

中医药领域信息数据元标准化与规范化工作研究^{*}

徐倩 熊振芳 毛树松

周琼

(湖北中医药大学 武汉 430061)

(湖北省中医院 武汉 430061)

〔摘要〕 介绍中医药领域信息数据元标准化工作的思路与步骤,制定标准化工作方案,梳理归纳标准化工作中出现的常见问题及解决方法,为中医药信息领域数据元标准化工作提供借鉴。

〔关键词〕 中医药信息;数据元;标准

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2017.09.012

Study on Standardization and Normalization Work of Information Data Elements in the Field of Traditional Chinese Medicine

XU Qian, XIONG Zhen-fang, MAO Shu-song, Hubei University of Chinese Medicine, Wuhan 430061, China; ZHOU Qiong, Hubei Provincial Hospital of TCM, Wuhan 430061, China

〔Abstract〕 The paper introduces the thoughts and steps of standardization work of information data elements in the field of Traditional Chinese Medicine (TCM), formulates the schemes for standardization work, sorts out common problems and solutions in the standardization work, and provides references for standardization work of information data elements in the field of TCM.

〔Keywords〕 Traditional Chinese Medicine (TCM) information; Data element; Standard

1 引言

数据元标准化是信息技术标准化的重要组成部分,是信息化的基础和信息运行的保证^[1]。中医药作为我国独具特色的资源,在信息化发展历程中,也面临着标准化程度不高、数据定义不一致造成的信息孤岛、信息烟囱、数据鸿沟、数据垃圾和数据

冲突等问题^[2]。2015 年国家中医药管理局拨出专项资金支持开展 101 项中医药信息标准制修订项目,以建立统一开放的中医药信息化标准体系,按照其内在联系形成一个科学的有机整体,促进中医药信息化建设有序、高效、快速和健康地发展^[3],逐步完成中医药信息资源共享和交换。在中医药信息标准研究与制定项目中有 39 个涉及基本数据集、数据元目录类标准的项目,共 18 041 个数据元,但由于各项目承担单位不同,项目前期提取完成的数据元还存在着编码不统一,不同项目单位交叉都有的数据元名称不一致,定义不规范等问题。在此背景下中医药信息研究会专门组织成立中医药信息标准数据元规范化工作小组,开展中医药信息数据元规范化研究。本文就中医药信息数据元及基本数据集类标准项目中存在的问题及解决方法、研究思路、

〔修回日期〕 2017-06-12

〔作者简介〕 徐倩,硕士研究生;通讯作者:熊振芳,博士,教授。

〔基金项目〕 国家中医药管理局中医药信息标准研究与制定项目“中医医院护理管理信息基本数据集”(项目编号: SATCM-2015-XXBZ [012])。

工作步骤进行介绍。

2 相关概述与研究思路

2.1 概述

中医药信息标准是指中医药事物处理过程中，对其信息采集、传输、交换和利用时所采用的统一的规则、概念、名称、术语、代码和技术。目前中医药信息标准主要包括术语标准和数据交换标准^[4]。信息数据元目录及值域代码、基本数据集即属于数据交换类标准。数据元是用一组属性描述其定义、标识、表示和允许值的数据单元（GB/T 18391.1—2009）。数据元属性根据重复程度分为公用属性和专用属性两类。其中，公用属性有 7 项，分别是版本、注册机构、相关环境、分类模式、主管机构、注册状态和提交机构；专用属性有 6 项，分别为数据元标识符（DE）、数据元名称、数据元定义、数据元值的数据类型、表示格式、数据元允许值^[5]。

2.2 研究思路

中医药信息数据元是卫生信息数据元的重要组成部分，其标准化工作主要依据卫生信息数据元目录^[6]编制原则，按照“统一思想、统一方法、统一标准”的思路开展^[2]。数据元属性设置和属性描述、数据元标识符编制、分类部分等方法参考 WS/T 303 - 2009《卫生信息数据元标准化规则》^[7]、《卫生信息数据模式描述指南》、WS/T 305 - 2009《卫生信息数据集元数据规范》^[8]、国家标准 GB/T 18391.3 - 2001（信息技术/数据元的规范与标准化/第 3 部分：数据元的基本属性）^[9]等标准文件。

3 研究步骤

3.1 成立中医药信息标准数据元规范化工作组及任务分工

根据工作任务将 39 个标准项目分为电子政务、中医临床（医）、中医临床（药护技）、医院管理和中医馆等 5 个专题，组织 5 个工作小组完成。各工作组成员负责与本组项目单位负责人的沟通、协调及数据元汇总、整理、反馈等工作，数据元的编

码、名称规范、定义规范等工作由整个工作组共同完成，具体项目分组，见表 1。

表 1 标准项目名称及专题分组

专题组	项目名称
I 电子政务	中医药综合统计信息数据元目录及值域代码
	中医药综合统计信息基本数据集
	中医医院经济管理绩效考评信息数据元目录及值域代码
	中医药部门公共卫生服务补助资金项目管理信息基本数据集
II 中医临床（医）	中医医院皮肤科信息数据元目录及值域代码
	中医医院心血管科信息数据元目录及值域代码
	中医电子病历基本数据集
	中医医院推拿科电子病历基本数据集
	中医医院急诊科信息数据元目录及值域代码
	中医医院外科信息数据元目录及值域代码
	中医医院骨伤科电子病历基本数据集
	中医医院针灸科电子病历基本数据集
	中医辅助诊疗信息基本数据集
	中医医院移动医疗信息基本数据集
	中医临床路径信息数据元目录及值域代码
	中医临床路径信息基本数据集
III 中医临床（药护技）	中医医院医疗质量控制信息数据元目录及值域代码
	中药煎药管理与质量控制信息基本数据集
	中医医院护理管理信息数据元目录及值域代码
	中医医院护理管理信息基本数据集
	中医医院临床护理信息基本数据集
IV 医院管理	中医医院医技检查项目信息基本数据集
	中医院协同办公信息数据元目录
	中医院协同办公信息数据元值域代码
	中医医院协同办公信息基本数据集
	中医医院资源管理信息数据元目录
	中医医院数据资源管理数据元值域代码
	中医医院资源管理信息基本数据集
	中医医院伦理审查信息基本数据集
	中医医院临床药事管理信息基本数据集
	中医医院医务管理信息基本数据集
V 中医馆	中医馆健康信息云平台信息数据元目录及值域代码
	中医馆健康信息云平台知识库信息基本数据集
	中医馆健康信息云平台电子病历信息基本数据集
	中医馆健康信息云平台业务监管信息基本数据集
	中医馆健康信息云平台辅助开方信息基本数据集
	中医馆健康信息云平台信息管理信息基本数据集
	中医馆健康信息云平台治未病信息基本数据集
	中医馆健康信息云平台远程会诊信息基本数据集

3.2 确定工作方案

工作组通过组织召开方案讨论会，在咨询专家意见及各项目单位意见的基础上，经过两轮的预实验形成以下工作方案，见图 1。

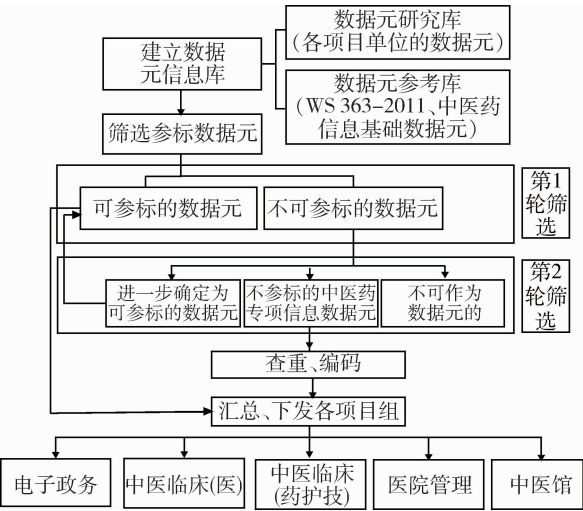


图 1 数据元规范化工作方案

3.3 组织实施

3.3.1 建立数据元信息库 对标准项目组提交的标准文本进行分类整理，提取标准中的数据元信息，形成包含数据元名称、标识符、定义、数据类型、表示格式、允许值及数据元来源和出现频次等内容的 Excel 电子表（简称数据元研究库），对各项目组的每个数据元做好相应标识，以便后期有效溯源和反馈。同时将卫生信息数据元目录^[6]及已编制完成的中医药信息数据元^[2]标准按照卫生信息 17 个部分整理成数据元参考库 Excel 电子表。

3.3.2 筛选可参标数据元 运用 VLOOKUP 函数，对比数据元研究库与数据元参考库，初步筛选出数据元研究库中属于参标的数据元，然后以双人核对形式进行 2 轮人工甄别，进一步确定可参标数据元，最后将数据元研究库中的数据元分为可参标数据元（卫生信息数据元和中医药基础信息数据元）、中医药专项信息数据元和不可作为数据元的数据元 3 个类别。

3.3.3 规范可参标数据元 汇总参标的数据元后，将标准的数据元标识符、数据元名称、定义、数据元值的数据类型、表示格式、数据元允许值添加到原数据元列表中以供修改和参考。对于名称相同、语义不同的数据元，工作组给予保留原有定义，但也将标准的数据元定义、表示格式等信息列出，供项目单位参考。

3.3.4 处理不可作为数据元的数据元 数据元是一组可识别、可定义的数据基本单元，一般由名称、属性、表示 3 部分组成，在一定语境下，通常用于构建一个语义正确、独立且无歧义的特定概念的信息单元^[10]。工作组在处理数据元的过程中发现如“生命体征”、“必选项”、“选择框提示”、“简要病情”、“营养护理”等不可作为数据元名称的数据元，均在 Excel 数据表中标记，给予“建议修改或删除”意见。

3.3.5 规范中医药专项信息数据元 中医药各业务范畴信息具有灵活性和的特殊性等特点，统一并规范中医药专项信息数据元是本次标准化工作的难点和重点。具体工作方法如下：（1）统计频次：统计数据元研究库中的中医药专项信息数据元名称的频次，以频次 ≥ 2 为原则，筛选出数据元名称相同或相近且定义相同或相近的数据元。（2）统一数据元名称、定义：与项目组沟通，征求各项目组单位意见，通过专家咨询最终对数据元名称、定义进行统一。

3.3.6 编制中医药专项信息数据元标识符 中医药领域信息数据元的编码标准依据《卫生信息数据元标准化规则》^[7]制定。首先将数据元研究库中的中医药专项信息数据元统一查重，按卫生信息标准的 17 部分进行归类整理，形成一个不含重复项且按 17 部分进行分类的数据元库。然后按照数据元分类架构，采用字母数字混合码，包含 5 段，分别为主题分类代码、大类码、小类码、顺序码和附加码，14 位编码。数据标识符（DI）结构，见图 2。其中“附加码”代表一组数据元的连用关系编码，无连用关系的数据元其附加码为“00”，有连用关系的数据元附加码从 01 开始顺序编码^[11]。

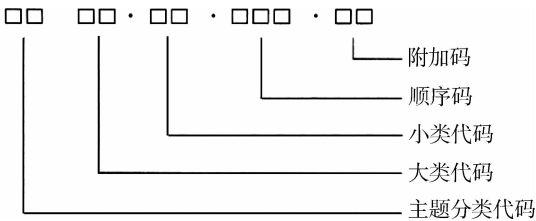


图2 数据标识符 (DI) 结构

中医药专题信息数据元与卫生信息数据元的主

题分类代码、大类代码、小类代码保持一致。中医药基础信息数据元标识符的顺序码从 401 开始编，中医药专项信息数据元标识符的顺序码用 1 位大写字母和 2 位阿拉伯数字表示，从大写字母 A、数字 01 开始编码，依数字自然顺序递增，编至 99 时，则再依字母顺序从 B01 继续编。卫生信息数据元、中医药信息数据元及中医院专项信息数据元的编码，见表 2。

表 2 数据元标识符编码段

部分	名称	卫生信息数据元标识符	中医药信息基础数据元标识符	中医药信息专项数据元标识符
2	标识	DE01.00.001.00	DE01.00.401.00	DE01.00.A01.00
3	人口学及社会学特征	DE02.01.001.00	DE02.01.401.00	DE02.01.A01.00
4	健康史	DE02.10.001.00	DE02.10.401.00	DE02.10.A01.00
5	健康危险因素	DE03.00.001.00	DE03.00.401.00	DE03.00.A01.00
6	主诉与症状	DE04.01.001.00	DE04.01.401.00	DE04.01.A01.00
7	体格检查	DE04.10.001.00	DE04.10.401.00	DE04.10.A01.00
8	临床辅助检查	DE04.30.001.00	DE04.30.401.00	DE04.30.A01.00
9	实验室检查	DE04.50.001.00	DE04.50.401.00	DE04.50.A01.00
10	医学诊断	DE05.01.001.00	DE05.01.401.00	DE05.01.A01.00
11	医学评估	DE05.10.001.00	DE05.10.401.00	DE05.10.A01.00
12	计划与干预	DE06.00.001.00	DE06.00.401.00	DE06.00.A01.00
13	卫生费用	DE07.00.001.00	DE07.00.401.00	DE07.00.A01.00
14	卫生机构	DE08.10.001.00	DE08.10.401.00	DE08.10.A01.00
15	卫生人员	DE08.30.001.00	DE08.30.401.00	DE08.30.A01.00
16	药品、设备与材料	DE08.50.001.00	DE08.50.401.00	DE08.50.A01.00
17	卫生管理	DE09.00.001.00	DE09.00.401.00	DE09.00.A01.00

3.4 结果统计

本次规范化研究工作处理数据元共计 18 041 个，通过与数据库成果库比对，筛选出可参标数据元 7 523 个，其中卫生信息数据元 6 192 个；中医药信息基础数据元 1 331 个。未参标的数据元 10 518 个，其中可以作为中医药专项组数据元的有 8 570 个，不建议作为数据元的数据元共计 1 901 个；还存在一部分数据元，同时指向多条参标数据元共计 47 个。

3.5 问题总结及解决方法

3.5.1 问题总结 工作组在数据元整理的过程中，发现各项目组的数据元在名称、定义等属性描

述及 17 部分分类方面上还存在着以下问题。（1）数据元名称、定义不标准：如“药物用法”在卫生信息数据元目录^[6]中标准定义为“对治疗疾病所用药物的具体服用方法的描述”，而有的数据集中数据元名称为“中药用药方法”，有的数据集中数据元名称为“用药方法”，将其分别定义为“中药的用药方法的描述”、“药物（含中药）治疗疾病具体用法的描述”等。（2）数据元内涵相同但数据元名称、定义不一致：如出现在不同数据集中的数据元名称“使用医疗机构中药制剂标志”、“使用中药制剂标志”，虽数据元名称与定义不同，但实际表达的内涵相同。（3）数据元名称不规范：数据元是在特定的语义环境中被认为是不可再分的最小数据单元^[12]。但有项目标准中却出现了“生命体征”这

一数据元,其数据元名称同时指向“体温(℃)”、“心率(次/min)”、“呼吸次数(次/min)”、“收缩压(mmHg)”、“舒张压(mmHg)”等多条参标数据元。(4)数据元定义不准确:数据元定义表述含糊不够明确,如有数据集中将“中药使用途径名称”定义为“中药使用途径的代码对应的名称”。(5)数据元的数据类型选择不合适:如逻辑型数据元的数据类型为布尔型,即 T/F(是/否),而在有的数据集中数据元允许值却以可枚举的代码表形式表示。(6)其他:中医药信息标准制修订项目中《中医基本症状信息分类与代码》、《中医舌象诊断信息分类与代码》、《中医脉象诊断信息分类与代码》等中医基础信息标准代码已编制完成,中医单一症状如“寒战”“胸闷”等,应改为代码形式指向其中医基本症状代码,不建议单独作为数据元。

3.5.2 解决方法 工作组在数据元整理后针对每个数据元均给予参考意见,对可参标的数据元均在原来的数据元名称、定义后把标准的数据元,定义罗列出来以供参考;对不建议作为数据元的数据元进行标示并给予“建议修改或删除”意见;对中医药专题信息数据元的数据元名称、定义均给予参考的名称和定义。最后将各项目组数据元按原顺序还原,下发给各项目组,各项目组对反馈的信息进行逐一比对,仔细核实每条数据元(包括数据元标识符、数据元名称、定义、数据元值的数据类型、表示格式、数据元允许值及分类部分)的参考意见,对有异议、后期决定增、删、改的数据元,均标记好后统一反馈给标准化工作组进行下一步处理。

4 结语

目前,中医药信息标准研究与制定项目中的 39 个涉及基本数据集及数据元目录的项目数据元标准化研究工作已进行了两轮协调反馈,反馈给各单位的数据元参考意见及数据元标识符均以国家卫生和计划生育委员会及国家中医药管理局已颁发的

有关标准文件为依据,结合中医药信息基本数据集和数据元目录的特点,从中医药数据标准化的实际情况出发,通过召开培训会、多次向专家咨询及与项目单位沟通而确定。由于研究时间短,工作量较大,在数据元分类、核对、编码等数据元规范化过程中可能会出现个别疏漏和不合理的地方,另外给予的参考定义表述是否规范合理等,均需在实践中加以验证和完善。

参考文献

- 1 张爱,赵艳华,张铭,等.国际数据元标准化研究(2)[J].标准科学,1998,(5):26-27.
- 2 马红敏,常凯,孙静,等.中医药信息数据元标准编制思路与方法[J].医学信息学杂志,2014,35(7):46-49.
- 3 常凯.中医药信息化标准体系构建研究[D].武汉:湖北中医药大学,2012.
- 4 董燕,崔蒙.中医药信息标准研究现状与趋势[J].世界科学技术-中医药现代化,2009,11(4):612-616.
- 5 冯丹,周游,姚远,等.我国居民健康档案基本数据集数据元目录编制[J].中国卫生信息管理杂志,2012,9(1):71-74.
- 6 孟群,刘丽华,胡建平,等.WS/363 2011 卫生信息数据元目录[S].2011.
- 7 刘丽华,王才有,金水高,等.WS/T303 2009 卫生信息数据元标准化规则[S].2009.
- 8 原卫生部政策法规司主编.WS/T 305-2009 卫生信息数据集元数据规范[S].2009.
- 9 GB/T 18391-2001 信息技术数据元的规范与标准化.第 1-6 部分:数据元的规范与标准化框架(idt ISO/IEC [1] 79:1994-2000)[S].2001.
- 10 刘丽华,高筠,代伟,等.医院统计指标公用数据元目录设计[J].中国医院.2007,12(11):15-17.
- 11 冯丹,周游,姚远,等.我国居民健康档案基本数据集数据元目录编制[J].中国卫生信息管理杂志,2012,9(1):71-74.
- 12 谢曼,王晓芬,吴燕,等.体育信息基础数据元的现状分析[C].上海:全国体育信息科技学术大会,2016.