

健康医疗科学数据共享标准体系框架构建^{*}

李赞梅 钱 庆 李 姣 吴思竹

(中国医学科学院医学信息研究所 北京 100020)

〔摘要〕 调研国内外相关领域标准体系现状、框架和内容,介绍标准体系构建方法,提出由基础、数据、管理、技术、服务与应用以及专业标准 6 大类 23 小类组成的健康医疗科学数据共享标准体系框架,以期对相关数据共享标准化工作及标准制定提供参考。

〔关键词〕 健康医疗;科学数据共享;标准体系框架

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.11.011

Building of the Framework of Health Care Scientific Data Sharing Standard System LI Zanmei, QIAN Qing, LI Jiao, WU Sizhu, Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100020, China

〔Abstract〕 The current situation, frameworks and contents of the standard systems of related fields in and outside China has been investigated. The paper introduces the building methods of the standard system, puts forward the framework of health care scientific data sharing standard system that is comprised of the 6 categorizes including basis, data, management, technology, service and application and professional standards, which are consisted of 23 subcategories, in the hope of providing reference for the standardization work of related data sharing and formulation of standards.

〔Keywords〕 health care; scientific data sharing; standard system framework

1 引言

科学数据是国家科技创新和经济社会发展的重

要基础性战略资源。医疗领域科学数据共享活跃,欧美发达国家早已建设形成一批高水平的健康医疗科学数据共享平台,如目前世界上最权威的基因序列登记数据库 GenBank^[1]、国际上最重要的临床试验注册网站 ClinicalTrials.gov^[2]、涵盖英国国家医疗卫生数据的 NHS digital^[3]等。自 21 世纪初我国健康医疗科学数据共享得到快速发展,大批国家、领域、机构或项目级别的科学数据中心纷纷开展共享服务,如国家人口与健康科学数据共享平台(National Scientific Data Sharing Platform for Population and Health, NCMI)^[4]、组学原始数据存储归档系统 GSA^[5]等。标准化是科学数据有效管理与共享的前

〔收稿日期〕 2018-09-01

〔作者简介〕 李赞梅,助理研究员,发表论文 10 余篇;吴思竹,副研究员,博士,发表论文 20 余篇。

〔基金项目〕 国家科技基础条件平台后补助项目“国家人口与健康科学数据共享平台标准宣贯实施与推广”(项目编号:NCMI-MCEOI-201706);国家社科项目“国家大数据战略背景下健康医疗开放数据资源分类体系构建与应用研究”(项目编号:17CTQ020)。

提和重要基础,贯穿健康医疗科学数据共享全生命周期。科学数据管理与共享实践普遍重视标准化工作与标准制定,如 NCMI 以自身汇聚的数据为基础,先后制定资源元数据、分类与编码、资源标识、资源遴选与评价等 10 余项平台标准及为数众多的专业领域标准。

标准体系是一定范围内的标准按其内在联系形成的科学有机整体^[6],是编制标准制修订计划的主要依据,构建标准体系是推动科学数据共享发展的重要保障。然而纵观各数据共享标准化项目与实践发现国内外健康医疗科学数据共享尚缺少统一的标准体系指导^[7],阻碍了不同来源数据交换与共享。因此本文在对相关领域标准体系调研基础上分析健康医疗科学数据共享全生命周期过程的标准化需求,研究并构建健康医疗科学数据共享标准体系框架,以期对相关标准化工作及标准规范制定提供参考。

2 健康医疗科学数据共享标准体系建设现状

2.1 相关科学数据共享组织标准体系建设初具规模

随着科学数据资源建设与共享的“井喷式”发展,加强标准研制已成为国际国内各科学数据共享组织关注的热点。如国际 3 大科学数据组织之一的研究数据联盟 (Research Data Alliance, RDA)^[8],与诸多标准开发组织建立合作,在元数据标准数据永久标识、分类编码、互操作、质量控制、基础与术语、安全与隐私、出版、引用、再利用等标准领域进展显著。全球最大的跨学科科技数据组织国际科技数据委员会 (The Committee on Data for Science and Technology, CODATA)^[9]组建标准协调任务组,探讨不同科学团体间数据标准的协调问题。科技部科学数据共享工程为更好推进离散数据整合与共享,提出由基础、公用、技术和学科领域标准组成的体系框架^[10]。在该框架指导下水利^[11]、交通^[12]、气象^[13]、地球系统^[14]等领域纷纷研制本行业科学数据共享标准体系并加以实施。作为科学数据共享工程支持的国家人口健康领域唯一科学数据共享项目,国家人口与健康科学数据共享平台自建

设初期就建立标准化工作组,围绕平台资源建设和共享管理需要,制定以促进数据共享为目的的各类标准规范,初步形成从数据采集、处理、加工、存储、质量控制到发布的一套较为完整的标准体系。中国标准化研究院王志强等人^[15]提出科技资源开放共享标准化的解决方案,构建由总体标准、元数据、分类编码、科技资源标识规范、科技资源管理标准、科技资源安全标准等子体系构成的基本框架。

2.2 卫生健康信息及大数据领域标准开发组织相关标准体系具有一定参考和借鉴价值

国际健康信息学标准化权威机构 ISO/TC 215^[16]开展卫生健康信息系统建设与应用标准研发,已发布和正在研制的 230 余项国际标准主要涉及架构-框架和模型类、信息安全类、系统和设备互操作类、语义内容类、质量和单位类标准等。我国国家卫生标准委员会信息标准专业委员会提出以互联互通标准为重点的医疗健康信息标准体系概念模型^[17],从业务领域、标准内容和标准级别 3 个维度划分出基础类、数据类、技术类、安全与隐私类、管理类 5 大类标准。国际大数据标准化工作主要集中在 ISO/IEC JTC1/WG9 大数据工作组^[18]以及 JTC1/SC32 数据管理和交换技术委员会^[19]。WG9 编制的两项大数据基础性国际标准 (参考架构、概念和术语) 以及 SC32/WG2 元数据工作组对于相关行业标准制定有重要参考和借鉴意义。美国国家标准与技术研究院大数据公共工作组 NBD-PWG^[20]致力于构建可扩展的大数据互操作框架,现已发布术语和定义、大数据分类、安全和隐私、参考体系结构、标准路线图等标准。全国信标委大数据标准工作组专门负责制定和完善我国大数据领域相关标准^[21],提出结合数据全周期管理的大数据标准体系框架,包括基础、数据、技术、数据安全、数据质量、产品和平台以及应用和服务标准等方面^[22]。福州、湖北、四川、广东、天津等地为进一步贯彻落实国家健康医疗大数据融合共享与开放应用发展,相继提出构建包括基础类、数据类、技术类、应用类、管理类以及安全与隐私保护类标准等在内的省

(市)级人口健康信息化和健康医疗大数据标准体系。全国科技平台标准化技术委员会 SAC/TC 486^[23]是与我国科学数据共享关系最为密切的标准开发组织,已发布和立项科技平台标准化指南、元数据标准化基本原则与方法、通用术语、科技资源标识、元数据注册与管理、核心元数据、统一身份认证等 30 余项国家及行业标准,相关标准为科学数据资源共享利用发挥重要作用。综合上述分析,健康医疗科学数据共享相关标准化工作取得了长足的进步,较好地满足健康医疗科学数据采集、加工、描述、汇交、传输、存储、整合及再利用等的标准化需求。然而,健康医疗科学数据共享标准研制缺乏顶层设计,也尚未形成统一的标准体系。

3 标准体系构建过程与方法

主要遵循国家《标准体系构建原则和要求》的有关要求和内容,大体包括以下环节:一是在构建标准体系之前对健康医疗科学数据共享的业务进行全生命周期分析,了解标准化所支撑的业务内容,明确标准体系建设要达到的目标,确定标准体系构建的指导思想和基本原则。二是开展健康医疗科学数据共享相关标准及标准体系的调查研究。调查与分析的内容包括国内外相关标准及标准体系建设情况、健康医疗科学数据共享现有的标准化基础以及存在的问题等。三是根据标准化业务需求,充分借鉴和吸收国内外现有的相关标准体系结构框架建设经验,以科学数据共享生命周期为主线,梳理健康医疗科学数据共享标准化工作关键问题。从标准级别、类型、类别、业务生命周期等不同角度对健康医疗科学数据共享标准体系结构进行分析,确定标准体系框架参考模型。四是编制标准体系框架或结构图及标准明细表。以标准体系框架参考模型为基础,确定标准体系基本框架,基于收集整理的拟采用的各类外部标准和本领域已有的内部标准提出近期和将来规划制定的标准列表,编制标准明细表。同时由于标准体系是为业务目标服务,随着健康医疗科学数据共享业务需求和技术的发展,标准体系在使用过程中也应不断优化完善。采用的研究方法

主要包括调查法、文献研究法和系统科学方法。健康医疗科学数据共享标准体系构建思路 and 过程,见图 1。

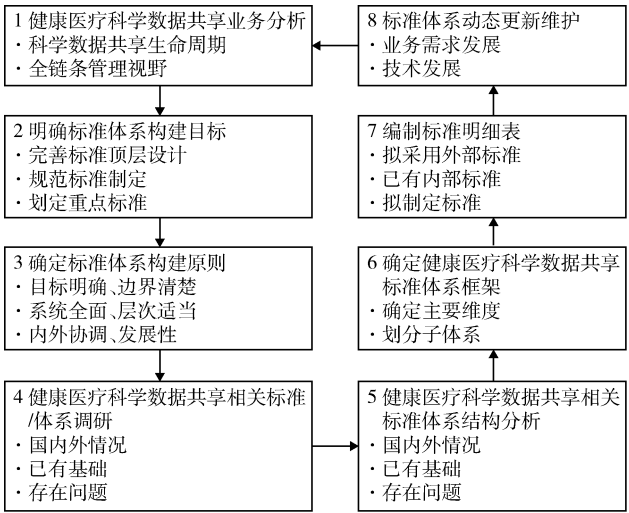


图 1 健康医疗科学数据共享标准体系构建思路 and 过程

4 健康医疗科学数据共享标准体系框架

4.1 参考模型

健康医疗科学数据共享标准体系框架参考模型由标准类型、标准类别和科学数据共享业务生命周期 3 个主维度构成,见图 2。相关维度互相交叉,派生出子体系或具体标准。根据标准类型属性可以将标准分为基础、通用和专用标准。根据标准类别属性可以将标准分为数据类、管理类、技术类、服务和应用类。根据健康医疗科学数据共享生命周期各阶段可以将标准分为数据规划类、采集与汇交类、传输与交换类、存储类、处理类、发布与服务类、再利用类等。

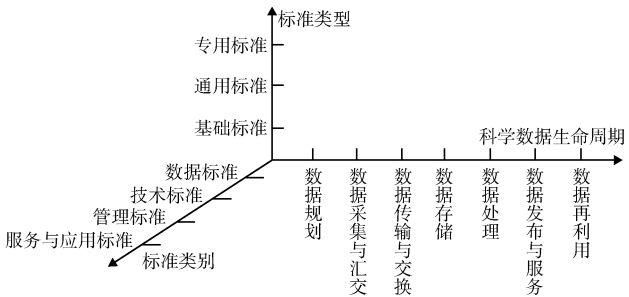


图 2 健康医疗科学数据共享标准体系框架参考模型

4.2 具体框架

4.2.1 概述 标准体系框架（或结构图）用于表达标准体系的范围、边界、内部结构以及意图，主要反映标准体系的标准组成和层次结构关系。基于标准体系框架参考模型将健康医疗科学数据共享标准体系框架划分为基础、数据、管理、技术、服务与应用以及专用标准 6 大类以及 23 小类，见图 3。

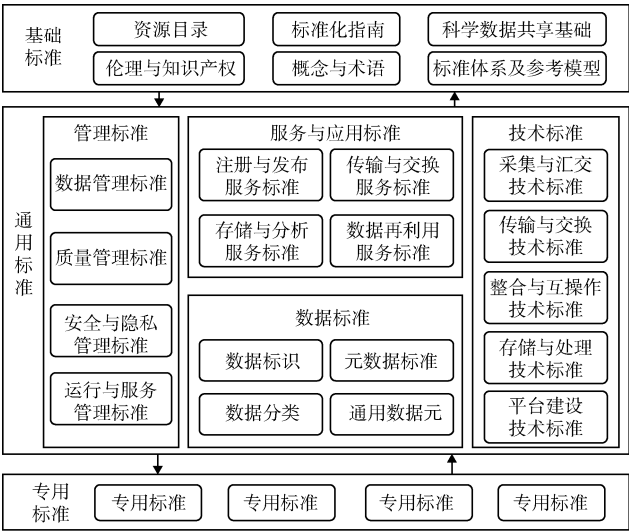


图 3 健康医疗科学数据共享标准体系框架

4.2.2 基础标准 是健康医疗科学数据共享各项标准制定、协调与应用普遍要遵循的基础，主要包括标准化指南、基础概念与术语、科学数据共享基础、资源目录、标准体系与参考模型以及数据共享伦理与知识产权 6 个子类。

4.2.3 通用标准 是健康医疗科学数据共享与服务过程中具有共性的相关标准，可划分为数据、管理、技术、服务与应用标准 4 个主要类别。（1）数据标准。是健康医疗科学数据共享标准体系的核心，主要针对数据内容相关要素进行规范，包括数据标识规范、元数据标准、通用数据元标准以及数据分类与编码标准 4 个子类。数据标准与健康医疗科学数据共享业务系统实现无关。（2）技术标准。对应健康医疗科学数据共享相关活动所需采用的各类软硬件技术规范。围绕科学数据共享全生命周期关键环节，可以将技术标准分为数据采集与汇交技术标准、数据传输与交换技术标准、数据整合与互

操作技术标准、数据存储与保存技术标准、数据处理技术标准以及平台建设技术标准 5 个方面。（3）管理标准。为健康医疗科学数据共享提供管理手段和措施，贯穿数据共享生命周期各阶段，是实现数据共享科学管理和有效运行的重要保障，主要包括数据管理标准、质量管理标准、安全与隐私管理标准、运行与服务管理标准 4 个方面。（4）服务与应用标准。是提供健康医疗科学数据共享服务和相关应用的标准总称，涉及针对数据共享各相关方（数据发布者、使用者、管理者及第 3 方等）的各类型服务，主要包括数据注册与发布服务标准、数据传输与交换服务标准、数据存储与分析服务标准、数据再利用服务标准等 4 个方面。

4.2.4 专用标准 是指根据通用标准制定出来、以满足特定领域数据共享需求的标准。健康医疗科学数据共享标准体系中的专用标准有临床医学术语标准、公共卫生数据元设计和管理规范、药品名称规范、疾病分类与代码标准等。

5 结语

标准体系构建是科学数据共享的一项重要基础性工作。本文以健康医疗科学数据共享全生命周期对标准化的需求为基础，参考与借鉴健康医疗领域相关标准以及其他领域标准体系建设成果和经验，结合现有的标准规范，提出健康医疗科学数据共享标准体系框架并对主要内容进行描述，以期对健康医疗科学数据共享标准体系确立及后续有关工作提供基础和参考。下阶段将以本研究构建的标准体系框架为基础，以国家人口与健康科学数据共享平台标准化工作为切入点，编制针对该平台科学数据共享的标准体系结构图与标准明细表，科学地指导相关标准的制修订计划、规划编制和实施。

参考文献

1 GenBank. GenBank Overview [EB/OL]. [2018 - 04 - 15]. <https://www.ncbi.nlm.nih.gov/genbank/>.
2 U. S. National Library of Medicine. ClinicalTrials. gov [EB/OL]. [2018 - 04 - 15]. <https://clinicaltrials.gov>.

- gov/.
- 3 National Health Service. NHS digital [EB/OL]. [2018 - 04 - 15]. <https://digital.nhs.uk/>.
 - 4 国家人口与健康科学数据共享平台 [EB/OL]. [2018 - 04 - 16]. <http://www.ncmi.cn/1>.
 - 5 BIG data center. BIG Standards for Big Omics Data [EB/OL]. [2018 - 04 - 16]. <http://bigd.big.ac.cn/standards>.
 - 6 中国标准化研究院. GB/T 13016 - 2018 标准体系构建原则和要求 [M]. 北京: 中国标准出版社, 2018: 2.
 - 7 司莉, 贾欢. 科学数据的标准规范体系框架研究 [J]. 图书馆, 2016 (5): 5 - 9.
 - 8 Research Data Alliance. Research Data Sharing without Boundaries [EB/OL]. [2018 - 04 - 18]. <https://rd-alliance.org/>.
 - 9 CODATA, The Committee on Data for Science and Technology. Coordinating Data Standards Research Data Sharing without Boundaries [EB/OL]. [2018 - 04 - 18]. <http://www.codata.org/task-groups/coordinating-data-standards>.
 - 10 徐枫. 科学数据共享标准体系框架 [J]. 中国基础科学, 2003 (1): 46 - 51.
 - 11 耿庆斋, 朱星明. 水利科学数据共享标准体系研究与构建 [J]. 水利学报, 2007, 38 (2): 233 - 238.
 - 12 周紫君, 王辉, 林焱, 等. 交通科学数据共享标准规范体系框架研究 [J]. 交通与计算机, 2008, 26 (5): 152 - 154, 130.
 - 13 王国复, 张钛仁, 丁海芳. 气象科学数据共享标准体系框架研究 [C]. 西宁: 中国气象学会气象通信与信息技术委员会 2009 年度学术研讨会论文集, 2009: 1 - 7.
 - 14 王卷乐, 孙九林. 地球系统科学数据共享标准规范体系研究与应用 [J]. 地理科学进展, 2009, 28 (6): 839 - 847.
 - 15 王志强, 杨青海. 科技资源开放共享标准体系研究 [J]. 中国科技资源导刊, 2016, 48 (4): 19 - 23, 61.
 - 16 International Organization for Standardization. ISO/TC 215 Health informatics [EB/OL]. [2018 - 04 - 25]. <https://www.iso.org/committee/54960.html>.
 - 17 汤学军, 董方杰, 张黎黎, 等. 我国医疗健康信息标准体系建设实践与思考 [J]. 中国卫生信息管理杂志, 2016, 13 (1): 31 - 36.
 - 18 光亮, 张群. ISO/IEC JTC1/WG9 大数据国际标准研究及对中国大数据标准化的影响 [J]. 大数据, 2017, 3 (4): 20 - 28.
 - 19 ISO/IEC JTC1 SC32. Standards for Data Management and Interchange [EB/OL]. [2018 - 05 - 10]. <http://jtc1sc32.org/>.
 - 20 NIST Big Data Public Working Group. Big Data at NIST [EB/OL]. [2018 - 05 - 15]. <https://bigdata.wg.nist.gov/>.
 - 21 全国信息技术标准化技术委员会. 全国信息技术标准化网 [EB/OL]. [2018 - 05 - 25]. <http://www.nits.org.cn/>.
 - 22 《大数据标准化白皮书 v2.0》发布大数据标准体系框架 [J]. 中国标准导报, 2016 (1): 9 - 10.
 - 23 中华人民共和国科学技术部. 全国科技平台标准化技术委员会工作会议在京召开 [EB/OL]. [2018 - 04 - 23]. http://www.most.gov.cn/kjbgz/201402/t20140220_111901.htm.

《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