

大数据时代医疗健康数据治理方法研究

常朝娣^① 陈敏*

摘 要 目的: 探讨大数据环境下医疗健康数据治理方法。方法: 通过阐述大数据治理现状, 分析医疗健康大数据资源特性及治理问题, 并基于现有大数据治理方法论和模型, 探索大数据环境下医疗健康数据治理体系和实施步骤。结果: 建立医疗健康大数据治理体系, 以原则为驱动, 以核心指标为实施目标, 以大数据等相关技术为支撑, 从组织、标准、业务策略和流程等方面开展大数据治理工作。并对治理过程和治理效果进行评估, 及时调整现有实施方案。结论: 医疗健康大数据治理对提高医疗健康大数据质量、拓展医疗健康大数据应用具有重要作用。

关键词 医疗健康 大数据治理 治理体系

Doi:10.3969/j.issn.1673-7571.2016.09.001

[中图分类号] R-012 [文献标识码] A

Research on Governance Methods of Healthcare Resources in Big Data Era / CHANG Zhao-di, CHEN Min//China Digital Medicine.-2016 11(9): 02 to 05

Abstract Objective: To investigate the governance methods of healthcare resources under the environment of large data. Method: By describing the status of big data governance, analyses the characteristics and governance issues of healthcare big data resources. Based on the existing methodology and the model of big data governance, discusses the governance system and implement process. Result: Establishing the governance system of healthcare big data resources. With the principle for driving, the core indicators for the target, the big data and other related technologies as the support, from the organization, standards, business strategies and processes and other aspects to start the healthcare big data governance. It also should evaluate the governance process and effect to adjust the existing implementation measures. Conclusion: Healthcare big data governance has an important role in improving the quality and expanding the application of healthcare big data resources.

Keywords healthcare, big data governance, governance system

Fund project Research on the Healthcare Big Data Analysis Platform and Its Application (No. 0231516112)

Corresponding author School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, Hubei Province, P.R.C.

随着大数据时代的到来, 移动健康App的普及应用、互联网与医疗的深度融合使得医疗数据呈指数级增长趋势, 基于基因测序、生物分析、云计算等技术的进步与发展, 医疗健康大数据被广泛应用于临床决策支持、药物研发、远程病人数据分析、公共卫生领域等方面^[1]。由于医疗数据分布广而有序、医学信息的极度不对称及相关数据标准的缺乏, 并不能确保数据采集的准确性、完整性和及时性, 对于医疗健康大数据应用的关键要素—数据质量也就无法保证。在现阶段医疗数据剧增时期, 大数据治理显得尤为重要。

①基金项目: 医疗健康大数据分析平台及应用研究(编号: 0231516112)

*通讯作者: 华中科技大学同济医学院医药卫生管理学院, 430000, 湖北省武汉市硚口区航空路13号

①华中科技大学同济医学院医药卫生管理学院, 430000, 湖北省武汉市硚口区航空路13号

2 • China Digital Medicine, Eec.2016, Vol.11, No.9

1 大数据治理概述

《DAMA 数据管理知识体系指南》中将数据治理 (Data Governance) 定义为数据资产管理的权威性和控制性活动, 是对数据管理的高层计划与控制, 包括在数据管理和使用层面进行规划、监督和强制执行^[2]。我国ITSS WG1则认为数据治理包含对信息利益相关者的评估、有效助力业务的决策机制和方向及对合规绩效的监督^[3]。数据治理多关注过程, 即管理职责的范围和对数据进行有效管理所采取的政策和方法。大数据治理则是广义信息治理计划的一部分, 制定与大数据相关的数据优化、隐私保护或数据变现的政策和目标^[4]。数据治理提供相关政策、应用方法等保证数据的有效管理; 而大数据治理更关注通过运用不同技术工具挖掘大数据应用价值, 映射到医疗领域则是通过大数据治理为医疗健康大数据的融合、共享营造安全和谐发展环境, 从而推进医疗健康大数据各业务领域和层级的创新应用。

1.1 国外现状 国际相关组织在数据治理领域进行相关研究与实践, 如ISO 38500的IT治理模型、国际数据管理协会 (DAMA) 关于数据治理的七大环境要素、数据治理协会 (DGI) 的数据治理框架模型、国际商业机器公司 (IBM) 提出的数据治理成熟度模型等^[5]。美国一项调查显示, 大多数医疗保健机构并未开展数据治理相关项目的研究; 已开展的研究则多关注于数据传输标准制定和患者隐私安全, 较局限于具有直接强制性、合规性的领域; 并提出有效的数据治理框架应包括原则、政策、标准和过程、技术等元素^[6]。也有学者从IT数据治理出发提出数据治理的五大决策域, 即数据原

则、数据质量、元数据、数据获取的便捷性和数据生命周期^[7]。此外, 还有机构通过实施数据仓库治理计划确保数据的完整性和患者的隐私安全^[8]。

1.2 国内现状 我国数据治理相关研究较多, 且多应用于通信、金融、互联网等领域, 在医疗领域研究较少, 大多只停留在理论层面。有学者认为战略、机制、专题及实施构成数据治理体系的完整框架^[9]; 廖宇峰提出大数据治理参考模型的4大要素: 原则、关键域、实施和评估, 涵盖大数据治理工作的方方面面^[6]。我国ITSS WG1组织曾提出数据治理三大关键要素: 数据治理准则、数据治理域和数据治理实施方法论, 并在国际上形成共识^[3]。此外, 以时间、空间、语义和语境构成的全局数据也被认为是大数据时代数据治理的新范式, 全局数据下的数据治理是关注信息系统执行层面的体系, 通过流程、策略、标准和组织的有效组合来保证信息化建设的全过程监管, 从而解决提高数据质量^[10]。

2 医疗健康大数据资源特性及治理问题

2.1 医疗健康大数据来源及特征 根据医疗健康大数据的来源及作用效果不同, 可分为四大类: 诊疗辅助类: 包括亚健康及患病人群通过医疗机构、第三方检验机构或网络平台参与病情的咨询、预约、诊断、治疗等过程所产生的医疗数据; 健康监测类: 指基于移动物联网对个人身体体征及日常行为进行监测的生命体征类数据, 常被应用于慢性病患者的自我管理; 公共卫生类: 主要指区域性的医疗服务平台、公共卫生信息系统等产生的医疗数据; 定向生物医学类: 主要是关于新药研发、生物标本和基因测序

的信息, 多应用于个性诊疗、精准医疗、临床药物实验等医学研究。

医疗健康大数据除具有大数据5V (volume、variety、value、velocity、veracity) 特征外, 还具备自身鲜明特性, 即阶段性、时效性和冗余性: 阶段性: 由于医疗数据种类繁多、收集较为广泛, 加之部分医疗数据带有主观性质, 医疗信息存在明显的阶段性特征; 时效性: 疾病的发生和发展过程及医学影像、病理等信息有很强时间维度属性, 床旁监护等仪器产生的数据也具有很强的时效性; 冗余性: 信息孤岛、信息烟囱及疾病的多发性常导致医疗数据存在大量垃圾数据, 如常见病病情描述、慢性病患者的重复性检查等。

2.2 面临的问题 现阶段我国医疗行业数据治理主要存在以下问题: 一是缺乏有效管理规范体系: 现有医疗数据管理办法较散, 缺少较为完善的针对医疗健康大数据所制定的治理体系, 且多数医疗机构并未采取必要的治理措施。二是应用数据质量较低: 医疗健康大数据多为非结构/半结构数据, 如互联网云端数据、移动健康设备采集数据等, 各医疗卫生应用平台存在大量的异构数据, 导致挖掘医疗健康大数据时采集的原始数据质量差, 且数据处理过程繁琐, 影响最终的应用数据质量。三是数据共享困难: 多机构/企业/组织/政府间医疗信息的共享是开展数据挖掘项目的必要条件, 但目前医疗数据开放程度小, 作用并不显著。四是数据安全隐患较大: 网上预约诊疗、健康监测设备的普及应用必然加大医疗健康大数据泄露风险, 虽已有相关的隐私保护条例, 但仍需从大数据角度出发, 系统化建立医疗健康大数据的安全防范措施。

3 医疗健康大数据治理体系

医疗健康大数据治理体系是实施大数据治理的基础,应针对医疗健康大数据特性及数据管理现存问题,从大数据视角出发,系统化、规范化、科学性的建立医疗健康大数据治理体系。完整的医疗健康大数据治理体系应包含五方面:原则、核心指标、关键组件、技术及大数据治理评估,具体如图1所示。原则指导着大数据治理全过程,是实现治理目标的基本指导思想,包括战略一致、风险可控、运营合规和绩效提升^[5];核心指标是医疗健康大数据治理应实现的目标,组织关于大数据治理的所有活动都应围绕该目标进行实施,其中应用服务、数据质量、数据认责、数据安全性与合规性为大数据治理的重要参考指标;关键组件是实施大数据治理的重点内容,包括组织、标准、业务策略和流程;此外,为确保高的数据质量、可用性及安全性,组织急需建立大数据环境下的IT基础架构,从而为大数据治理提供技术支撑;对大数据治理实施过程及治理效果进行评估可深入了解组织当前治理现状,便于为及时调整、优化实施方案提供管理决策。

4 医疗健康大数据治理实施步骤

4.1 成立大数据治理组织 机构启动大数据治理项目的首要步骤是设计健全的大数据治理组织结构,它是全面开展治理工作的基础。定义角色与职责:治理团队成员应来源于不同层次,组织需严格划分成员类别并分配特定职责,执行个人参与数据治理问责制。团队成员包括管理人员、审计人员、制度/标准制定人员、业务人员、技术人员和评估人员;组织角色可分为

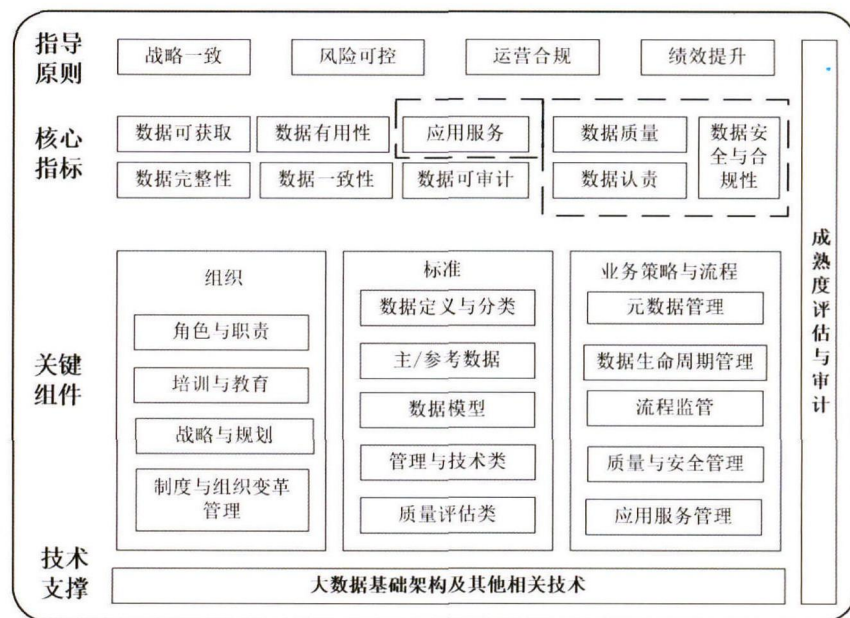


图1 医疗健康大数据治理体系

治理委员会、管理委员会、业务组、技术组、评估组。培训与教育:组织各成员需定期参与培训及继续教育,目的是将大数据治理提升到战略层面,使各类人员树立大数据治理意识及权责意识,提高大数据治理水平。战略与规划:通过评估现阶段大数据治理现状,依照四大原则制定或调整大数据治理的指导纲领及战略规划;并对机构相关业务、管理、决策等数据的全生命周期各阶段进行分析,识别各自需求,设计规划方案,其中规划需与机构总体发展战略、大数据发展战略相匹配,且各阶段参与人员也需要参与大数据规划制定。制度与组织变革管理:制定大数据治理各阶段管理制度,并根据治理环境变化对组织制度或流程进行适当变革。

4.2 建立大数据标准体系 医疗健康大数据必须遵循统一标准进行管理,才能有效保证各业务部门、系统间数据的规范性、流通性及共享性,是机构进行大数据治理的重要环节。大数据治理相关标准分为基础性和应用性

标准,包括数据定义与分类(元数据)、主数据、参考数据(数据字典)、数据模型、管理和技术类、质量评估类等内容,管理委员会及业务组人员可参考已有标准,并与现有医疗系统、业务流程相结合,开展医疗健康大数据标准体系的建立、实施、修改等工作。

4.3 制定业务策略和流程管理 建立和执行围绕医疗健康大数据生命周期的业务策略及流程管理是大数据治理实践基础,组织需定义与数据相关的业务规则,控制数据的访问与交换,建立持续的监测机制,并对业务流程进行监管。元数据管理:明确元数据管理策略,确定实现元数据管理策略所需的元数据集成体系结构,结合业务需求选择合适的业务元数据和技术元数据管理工具,并制定相应的元数据管理策略进行全面的元数据管理。可参考元数据管理成熟度模型,优化、提升元数据管理水平。数据生命周期管理:大数据时代管理重点是以相对可控的成本创造更多价值,一般包括

数据生成、传输、存储、清洗、集成、应用和数据销毁等过程。基于大数据治理战略规划,根据数据质量及应用情况,对大数据生命周期进行管理,是实现医疗健康大数据治理目标的重要过程,包括主数据、参考数据的管理等。流程监管:是对整个大数据治理流程进行监控和管理,保证治理项目实施进度及评估质量,包括流程目标、流程任务、流程分级的管理。质量与安全管理:组织需依据风险可控和数据的合规性原则,建立相应的大数据质量管控和安全规范及策略,定期检查、评估各阶段数据质量及安全可靠程度,保证大数据的质量和安。质量管理可分为元数据质量管理(真实性/完整性等)和过程质量管理(存储/传输/集成/分析/应用质量等);安全管理包括基础数据的使用安全、终端访问安全等。应用服务管理:应用服务管理是对医疗健康大数据深度加工、分析和应用的管理,从而提供应用创新服务,使之更好地服务于公众、管理决策人员等。

4.4 建立大数据基础架构 大数据环境下,为保障治理工作的顺利实施,除各阶段必要的信息技术或工具外,如质量控制、数据保护/访问、审计/安全工具等,组织还需搭建大数据基础架构,包括大数据集成基础设施、大数据分析平台、大数据处理平台、大数据质量监控平台等。

4.5 医疗健康大数据治理评估 大数据治理实施需与评估交叉配合完成,组织需对治理实施过程及效果进行评估^[5]。医疗健康大数据治理评估方法可分为成熟度评估和审计,前者反映组织在大数据治理规划设计、流程管理、保障机制(制度/标准/技术)等方面的水平,可采用IBM数据治理成熟度模型

度量数据治理能力;后者是组织内部审计人员基于全局的视角,依据相关法律、法规、政策、制度及标准对大数据治理环境、治理现状进行独立、客观的监督、风险分析、评价和咨询,利于提升组织大数据治理能力,保证治理目标的实现。

5 小结

个人医疗健康信息的激增、医疗数据互联融合的发展趋势,对隐私安全和数据质量提出新的挑战,对医疗健康大数据进行治理必然引发医疗行业各机构/组织的共鸣。医疗健康大数据治理是提高医疗行业大数据质量、应用水平及信息化管理水平的有效手段,加速“数据即服务”理念的深入理解和普及,同时也是医疗卫生事业健康、稳步发展的重要环节。技术的进步与发展为医疗健康大数据治理提供可能,但治理意识的培育和强化、参与主体的能动性才是实现医疗健康大数据治理的根本保障。

参考文献

- [1] 郭晓明,周明江.大数据分析在医疗行业的应用初探[J].中国数字医学,2015,10(8):84-85,111.
- [2] DAMA International.The DAMA Guide to the Data Management Body of Knowledge[M].USA:Technics Publications,2009.
- [3] 张明英,潘蓉.《数据治理白皮书》国际标准研究报告要点解读[J].信息技术与标准化,2015(6):54-57.
- [4] 桑尼尔·索雷斯.大数据治理[M].匡斌,译.北京:清华大学出版社.
- [5] 廖宇峰.大数据环境下的数据治理模型探析[J].农业图书情报学刊,2016,28(6):14-18.
- [6] Andronis,Moysey KK.Data governance for health care providers[J].Studies in

Health Technology & Informatics, 2013,193(193):299-315.

[7] Khatri V,Brown CV.Designing Data Governance[J].Communications of the Acm,2010, 53(1):148-152.

[8] Elliott TE,Holmes JH,Davidson AJ,et al.Data Warehouse Governance Programs in Healthcare Settings:A Literature Review and a Call to Action[J].eGEMs (Generating Evidence & Methods to improve patient outcomes),2013,1(1).

[9] 高亮.数据治理:让数据治理更好[J].中国教育网络,2014(12):64-66.

[10] 朱琳,赵涵菁,王永坤,等.全局数据:大数据时代数据治理的新范式[J].电子政务,2016(1):34-42.

【收稿日期:2016-07-28】

【修回日期:2016-08-06】

(责任编辑:赵士洁)

(下接第07页) 大数据时代的卫生信息化建设思考[J].中国卫生信息管理杂志,2014,11(1):43-46.

[5] 李萍,刘鹏.加强医院精细化管理的思考[J].现代医院管理,2011,9(5):28-30.

[6] Ginsburg GS,Willard HF.Genomic and personalized medicine:foundations and applications[J].Transl Res,2009,154(6):277-287.

[7] 李国杰,程学旗.大数据研究:未来科技及经济社会发展的重大战略领域—大数据的研究现状与科学思考[J].中国科学院院刊,2012,27(6):647-656.

[8] Tony Hey,Stewart Tansley,Kristin Tolle.The Fourth Paradigm:Data-Intensive Scientific Discovery[J].Microsoft,2009.

【收稿日期:2016-03-18】

【修回日期:2016-07-21】

(责任编辑:赵士洁)