

基于三角互证的医院数据治理水平指标体系设计^{*}

孙海珍¹ 时生辉² 郑秋莹¹

(¹ 北京中医药大学管理学院 北京 100029 ² 北京中医药大学科技园管理办公室 北京 100102)

〔摘要〕 目的/意义 构建医院数据治理水平指标体系，助力医院提高数据价值。方法/过程 基于现有数据治理及医院数据治理相关研究，运用专家咨询法、三角互证法和技术－组织－环境分析框架，通过文献分析、专家咨询、三角互证、评价指标计算 4 个环节，探讨指标体系构建路径。结果/结论 构建含 3 个一级指标、11 个二级指标和 32 个三级指标的医院数据治理水平指标体系，与现有智慧医院评价体系互补，反映医院数据治理建设基础与发展目标，为数字化医院建设提供参考。

〔关键词〕 数据治理；技术－组织－环境框架；三角互证；词频－逆文档频率

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2025.03.009

Design of the Hospital Data Governance Level Index System Based on Triangle Mutual Certificate

SUN Haizhen¹, SHI Shenghui², ZHENG Qiuying¹

¹School of Management, Beijing University of Chinese Medicine, Beijing 100029, China; ²Management Office of Beijing University of Chinese Medicine, Beijing 100102, China

〔Abstract〕 **Purpose/Significance** To build the hospital data governance level index system, and to help hospitals enhance the value of data. **Method/Process** Based on the existing research on data governance and hospital data governance, the study uses the expert consultation method, triangle mutual certificate method and technology－organization－environment analysis framework to discuss the construction path of the index system through 4 links; literature analysis, expert consultation, triangle mutual certificate and evaluation index calculation. **Result/Conclusion** The hospital data governance level index system with 3 first－level indicators, 11 second－level indicators and 32 third－level indicators is constructed, which is complementary to the existing smart hospital evaluation system, reflect the foundation and development goals of the current hospital data governance construction, and provide references for the digital hospital construction.

〔Keywords〕 data governance; technology－organization－environment framework; triangle mutual certificate; TF－IDF

〔修回日期〕 2024-12-10

〔作者简介〕 孙海珍，硕士研究生；通信作者：郑秋莹，博士，教授。

〔基金项目〕 北京市科学技术委员会产业研究院建设项目（项目编号：20230481128）。

1 引言

在健康中国与数字中国两大战略引领下，为促进优质医疗资源扩容和区域均衡布局，应将数据要素全面融入医院高质量发展，实现降本、提质、增效，医院数据治理工作因此变得更为重要。国内学

界围绕公共领域数据治理的理论框架与技术路径展开了系统性探索,重点关注治理范式创新及制度体系构建。然而,相较于其他领域,医疗卫生数据治理研究尚存显著局限:其一,该领域研究起步相对滞后,现有成果多聚焦技术实现与平台架构等微观层面,主要探讨医疗机构通过数据治理实现精细化管理的实践路径;其二,既有研究普遍缺乏跨区域、多维度的大规模实证分析,针对全国层面的宏观评估体系仍属空白。为保证指标体系的科学性,明确后续指标选用的筛选方式、理论基础,邀请 3 位领域内专家,在正式开展指标体系构建前进行预咨询,基于专家建议选择三角互证法进行指标趋同性分析,通过文本内容分析法引入词频-逆文档频率(term frequency-inverse document frequency, TF-IDF)值筛选医院数据治理水平评价指标,打造医院数据治理水平指标体系,赋能医院高质量运用数据要素。

2 数据治理研究概述

国内数据治理相关研究主要从技术与制度两个层面出发^[1-3],既关注政府及行业领域的治理框架构建,包括治理方法设计、实施策略制定和运行路径规划,又深入探讨制度层面的核心要素,如数据权利主体界定、权责分配机制、主体控制权实现、个人信息保护体系、数据跨境流动规则等^[4]。国外相关研究集中在特定领域治理框架构建、关键要素和原则、法律问题与监管等^[5-7],涉及技术、伦理、法律和政策等多方面^[8-9],强调治理框架应包含数据安全、数据质量、隐私保护等关键要素^[10-12]。

目前数据治理概念尚未达成共识。数据治理研究所(Data Governance Institute, DGI)和数据管理协会(Data Management Association, DAMA)分别从数据相关决策活动及数据资产活动两个角度定义。相关学者多认为数据治理有双层含义,一指“数据与数据技术的治理”,即通过数据与数据技术实现治理体系与治理能力现代化^[13];二指“数据与数据化治理”,即对数据与数据化社会关系进行治理^[14-15]。医院数据治理应遵循《国家健康医疗

大数据标准、安全和服务管理办法(试行)》《医疗卫生机构网络安全管理办法》《医疗数据治理规范》《网络安全法》《数据安全法》《个人信息保护法》等相关规定,确保数据准确性和安全性。结合既往研究成果,本研究将医院数据治理定义为:基于数据治理基础框架与标准规范,针对医疗服务及管理场景中多源异构数据的系统性管理过程,其核心任务在于通过数据收集、管理、应用、共享、开放、监督与安全等环节的管控与处理,最终实现数据价值最大化。

3 医院数据治理水平评价指标体系构建研究设计

3.1 TOE 理论框架

技术-组织-环境(technology-organization-environment, TOE)框架多用来解释不同领域和情境下的复杂现象和影响因素^[16-17]。技术包括机构现有技术及未引进技术,组织涵盖机构范围、规模、管理结构特性和人力资源,环境指机构所在行业利益相关者影响机构活动的行为。既有研究采用生命周期理论、价值链理论、公共价值理论、协调创新理论以及其他经济学相关理论解释数据治理。多中心治理理论常应用于思考在线平台所产生数据的治理模式,强调多主体共治对平台的重要性;生命周期理论关注数据治理及流通;价值链理论关注数据资源及公共价值实现;公共价值理论针对公共管理者;协调创新理论强调参与主体;经济学理论关注成本效益。这些理论不适用于医院数据治理水平评价。

医院数据治理要考虑数据本身及数据化治理。医院数据多来源于临床诊疗和患者管理,应确保其隐私安全。医院数据治理涉及技术、管理、政策及患者权益等多方面。因此,以 TOE 为理论基础,将技术、组织、环境作为医院数据治理水平评价指标体系的一级指标。技术维度关注医院数据收集、处理、存储、分析能力;组织维度考察医院数据治理组织架构、人员配置、管理流程等;环境维度包括政策、市场、社会环境等。

3.2 三角互证法与 TF-IDF 值

由于不同学者对数据治理的认知存在差别，为提出全面的医院数据治理水平评价指标体系，使用三角互证法进行对比，见图 1。三角互证法是保证质性研究质量的常用方法，结合 3 种及以上不同资料来源数据与理论视角，通过验证提高研究信度与效度。其原理为从不同角度观察、解释和验证同一事物或现象，有助于更好地理解。主要分为数据互证、研究者互证、理论互证与方法互证，本研究采用数据互证。词频-逆文档频率（term frequency-inverse document frequency，TF-IDF）是广泛认可的文本挖掘技术，在信息检索和文本挖掘领域，通过计算词频（TF）和逆文档频率（IDF）的乘积，评估词对文件集或语料库中文件的重要程度。词语在越多文档中出现，IDF 值越低；相反，词语在越少文档中出现，IDF 值越高。本研究先利用三角互证法抽样，根据研究层次与视角抽取 3 组资料：数据治理层面、医院数据治理层面、数据治理政策层面，对比构建医院数据治理水平评价指标体系。各指标在同一层面资料的趋同性通过 IDF 值表现，IDF 值越低的指标趋同性越强，对数据治理工作越重要。

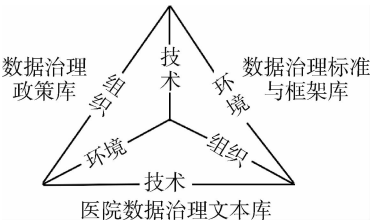


图 1 医院数据治理指标体系三角互证设计

4 指标体系构建过程及结果

4.1 基于三角互证法的文献语料集整理

4.1.1 数据治理标准与框架库 梳理数据治理标准与框架文献，构建数据治理水平评价指标体系应遵循 ISO/IEC 38500、ISO/IEC TS 38501、ISO/IEC RT 38502、ISO/IEC 38503、《信息技术大数据术语》《数据治理白皮书》《国家数据标准体系建设指

南》等标准。结合既有研究，保留高频内容^[18-21]，剔除 3 个语料库中 TF、IDF 值均偏低的“数据价值”“设计”两项内容，保留“资金”“组织内部构建”“相关人员的思想共识”“人力资源”“数据标准”“数据质量”“数据安全”“统筹规划”“数据更新”等内容，并将监控优化、数据利益相关者、数据主动/被动和持续的治理流程等内容分解于数据采集完整性、数据更新时效性、数据处理、数据深度加工、医疗网络安全防护设备、数字身份管理、隐私保护、机构设置、人才培养等二级、三级指标中。在数据治理内容方面，目前使用最广泛的是 DGI 和 DAMA 的数据治理框架。DAMA 数据管理框架、《数据管理能力成熟度评估模型》等框架虽涉及数据治理，但系统化不足，偏重数据管理。DGI 数据治理体系是唯一系统化诠释数据治理的框架，因此选择参考 DGI 数据治理体系构建医院数据治理水平指标体系。

4.1.2 医院数据治理文本库 在中国知网以“篇名=数据治理”为检索式，时间区间选定 2020—2024 年，文献类型为期刊，将期刊来源类别限定为“SCI”“CSSCI”“北大核心”，得到 440 条结果，其中相关文献 42 条。在 Web of Science 中，针对 2020—2024 年期刊，以标题中含有“Data Governance Framework”为检索条件得到 20 条结果，以标题中含有“Data Governance”为检索条件得到 375 条结果。梳理相关文献，多将数据治理分为管理、数据、基础支撑、应用 4 个方面^[22-29]，涉及 10 项内容。保留“专业人才”“基础设施”“治理规划与制度”“数据质量控制”“数据管理与应用”“数据治理保障体系”“数据处理”7 项内容，将“基于大数据的医疗服务能力”“外部响应能力”“大数据创新应用”3 项内容归于“数据管理与应用”。数据方面多为治理流程，包括数据标准化、数据资源存储、数据交互、跨院资源协同、数据质量控制、数据安全等内容。管理方面主要为数据的管理和使用。基础支撑方面包括但不限于政策、技术、资金、平台、专业人才等。应用方面多表现在业务信息系统中，即临床服务类系统、护理服务类系统、医技管理类系统、医疗与药品管理类系统、运

营管理类系统等。基于对医院数据治理文本库的分析发现，医院数据治理指标体系建设应包含诊疗前、中、后流程或管理、医疗、服务 3 部分内容，谢刚等^[30]提出的多科性医院大数据治理能力评价指标体系被多位学者引用。

4.1.3 数据治理政策文本库 在中国知网“政府文件”“法律法规”模块，发布机关选择“中央单位”，时间限定为 2020—2024 年，针对全文检索得到 96 条结果。数据治理应基于“安全是前提，标准是保障，服务是目的”的原则开展活动。政策中数据治理工作主要任务为数据资产管理、数据融合共享、数据互认、数据质量监控、数据应用、数据安全管控，从政策文本库抽出相关指标，得到“基础设施”“技术”“检查检验结果数据共享互认”“数据安全保障”“标准规范”“数据应用”“数据采集”“数据存储”“数据处理”“数据质量监控”“人才队伍建设”“数据完整性”“数据挖掘分析”“互联互通”等指标。

4.2 指标体系构建结果

基于三角互证法中的 3 组资料，剔除 TF、IDF

值低的指标后，对剩余指标进行分类归纳，得到医院数据治理水平评价指标体系，见表 1。技术层细化为数据质量控制能力、数据共享与流通能力、数据应用能力、数据安全保障技术能力 4 个二级指标。涵盖医院现有及未引进的技术，具体指医院在各类活动中对数据进行收集、处理、加工、共享、流通与应用的技术能力，主要用于评估医院在数据全生命周期中运用的技术。组织层细化为机构设置、基础硬件设施、人才队伍 3 个二级指标。主要指医院的组织特性，涵盖医院规模、管理结构、人力资源状况和信息技术基础设施，考查医院的数据治理组织架构、人员配置、管理流程等。环境层包括内部与外部环境，需考虑医院所处的经济、文化、政策等因素影响，其中相关政策为医院数据治理工作指明方向，政府财政支出为医院提供稳定的物质基础，在利好的外部环境中医院内部工作人员对数据治理的认知与执行对医院数据治理工作的实施至关重要，因此该一级指标下选择政策环境、资金环境、制度环境、理念环境 4 个二级指标，前两个二级指标为外部环境，后两个为内部环境。

表 1 医院数据治理水平评价指标体系及文本库 TF、IDF 值

一级指标	二级指标	三级指标	数据治理标准		医疗机构数据		数据治理	
			与框架库		治理文本库		政策文本库	
			TF 值	IDF 值	TF 值	IDF 值	TF 值	IDF 值
技术	数据质量控制能力	诊疗和健康数据采集完整性	351	0.215	355	0.010	37	0.176
		诊疗和健康数据处理方法多样性	777	0.011	421	0.010	45	0.214
		诊疗和健康数据深度加工精准性	147	0.312	41	0.393	22	0.301
		诊疗和健康数据更新时效性	57	0.422	38	0.301	22	0.255
	数据共享与流通能力	医院数据共享共用信息化建设水平	627	0.039	230	0.043	91	0.141
		医院平台互联互通标准化成熟度水平	26	0.659	57	0.208	17	0.255
		医院互联互通标准化成熟度水平	18	0.738	50	0.281	19	0.301
	数据应用能力	医院新信息技术应用率	2327	0.005	253	0.010	491	0.051
		医院医疗业务智慧化成熟度	964	0.044	472	0.010	120	0.079
		医院医疗服务高效性	0	1.914	108	0.176	126	0.176
	数据安全保障技术能力	医疗网络安全防护设备数量	94	0.451	15	0.067	78	0.141
		医疗卫生人员数字身份管理	330	0.173	29	0.118	26	0.301
		医疗健康数据隐私保护	470	0.121	92	0.243	46	0.141

续表 1

一级指标	二级指标	三级指标	数据治理标准		医疗机构数据		数据治理	
			与框架库		治理文本库		政策文本库	
			TF 值	IDF 值	TF 值	IDF 值	TF 值	IDF 值
组织	机构设置	数据治理委员会设置	105	0.499	17	0.509	31	0.653
		数据治理办公室设置	14	0.835	9	0.669	0	1.255
	基础硬件设施	机房面积与规范化水平	4	1.437	2	0.010	1	0.954
		网络通路配置水平	54	0.516	36	0.368	8	0.556
		计算机终端及外设数量	64	0.467	16	0.669	3	0.778
		服务器及存储设备数量	12	0.914	13	0.924	56	0.653
	人才队伍	数据架构师数量	85	0.592	45	0.778	21	0.352
		数据资产管理专业人才数量	132	0.409	42	0.720	23	0.352
		医院信息化工作专业人员数量	119	0.346	86	0.393	70	0.109
		医护人员数据素养	72	0.499	15	0.778	3	0.653
环境	政策环境	医院数据治理相关法律法规出台量与相关度	670	0.081	21	0.778	169	0.176
		医院数据治理相关政府文件发布量与相关度	313	0.251	30	0.623	92	0.051
		医院数据治理相关数据标准制定量与认可度	308	0.136	211	0.032	40	0.255
	资金环境	医院数据治理工作院内自筹经费	98	0.552	18	0.623	7	0.653
		医院数据治理工作财政拨款经费	127	0.409	17	0.669	13	0.556
	制度环境	医院数据治理相关数据标准使用率	933	0.022	142	0.192	16	0.410
		数据治理管理制度制定量与精准性	931	0.022	142	0.192	91	0.176
	理念环境	医疗卫生人员数据治理工作接受度	423	0.075	35	0.322	5	0.778
		医疗卫生人员数据治理工作共识度	423	0.075	35	0.322	17	0.255

从词频来看，除“医疗服务高效性”“机房面积与规范化水平”“数据治理办公室设置”3个三级指标 TF 值偏低外，其余三级指标词频大多高于文本库文本数量，表明这些指标在文本库中均具有一定重要性。TF 值偏低的3个指标仅在某文本库中较低，主要受文本库类型影响。从 IDF 值来看，数据治理标准与框架库、数据治理政策文本库以及医院数据治理文本库中65%的指标，IDF 值均低于0.5。3组资料的TF、IDF 值显示，本指标体系中大部分指标在3个文本库中均多次被提及，语料库重要程度显著。

5 结语

本研究从整体性视角出发，指标体系全面覆盖技术、组织、环境3个维度，既为评估医院当前数据生产力水平提供有力工具，也为我国医院数据治理领域研究提供支撑。但作为探索性研究，未来应进一步拓

展其深度和广度。一方面，本研究仅探讨医院数据治理水平指标体系，后续研究应当考虑“公立医院”“非公立医院”在数据治理上的差异，并制定针对性策略。另一方面，我国医院数据治理工作稳步前行，正努力实现健康医疗数据价值的深度挖掘与有序释放，积极探索前沿领域的数据应用新范式与新业态。未来研究应聚焦于医院数据治理实践中的权责关系协调、伦理考量、法律法规遵循以及监督管理体系建设等复杂问题，促进医院更好地发挥数据要素作用，提供更为全面深入的洞见。

作者贡献：孙海珍负责论文撰写；时生辉负责研究设计、提供指导；郑秋莹负责提供指导、论文审核。

利益声明：所有作者均声明不存在利益冲突。

参考文献

1 刘银喜,王瑞娟,蔡毅臣. 数据治理共同体的内涵意蕴

- 及其构建路径——基于国家数据局的职责构成分析 [J]. 内蒙古社会科学, 2024, 45 (1): 76–83.
- 2 林伟, 周耀铭. 国内外数据治理研究述评 [J]. 数字图书馆论坛, 2022 (6): 65–72.
- 3 HE X. EU data governance act: new developments in data governance rules and implications for China [J]. *Modern law research*, 2024, 5 (1): 11–17.
- 4 裴莹蕾, 王娅. 构建区域医疗大数据治理体系的实践探索 [J]. 卫生经济研究, 2022, 39 (4): 71–74, 78.
- 5 LENCA M, FINS J J, JOX R J, et al. Towards a governance framework for brain data [J]. *Neuroethics*, 2022, 15 (2): 20.
- 6 TORABI F, SQUIRES E, ORTON C, et al. A common framework for health data governance standards [J]. *Nature medicine*, 2024, 30 (1): 26–29.
- 7 MARTA Z, JUAN Y. A reference framework for the implementation of data governance systems for industry 4.0 [J]. *Computer standards & interfaces*, 2022, 81 (4): 103595.
- 8 NIVEDITA S, RAHUL M, ANANT B, et al. Rebooting consent in the digital age: a governance framework for health data exchange [J]. *BMJ global health*, 2021, 6 (S5): e005057.
- 9 AKSHARA K, SHAMITA K, PRASHANT G, et al. Smart-AirQ: a big data governance framework for urban air quality management in smart cities [J]. *Frontiers in environmental science*, 2022, 10 (4): 785129.
- 10 DOVE E S. Biobanks, data sharing, and the drive for a global privacy governance framework [J]. *The journal of law, medicine & ethics*, 2015, 43 (4): 675–689.
- 11 ALEXANDER B, FRUZZINA M, MARIA K B. The international data governance landscape [J]. *Journal of law and the biosciences*, 2022, 9 (1): 5.
- 12 CHRISTIAN P, STEFAN B, MICHAEL B, et al. Regulatory spillovers and data governance: evidence from the GDPR [J]. *Marketing science*, 2022, 41 (4): 746–768.
- 13 CELIA A R, ALICIA M G, MÁXIMO B W, et al. Health data hubs: an analysis of existing data governance features for research [J]. *Health research policy and systems*, 2023, 21 (1): 70.
- 14 何依, 司莉. 国内外科学数据治理研究的比较分析 [J]. 情报科学, 2024, 42 (1): 153–161, 171.
- 15 张宁, 袁勤俭. 数据治理研究述评 [J]. 情报杂志, 2017, 36 (5): 129–134, 163.
- 16 张跃胜, 金文俊, 谭宇轩. 地方政府数字治理能力提升路径研究——基于 TOE 分析框架的组态分析 [J]. 经济与管理研究, 2024, 45 (4): 93–111.
- 17 刘春艳, 孙明阳. 基于 TOE 框架的公共数据开放利用水平影响因素及提升路径分析 [J]. 现代情报, 2024, 44 (3): 105–119.
- 18 程广明. 大数据治理模型与治理成熟度评估研究 [J]. 科技与创新, 2016 (9): 6–7.
- 19 林增榕, 涂晓贤. 医院绩效考核信息平台数据治理体系设计 [J]. 现代医院管理, 2024, 22 (1): 108–111.
- 20 郑跃平, 甘祺璇, 张采薇, 等. 地方政府数据治理的现状与问题——基于 43 个政务热线部门的实证研究 [J]. 电子政务, 2020 (7): 66–79.
- 21 李凯, 丁志虎. 数据治理在推动现代医院高质量发展中的研究与应用 [J]. 医学信息, 2024, 37 (8): 70–73, 82.
- 22 严昕. 公共图书馆数据治理框架构建研究 [J]. 图书馆, 2020 (5): 58–63.
- 23 韦苇, 任锦鸾, 杨青峰. 短视频平台数据治理框架和机制研究 [J]. 电子政务, 2022 (4): 64–72.
- 24 邓姚, 柴斌英, 钱方舟, 等. “国考”背景下医疗数据治理模式的探索与思考 [J]. 江苏卫生事业管理, 2022, 33 (3): 285–287, 291.
- 25 冯亚兰, 傅育红. 高质量发展视角下的医院数据治理探索实践 [J]. 现代医院管理, 2023, 21 (1): 108–110.
- 26 朱传安, 王霞, 唐佳骥, 等. 绩效考核背景下医院数据治理及开发应用实践 [J]. 医学信息学杂志, 2023, 44 (10): 86–90.
- 27 傅昊阳, 徐飞龙, 范美玉. 论医院健康医疗大数据治理及体系构建 [J]. 中国中医药图书情报杂志, 2019, 43 (3): 1–5.
- 28 阮彤, 邱加辉, 张知行, 等. 医疗数据治理——构建高质量医疗大数据智能分析数据基础 [J]. 大数据, 2019, 5 (1): 12–24.
- 29 邓军增. 医院健康医疗数据治理探讨 [J]. 医学信息学杂志, 2021, 42 (8): 14–17.
- 30 谢刚, 孙玉军, 李月云. 基于未确知测度理论的多科性医院医疗大数据治理能力成熟度评价 [J]. 技术经济, 2019, 38 (9): 89–96.