

基于深圳市全民健康信息平台的数据采集系统设计^{*}

王 爽 胡 琪 陈晓璇 黄 琮 郑 静

(深圳市卫生健康发展研究和数据管理中心 深圳 518001)

〔摘要〕 目的/意义 设计数据采集系统,以解决深圳市全民健康信息平台采集电子健康档案相关数据、电子病历相关数据和全员人口数据工作量大的问题,提升深圳市全民健康信息化水平。方法/过程 在深圳市各医疗卫生机构前置服务器部署数据采集软件,按照数据采集有关规范要求,通过数据采集、清洗、上传等一体化服务,将相关原始数据转换成标准格式后装载进前置机数据库。结果/结论 利用数据采集系统,建立 3 大基础数据库,使全市居民医疗健康信息得到完整保存。

〔关键词〕 深圳市全民健康信息平台;数据采集系统;数据质量

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2023.09.013

Design of the Data Acquisition System Based on Shenzhen National Health Information Platform

WANG Shuang, HU Qi, CHEN Xiaoxuan, HUANG Qiong, ZHENG Jing

Shenzhen Health Development Research Data Management Center, Shenzhen 518001, China

〔Abstract〕 **Purpose/Significance** The data acquisition system is designed to solve the problem of heavy workload in collecting electronic health record (EHR) data, electronic medical record (EMR) data and whole population data for Shenzhen National Health Information Platform, and to improve the level of public health informatization in Shenzhen. **Method/Process** Through the deployment of data acquisition software in the front server of various medical and health institutions in Shenzhen, the relevant original data is converted into standard format and loaded into the front database through integrated services such as data collection, cleaning and uploading in accordance with the relevant specifications of data collection. **Result/Conclusion** Three basic databases are established by using the data acquisition system, so that the medical and health information of the city residents could be preserved completely.

〔Keywords〕 Shenzhen National Health Information Platform; data acquisition system; data quality

1 引言

深圳市全民健康信息平台以保障公众健康为目标,充分利用网络、存储、计算等技术建立安全可

控和标准统一的平台。该平台要采集深圳市居民全生命周期各业务领域的完整健康信息,以建立全民健康信息数据中心。数据采集内容包含 6 大业务领域数十个专业业务应用系统数据、各区级全民健康信息平台上传数据,以及市属医院、社区健康服务

〔修回日期〕 2023-03-29

〔作者简介〕 王爽,高级工程师,发表论文 6 篇;通信作者:郑静,博士,教授级高级工程师。

〔基金项目〕 四川省区域合作重点项目(项目编号:2022YFQ0097);广东省深圳市基础研究专项(自然科学基金)面上项目(项目编号:JCYJ20210324134405016)

中心、社会办医疗卫生机构数据等。结合数据量大、类型多样复杂的特征,利用系统进行数据采集、存储和质控,在全市医疗机构统一使用,建立全员人口信息数据库、健康档案信息数据库、电子病历数据库 3 大基础数据库。利用系统采集和存储居民诊疗信息,实现全市居民医疗健康信息完整保存,是全市健康大数据产业体系形成的重要步骤^[1-5]。

目前已有数据采集系统均在特定环境下运行,采用传统方式将影响数据采集的准确性、完整性和效率。如何设计统一的数据采集系统完成所有医疗数据采集是目前平台亟待解决的关键问题^[6-8]。

2 系统设计准备

设计目标为:在全市建立统一数据采集系统,整合各医疗机构业务系统数据,建立 3 大基础数据库,为全民健康信息平台中不同应用系统提供统一数据支撑服务。在此目标基础上确定数据采集系统的范围、内容和设计原则^[9-12]。

2.1 数据采集范围

数据采集系统须完成全市医疗卫生计生行业内部数据采集工作,采集内容主要包括医院信息系统业务数据、社区健康服务与管理信息系统数据、深圳市卫生健康委员会直属机构业务系统数据。

2.2 数据采集内容

从全员人口信息数据库、健康档案信息数据库和电子病历数据库 3 个维度分析需要采集的内容。全员人口信息主体包括公民身份证号码、姓名、性别、民族、出生地、出生日期等基本信息,以及各部门业务系统在利用人口卫生计生基本信息过程中产生的其他存在共享需求的全员人口信息等。健康档案包含个人基本信息和卫生服务记录信息。其中,个人基本信息主要包括人口学信息、亲属信息、社会保障信息、基本健康信息、建档信息;卫生服务记录主要包括儿童保健、妇女保健、疾病预

防、疾病管理、医疗服务等信息。电子病历主要内容包包括病历概要、门(急)诊病历记录、住院病历记录、健康体检记录、转诊记录、法定医学证明及报告、医疗机构信息 7 个业务领域的基本医疗服务活动记录。

2.3 系统设计准则^[13-14]

2.3.1 统筹性 在针对深圳市医院、公共卫生机构及卫生管理机构信息数据进行采集、转换、传输、质量控制和监管过程中,应结合医疗卫生机构业务系统情况,逐步推进实施。

2.3.2 统一性 根据数据集、值域表、维护说明、数据源变更报备等一系列文件,规范实施路径,减少信息不对称导致的额外工作量。

2.3.3 安全性 采用科学的服务器备份策略,重要数据发生变更便全量备份至运维终端一次,数据库、应用系统重要数据每天全量备份至运维终端一次。

2.3.4 扩展性 采用 VUE + Spring Boot 等先进开发技术,具备良好的内部集成能力,通过单点登录实现与上级系统的联通性。

2.3.5 稳定性 系统前端采用 VUE,后端使用 Spring Boot、Redis,打包工具使用 webpack gulp lessc,具备良好的系统拓展功能,能够简便地进行二次功能开发。

3 系统设计方案

3.1 总体流程

系统首先通过数据采集程序将各医疗机构业务数据采集到缓存库,并在缓存库中进行质量控制(以下简称质控),如果质控通过则上传至深圳市全民健康平台数据库,如果质控不通过则通知医疗卫生机构进行数据整改,见图 1。



图 1 系统流程

3.2 数据采集程序

3.2.1 总体程序（图 2）

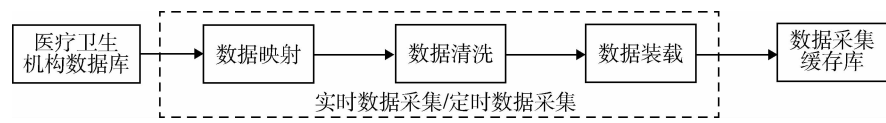


图 2 数据采集程序

3.2.2 采集方式 采集方式分为两种，一是定时采集（T+1 采集），二是实时采集。因为医疗业务的特殊性，各医疗机构业务系统之间需要实时调取数据。实时数据采集可以通过两种方式实现。第 1 种是根据业务需求内容，采用 HL 7 或 Web Service 等设计规范及深圳市全民健康信息平台自身业务需求制定的数据交换标准规范，开发相应组件，依照要求实时检索医疗机构业务系统后台数据库，根据相关检索结果信息，实时反馈推送至全民健康信息平台。第 2 种是全民健康信息平台建立相应服务协议接口及数据交互方式，由各医疗机构业务系统设定事件触发机制，当设定的事件被触发时，可直接将所产生数据推送至全民健康信息平台。采用提取、转换、加载（extract - transform - load, ETL）工具实现定时数据采集。医疗卫生机构业务系统厂商方开放业务系统数据库或数据存储只读权限并提供数据结构、技术支持。由采集方根据业务标准按需整合。按此种方式接入，不同业务系统采用的数据库系统可能不同，即使数据库相对应的数据表和字段也可能不同，因此各医疗卫生机构要开放数据采集相关业务系统后台数据库或所有存储数据只读权限，采集方根据业务需求，针对医疗卫生机构生产数据库进行分析整合，开发相应系统数据采集脚本，采集至中间库数据库，只要采集流程支持，数据清洗转换过程可以保持不变，因此数据清洗转换过程是较通用的模块。

3.2.3 数据映射 各医疗卫生机构有不同的数据库、数据格式、应用和操作平台，确定抽取的字段形成需求表，再与各业务系统数据库字段形成映射关系。数据采集程序根据预先定义的映射规则从医疗卫生机构数据库中抽取相应数据。

3.2.4 数据清洗 根据深圳市全民健康数据采集标准制定数据清洗规则。清洗模块首先过滤采集数据中的无用信息，避免读入大量无用信息影响系统性能。这一步主要是通过设置数据属性和条件完成性别设置、年龄设置等。在清洗完成之后，采集数据通常不会与标准数据模型相对应，仍然需要对数据进行清洗。清洗方式可分为对应清洗和自定义清洗。对应清洗有转码、字符截取、字符合并、常量设置、条件选择、关联表 and 对应转化的方式。根据数据类型不同采用任意组合的方式进行清洗，在清洗完成后存入数据清洗库。例如，“性别” 0 代表女、1 代表男，通过转化方式就可以将输入值中的 0 或 1 对应转化成女或男。自定义清洗是在清洗模块中留有接口，通过手工编写代码解决特殊清洗问题，如男性患者不会出现怀孕就诊记录。数据清洗可将每个步骤清洗结果分别存储至数据清洗库，用以监控、比较清洗前后的数据内容，以校验数据清洗结果准确性。根据不同清洗结果，及时修正清洗规则，达到数据清洗有据，清洗结果内容与原始内容统计分析类结果完全一致。

3.2.5 数据装载 数据装载模块主要是增量装载和整合装载。增量装载是数据的堆积，无须考虑数据的整体性，如医院挂号数据。整合装载是将数据进行整合再存入目标数据库，如患者历史就诊记录。在加载之前要将数据进行整合处理，一方面满足医疗机构内跨业务的专项操作需求，另一方面动态建立医疗机构居民健康档案全局视图。数据整合将采集到的业务数据分门别类组织好，并按设计要求分别存储到区域卫生资源中心，用以支持跨机构、跨级和跨业务的专线业务应用，例如新生儿随访、传染病管理、妇产幼保健、慢性病管理、转诊

管理、远程医疗等。在技术层面,因为不同类型数据有不同存储要求,所以数据装载提供多种存储格式,一是将数据存入关系型数据库中,并为平台方提供相应数据调用路径和权限。二是数据比对和入库,在数据装载时,根据标准表中主键或者业务主键判断是否已有相同记录,决定采用更新装载机制还是插入装载机制。当采用插入装载机制时,直接在目标库中插入需装载数据。当采用更新方式装载时,根据医疗机构业务系统中数据库表主键,结合

创建时间、审核时间、修改时间等字段内容判断数据是否需要更新操作。

3.3 数据质控程序

数据质控程序是对采集缓存库中的数据进行质控,如果通过则上传至深圳市全民健康平台数据库。在数据质控程序中有质量分析、质量评估和质控结果展示环节,见图3。

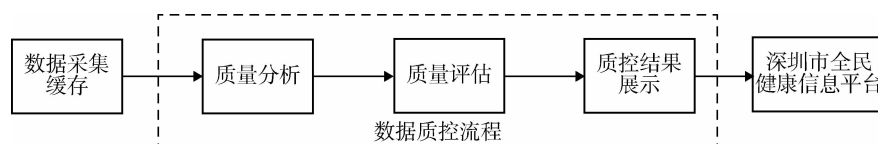


图3 数据质控流程

3.3.1 质量分析 质量分析是对原始数据经过采集程序到采集存储库中的传输情况进行跟踪,根据校验规则对原始数据进行字段级及表级关系验证,同时输出数据校验结果报表,将报表提供给各医疗机构,对不符合规则的数据进行排查。例如,医疗机构代码必须和上传机构一致,如果不一致则反馈给上传机构,让其修改重新上传。

3.3.2 质量评估 质量评估是数据监管系统的重要组成部分,包括数据的完整性、一致性、时效性、规范性4个维度。以业务信息为基础,将所有可监控、可计算的数据指标在质控平台中统一管理,为每项指标作出明确定义,包括指标类型、指标名称、计算公式、计算频次等。数据质量管理根据需要将所有相关指标分为两大类:原子指标、复合指标。指标类别分为监督指标、评估指标和考核指标。这些指标分类有重叠的部分,主要是根据目标不同在指标源中抽取不同的指标进行评估。

3.3.3 质控结果展示 对质控规则进行检查,形成质控结果,对质控结果进行统计展示,为提高数据质量提供数据支撑。数据质控程序实现了端到端的全流程数据监管,可以根据现场数据情况自定义校验规则,以构建完善的数据评分体系进而保证数据上传的有效性和质量。具体质控内容包括完整性、关联性、约束性、一致性、规范性和及时性。为提高数据采集质量,须提供完整的数据质量评估

体系,同时建立完善的数据监控机制,对医疗卫生机构数据采集情况进行综合展示,包括采集数量、采集成功率、质量评估结果等。

4 系统实施效果

采集系统可以根据深圳市全民健康数据采集标准要求,实现数据采集、清洗、上传等一体化服务。截至目前,已经完成全市所有医疗卫生机构对接。系统统一采集具有以下效果。一是从源头上减少人工填报造成的数据缺失、前后不一致等情况。由于之前大部分统计数据手动填报,不同报表中同一指标会出现数据不一致的情况。现将采集的数据共享给各个系统,控制数据质量,从源头上对数据进行规范和统一。二是有利于卫生行政部门作出科学决策和提高对医院的监管力度。通过对医疗数据的整合与挖掘,卫生行政部门可以加强对医院业务的管理以及制订更利于民生的卫生政策,从而提升深圳市整体医疗水平。三是提高居民看病就医幸福感。通过采集全市医疗数据形成居民健康档案和电子病历,医生可以查看就诊者历史就诊记录,避免就诊者做不必要的检查,减轻就医负担,从而提高医疗资源利用效率和降低就医费用。四是为科研工作者提供宝贵的医疗数据。这些医疗数据通过数据采集汇集到平台,为科研工作者对疾病的管理、预

防和干预等研究提供数据支持。

5 结语

数据采集系统采集全市各医疗卫生机构数据,统一汇聚到深圳市全民健康信息平台。平台统一分析和使用这些数据,采用可视化技术将个人健康档案、电子病历清晰展现,为医生或患者提供便利。数据采集系统为深圳市医疗管理、科研、教学提供重要数据源,可大幅度推进深圳市全民健康信息化水平^[15-17]。但是当前数据采集系统在质控方面只做基本质控,数据采集质量较差。后期考虑通过分析当前采集数据质量问题,提出更多质控条件,从而提高数据质量,满足平台需求。

参考文献

- 1 应桂英,郑建智,段占祺.基于业务人员的医院统计信息采集系统设计与实现[J].中国数字医学,2013,8(8):60-64.
- 2 吴乐乐,高白,张燕,等.临床科研一体化数据采集系统的设计与实现[J].中国数字医学,2019,14(2):82-84.
- 3 李万莎,谭富兵,刘晶哲.实验室生物安全智能管理系统设计[J].中国数字医学,2022,17(3):45-49.
- 4 程小恩,魏勇.数据采集系统在中医药健康大数据中的应用[J].中国数字医学,2016,11(9):18-20.
- 5 周莉莉,徐进,孙润康,等.基于省级健康信息平台的数据采集及治理研究[J].中国卫生质量管理,2022,29(3):65-68.

- 6 徐向东,胡建平,张学高.从技术架构演变看医院信息化发展历程[J].中国卫生信息管理杂志,2019,16(4):389-394.
- 7 李岳峰,胡建平,庾兵兵,等.医院信息化功能与技术标准框架设计与探讨[J].中国卫生信息管理杂志,2022,19(3):390-394.
- 8 李华才.加强医院信息化建设提升信息便民服务水平[J].中国数字医学,2015,10(3):1.
- 9 黄开峰,谢晓兰.医药信息数据采集系统研究[J].科技视界,2013,86(35):51-52.
- 10 常健,李萍,滕知轶,等.“智慧医疗”助力一站式住院服务模式实践与思考[J].中国医院,2017,21(12):1-3.
- 11 袁保锋,陈晗梅,李郑,等.CT数据采集系统故障伪影常见原因分析及处理[J].中国医疗设备,2021,36(1):175-178.
- 12 李炳南.磁感应断层成像数据采集系统的设计与研究[D].天津:天津工业大学,2019.
- 13 耿辉,贺海蓉,曾宪涛,等.多中心临床数据采集系统 REDCap 系统应用及架设[J].中国循证心血管医学杂志,2017,9(9):1025-1028.
- 14 卢昊,彭虎.应用于超声成像的数据采集系统设计[J].北京生物医学工程,2011,30(4):396-400.
- 15 苏静.CT数据采集系统[J].医学信息(上旬刊),2010,23(8):2777-2778.
- 16 汤娟华,朱琴芬,赵光伟,等.区域全民健康信息平台居民就医数据采集和管理策略[J].医学信息学杂志,2019,40(11):19-22.
- 17 佚名.《关于加快推进人口健康信息化建设的指导意见》答问[J].首都公共卫生,2014,8(1):42-46.

《医学信息学杂志》开通微信公众号

《医学信息学杂志》微信公众号现已开通,作者可通过该平台查阅稿件状态;读者可阅览当期最新内容、过刊等;同时提供国内外最新医学信息研究动态、发展前沿等,搭建编者、作者、读者之间沟通、交流的平台。可在微信添加中找到公众号,输入“医学信息学杂志”进行确认,也可扫描右侧二维码添加,敬请关注!



《医学信息学杂志》编辑部