

关于加快健康医疗大数据向科学数据转化的思考与建议*

张胜发 罗 葳 马玉环 张 月 赵远志 张晓宇 王嘉阳 周 伟

(中国医学科学院国家人口健康科学数据中心 北京 100730)

〔摘要〕 介绍国内外健康医疗信息化与数据开放共享进展与成效, 分析我国健康医疗大数据开放共享面临的挑战和困境, 阐述健康医疗大数据向科学数据转化的必要性并提出建议, 包括建立健康医疗大数据向科学数据转化的机制、建立国家健康医疗科学数据资源目录、推动健康医疗科学数据规范化管理等。

〔关键词〕 科学数据; 健康医疗大数据; 数据资源; 数据共享

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2022.11.002

Considerations and Suggestions on Accelerating the Transformation of Health and Medical Big Data into Scientific Data

ZHANG Shengfa, LUO Wei, MA Yuhuan, ZHANG Yue, ZHAO Yuanzhi, ZHANG Xiaoyu, WANG Jiayang, ZHOU Wei, National Population Health Data Center, Chinese Academy of Medical Sciences, Beijing 100730, China

〔Abstract〕 The paper introduces the progress and results of domestic and foreign health and medical informatization and open sharing of data, analyzes the challenges and difficulties faced by open sharing of health and medical big data in China, expounds the necessity of transforming health and medical big data into scientific data, and puts forward some suggestions, including establishing a mechanism for transforming health and medical big data into scientific data, establishing a national health and medical science data resource catalog, promoting standardized management of health and medical science data, etc.

〔Keywords〕 scientific data; health and medical big data; data resources; data sharing

1 引言

健康医疗大数据是近几年随着数字浪潮和信息

现代化发展而出现的新概念, 目前尚无权威、统一的定义, 一般来说是指各种来源的大量非结构化或结构化的健康医疗数据^[1]。健康医疗大数据收集了大量健康、医疗、公共卫生保健等信息, 依靠传统的计算机统计分析技术是不够的, 需要找到新的方法更有效地进行海量数据分析^[2-3]。随着医药卫生科学技术和卫生健康信息化的快速发展, 环境因素、生活方式、遗传信息、健康表型、诊疗记录、疾病转归等深入融合、覆盖所有人群和整个生命周期的健康医疗大数据越来越成为国家重要的基础性战略资源^[4]。健康医疗大数据的充分、有效利用逐步成为国家安全战略、国家大数据战略、健康中国

〔修回日期〕 2022-10-12

〔作者简介〕 张胜发, 博士, 助理研究员, 发表论文 20 余篇; 通信作者: 周伟, 硕士, 高级工程师, 发表论文 10 余篇。

〔基金项目〕 中国工程院咨询项目“国家医学卫生健康与人口科学数据信息化发展模式研究”(项目编号: 2019-ZD-23)。

战略和创新驱动发展战略布局的重要内容^[5-7]。健康医疗大数据的应用与发展也成为国内外卫生健康信息化工作与医药卫生科技创新研究的重点^[8-9]。本文梳理了国内外医疗卫生信息化建设和健康医疗大数据应用进展,总结我国健康医疗大数据应用存在的问题和挑战,提出健康医疗大数据向科学数据转化的政策建议,以期为促进健康医疗大数据应用,释放我国健康医疗大数据资源价值提供参考。

2 国外健康医疗信息化与数据开放共享发展情况

2.1 总体情况

部分欧美发达国家很早就开始重视医疗卫生信息化建设和健康医疗大数据的管理、存储、共享及其应用,较早将其纳入国家数字战略,并搭建了较为成熟的健康医疗大数据服务平台,为健康医疗大数据的管理、开发与共享应用奠定了基础。

2.2 美国

美国是世界上较早推动医疗卫生信息化和健康医疗大数据应用的国家,于 20 世纪 70 年代就开始进行医疗卫生信息化工作,在 20 世纪 90 年代推进医疗卫生信息化标准的相关战略研究。2010 年,美国政府大力推动健康医疗大数据国家战略,并于 2014 年发布了《美国联邦政府医疗信息化战略规划:2015—2020》^[10-11]。目前美国已建立了国家级健康医疗大数据开放共享平台 healthdata.gov,公开数据内容包括医疗卫生服务信息、医疗卫生服务提供者信息、医疗卫生服务知识和科学知识,逐步实现健康医疗大数据共享,推动健康医疗大数据创新应用^[12-13]。

2.3 英国

英国早在 1994 年就开始建立覆盖全国医疗卫生机构的卫生信息网络,并于 2002 年将医疗卫生信息化正式纳入国家发展战略^[14]。当前英国已经建立了全国医疗卫生信息综合存储服务系统,该系统收集和存储了 2.3 万多个医疗卫生数据系统,覆盖

了 5 000 多万居民的健康医疗服务信息,能够为 130 多万医务人员和所有居民提供数据驱动的健康医疗服务^[12]。英国政府还专门建立了包含医疗健康在内的公共数据公开网站 data.gov.uk,以提高数据的公开度和透明度,促进健康医疗大数据创新应用。截至 2020 年 12 月,该网站汇集卫生健康领域的公开数据集 2 071 个,包括精神卫生、儿童健康、健康医疗资源与服务等^[15]。

2.4 日本

日本一直非常重视医疗卫生信息化产业发展,自 2001 年起在国家层面陆续出台了“电子日本战略(e-Japan)”“信息日本 2015 战略(i-Japan 2015)”等,有效推动了医疗卫生信息化发展,80% 的医院实现电子病历应用^[16-17]。近年来,日本基于疾病预防、临床诊疗、健康管理等大数据,不断探索应用云计算、人工智能等技术解决高龄化、医护资源不足及分布不均等问题^[18-19]。

3 我国健康医疗信息化进展与成效

3.1 发展情况概述

我国政府一直高度重视医疗卫生信息化和健康大数据的应用与发展。20 世纪 90 年代以来,我国实施了国家信息化重大应用项目建设,2000 年党的十五届五中全会将信息化纳入国家战略^[7]。医疗卫生信息化战略自 2006 年开始,《2006—2020 年国家信息化发展战略》明确提出加强医疗卫生信息化建设、推进医疗服务信息化等战略重点^[20]。此后,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》《关于印发促进大数据发展行动纲要的通知》《关于印发“十三五”全国人口健康信息化发展规划的通知》《关于加强全民健康信息标准化体系建设的意见》等一系列政策文件,从夯实基础平台建设、完善业务应用系统、健全健康医疗大数据标准、强化健康医疗大数据安全防护等方面不断加强医疗卫生信息化建设,为健康医疗大数据应用发展提供有力支撑^[21-23]。

3.2 不断加强基础平台建设

我国不断加强医疗卫生信息基础设施平台建设。各地区及有关部门正在协调推进国家卫生信息平台建设,加强全民、基本医疗、公共卫生、医疗保障等方面的数据收集,探索建立区域医疗卫生信息平台。当前,国家医疗健康大数据中心正处于探索建设阶段,省级医疗健康大数据中心和区域卫生信息平台建设初现成效,市、县级区域卫生信息平台建设也逐步推进。至 2020 年 12 月,我国 30 个省份已经建成了区域卫生信息平台,建设率达到 100%,市、县区域卫生信息平台建设率分别达到 62.8% 和 46.4%。通过区域卫生信息平台,省级平台采集医疗机构数据和基层医疗卫生机构数据的比例高达 100% 和 96.7%,市级平台采集医疗机构数据和基层医疗卫生机构数据的比例为 63.1% 和 59.3%,县级平台采集医疗机构数据和基层医疗卫生机构数据的比例为 42.5% 和 44.7%。平台基本覆盖数据上报、平台主索引、注册服务、数据采集与交换等 12 项基础功能。卫生健康、医疗保障、中医药与教育、科技、工业和信息化等跨部门健康医疗数据共享机制正在逐步探索建立。

3.3 完善业务应用系统

我国不断强化健康医疗服务相关机构业务应用系统建设,医疗卫生服务、公共卫生等机构或部门的业务应用系统功能不断完善,覆盖范围不断扩大。“十三五”期间我国计划生育应用信息系统已覆盖 13.7 亿人口,以电子病历为核心的医疗机构和基层医疗卫生机构信息化建设基本实现了全覆盖,公共卫生信息体系基本建立。“十三五”期间,我国建成病案首页库(5 亿条)、全员人口库(14 亿条)、死亡信息库(每年 600 万条)、药品编码库(17 万条)、医疗卫生机构库(98 万家)等数据资源。仅上海市医联工程建立的医疗健康信息库就汇聚了就诊记录 4.26 亿份,处方 21.09 亿条,检验检查报告 2.78 亿份,影像数据量累积达到 1 210TB。中国疾控中心系统已累计储存病历信息 1.5 亿人次,每年新增病历信息 1 000 万人次。

3.4 健全人口健康信息标准

我国人口健康信息标准体系不断健全,截至 2020 年 12 月,我国已制定《卫生信息基本数据集编制规范》《卫生信息共享文档编制规范》《健康档案共享文档规范》《电子病历基本数据集》《信息安全技术 健康医疗数据安全指南》等 200 多项健康医疗相关标准与规范。制定了国家卫生信息平台、医院信息化和国家公共卫生信息化建设等相关系统或平台标准和规范,制定印发了省统筹区域人口健康信息平台 and 医院信息平台的应用功能指引,有力支撑了健康医疗大数据的发展与应用。

3.5 强化健康医疗大数据安全防护

我国不断完善健康医疗大数据安全防护相关制度建设,相继出台了《网络安全法》《网络安全等级保护条例》《国家健康医疗大数据标准、安全和服务管理办法(试行)》《数据安全管理办法(征求意见稿)》《数据安全法》《人类遗传资源管理条例》《个人信息保护法》等相关法律法规,强化信息安全防护体系建设和数据安全保护,对个人信息安全和健康医疗数据安全等做出明确规定,有利于我国建立和完善个人隐私信息保护制度,推动健康医疗大数据的规范应用与发展。

4 我国健康医疗大数据开放共享面临的挑战和困境

4.1 数据壁垒现象严重

健康医疗大数据采自不同的机构、部门和业务系统,由于缺乏国家层面健康数据管理体制机制和常态化的信息联通机制,各地区、各部门的基础数据库、数据中心、平台存在重复建设、分散建设、多头管理、多头采集、多系统共存等问题。不同类型机构间标准、接口规范不统一、数据结构多样,导致底层数据采集、处理和整合困难。政府相关部门、医疗卫生服务机构、高校、科研院所、第三方检测中心等的数据资源分散在不同的机构信息化系统及数据库中,产生了“信息孤岛”和“数据壁垒”。

4.2 安全保护问题突出

健康医疗大数据含有大量隐私信息,数据的收集、存储、使用、加工、传输、提供、公开等各个环节处理环节存在着诸多不确定性,传统的数据安全保护策略和手段暴露出众多漏洞,内部失误、商业买卖、网络攻击等带来的数据泄露、数据丢失等问题,给数据安全保护带来了更大的挑战。同时,我国隐私保护、数据应用和相关法律法规和制度不健全,导致出现未经同意或非法收集、窃取、处理、使用和交易个人卫生健康与人口数据等现象,大量个人隐私和健康信息被泄露,严重损害社会利益和公民个人合法权益。与此同时,大部分机构、部门缺乏专业的数据安全管理和隐私保护能力,很难在保障数据安全的情况下对外提供或公布数据,导致卫生健康与人口数据在各个机构、部门和业务系统海量沉积,数据利用率极低,数据资源价值得不到有效挖掘。

4.3 标准应用不足

随着健康医疗大数据应用程度的不断深入和开放共享需求的急剧增加,健康医疗数据术语代码类标准不健全、相关标准执行不到位、医疗健康数据内容与格式不统一、数据质量良莠不齐等问题突出。卫生健康与人口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得到应用。

4.4 共享机制不健全

我国健康医疗大数据控制权、使用权、应用准入和退出机制等有关法律法规、配套政策及治理机制缺失,大部分拥有数据资源的组织无法合规合法地共享利用数据,造成数据资源极大浪费。健康医疗数据通常产生于医疗卫生服务机构、公共卫生机构、健康体检机构等主体,但以上主体普遍缺乏专业的医疗卫生大数据分析能力。同时,出于数据安全性考虑,健康医疗大数据生产机构对数据共享和应用多保持审慎的态度。有价值的健康医疗大数据不能合规合理地开放应用,严重影响了数据潜在价

值的开发。同时,拥有数据资源控制权的部分人员存在违规私自买卖或交易、滥用数据等行为,极大危害国家安全和经济利益。

4.5 缺乏权威数据管理平台

数据共享业务流程由于涉及收集、汇交、管理、存储、共享和利用等多个环节,容易出现数据内容、数据质量、数据格式不统一的现象,迫切需要加快建设统一、权威、互联的国家和地方医疗卫生数据管理中心和管理平台,建立和完善数据资源共享的开放系统,通过统一的数据管理平台建立数据资源目录,从而实现跨机构、跨部门、跨地区的数据开放共享和协同应用。

5 健康医疗大数据向科学数据转化的必要性

5.1 实现向科学数据转化有助于保护隐私安全、打通数据壁垒

健康医疗科学数据是通过基础研究、应用研究(临床医学、公共卫生、健康管理)等产生的,是经过过去隐私化、结构化、标准化处理,可共享利用的数据。国家已发布的《科学数据管理办法》对科学数据生产、安全、汇交、存储及共享等方面做出了明确、具有可操作性的规定。将健康医疗大数据转化为科学数据,可以通过去隐私化等方式保护隐私安全,利用结构化和标准化等技术打通数据壁垒。从技术、标准、法律等方面对科学数据进行规范管理,推动卫生健康与人口数据的整合汇交和加工整理。在规范化管理的基础上,采用已有的科学数据共享机制实现健康医疗科学数据的充分、有效共享利用和数据挖掘。

5.2 我国健康医疗大数据资源丰富,数据价值有待进一步发掘

根据已有数据预测,我国 3 万家综合性医院每年新增数据量可达 EB 级,2020 年我国健康医疗大数据量估计已达 2.26ZB,数据呈指数型增长,到 2030 年将达到 10ZB^[26]。按照英国国民健康系统 5 500 万患者记录数据价值 96 亿英镑显性商业价值

推算,我国健康医疗大数据潜在价值超 2 万亿元^[27]。由于信息的时效性,大部分健康医疗大数据的价值随着时间的推移而降低甚至消失,及时将健康医疗大数据转化为科学数据,可以明确数据使用权,有效解决数据孤岛、数据流失、隐私保护和数据安全等方面的问题,在保障国家和数据安全的条件下,最大限度地发挥健康医疗大数据资源价值,更好地支撑国家科技创新、经济社会发展和国家安全。

5.3 国家科学数据中心亟须获取高价值数据资源

目前,我国已经建立了包括健康医疗在内的相关领域国家科学数据中心,汇聚了健康医疗领域的业务数据、监测数据、研究数据等,覆盖健康管理、公共卫生、生物医学、临床医学、药学等数据类型,为健康医疗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。但是,当前国家科学数据中心的数据资源总量有限,国家科学数据中心对医药卫生科技创新与前沿性研究支撑作用有限,制约了健康中国战略和国家创新驱动发展战略迈向纵深,亟须有高价值的健康医疗大数据进一步转化为健康医疗科学数据,挖掘数据资源价值。

6 健康医疗大数据向科学数据转化的政策建议

6.1 建立健康医疗大数据向科学数据转化的机制

建议由国家行政主管部门负责制定健康医疗大数据向科学数据转化的制度。通过科研立项、委托服务等方式,全面推动健康医疗大数据向科学数据转化,通过去隐私化等方式保护隐私安全,利用结构化和标准化等技术打通数据壁垒,采用已发布的科学数据共享机制实现健康医疗科学数据的充分、有效共享利用。

6.2 建立健康医疗科学数据资源目录并出版发布

建议由国家行政主管部门负责,依托国家科学数据中心,建立健康医疗科学数据服务体系,对健康医疗科学数据进行注册登记,建立国家健康医疗科学数据资源目录并按规定发布,提高健康医疗科

学数据的可发现性和可访问性。完善健康医疗科学数据标识赋码规则,建立数据标识、引用规范和数据利用等相关制度。鼓励科研人员充分使用已有数据资源,对科研立项过程中的科学数据按照已出版的数据目录进行查重,减少数据的重复采集和存储。

6.3 加快推进健康医疗科学数据开放共享

建议完善已注册至国家科学数据中心的健康医疗科学数据资源开放共享机制,制定健康医疗科学数据标识赋码规则,建立数据标识、引用规范和数据权益归属制度,推进数据出版和共享保护。鼓励科研人员开展科学数据汇交与开放共享,建立对数据出版的保护与激励机制,将数据出版纳入科研成果体系和科研人员贡献度与影响力评价。建立科学数据共享激励机制,鼓励科研人员主动共享健康医疗科学数据,将共享数量和使用率纳入评审标准,通过资金、数据优先使用权、科研业绩等方式进行奖励。

6.4 完善健康医疗科学数据安全管理体系

建议由国家行政主管部门负责,推动各相关部门和单位开展健康医疗科学数据安全、标准和共享应用等方面的规范化管理。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,规范健康医疗科学数据管理,保护数据安全,避免数据泄露、窃取、篡改、非法使用等安全风险。通过建立由数据标准、技术标准、服务与应用标准等组成的健康医疗科学数据标准体系和质量控制体系,提高健康医疗科学数据资源规范性和可用性。

6.5 加强健康医疗科学数据标准与质量管理

由国家科学数据中心负责,推动建立由基础标准、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国家卫生健康与人口科学数据标准体系框架,建立国家卫生健康与人口科学数据质量控制体系,提高健康医疗科学数据质量和价值。

参考文献

- 1 Huang T, Lan L, Fang X, et al. Promises and Challenges of Big Data Computing in Health Sciences [J]. Big Data Research, 2015, 2 (1): 2-11.
- 2 De la Torre Díez Isabel, Cosgaya H M, Garcia-Zapirain B, et al. Big Data in Health: a Literature Review from the Year 2005 [J]. Journal of Medical Systems, 2016, 40 (9): 1-6.
- 3 Wang X, Carolyn W, Liu Z H, et al. Big Data Management Challenges in Health Research - a Literature Review [J]. Briefings in Bioinformatics, 2019 (1): 1-12.
- 4 Ouyang Z, Sargeant J, Thomas A, et al. A Scoping Review of "Big Data", "Informatics", and "Bioinformatics" in the Animal Health and Veterinary Medical Literature [J]. Animal Health Research Reviews, 2020, 20 (1): 1-18.
- 5 国务院. 关于印发促进大数据发展行动纲要的通知 [J]. 中华人民共和国国务院公报, 2015 (26): 26-35.
- 6 国务院. "健康中国 2030" 规划纲要 [J]. 中华人民共和国国务院公报, 2016 (32): 5-20.
- 7 中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要 [N]. 人民日报, 2021-03-13 (001).
- 8 代涛. 健康医疗大数据发展应用的思考 [J]. 医学信息学杂志, 2016, 37 (2): 2-8.
- 9 孙烨祥, 吕筠, 沈鹏, 等. 健康医疗大数据驱动下的疾病防控新模式 [J]. 中华流行病学杂志, 2021, 42 (8): 1325-1329.
- 10 董晓彤, 王香, 刘福蓬, 等. 国内外卫生信息化发展研究比较 [J]. 中国卫生事业管理, 2019, 36 (2): 110-112.
- 11 朱彦, 徐俊, 朱玲, 等. 主要发达国家医疗健康大数据政策分析 [J]. 中华医学图书情报杂志, 2015, 24 (10): 13-17.
- 12 舒影岚, 陈艳萍, 吉臻宇, 等. 健康医疗大数据研究进展 [J]. 中国医学装备, 2019, 16 (1): 143-147.
- 13 吴敏, 甄天民, 谷景亮, 等. 健康医疗大数据国内外发展及在卫生决策支持中的应用展望 [J]. 卫生软科学, 2019, 33 (2): 76-79.
- 14 胡红濮, 代涛, 刘硕, 等. 英国卫生信息化建设经验及启示 [J]. 中国数字医学, 2015, 10 (7): 10-14.
- 15 辛璐璐. 国际数字政府建设的实践经验及中国的战略选择 [J]. 经济体制改革, 2021 (6): 164-170.
- 16 刘彩霞, 缪园. 日本的 "IT 基本法" 和 "电子日本战略" [J]. 中国信息导报, 2002 (11): 40-42.
- 17 钱旦敏, 余苏, 吴辉群, 等. 基于 WIOD 的国内外医疗卫生信息化比较研究 [J]. 中华医学图书情报杂志, 2016, 25 (12): 1-6.
- 18 吴敏, 甄天民, 谷景亮, 等. 健康医疗大数据国内外发展及在卫生决策支持中的应用展望 [J]. 卫生软科学, 2019, 33 (2): 76-79, 89.
- 19 赵飞, 武琼, 曾良怀, 等. 健康医疗大数据共享开放模式研究 [J]. 中国卫生信息管理杂志, 2019, 16 (6): 649-654.
- 20 中共中央办公厅, 国务院办公厅. 关于印发《2006—2020 年国家信息化发展战略》的通知 [J]. 中华人民共和国国务院公报, 2006 (18): 6-13.
- 21 原国家卫生计划生育委员会, 国家中医药管理局. 关于加快推进人口健康信息化建设的指导意见 (国卫规划发〔2013〕32 号) [EB/OL]. [2021-12-01]. <http://www.nhc.gov.cn/guihuaxxs/s10741/201312/09bce5f480e84747aa130428ca7fc8ad.shtml>.
- 22 国务院. 关于印发促进大数据发展行动纲要的通知 [J]. 中华人民共和国国务院公报, 2015 (26): 26-35.
- 23 国务院办公厅. 关于促进和规范健康医疗大数据应用发展的指导意见 [J]. 中华人民共和国国务院公报, 2016 (19): 24-28.
- 24 原国家卫生计划生育委员会. 关于印发 "十三五" 全国人口健康信息化发展规划的通知 [J]. 中华人民共和国国家卫生和计划生育委员会公报, 2017 (2): 36-43.
- 25 国家卫生健康委员会. 关于加强全民健康信息标准化体系建设的意见 [EB/OL]. [2021-12-01]. <http://www.nhc.gov.cn/cms-search/xxgk/getManuscriptXxgk.htm?id=4114443b613546148b275f191da4662b>.
- 26 金小桃. 健康医疗大数据 [M]. 北京: 人民卫生出版社, 2018.
- 27 Ernst & Young. Realising the Value of Health Care Data: a Framework for the Future [EB/OL]. [2021-12-09]. https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/life-sciences/life-sciences-pdfs/ey-value-of-health-care-data-v20-final.pdf.