

基于双闭环模式的医院数据质量提升方法 研究与实践^{*}

何仲廉 刘少堃 冯晨阳

刘迷迷 周毅

李超峰

(中山大学附属肿瘤医院信息中心
广州 510080)

(中山大学中山医学院
广州 510080)

(中山大学附属肿瘤医院信息中心
广州 510080)

[摘要] 介绍医院数据治理研究现状, 分析医院数据质量的实践要求, 阐述业务闭环与数据闭环双闭环模式下的数据治理方法及应用情况, 指出该模式具有实时、源数据问题快速定位等特点, 有效帮助医院提高数据质量, 同时提供一种长期的可持续的数据质量监控方案。

[关键词] 双闭环; 数据治理; 实时数据质量监控

[中图分类号] R-058 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2022.06.005

Study and Practice of the Improvement Method of Hospital Data Quality Based on Double Closed Loop Mode HE Zhonglian, LIU Shaokun, FENG Chenyang, Information Center, Sun Yat-sen University Cancer Center, Guangzhou 510080, China; LIU Mimi, ZHOU Yi, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China; LI Chaofeng, Information Center, Sun Yat-sen University Cancer Center, Guangzhou 510080, China

[Abstract] The paper introduces the research status of hospital data governance, analyzes the practical requirements of hospital data quality, expounds the data governance method and its application under double closed loop mode of business closed loop and data closed loop, and points out that this mode has the characteristics of real-time, quick location of source data problems, etc., which can effectively help hospitals improve data quality and provide a long-term and sustainable data quality monitoring scheme.

[Keywords] double closed loop; data governance; real-time data quality monitoring

[修回日期] 2022-06-07

[作者简介] 何仲廉, 副主任, 发表论文 10 余篇; 通信作者: 李超峰, 主任, 发表论文 30 余篇。

[基金项目] 中山大学肿瘤防治中心“青年优创”计划项目“医院信息系统数据质量监测管理体系与实践研究”。

1 引言

1.1 数据治理相关研究和实践

随着公立医院转向高质量发展并进入了以智慧医院为导向的建设新阶段, 医院日益注重健康医疗数据及其质量, 通过数据治理开展质量提升工作并逐渐系统化与常态化。国际数据管理协会 (Data Management Association, DAMA) 制定的《DAMA 数据管理的知识体系指南》较为全面地阐述了数据管理相关知识与技术体系, 是各行业领域数据治理

的权威理论。数据治理体系在金融、通信等行业应用都已比较成熟^[1]。虽然“数据治理”理念已经涉及医疗领域，国内外目前仅对其政策方面展开大量研究，应用方面关于数据质量的研究较少^[2]。目前医院开展的数据治理工作一般基于大数据平台或数据仓库建设，使用面向大数据的治理模型^[3]，如中山大学附属第一医院利用“数据中台”技术，以主题数据为线索进行数据治理，取得一定成效^[4]；广州市妇女儿童医疗中心基于专病数据库建设，运用数据治理循环计划-执行-检查-改进（Plan-Do-Check-Action, PDCA）框架进行多层次数据治理，形成专病数据资产和科研数据平台^[5]。

1.2 存在的问题

上述基于数据仓库集成发现数据问题的做法存在以下问题：数据质量问题的发现相对滞后，缺乏实时性；数据集成后需要重新与业务数据表映射才能对生产系统中的数据问题进行定位，增加了整改阶段的难度与成本。为解决以上问题，本文重点关注在数据产生过程中发现数据质量问题，并从数据业务规范角度出发进行实时数据监控。从电子病历高级别医院评级实践经验出发，提出业务闭环与数据闭环双闭环模式下的数据治理方法，在传统数据获取、分析、监测、处理、验证的闭环数据治理监控流程上，融合医院业务系统的闭环流程进行闭环数据监控。

2 医院数据质量实践要求

2.1 电子病历评审的应用与数据

国家卫生健康委员会将电子病历评级与医院等级评审挂钩，针对电子病历系统应用与数据质量分别提出明确要求。电子病历评级评审对各应用功能点要求较细致，数据质量方面要求最近3个月的数据达标。数据质量提升是长期工程，不少医疗机构在建设信息系统时更看重应用及功能，对数据重视程度不够，无法形成常态化的数据治理^[6]。从高级别电子病历评审实践经验来看，应用性要求与数据质量要求在本质上是从评级整体要求的不同角度考

量。以单体医院最高级别6级为例，其中应用整体要求各业务应用在系统中实现闭环管理，而在数据质量要求层面，除了1~5级的一致性、完整性、整合性要求外，增加了对各业务闭环中数据及时性的考察。有机结合评审过程中的应用性建设与数据质量治理，应用性建设以数据质量为验证，数据质量治理落实在各应用系统功能点上，两者互为促进，这是本文研究的出发点。在新模式下应全面提升医院电子病历系统应用的深度与质量，从而为医院医疗质量与安全提供信息化支撑。

2.2 业务闭环

2.2.1 概况 在信息治理大策略框架下，数据治理和流程治理是主线，相辅相成^[7]。从医疗核心制度出发，大多数医院的业务闭环具有共通性，如药物闭环、检查闭环、手麻闭环等。以中山大学附属肿瘤医院为例，评审过程中共识别出15个业务闭环，包含静脉药物配置中心（Pharmacy Intravenous Admixture Services, PIVAS）闭环、口服药/静脉注射/肌肉注射闭环、科室配液闭环、检验闭环、检查闭环、输血闭环、手麻闭环、会诊闭环、放疗闭环、检验危急值闭环、检查危急值闭环、日间化疗闭环、门诊口服药闭环、病理闭环等。实际上业务节点的有无和多少需要通过现场观摩与系统调研方法对医院实际业务进行确认，对分解出的每个业务点进行系统与数据源的映射确定。

2.2.2 应用实例 以手麻闭环业务为例，流程中每个业务点的数据来源均可对应数据质量要求中涉及的该节点所有字段，也可以结合实际需要添加其他需要进行数据治理的数据字段；要找准业务的唯一索引，这是业务闭环中串联各系统与各业务点的关键数据。值得注意的是，业务索引需要具体到每一次执行的业务并具有唯一性，如长期医嘱存在一次开立多次执行的情况，需要使用“医嘱号”+“医嘱执行号”，以单独的“医嘱号”作为业务唯一索引，见表1。通过以上方法与要点对全院关键医疗业务流程进行闭环整理，可以梳理出院内业务与系统的闭环映射，往往也能发现系统接口质量问题，为进一步进行数据治理做好准备。

表 1 手术麻醉闭环梳理示例

序号	业务节点	系统来源	数据来源 - 表	数据来源 - 字段	索引字段
01	手术申请	医生工作站	MET_ OPS_ APPLY	APPLY_ DOC, OPER_ DATE	OPERATIONNO
02	手术安排	HIS	MET_ OPS_ ARRANGE	OPER_ CODE, OPER_ DATE	OPERATIONNO
03	患者交接 - 离开病区	PDA 交接系统	PDA_ SHIFT_ DATA	略	RECORDID
04	患者交接 - 进入手术室	PDA 交接系统	PDA_ SHIFT_ DATA	略	RECORDID
05	进入术间	手麻系统	MED_ OPERATION_ MASTER	WASH_ NURSE, IN_ DATE	RESERVED5
06	手术安全核查 (一)	手麻系统	MED_ CUSTOM_ DATA	略	ID
07	麻醉开始	手麻系统	MED_ OPERATION_ MASTER	ANES_ DOC, ANESSTARTDATE	RESERVED5
08	手术安全核查 (二)	手麻系统	MED_ CUSTOM_ DATA	略	ID
09	手术开始	手麻系统	MED_ OPERATION_ MASTER	OPS_ DOC, OPERATIONDATE	RESERVED5
10	手术结束	手麻系统	MED_ OPERATION_ MASTER	OPS_ DOC, OPS_ ENDDATE	RESERVED5
11	麻醉结束	手麻系统	MED_ OPERATION_ MASTER	ANES_ DOC, ANESENDDATE	RESERVED5
12	手术安全核查 (三)	手麻系统	MED_ CUSTOM_ DATA	略	ID
13	离开手术间	手麻系统	MED_ OPERATION_ MASTER	WASH_ NURSE, OUT_ DATE	RESERVED5
14	进入复苏室	手麻系统	MED_ OPERATION_ MASTER	FS_ NURSE, IN_ PACU_ DATE	RESERVED5
15	离开复苏室	手麻系统	MED_ OPERATION_ MASTER	FS_ NURSE, OUT_ PACT_ DATE	RESERVED5
16	患者交接 - 回到病区	PDA 交接系统	PDA_ SHIFT_ DATA	略	RECORDID

注：篇幅所限，仅列出操作人与操作时间字段。

2.3 数据治理闭环

数据治理是医疗业务过程中的一项长期任务，是采集、运用、发现、评价、更新等环节不断运转的闭环过程。数据治理方法一般遵照 PDCA 闭环理念开展^[8]。一般数据治理闭环步骤如下：一是制定数据质量监控检核方案，对数据质量进行监控检核，制定数据治理目标。二是制定数据质量规则库，对数据质量需求进行收集，通过发现数据使用质量问题、数据过程质量问题和数据总体质量问题并进行整理，对质量总体要求进行整合，确认数据质量总体目标。三是按照数据质量规则库执行数据质量管控，得到数据质量问题。四是对数据质量问题进行管理，包括数据质量循环管理、问题反馈、原因分析、问题修正、问题整改。五是数据质量评估，返回第 1 步。包括从核心指标、模式和管理流程进行评估。

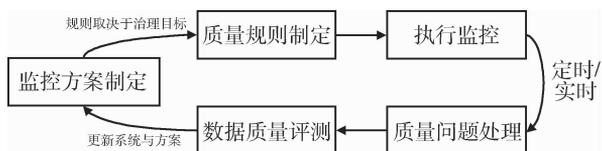


图 1 一般数据治理闭环示意

3 双闭环数据质量监测模式实现

3.1 概述

从评审实践出发，结合业务闭环与数据闭环的双闭环模式进行数据治理。在应用建设业务闭环的基础上同时关注业务闭环中每个业务点在后台对应的数据质量问题，使用统一数据治理闭环实时获取关键业务数据，并按预定义数据规则进行闭环数据监控，持续提升应用与数据的质量。

3.2 架构

3.2.1 关键步骤 以手麻闭环为例，关键步骤如下：一是定义业务闭环。将闭环业务点对应的系统、表、字段进行梳理与映射。二是定义数据规则。将文字描述的数据规则与数据质量要求转化成可执行的结构化查询语言（Structured Query Language, SQL）。三是数据监控过程。使用监控引擎与数据规则库对闭环进行实时监控，记录相关数据问题。四是问题分析整改。分为主动整改与通知整改，管理员与业务负责人可随时主动查询数据质量趋势与数据质量问题报告，系统负责

人会定时收到数据质量问题通知。五是系统完成整改更新后，进入下一个闭环周期，至此完成数据治理闭环。通过以上步骤在业务闭环上实施数据治理闭环，采取多重后台数据库连接业务闭环的各系统数据库的架构，具有异构、实时的特点，见图 2。

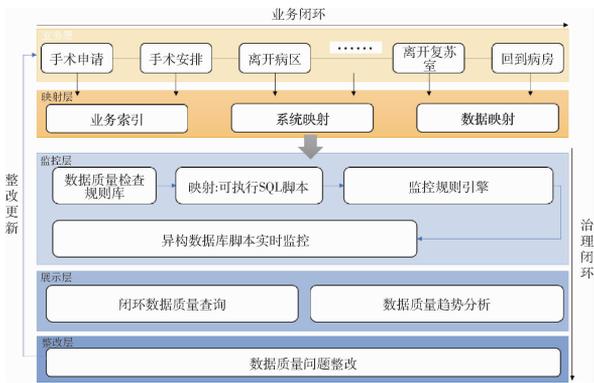


图 2 基于双闭环模式的数据治理架构

3.2.2 同义属性匹配模块技术应用 由于医院

信息架构中存在异构数据库，实践中该架构业务的逻辑层（闭环）到物理层（数据库）的映射确立是实现难点。本研究采取一种同义属性匹配模块技术，该模块从各异构数据库的字典、模式、内容 3 个级别提取属性、特征值，形成规范化特征向量，利用优化的 sigmoid 函数计算属性列的相似性系数。利用该技术在数据规则层上实现规则时只需要考虑逻辑层，而屏蔽异构的物理层。

3.3 数据质量监控规则库

3.3.1 概述 在对业务点实施数据闭环监控之前，需要根据数据治理目标建立数据质量监控规则库，这是实现数据自动监控的核心。根据数据质量标准（如数据质量评审要求中关于完整性、一致性、整合性和及时性的规则）映射成可执行的 SQL 脚本规则，供监控引擎读取并进行数据问题发现，同时将检查产生结果存放到监控结果表中，见表 2。

表 2 数据质量监控规则库示例

ID	业务点	分类	监控规则	规则描述	源表	源字段	来源	级别
1	检验申请	完整性	SELECT 1 FROM P_ PATIENTS WHERE BC_ DATE IS NULL	检验申请日期不为空	P_ PATIENTS	BC_ DATE	检验信	一般
2	检查申请	一致性	SELECT 1 FROM EMAR_ PACS _ REQUEST WHERE EXIST (SELECT 1 FROM EMAR_ DICT WHERE APPLYDOC = EMAR_ DICT. NAME)	申请医生来源于员工表	EMAR_ PACS _ REQUEST	APPLYDOC	影像存	高
3	检查申请	整合性	SELECT 1 FROM EMAR_ PACS _ REQUEST WHERE EXIST (SELECT 1 FROM EMAR_ DICT WHERE APPLYDOC = EMAR_ DICT. NAME)	申请系统与报告系统的项目名称一致	EMAR_ PACS _ REQUEST	ITEM_ NAME	超声	高
4	检查申请	及时性	SELECT 1 FROM EMAR_ PACS _ REQUEST WHERE EXIST (SELECT 1 FROM EMAR_ DICT WHERE APPLYDOC = EMAR_ DICT. NAME)	报告时间大于检查时间	EMAR_ PACS _ REPORT	REPORT_ DATE	心电	高

注：仅显示核心字段描述。

3.3.2 规则表达准则 规则表达以能识别和校验数据为准，没有固定准则。除评审相关数据质量要求外，可扩展加入各业务点需要监控的其他数据字段，如麻醉记录单上的出入量等业务上需要重点关注的字段等。

3.3.3 数据规则转化 实践中数据规则由业务描

述转化为程序可执行规则是实现的难点。一般较为简单的做法为使用 PL/SQL 语法表达实现，如完整性使用 is not null 语法即可判断，一致性使用 exists 等语法即可完成。而涉及多表多字段的复杂规则不能直接采用 SQL 执行规则校验，而是采用缓存技术将规则在程序内存中通过实体拼装，完成验证后释

放。另外多表多字段的复杂规则中连接字段的选择对效率影响尤为重要。

3.4 监控引擎

监控引擎是数据质量监控平台的发动机，负责执行监控脚本并产生监控结果，由可供调度程序按需执行存储，需要部署在一个能够读取其他业务库的数据库用户下。数据质量监控引擎执行流程如下：一是通过调度程序检测新的业务数据，实时触发监控引擎执行。二是监控引擎先根据数据监测字段表判断业务数据是否在监控范围内，再顺序读取规则库中的数据质量规则，并判断规则是否有效。满足条件后执行检查规则并将检查结果输出到结果表中。三是监控规则执行过程可能会失败，因此需要日志表记录执行情况，针对执行失败的规则发送日志给系统管理员，以便及时修复问题。四是一条规则执行完成后，如无失败则继续读取数据质量规则库下一条相关规则进行校验，直至完全读取。五是执行完最后一条规则结束监控引擎的一次运行，同时将数据问题记录以报告形式发送给相关业务人员与工程师。由于一般情况下数据问题在整改开始时会比较，可合并同类问题形成汇总报告发送，见图 3。

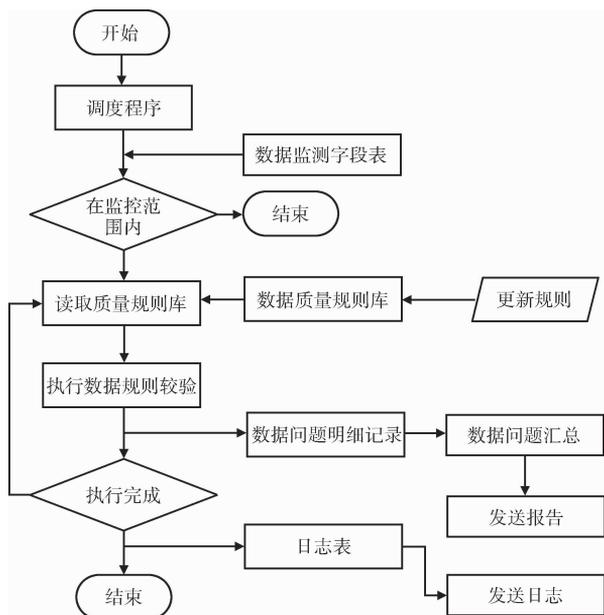


图 3 数据质量监控引擎执行流程

4 持续数据治理方案

4.1 应用成效

基于以上双闭环模式的应用实践，目前中山大学附属肿瘤医院已实施 15 个核心业务闭环的数据治理，共设 124 个闭环监控点，涉及 14 个系统，数据库类型涵盖 Oracle、Mysql、Sqlserver；在数据质量规则方面，根据评审要求与医院实际情况，共建立数据规则 926 条，其中完整性 325 条，一致性 104 条，整合性 363 条，及时性 134 条；平台运行 3 个月有质量问题的治理数据共 782 条，在持续的系统整改更新下，业务闭环的完全闭环率（实现闭环业务数/总业务数）从实施前的 60% 提升至 98% 以上。应用双闭环的数据治理体系后医院数据质量获得显著提升，具体表现为以下几方面：一是建立可视化业务闭环质量监控平台，提高全院临床人员、管理人员对闭环数据的感知度，增强其对业务数据的认识与理解。二是开发基于闭环的电子病历高级别评审数据质量评估监测平台，在一段时间的持续应用下医院电子病历完全满足 6 级数据质量要求。三是与医院基于大数据平台的数据集中分析治理方法相互补充，在更加关注临床科研与单病种数据的治理基础上，补充了业务运行流程的数据质量治理方法。四是形成医院双闭环特色的数据质量监测技术体系，为医院后续的绩效考核评价、三甲复审等提供数据治理框架参考，为医院未来发展提供有价值的数

4.2 经验与改进途径

4.2.1 经验 通过以上实践与应用，总结数据治理经验如下：一是业务闭环是双闭环数据治理的基础，需对医院实际业务与信息基础进行梳理，闭环监控的关键除了业务节点与数据源确定外，还要重视业务索引的明确与统一。二是数据治理闭环应用、发现、定位的数据问题，往往是某个系统或程序上的设计缺陷或错误导致，而对系统问题的定位与修复则比较依赖系统工程师的经验与能力，也是整个闭环中提升问题整改效率中需要特别关注的部

分。三是数据治理着力点并非在于数据本身,而是业务流程、决策以及多方组织机构的相互协作^[9]。在以技术驱动的基础上需要全院各部门人员的共同参与,特别是数据使用部门与相关业务部门的全程参与,提高全院的数据治理意识与水平。

4.2.2 改进途径 下一步可就性能影响做进一步的研究与改进,减少方案对业务可能带来的影响。如通过实时数据库备份技术,包括 Oracle 的 Data Guard (甲骨文推出的一种高可用性数据库方案)技术、Sql Server 的数据库订阅技术等,或者通过网络旁读分流等技术方法,在减少生产数据库压力的同时获得准确、无延时的生产数据,持续进行实时数据治理。另外根据医院实际需求扩展数据规则库,除电子病历评审相关数据质量规则外,可加入更多的有用数据规则,扩大数据质量监控覆盖面。

5 结语

本文在电子病历评审过程中总结应用性建设与数据质量治理的实践经验,提出以业务闭环为基础的双闭环数据治理模式,将电子病历评级中高级别的业务闭环与数据质量要求结合,治理过程覆盖医院全流程的关键业务节点,更加实时与全面地提升医院整体业务与医疗质量,同时也改善了患者在院

就医体验,为医院长期、可持续的数据治理提供一种可操作的技术方案,推动医院数据不断迭代改善的正向反馈,促进医疗机构业务和数据质量同时提升。

参考文献

- 1 苏洁. 金融数据治理趋严 [N]. 中国银行保险报, 2022-03-07 (7).
- 2 叶琳, 罗铁清. 医疗数据治理综述 [J]. 信息时代, 2021 (5): 10-12.
- 3 俞鹏飞, 罗颖文, 刘建模, 等. 面向医院的大数据治理模型设计 [J]. 医学信息, 2021, 34 (10): 18-20.
- 4 龙思哲. 基于数据中台的医院信息系统数据治理方案的探讨 [J]. 当代医学, 2021, 27 (29): 193-194.
- 5 曹晓均, 韦晓燕, 毛铃镶. 医院专病数据治理实践 [J]. 中国数字医学, 2021, 16 (11): 17-19.
- 6 王韶锋, 赵善斌, 杨静. 医院数据治理与数据质量提升研究 [J]. 现代医院, 2021, 21 (11): 1761-1763.
- 7 费晓璐, 李嘉, 黄跃, 等. 医疗大数据应用中的数据治理实践 [J]. 中国卫生信息管理杂志, 2018, 15 (5): 554-558.
- 8 张学高, 胡建平. 医院数据治理框架、技术与实现 [M]. 北京: 人民卫生出版社, 2019.
- 9 姬卫东, 李琳, 张振, 等. 互联互通背景下医疗数据治理面临的问题与对策 [J]. 中国数字医学, 2021, 16 (11): 6-10.

《医学信息学杂志》版权声明

(1) 作者所投稿件无“抄袭”“剽窃”“一稿两投或多投”等学术不端行为,对于署名无异议,不涉及保密与知识产权的侵权等问题,文责自负。对于因上述问题引起的一切法律纠纷,完全由全体署名作者负责,无需编辑部承担连带责任。(2) 来稿刊用后,该稿包括印刷出版和电子出版在内的出版权、复制权、发行权、汇编权、翻译权及信息网络传播权已经转让给《医学信息学杂志》编辑部。除以纸载体形式出版外,本刊有权以光盘、网络期刊等其他方式刊登文稿,本刊已加入万方数据“数字化期刊群”、重庆维普“中文科技期刊数据库”、清华同方“中国期刊全文数据库”、中邮阅读网。(3) 作者著作权使用费与本刊稿酬一次性给付,不再另行发放。作者如不同意文章入编,投稿时敬请说明。

《医学信息学杂志》编辑部