• 医学信息研究 •

医疗数据信息安全政策研究*

郭 强 王乐子 母健康 朱 翀

（中国医学科学院肿瘤医院 北京 100021） （神州数码医疗科技股份有限公司 北京 100080）

胡新龙 张渝翔 路正鹏 马建辉

（中国医学科学院肿瘤医院 北京 100021） （神州数码医疗科技股份有限公司 北京 100080） （中国医学科学院肿瘤医院 北京 100021）

〔摘要〕 通过文本处理模型对医疗数据隐私相关法案进行关键词提取，通过统计分析词频的方式对相关法案发展做出分析并提出相应建议，包括发展保护个人数据隐私技术、推动医疗数据应用全流程管理等方面，为相关研究提供参考。

〔关键词〕 大数据；模型；隐私；词频；法案

〔中图分类号〕R－056 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2020.01.004

Study on Security Policy of Medical Privacy Data Information *GUO Qiang, Cancer Hospital Chinese Academy of Medical Sciences, Beijing 100021, China; WANG Lezi, MU Jiankang, ZHU Chong, Digital China Health Technologies Corporation Limited, Beijing 100080, China; HU Xinlong, ZHANG Yuxiang, Cancer Hospital Chinese Academy of Medical Sciences, Beijing 100021, China; LU Zhengpeng, Digital China Health Technologies Corporation Limited, Beijing 100080, China; MA Jianhui, Cancer Hospital Chinese Academy of Medical Sciences, Beijing 100021, China*

〔**Abstract**〕 The paper extracts key words of related bills of medical data privacy through text processing model, analyzes the development of relevant bills through statistical analysis of word frequency, and puts forward corresponding suggestions, including develop and protect personal data privacy technologies, promote the entire process management of medical data application, to provide references for related study.

〔**Keywords**〕 big data; model; privacy; word frequency; act

1 引言

在当今信息化迅速发展的时代，数据爆发性的增长给互联网、金融、医疗等行业带来了机遇和变革，与此同时也面临数据安全的挑战。在医疗健康领域，新式健康医疗智能仪器的广泛应用成为威胁

〔修回日期〕 2019－11－07

〔作者简介〕 郭强，室主任；通讯作者：马建辉。

〔基金项目〕 中国医学科学院医学与健康科技创新工程项目（项目编号：2017－I2M－2－003）。

数据安全的主要因素之一，移动互联网正成为医疗数据安全攻防的主战场，公民个人隐私和医疗数据应用安全的矛盾日益显露[1]。在隐私权保护方面，相比于欧盟和美国，现阶段我国没有专门的法律、法规。涉及个人隐私保护有关的法律法规分散于近 70 部法律、行政法规和 200 部规章中[2]，主要有宪法、民法、行政法、诉讼法和刑法等。由于我国民事立法研究较晚，对人格权研究深度不够，隐私权与隐私的分界线不甚明显[3]，与此同时又深受我国传统人文的影响，立法者没有给予足够的重视和保护。由于我国没有明确保护个人隐私权的成文法律法规。在隐私权相关案件中司法机构是依靠名誉权的形式来保护隐私权。2013 年 9 月工信部出台《电信和互联网用户个人信息保护规定》，根据《全国人民代表大会常务委员会关于加强网络信息保护的决定》（2012 年），进一步深化个人信息的覆盖范围，提出个人隐私信息的采集、储存和应用规则及安全保障等要求，为大数据应用的个人信息保护设立法律法规保障。这是目前国内最为完备和明确的关于个人数据隐私保护的法律[4-5]。

关于隐私权的安全保护措施是对人们个人信息负责，但也会妨碍其收集并影响数据的研究应用，进而限制大数据的使用和发展。随着近年来科技的发展，个人信息采集与隐私保护的矛盾越来越突出。根据对医疗健康大数据产业的研究，从数据安全合规的视角，结合国内出台的有关法律法规对医疗健康大数据行业个人隐私保护与信息共享安全问题进行简明分析。

2 医疗信息安全相关法案

2.1 分析方法

通过 Python 对每部法案进行全模式切词分词[6]，然后去除停用词并设置高频词、低频词阈值[7-8]，最后统计法案中的关键词词频，结合法案原文达到分析的目的。医疗数据的安全主要体现在医疗数据的采集（收集）、传输、存储、使用（应用）的整个流程中，而与数据共享相关主要有“共享”、“标准”、“规范”、“开放”等词；所以分析法案特征词时将“采集（收集）”、“传输”、“存储”、“使用（应用）”、“共享”、“标准”、“规范”、“开放”等作为特征词进行分析。由于分析时对涉及的每部法案进行独立分析，因此未对特征词做权重处理，仅用法案中的词频结合原文进行解读。本文主要分析多部律法条文，这些律法在医疗数据隐私安全以及共享方面具有代表性[9]。

2.2 医疗数据安全政策

2.2.1 《网络安全法》 2016 年 7 月《网络安全法（草案）》人大常委会二次审议稿在人大网对外发布征求意见。该草案共 7 章 68 条，要点在于保障网络空间主权安全、产品应用和服务网络安全、网络数据信息安全、网络安全监测预警以及突发事件应急处置预案等[10]。《网络安全法（草案）》的亮点：一是明确提出维护国家网络空间主权，将网络空间安全上升到国家主权安全的高度。二是将已有的有效规章法规上升为法律，与现有的法律法规实现较好的衔接，如《计算机信息系统安全保护条例》、《互联网信息服务管理办法》、《关于加强网络信息保护的决定》等，在草案中这些法案均被上升为法律。三是确立对重要信息相关的基础设施保护方略，参考国外法律法规明确重要信息基础设施的安全保护范围，另外还规定购买网络相关产品和服务要通过安全部门审查、向境外输出数据要进行信息安全评估和风险检测等，构建重要信息基础设施的安全保护体系。四是纳入网络安全战略规划、人才培养、技术研发、标准制定等综合性内容，草案将以上要点纳入网络安全基本法的范畴，依靠法律手段支持网络安全相关工作[11]。然而草案有其不足的方面：关于信息安全的范围表达不够准确；几乎没有涉及政府机关相应信息系统安全；主要是个人信息保护，未反映大数据时代的特点；一些条款较笼统、操作性较差。为弥补其不足之处，2017 年 6 月颁布的《网络安全法》增加 11 条修进，至此我国就网络安全领域的相关要求上升至国家法律层面[12]。在个人隐私信息保护方面，《网络安全法》有一个十分明显的特点，即个人对其信息有很强的控制权[13]。如相关机构在采集个人信息时需经被采

集者同意；应用和储存个人信息时必须严格按照与用户的协议；如果在使用或储存个人信息时违背与用户的协议，用户有权对其进行删除和更改。该法案中出现“安全”182 次、“共享”1 次、“开放”2 次、“保护”39 次。该法案规定了我国个人信息保护的基本框架，即“公布信息采集和应用的准则，明确信息采集、应用的目的、形式和范围并且需要经过被收集者同意”[14]，而在个人信息的开放共享方面未做过多阐述。另外 2017 年全国信息安全标准化技术委员会发布《个人信息安全规范》，为企业及研究机构保护和使用个人信息提供更加详细的操作规范[15-19]，自其正式发布以来被广泛应用于各行各业的合规实践中。虽然《个人信息安全规范》是处理信息安全问题方面国家推荐的标准，但对监管部门来说该规范是其在网络安全管理和执法过程中重要的参考准则。建议企业依据《个人信息安全规范》逐步提高个人信息保护水平，为其产品与服务保驾护航[20]。

2.2.2 《促进大数据发展行动纲要》（以下简称《行动刚要》）和《关于促进和规范健康医疗大数据应用发展的指导意见》 为推动各行业、各领域中的数据资源安全共享，国务院于 2015 年 9 月 5 日发布《促进大数据发展行动纲要》。其中“共享”、“开放”、“安全”和“隐私”分别出现 59、36、74、5 次，以此可以看出国家大数据发展中数据安全共享的重要性。该纲要是当前我国为推动信息化发展公布的第 1 份系统性的权威文件，是从国家信息化发展大局出发指导未来大数据发展的纲领性文件。近年来随着信息技术的发展，各行各业的数据呈爆炸式增长，已经成为与国家能源和物质同等重要的战略型储备资源。数据资源虽然具有增长速度快以及种类丰富等特点，但是来源极为分散且价值密度低，要想使数据资源造福国家及人民大众就必须打破地区和行业的壁垒，实现共享开放[21]。总的来看，《行动纲要》是针对我国大数据发展过程中“不愿、不敢、不会共享开放”情况的总体布局规划。因此提出建立国家级统一数据应用平台并且不同部门之间实现数据共享；构建国家级网络数据汇集开放共享和分析平台、国家统一的信誉信息共享开放平台、地市级及其以上政府统一的政务和惠民服务信息平台、全国统一的不同企业间的公共服务大数据共享开放平台。《行动纲要》设置的未来发展目标及任务是我国大数据发展的前瞻性战略集成，同时也面临不同区域、行业、部门之间的壁垒和利益分配以及信息安全等困难挑战，因此建立健全有效的法律保护政策是实现《行动纲要》的关键。为贯彻落实《行动纲要》的要求，国务院办公厅于 2016 年 6 月发布《关于促进和规范健康医疗大数据应用发展的指导意见》，该法规中“共享”、“开放”、“安全”、“隐私”分别出现 25、13、33、6 次，可见其重点在于数据的开放和共享。其基本原则为以人为本，创新驱动；规范有序，安全可控；开放融合，共建共享[22]。总的来说该法规的要点在于加速建立统一权威、相互联通的国家级信息健康共享平台[23]，推动医疗行业健康数据资源共享，进一步深化“互联网 + 健康医疗”服务。

2.2.3 《卫生行业信息安全等级保护工作的指导意见》（以下简称《指导意见》） 该《指导意见》中“等级保护”、“安全”、“信息安全”、“开放”、“共享”、“隐私”分别出现 43、89、63、0、0、0 次。通过对相关词频提取分析可知该《指导意见》的重点是信息安全等级保护方面。关于信息安全等级保护，国内是从 2007 年 6 月《信息安全等级保护管理办法》开始明确相关单位和部门的职责、任务；自 2010 年 4 月开始为完成安全等级测评标准体系建设以及信息系统三级安全等级测评相关工作，公安部发布《关于推动信息安全等级保护测评体系建设和开展等级测评工作的通知》。2011 年为贯彻执行定级、整改、测评相关国家标准，原国家卫生部发布《卫生行业信息安全等级保护工作的指导意见》、《关于全面开展卫生行业信息安全等级保护工作的通知》。在信息系统定级工作中相关国家标准是《计算机信息系统安全等级保护划分准则》、《信息系统安全保护等级定级指南》；系统建设整改工作中相关国家标准是《信息系统安全等级保护基本要求》、《信息系统安全等级保护实施指南》、《信息系统等级保护安全设计技术要求》；信息系统测评工作中相关国家标准是《信息系统安全等级保护测

评要求》、《信息系统安全等级保护测评过程指南》。另外还有几十个国家和相关部门颁布和标准、指南等技术指导性文件，构成信息安全等级保护的初级标准体系，基本可以满足国家信息安全等级保护制度施行的要求。

2.2.4 《信息安全技术 个人信息安全规范》（以下简称《安全规范》） 出现“个人信息”607 次、“安全”139 次“收集”91 次、“共享”52 次、“转让”46 次、“公开”56 次、“隐私”41 次、“隐私政策”37 次、“去标识化”11 次、“传输”11 次、“存储”16 次，由以上词频结合原文进行分析，可以看出《安全规范》是以国家制定的准则来明确个人数据信息的采集、储存、使用、共享规范操作，为个人隐私保护政策的制定提供方向。《安全规范》是在《网络安全法》的法律条文规定基础之上明确了网络信息安全中具体问题的定义。首先，对于个人敏感信息的定义是在《信息安全技术 公共及商用服务信息系统个人信息保护指南》中提出的，是指个人信息遭受暴露和更改后对个人信息主体产生恶劣的影响。《安全规范》则进一步指出个人敏感信息被泄密、被恶意使用会危及个人信息主体的人身、财产安全，致使其声誉和身心健康受到侵害等不良后果。其次，《安全规范》给出个人信息收集的定义，即个人信息主体自发给出的、网络运营商通过日志记录等其他手段自动采集的、数据使用者从其他机构间接得到的称为“收集”，另外规定在终端得到的个人信息如果没有传回相应的储存设备，不在收集的定义范围内。最后，《安全规范》对于个人信息处理时的匿名化与去标识化有不同定义，即匿名化处理后的数据信息不能还原，不再属于个人信息的范畴；而去标识化处理后的数据信息仅能在没有其他关联信息的情况下保证信息主体的安全，但依然存在潜在被关联分析识别的隐患。2017 年 8 月发布的《信息安全技术 个人信息去标识化指南》征求意见稿中提出去标识化的流程以及相关处理技术等，相关企业机构在进行去标识化数据处理时可以借鉴其具体方法和流程。

2.2.5 《人口健康信息管理办法（试行）》 2014 年国家发布《人口健康信息管理办法》。该法规一是给出人口健康信息包含范围，明确指出人口健康信息安全管理制度；二是明确个人信息采集时信息主体的知情同意权；三是明确个人健康信息的相应管理原则，明确规定个人信息要分级保存、定期更新维护；四是推行使用信息备案登记制度，给出信息使用的条件和模式；五是明确个人信息隐私安全保护要求，对个人健康数据进行安全等级保护和审查；六是明确相关法律责任。该法规中“共享”、“开放”、“安全”、“隐私”分别出现 4、0、15、6 次，该律法条文中涉及数据采集、存储、管理及共享等敏感词，然而法案原文中没有涉及数据安全传输的相关条文规定，结合法律原文第 13、14 条，可以看出该法规只对数据共享略作描述，没有解释实施细则，因此《人口健康信息管理办法》尽管是现阶段医疗健康行业的权威法规，但面对日益增长且复杂的数据，在医疗数据方面仍显不足。

2.2.6 《国家健康医疗大数据标准、安全和服务管理办法（试行）》 该法规中出现“健康医疗”75 次、“大数据”76 次、“安全”54 次、“管理”60 次、“标准”37 次、“规范”15 次、“共享”7 次、“采集”3 次、“存储”6 次、“传输”3 次，从该法规提取的关键词词频可以看出该法规涉及医疗数据的安全应用和共享以及安全管理的相关标准。结合原文从以下 4 个方面对该法规进行分析：一是标准管理方面。对于医疗数据安全管理倡导多方参与，建设完备的医疗健康数据标准化管理平台，严格制定标准化管理规范、鼓励制约制度以及风险评估等实施措施。二是安全管理方面。落实“一把手”负责制，完善数据安全人才培育制度，提出数据分级分类储存要求并对网络安全、数据传输、监测以及相关基础设施安全等关键管理环节提出确切的操作要求。三是在服务管理方面。实行统一分级授权、分类应用管理、权责一致的管理制度，明确医疗大数据从形成、采集、保存、应用、共享及销毁等关键环节中相关部门的职责定位，进一步加强对医疗健康数据的安全共享和开放。四是管理监督方面。呼吁相关机构医疗健康信息平台端口能够对卫生监管部门开放。定期开展对含有医疗数据的信息平台的安全评估检测，建立安全管理责任制。以

上几部法律法规均有自身的不足之处，或注重网络安全，或注重数据共享、系统平台建设应用等。《国家健康医疗大数据标准、安全和服务管理办法（试行）》的发布明确提出健康医疗大数据的含义，制定实施范围及总体标准流程，清晰地划分各级卫生行政部门权力和责任。从提取的关键词词频也可以看出该法规涉及医疗数据的安全应用和共享，是我国现阶段法律法规中关于医疗数据较为完善的一部法规。

2.3 医疗数据共享

《网络安全法》主要是国家基于整个网络空间领域安全的法律；《促进大数据发展行动纲要》和《关于促进和规范健康医疗大数据应用发展的指导意见》是为推动各领域数据共享制定的法规；《“健康中国 2030”规划纲要》是为促进互联网和医疗健康行业的融合制定的法规；《卫生行业信息安全等级保护工作的指导意见》是通过等级保护来保障医疗卫生行业信息安全制定的法规；《信息安全技术 个人信息安全规范》和《信息安全技术 个人信息去标识化指南》是为保护个人信息安全制定的法规；《人口健康信息管理办法（试行）》是关于保证人口健康信息采集、存储以及管理数据安全制定的法规；《国家健康医疗大数据标准、安全和服务管理办法（试行）》涵盖范围极广，是以上多部法律法规的集成和推广，是目前国内医疗健康领域较为权威的一部法规。在大数据隐私的探讨中对于大数据的共享应用和隐私保护一直是被争论的问题。无论业内人士如何讨论，立法者都应考量数据的隐私安全及其经济价值，在发展大数据的同时也应发展检测技术和监管机制以应对日益增加的数据量和复杂度。企业本身是趋利的，只有在法律层面对其做出规定才能有效解决大数据隐私安全和共享问题。加强建设适合于人工智能应用相关的数据安全共享法律法规，是探索人工智能在远程诊疗、精准医疗、临床诊疗、公共卫生服务等众多情景中应用的基石。人工智能技术发展如火如荼，然而此过程必须以人为本，保证医疗健康数据安全，规范医疗数据信息的归属权、使用权。加强数据管理及其安全性，对医疗健康敏感数据信息进行分类分级，建立强大有效的安全防护体系。医疗健康大数据是新近发展起来的事物，在应用处理的过程中会遇到各种问题，因此与时俱进地完善管理规范标准是常态，也是必要手段。

3 建议

3.1 概述

不论是在欧美发达国家还是在我国，医疗机构、保险公司以及医护人员均有保护就医者隐私的法定义务。但医疗机构内部管理存在一定漏洞，有可能导致患者隐私数据泄露。另外我国正在致力于建立标准化电子病历，以达到区域或全国医疗信息共享，也会面临隐私数据泄露的问题。医疗领域的研究人员往往对于去标识化相关知识比较欠缺，仅能根据常规手段处理直接标识符。需要注意的是经过常规手段去标识化后的个人数据也不是肯定安全的。因为随着信息技术的发展，大数据分析能力越来越强大，尤其是医疗行业，正朝着多源交叉分析的方向发展，分析人员更容易通过关联分析挖掘出更多的个人信息，从而增加隐私信息泄露的危险。通过学习和借鉴其他国家相关法律及经验并结合国内相关法律法规建设情况提出相应建议。

3.2 以人为本，完善个人隐私法律法规

公立医院改革是我国医药卫生领域改革的重点，医疗数据的信息化是改革的有效手段。因此应尽快对电子病历立法，明确医疗数据的归属权、使用权，确保其应用过程中的有效性、认可性、安全性，规范相关实体的权利和义务，明确对侵犯个人隐私数据所要承担的法律后果。

3.3 发展保护个人数据隐私的技术

随着国内医疗信息的发展，可以通过设置医疗领域的数据失真、数据加密、限制发布等政策加强对隐私数据的防护。如在信息去标识化方面，相关加密算法可以考虑使用深度学习、区块链等。同时还应加强去标识化和再识别风险相关的算法研究，

在共享之前针对已经去标识化的数据计算再识别风险，以确保共享数据的安全。总之，在敏感信息处理方面要跟上信息技术的发展，以确保医疗健康数据的安全。

3.4 推动医疗数据应用的全流程管理

虽然我国正在推动医疗领域的信息披露工作，但是对个人数据的使用和披露没有确切的规定，也没有完善的数据信息保护机制。国家应加强医疗隐私数据的安全风险评估和保护，尽量做到医疗数据使用和信息保护的平衡。可以参考美国健康保险流通与责任法案（Health Insurance Portability and Accountability Act，HIPAA）或者欧盟通用数据保护条例（General Data Protection Regulation，GDPR），建立医疗隐私数据安全管理机构，包括医疗数据管理委员会、隐私数据业务部门、隐私数据技术部门以及风险安全评估部门。完善数据安全共享相关法律法规，培训数据安全专业人才。相关企业要定期组织员工进行信息安全培训指导，增强保护医疗隐私数据安全的意识，并做好隐私数据安全评估，预防可能出现的风险。

4 结语

总之科技和法律的完善是保证数据安全应用的基石。虽然现阶段人们对个人隐私权不是特别重视，但随着时间的推移会逐渐察觉隐私权的重要性。我国迫切需要通过制定确切的医疗数据标准和流程，使人们真正认识到医疗行业中信息安全的重要性，这是我国医疗行业持续健康发展的保障。

参考文献

1 李玲娟，郑少飞. 基于数据处理的数据挖掘隐私保护技术分析［J］. 计算机技术与发展，2011，21（3）：94－97.

2 安敏. 论网络隐私权的法律保护问题［J］. 楚雄师范学院学报，2003，18（1）：72－75.

3 杨金丹. 网络隐私权的私法保护［D］. 长春：吉林大学，2010.

4 梁玲. 计算机网络信息安全技术研究［J］. 电子设计工程，2010，18（7）：209－210.

5 徐乐. 大数据时代隐私安全问题研究［D］. 成都：成都理工大学，2016.

6 郑贤君. 宪法文本分析：一种解释方法［J］. 法律科学，2008，26（2）：38－46.

7 黄萃，苏竣，施丽萍，等. 政策工具视角的中国风能政策文本量化研究［J］. 科学学研究，2011，29（6）：876－882.

8 徐国海. 面向中文医疗文本的命名实体识别研究［D］. 上海：华东师范大学，2019.

9 管丽莹，黄小蓉. 医院计算机网络及信息安全管理［J］. 现代医院，2006，6（8）：144－144.

10 杨海平. 网络信息安全研究［J］. 情报科学，2000，18（10）：944－947.

11 王兰成，李超. 论档案信息共享中的隐私保护及新技术［J］. 档案学研究，2017，24（4）：41－45.

12 王冠. 完善我国互联网有害信息监管法律制度的若干思考［D］. 南昌：南昌大学，2012.

13 王伟. 我国信息网络安全立法问题研究［D］. 西安：西安理工大学，2006.

14 张震江. 医院网络安全现状分析及研究［J］. 计算机系统应用，2006，15（7）：87－89.

15 刘逸敏. 基于访问目的的隐私数据访问控制机制研究［D］. 上海：复旦大学，2012.

16 于靓. 论被遗忘权的法律保护［D］. 长春：吉林大学，2018.

17 黄小燕. 欧盟网络个人信息法律保护研究［D］. 暨南大学，2018.

18 许怀湘. 美国区域卫生信息化，国家卫生信息网和医疗改革［J］. 中国数字医学，2009，4（10）：89－92.

19 朱晓勃. 我国医院信息化建设现状与发展对策研究［J］. 现代仪器，2015，21（1）：76－79.

20 许怀湘. 美国区域卫生信息化，国家卫生信息网和医疗改革［J］. 中国数字医学，2009，4（10）：89－92.

21 沈昌祥. 加快推进信息安全等级保护工作［J］. 信息网络安全，2008，8（5）：4－5.

22 王倩，朱宏峰，刘天华. 大数据安全的现状与发展［J］. 计算机与网络，2013，39（16）：66－69.

23 王红梅，宗慧娟，王爱民. 计算机网络信息安全及防护策略研究［J］. 价值工程，2015，34（1）：209－210.