

健康医疗大数据中心建设实践与思考

许文韵

(福建医科大学附属协和医院信息管理处 福州 350001)

〔摘要〕 介绍福州健康医疗大数据中心建设背景、历程,从平台架构、基础设施、数据采集、系统安全等方面总结中心建设成果,分析存在的问题,包括数据来源有限和存在数据壁垒、缺乏统一技术标准和基础数据治理成效慢等,提出应对措施和建议。

〔关键词〕 健康医疗大数据;大数据中心;建设;思考

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2020.08.010

Practice and Consideration on the Building of Health and Medical Big Data Center XU Wenyun, Information Management Department, Fujian Medical Union Hospital, Fuzhou 350001, China

〔Abstract〕 The paper introduces the building background and process of Fuzhou Health and Medical Big Data Center, summarizes the building achievements of the center from the aspects of platform architecture, infrastructure, data acquisition, system security, etc., analyzes the problems, including limited data sources and data barriers, lack of unified technical standards and slow effect of basic data governance, and puts forward countermeasures and suggestions.

〔Keywords〕 health and medical big data; big data center; building; consideration

1 引言

2016年6月国务院印发《关于促进和规范健康医疗大数据应用发展的指导意见》(国办发〔2016〕47号),指出要夯实健康医疗大数据应用基础,推进公共卫生大数据应用,认为健康医疗大数据应用发展将带来健康医疗模式的深刻变化,有利于激发医药卫生体制改革的活力,提升健康医疗服务效率和质量^[1]。2016年10月原国家卫计委启动第1批健康医疗大数据中心与产业园建设国家试点工程,确定福建省、江苏省以及福州、厦门、南京、常州

为第1批试点省市,其中福州立即启动国家健康医疗大数据中心建设工作。该项目是国家健康医疗大数据南方中心重要组成部分,旨在整合汇聚福州市乃至福建省人口健康医疗数据,加快“互联网+医疗”的发展,深化医疗卫生体制改革,引领福州市健康医疗产业链和数字经济发展。

2 平台建设情况

2.1 建设历程

2016年10月21日原国家卫计委启动第1批健康医疗大数据中心与产业园建设国家试点工程。2017年1月与中电数据签订合作框架协议,确定中电数据为健康医疗大数据中心与产业园建设国家试点工程(福州)的建设和运营主体单位。2017年4

〔修回日期〕 2020-03-12

〔作者简介〕 许文韵,工程师,发表论文1篇。

月确定“一个办法，两大平台，一个中心，两个基地”的建设方案，即《福州市健康医疗大数据资源管理暂行办法》（“一个办法”）、国家健康医疗大数据平台（福州）和国家健康医疗大数据安全服务平台（福州）（“两个平台”）。2018 年 6 月作为健康医疗大数据中心与产业园建设国家试点工程（福州）的核心基础设施，东南健康医疗大数据中心奠基动工，标志着试点工程进入实质性建设阶段。2018 年 11 月 29 日健康医疗大数据中心与产业园建设国家试点工程（福州）成果发布暨在榕生态企业投资签约活动在福州举行，福州市人民政府向中电（福建）健康医疗大数据运营服务有限公司颁发数据授权证书，正式授权该公司运营国家健康医疗大

数据平台（福州）。

2.2 平台架构

健康医疗大数据中心的物理基础设施选址于福州滨海新城中国东南大数据产业园。中心占地 60 亩，规划建成投产 15 000 个机柜，用于存放福州市以及福建省内部分县市的健康医疗数据。园区拥有福州国家级互联网骨干直联点及超算中心，为大数据开发应用提供良好的硬件条件。其架构包括统一运维平台、安全运维平台、分布式云数据中心、数据中心操作系统、大数据服务支持平台、服务运营平台，见图 1。

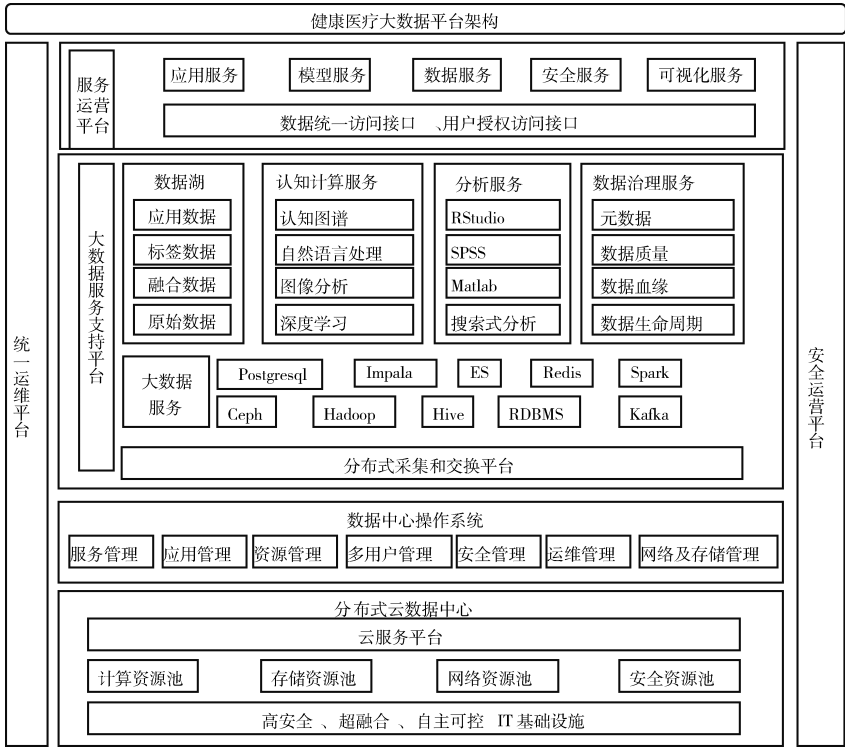


图 1 健康医疗大数据平台架构

2.3 基础设施

作为平台底层的基础设施，中心选用高安全、超融合并且技术自主可控的基础硬件设备，提供大量计算、存储、网络资源，具有可靠、高效的基础支撑作用。在硬件设施基础上形成分布式云数据中心，除支撑健康医疗大数据上层应用需求外，还可

综合对外提供计算、存储、网络、安全等基础设施即服务（Infrastructure as a Service, IaaS）。平台数据汇集采用大规模并行处理，分布式地采集、处理和交换数据，按照对应业务数据模型抽取、转换、加载（Extract – Transform – Load, ETL）到目标数据库中。在此基础上对外提供多种大数据分析架构服务接口，支持 Hadoop、Spark 等主流大数据开发

架构,具有数据湖、认知计算、分析、数据治理等大数据服务。统一对运维平台和安全运营平台进行安全准入、身份甄别、授权管理、安全保障等管理工作,为数据运营保驾护航,切实保障数据安全。

2.4 数据采集

福州市政府出台《福州市健康医疗大数据资源管理暂行办法》,为数据采集提供政策依据。平台采用分布式采集数据模式,在福州市各级卫生医疗机构设置前置机,在前期批量采集数据的基础上,每日增量采集医疗机构中医院信息系统(Hospital Information System, HIS),电子病历系统(Electronic Medical Record, EMR),实验室信息管理系统(Laboratory Information Management System, LIMS),医学影像存储与传输系统(Picture Archiving and Communication System, PACS) 4 个主要业务系统数据,传输至大数据中心的临时数据库中进行 ETL 处理。针对不同医疗机构数据分别进行清洗、标注工作。

2.5 数据安全

该平台通过了信息安全等级保护 3 级测评,执行相关安全规章制度。数据仅对经过授权使用的机构根据授权范围开放。目前只允许东南大数据产业园园区内经授权机构接入平台使用授权数据,数据不出园区,确保安全。

2.6 平台建设成效

经过几年发展,福州市健康医疗大数据平台已完成 37 家市级机构 HIS、EMR、LIS、PACS 4 个主要业务系统数据采集、汇聚和治理以及 34 家医院归一化处理工作,共涉及数据表 67 751 张,结构化存量数量总量 10TB (200 亿条),X 线,电子计算机断层扫描(Computed Tomography, CT),磁共振成像(Magnetic Resonance Imaging, MRI),B 超,心电等非机构化数据已入库 140TB 左右。共制定元数据规范 10 000 条及数据字典 500 个,近 300 个临床数据集,基本完成福州市数据采集及标注工作,为下一步数据深化治理及生态企业数据服务打下坚实基础。

2.7 数据利用和科研

福州健康医疗大数据在医学科研、精准医疗、智能影像、慢病管理等方面的应用都取得一定成果。例如依托福州健康医疗大数据中心,联合肝病、肝癌大数据联盟,将全国主要肝病专科医院、肝病医学中心等的相关数据汇聚到福州,建立全国肝病和肝癌大数据平台,通过辅助诊断和智能分析进行肝癌早筛。在研究健康医疗大数据基础上,结合人工智能技术实现 CT 影像肺结节筛查服务,可快速精确发现早期肺癌可疑影像。部分科研成果处于测试阶段,有望很快进入临床应用。

3 存在的问题及思考

3.1 数据来源有限,存在壁垒

3.1.1 数据局限性 4 年来国家健康医疗大数据中心(福州)对其下属 37 家市级医疗机构以及基层卫生院数据进行采集。目前中心获取数据的来源途径较为单一,主要限于福州市市级医疗机构,尚未涉及省属医院医疗数据。而省属医院尤其是省属三甲医院的医疗数据对于疾病诊疗和科研具有重要价值。

3.1.2 数据壁垒存在原因 (1) 医疗数据所有权问题。患者就诊产生的医疗数据所有权归属是一个复杂问题。现行法律法规对此并无明确规定,可能成为后续大数据开发和应用过程中的法律隐患。鉴于数据所有权不确定性,医疗机构向第 3 方机构提供所保管的患者医疗数据是否会侵犯患者隐私权等合法权利无法得到确认。而健康医疗数据涉及居民个人隐私,其对外发送或开放面临个人敏感信息泄露风险较大。要进一步推动国家个人数据信息属权立法工作,不仅能保障国家安全和公民隐私,界定医疗机构合法权利,还能为大数据发展过程中价值流通和放大创造条件。(2) 医疗机构信息化资源限制。随着医疗信息化发展,机构内部对信息化资源需求日益增加。而医疗机构对信息化的投入远远不及资源需求增速。对医疗机构进行数据采集会额外占用信息化资源。事实上卫生医疗主管部门已建立各类医疗监测、管理平台,例如居民健康档案,

医院质量监测系统 (Hospital Quality Monitoring System, HQMS), 医保体检审核付费系统, 孕产妇信息监管平台, 采集医院内病人诊疗和影像数据、病案首页数据、体检数据、孕产妇孕检分娩信息等, 涵盖大部分院内医疗信息系统。多个对外接口数据传输已经占用闲时信息化资源比例较大。建议各级卫生主管部门理清各相关部门数据概况, 开展数据资源梳理、资产目录编制和登记工作, 以此为基础做好顶层设计, 统筹兼顾, 整合多个数据采集接口和平台, 实现数据统筹, 争取做到“一数一采”, 充分重复利用, 减少信息化资源投入。统筹健康医疗大数据信息系统建设, 防止重复建设造成资源浪费^[2]。同时要合理设计数据汇聚范围和方式, 界定对医疗机构进行数据汇聚的时间跨度、内容和范围, 尽量做到“最小够用”。避免“一锅端”模式, 降低大数据中心与被汇聚医疗机构间数据流动风险。合理设计汇聚方式, 平衡采集和接收端数据处理成本。

3.1.3 非临床健康数据发展滞后^[3] 随着科技的发展, 智能可穿戴设备的涌现、健康管理医疗机构的兴起、基因检测的普及, 很多“互联网+”健康企业、生物高新技术企业拥有一定数量人口生理健康数据, 例如居家自我检测的医学数据, 基因序列、蛋白质组等生物学数据。这些数据对于医疗科研、疾病防控同样具有较大价值。鉴于各类企业采集数据可靠性以及信息孤岛普遍存在, 目前开放整合相关产业采集的健康数据较困难。未来这些数据也应纳入健康医疗大数据体系。推动实现医疗健康数据互联互通, 消除数据壁垒, 联通数据孤岛, 真正推动健康医疗大数据的充分利用。

3.2 缺乏统一技术标准, 基础数据治理成效慢

3.2.1 问题分析 不同医疗机构所使用管理信息系统结构不一, 导致医疗数据标准、格式差异较大。目前采用的方式是直接采集各医疗机构基础数据, 传输至中心后再根据各医院系统数据结构有针对性地进行清洗标注。随着采集来源增多, 依赖行业专家人工清洗海量数据耗时较长, 且对原始数据关系梳理进一步加重工作负担。这不仅提高处理成本, 也增加 ETL 复杂性。

3.2.2 应对措施 加强各医疗机构的信息化建设以及标准化改造。信息化建设要以电子病历为核心, 按照电子病历等级评审要求进行医院管理信息系统建设, 以满足数据汇聚接口要求。加大技术研究, 制定统一标准和技术要求, 各医疗机构按照统一的数据汇聚接口要求提供数据, 这样只需按照医疗机构类别进行处理, 提高治理效率, 降低处理成本。

3.3 数据利用率不高, 应用成果有待丰富

3.3.1 数据利用率较低 主要原因包括以下两方面: 一方面, 数据权属立法空白, 出于严格保护患者个人隐私需要, 对于所提供数据的处理较谨慎; 另一方面, 为保证数据安全, 目前只授权健康医疗大数据产业园内企业和科研机构使用中心数据, 产业园外机构无法接入, 数据利用率不高。

3.3.2 应用成果有限 基于现有数据已经产出部分科研成果和大数据应用, 但数量有限, 很多成果尚处于研究阶段未投入临床验证。运用医疗大数据的渠道和方式较为单一, 如电子病历数据共享、图像检测等应用, 远未充分其应有的价值。2016 年国务院发布《“十三五”深化医药卫生体制改革规划》, 提出健全基于互联网、大数据技术的分级诊疗信息系统; 应用药品流通大数据, 拓展增值服务深度和广度^[4]。这意味着健康医疗大数据应用不仅限于医院诊疗过程中, 还应贯穿整个人类健康中, 包括药物研发、临床诊断与治疗、保险支付和商业保险设计以及健康管理和公共卫生服务等环节。

3.4 建议

加强区块链等新技术应用, 建立可信、透明、可追溯的数据交换业务协同体系。增加数据安全保障机制, 在保证患者隐私和数据安全前提下, 开放接入地域限制, 增加接入科研机构数量, 扩大数据利用范围, 提高汇聚数据开发利用率。促进相关研究, 推动医疗行业大数据应用向更深、更广层面发展, 使科研成果尽快投入临床验证和使用并以此反哺于数据来源机构, 应用于基层医疗服务, 形成良性循环, 进一步加强医疗机构数据汇聚积极性。

(下转第 56 页)

入探针流量数据服务器。(2) 安装威胁预警分析服务器。分析探针流量数据并汇总已标记的应用数据,通过互联网大数据平台特征库进行数据匹配,从而分析流量是否符合恶意攻击进而进行记录和报警。(3) 部署下一代防火墙联动平台。主动控制其安全策略,接收来自安全威胁分析服务器的报警,按照管理人员预先配置的处置流程下发安全策略,实现探针和防火墙的实时联动,进行主动防御。(4) 流量探针。不仅可标记网络攻击流量,还能标记数据库、网页浏览及应用协议流量,经预警分析服务器收集由人工进行二次分析,可为运维人员进行应用层面预警,应及时修改应用代码防止漏洞被利用。随着安全设备发展,部分产品已应用人工智能技术进行防护,利用人工智能的学习能力自动对安全设备下发策略,自动阻断攻击源,配合人工配置,能够高精度识别正常和非法行为,在保证正常业务不中断的情况下阻断网络中的非法行为。

4 结语

互联网技术飞速发展,安全防护有待进一步加

强。使用零信任网络安全架构,使原本封闭环境内的资源得以安全共享并及时发现高危漏洞,提高防控程度,避免原本安全区域内某个节点被攻破后整个区域都陷入风险状态的情况。“互联网+医疗”模式在提高医疗信息化水平的同时,为患者提供方便快捷的医疗服务,但应重视信息安全,需要一套成熟且能够应对未来“互联网+”趋势下逐渐进化的架构。零信任架构将成为未来网络安全流行架构之一,网络安全将进入全新时代。

参考文献

- 1 方山,李华. 计算机网络边界安全防护管理方法研究[J]. 信息通信, 2017 (11): 155-157.
- 2 李放,韩亚坤,李曙光. 针对勒索病毒的医院安全防护体系设计[J]. 中国数字医学, 2019, 14 (11): 85-87.
- 3 廖莹璐,杭菲璐,谢林江. 浅谈企业网络终端安全与APT防御[J]. 自动化应用, 2018 (5): 75-77.
- 4 焦汀. 为什么说“零信任”将成为网络安全流行框架之一[J]. 计算机与网络, 2018, 44 (4): 54-55.
- 5 林辉玉,赵长林. 零信任安全呼唤新思维和新举措[J]. 网络安全和信息化, 2019 (2): 102-103.

(上接第 51 页)

4 结语

健康医疗大数据已纳入国家大数据战略布局,加快健康医疗大数据中心建设能够促进医疗服务发展和医疗资源合理流动,不断满足人民群众多层次、多样化健康需求。加强政策支持,继续汇聚更多医疗机构数据,加快大数据科研和科研成果转化,培育健康经济新动能是健康医疗大数据中心发展方向。

参考文献

- 1 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(国办发〔2016〕47号) [EB/OL]. [2016-06-24]. http://www.gov.cn/zhengce/content/2016-06/24/content_5085091.htm.
- 2 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.
 - 3 许培海,黄匡时. 我国健康医疗大数据的现状、问题及对策[J]. 中国数字医学, 2017, 12 (5): 24-26.
 - 4 国务院办公厅. “十三五”深化医药卫生体制改革规划(国办发〔2016〕78号) [EB/OL]. [2017-01-09]. http://www.gov.cn/zhengce/content/2017-01/09/content_5158053.htm.