Initiative, China's health status has made remarkable achievements, with a continuous increase in per capita life expectancy and a continuous decrease in the infant and child mortality rate, under-five mortality rate, and maternal mortality rate. China's medical and health service system is constantly improving, and service accessibility continues to improve. At the same time, China has regarded biomedical big data as an essential strategic resource and a prerequisite for implementing the Healthy China Initiative, and it attaches great importance to developing and applying biomedical big data.

In recent years, biomedical big data has played a crucial role in medical and health technology innovation, public health event response, medical and health service reform, and resident health services. However, the development and application of biomedical big data still face the following problems: (1) Biomedical big data resources are rich, but integration and utilization still need to be improved. (2) The foundation of the biomedical big data platform is weak, and the system functions still need to be improved. (3) The loss of biomedical big data overseas endangers national security. (4) Biomedical big data security and privacy protection are challenging to balance with data-sharing applications. (5) Standards for biomedical big data are lacking, and the application of such standards needs to be improved. (6) There needs to be more composite data talent.

Finally, this study proposes to strengthen the development and application of biomedical big data to promote the implementation of the Healthy China Initiative as follows: (1) Promote the transformation of business data to scientific data, and improve the construction of biomedical data resources. (2) Increase investment in data center construction, and improve the construction of research data centers. (3) Innovate the system and mechanism of biomedical big data construction to prevent data loss overseas. (4) Strengthen the data security system and technological breakthroughs, and promote the sharing of biomedical data security. (5) Improve biomedical data standards and norms, and strengthen the application of such standards and norms. (6) Increase the cultivation of interdisciplinary talent in biomedical data.

Healthy China Initiative, biomedical big data, scientific data, data sharing, data application

doi: [10.1360/TB-2023-0441](https://doi.org/10.1360/TB-2023-0441)