

中医药数据元和数据集标准规范化研究^{*}

舒亚玲 潘丹妹 赵移畛 金 琰

肖 勇

(河南省洛阳正骨医院 (河南省骨科医院) 洛阳 471002)

(湖北中医药大学 武汉 430065)

〔摘要〕 详细阐述借助 EXCEL 进行中医药数据元、数据集标准规范化的具体流程, 指出其应用能够提高工作效率, 降低出错率, 为相关研究提供参考。

〔关键词〕 数据元; 数据集; 标准规范

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2021.02.008

Study on Standardization of Data Element and Dataset of Traditional Chinese Medicine SHU Yaling, PAN Danmei, ZHAO Yizhen, JIN Yan, Luoyang Orthopedic Hospital of Henan Province (Orthopedic Hospital of Henan Province), Luoyang 471002, China; XIAO Yong, Hubei University of Chinese Medicine, Wuhan 430065, China

〔Abstract〕 The paper elaborates the standardization of data element and dataset of Traditional Chinese Medicine (TCM) with EXCEL, points out that its application can improve work efficiency and reduce error rate, so as to provide references for relevant study.

〔Keywords〕 data element; dataset; standard specification

1 引言

中医药发展“十一五”、“十二五”、“十三五”规划均将完善标准体系作为重点任务^[1]列入其中。2015 年国家中医药管理局拨付专项资金开展信息标准制修订研究, 共立项 101 项中医药信息标准, 项目承担单位涵盖北京、湖北等全国 13 个省 (市) 36 家中医药机构, 其中中医药大学 6 所、三级甲等中医医院 27 家、中医科研院所 3 家, 直接参与项目

研究人员达 640 余人^[2]。历时 3 年已发布第 1 批团体标准 57 项并于 2019 年 5 月 1 日正式实施, 其中数据元、数据集类标准 26 项, 功能规范类标准 13 项, 建设指南类标准 7 项, 分类代码及其他规范类标准 11 项。本文对上述 26 项数据元、数据集标准发布前最后的规范化整理进行总结和思考, 应用办公软件 EXCEL 2010 为数据元、数据集类标准规范化提供常规方法和思路。

2 研究思路与前期准备

2.1 研究思路

前期已组织专人对 101 项标准的名称术语^[3]以及数据元标识符、数据集内部标识符等进行规范化整理^[4], 项目组根据专家意见修改标准文本过程中增删数据元的情况, 故本次数据元、数据集标准规范化的重点主要是对 9 项数据元 (目录或值域代

〔收稿日期〕 2020-04-25

〔作者简介〕 舒亚玲, 硕士, 发表论文 8 篇; 通讯作者: 潘丹妹, 高级工程师。

〔基金项目〕 国家中医药管理局 2015 年中医药信息标准研究与制定项目“基层医疗卫生机构中医诊疗区 (中医馆) 健康信息平台信息数据元目录及值域代码” (项目编号: SATCM-2015-XXBZ [80])。

码) 和 17 项数据集的数据元标识符和数据集内部标识符进行统一, 此外还包括标准报批稿格式, 标点符号、语言文字及译文等的审查。研究具体思路, 见图 1。

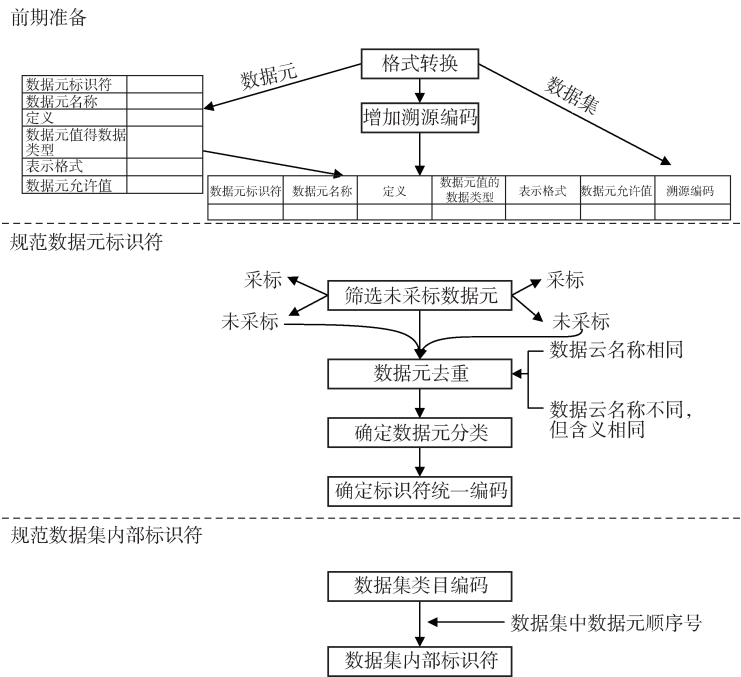


图 1 中医药数据元、数据集标准规范化研究思路

2.2 前期准备

2.2.1 格式转换 由于提交的中医药数据元、数据集标准均是 WORD 版本, 为便于开展规范化工作, 首先需要将 WORD 版本的数据元、数据集转换成 EXCEL 表格形式。数据集标准的表示格式是表格形式, 转换直接采用表格复制粘贴即可。但是在数据元目录中, 数据元的表示格式为目录式, 采用了《卫生信息数据元目录 第 1 部分: 总则》(WS363.1-2011) 中“数据元专用属性摘要式目录的通用格式”进行表示, 需利用公式 = IF (TRANSPOSE (OFFSET (\$ B \$ 1, (COLUMN (A1) - 1) + (ROW (A1) - 1) * 6, ,)) = 0, "", TRANSPOSE (OFFSET (\$ B \$ 1, (COLUMN (A1) - 1) + (ROW (A1) - 1) * 6, ,))) 将其转换成 EXCEL 中的表格格式。应用该公式前先检查目录式数据元中不存在一个属性占多行单元格的情况, 否则转换后的属性值会错位。另外应用时首先在一个单元格中输入此公式, 然后横向填充完一个数据元的 6 个属性 (因

为本次统稿的所有数据元类标准都是选用 6 个专用属性, 若属性值有增加需拖动相应横着的单元格数, 公式中对应的跨行的“6”也需做相应修改), 最后再选中一整条数据元包含的 6 个属性值竖直拖动, 填充完每一个数据元为 EXCEL 中的一行进行表示。

2.2.2 增加溯源编码 数据元和数据集类标准中的数据元在 EXCEL 中以表格显示, 为了规范完成后对标准文本做相应修改, 以及后期能准确定位每个数据元的来源, 在数据元标准的 6 个属性后面添加数据元溯源编码列, 溯源编码规则, 见图 2。标准规范负责人编号: 1 位字符, 可以是标准规范负责人姓的拼音大写首字母, 也可以是约定的任意方便记忆和识别的字符。项目编号: 3 位数字, 项目立项后的数字编号或者是预设定的标准号。顺序号: 4 位数字 (具体位数以统稿中单个标准中所含数据元条数最大值确定), 每个标准中的数据元在 EXCEL 中顺序填充数字。

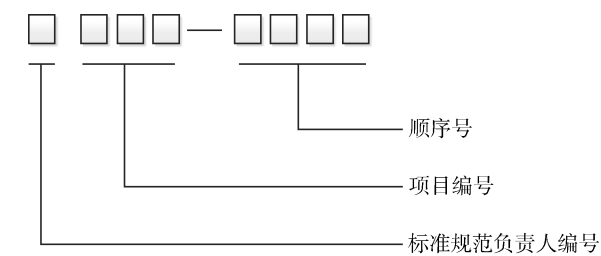


图 2 数据元溯源编码规则

3 规范数据元标识符

3.1 筛选未采标数据元

3.1.1 概述 将各个标准的数据元合并在一个 sheet 表中，共采集初始数据元 11 915 条（包含单个数据集内部重复的数据元和标准间重复的数据元）。对数据元标识符进行统一。

3.1.2 数据元名称与卫生信息数据元、中医药信息数据元名称相同的情况 完全采用卫生信息数据元、中医药信息数据元的标识符、名称、定义、数据元值的数据类型、表示格式、允许值范围所有属性。

3.1.3 数据元名称与卫生信息数据元、中医药信息数据元名称不同的情况 其数据元标识符以卫生信息数据元标识符编码方法为基础，数据元标识符的主题分类代码、大类代码、小类代码遵循 WS 363.1－2011，顺序码由项目组自行选择 1 位字母加两位数字的组合码来表示。首先需区分是否采标，对于采标的数据元，只需核对采标正确性，而对于未采标部分的数据元标识符，需要规划和统一编码。用 MID（B2，9，1）筛出数据元标识符顺序码段的首位字符；用 IF（ISNONTTEXT（J2），1，2）判断是否采标，以及筛出来的字符是否为数字，如果是赋值为 1（采标），反之则赋值为 2（未采标）。在 11 915 条数据元中筛选出采标数据元 5 562 条，未采标数据元 6 353 条，采标和未采标均包含数据

集内部和标准间重复的数据元。

3.2 合并重复数据元

对未采标的 6 353 条数据元去重，以便于对数据元标识符编码，这是保障数据元唯一性的前提。数据元名称排序后，首先对名称完全一致的数据元进行去重，将相同数据元名称的多条数据元合并为一条数据元和多个溯源编码形式。由于本次需要规范化的数据元有的来源于数据集标准中，数据元被赋予特定语境，名称不完全一致，但表示的意思相同，这时需要对这类数据元进行抽象化提炼为一个数据元，再进行去重合并，保障数据元的唯一性，见表 1。数据元去重合并结果，见表 2。

表 1 数据元去重合并整理示例

初始数据元名称	去重整理后数据元名称
W 期间类型代码	期间类型代码
期间代码	同上
预算执行期间代码	同上
绩效期间	同上
科目性质	科目性质类别代码
科目性质类别代码	同上
养老保险缴纳金额	养老保险缴纳金额
基本养老保险缴纳金额	同上
物资用途	物资用途
物资使用用途	同上
工资实发额	工资实发额
职员工资实发额	同上
报销人银行名称	银行名称
代发工资银行名称	同上
付款人银行名称	同上
收款人银行名称	同上
报销人银行账号	银行账号
代发工资银行账号	同上
付款人银行账号	同上
收款人银行账号	同上

表2 数据元去重合并结果

数据元目录（小类）	所属分类 表示	数据元数量（个）	
		去重 之前	去重 之后
第1部分：总则	—	—	—
第2部分：标识	2	163	58
第3部分：人口学及社会经济学特征	3	47	18
第4部分：健康史	4	14	12
第5部分：健康危险因素	5	19	14
第6部分：主诉与症状	6	362	174
第7部分：体格检查	7	133	84
第8部分：临床辅助检查	8	35	36
第9部分：实验室检查	9	29	27
第10部分：医学诊断	10	122	72
第11部分：医学评估	11	493	308
第12部分：计划与干预	12	820	494
第13部分：卫生费用	13	234	136
第14部分：卫生机构	14	521	224
第15部分：卫生人员	15	548	217
第16部分：药品、设备与材料	16	1 068	479
第17部分：卫生管理	17	1 745	858
合计	—	6 353	3 211

3.3 确定数据元所属分类

卫生数据元的分类和数据元标识符的编码规则存在一定规律,见表3。根据数据元标识符的前8位可以确定数据元对应所属的分类,然后应用Vlookup函数,在数据元所属分类列输入公式=VLOOKUP(LEFT(B2,8),Sheet1!\$D\$3:\$E\$18,2,0)(sheet1为表3中“数据元标识符编码规则”列改为只要前8位区分类别),对所有数据元填充此公式,确定相应数据元所属分类。

表3 数据元分类及数据元标识符编码规则

数据元目录（小类）	数据元标 识符编码规则	所属分类 表示
第1 部分：总则	—	—
第2 部分：标识	DE01.00. ***. **	2
第3 部分：人口学及社会经济 学特征	DE02.01. ***. **	3
第4 部分：健康史	DE02.10. ***. **	4
第5 部分：健康危险因素	DE03.00. ***. **	5
第6 部分：主诉与症状	DE04.01. ***. **	6
第7 部分：体格检查	DE04.10. ***. **	7
第8 部分：临床辅助检查	DE04.30. ***. **	8

续表 3

第 9 部分: 实验室检查	DE04. 50.	***, **	9
第 10 部分: 医学诊断	DE05. 01.	***, **	10
第 11 部分: 医学评估	DE05. 10.	***, **	11
第 12 部分: 计划与干预	DE06. 00.	***, **	12
第 13 部分: 卫生费用	DE07. 00.	***, **	13
第 14 部分: 卫生机构	DE08. 10.	***, **	14
第 15 部分: 卫生人员	DE08. 30.	***, **	15
第 16 部分: 药品、设备与材料	DE08. 50.	***, **	16
第 17 部分: 卫生管理	DE09. 00.	***, **	17

3.4 编码数据元标识符

数据元标识符编码规则,见图3。在 EXCEL 中将每类数据元顺序码段从 A01 开始顺序编码(字母大写,字母 I 和 O 因不易辨识不列入编码),为便于数据元顺序码段编码以及后期有可能新增数据元或者数据元分类有调整,将每类数据元单独放在一个 sheet 表中,便于定位顺序码段的具体值,如“毕业日期”、“入党日期”,“入校日期”、“学位授予日期”等均属于“第3部分:人口学及社会学特征”,根据表2可确定数据元标识符前8位均应为“DE02.01.”,将其后3位顺序码段从 A01 在 EXCEL 中顺序填充得到编码辅助列,见表4。在数据元标识符列输入公式 = E2&" .00" (E2 列为表4中的编码辅助列)补充数据元标识符的附加码。

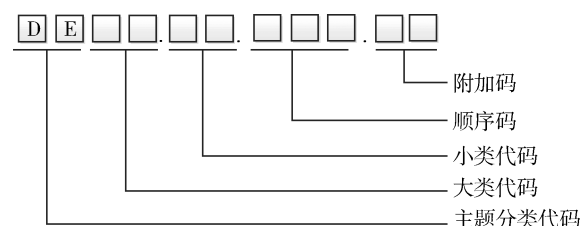


图3 数据元标识符编码规则

表 4 数据元标识符编码示例

编码辅助列	数据元标识符	数据元名称
DE02.01.A01	DE02.01.A01.00	毕业日期
DE02.01.A02	DE02.01.A02.00	入党日期
DE02.01.A03	DE02.01.A03.00	入校日期
DE02.01.A04	DE02.01.A04.00	学位授予日期

4 规范数据集内部标识符

4.1 确定数据集分类编码

数据集内部标识符由数据集分类编码（又称数据集类目编码）和数据元顺序号组成，数据集分类编码是对数据集本质内容按照卫生信息基本数据集编制规范（WS370－2012）进行分类，见图 4。将本次需进行规范化的 17 项数据集（包含子集在内

共 32 项）按照此分类编码规则分别编码数据集类目编码，见表 5。

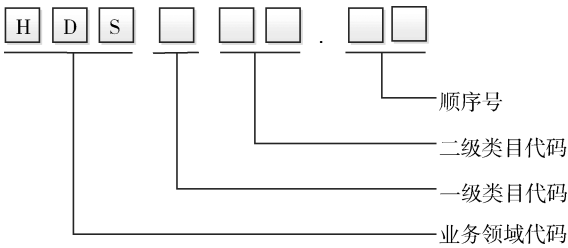


图 4 数据集分类编码

表 5 数据集分类编码

序号	数据集名称	数据集分类编码
1	中医药综合统计信息基本数据集	HDSC00. A1
2. 1	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 第 1 部分：预算精细化管理	HDSC00. A2
2. 2	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 第 2 部分：预算执行动态监控管理	HDSC00. A3
2. 3	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 第 3 部分：绩效考核管理	HDSC00. A4
3	中医医院协同办公信息基本数据集	HDSC00. A5
4. 1	中医医院资源管理信息基本数据集 第 1 部分 人力资源管理	HDSC00. A6
4. 2	中医医院资源管理信息基本数据集 第 2 部分 财力资源管理	HDSC00. A7
4. 3	中医医院资源管理信息基本数据集 第 3 部分 物力资源管理	HDSC00. A8
4. 4	中医医院资源管理信息基本数据集 第 4 部分 综合管理	HDSC00. A9
5	中医电子病历基本数据集	HDSD00. A1
6	推拿科电子病历基本数据集	HDSD00. A2
7	骨伤科电子病历基本数据集	HDSD00. A3
8	针灸科电子病历基本数据集	HDSD00. A4
9	中药煎药管理与质量控制信息基本数据集	HDSD00. A5
10. 1	中医医院护理管理信息基本数据集：第 1 部分 护理人力资源管理	HDSD00. A6
10. 2	中医医院护理管理信息基本数据集：第 2 部分 护理财力与物力资源管理	HDSD00. A7
10. 3	中医医院护理管理信息基本数据集：第 3 部分 护理业务资源管理	HDSD00. A8
11. 1	中医临床护理信息基本数据集：第 1 部分：护理评估记录	HDSD00. A9
11. 2	中医临床护理信息基本数据集：第 2 部分：护理记录	HDSD00. B1
11. 3	中医临床护理信息基本数据集：第 3 部分：ICU 护理记录	HDSD00. B2
11. 4	中医临床护理信息基本数据集：第 4 部分：NICU 护理记录	HDSD00. B3
11. 5	中医临床护理信息基本数据集：第 5 部分：急诊护理记录	HDSD00. B4
11. 6	中医临床护理信息基本数据集：第 6 部分：产科护理记录	HDSD00. B5
11. 7	中医临床护理信息基本数据集：第 7 部分：血液净化护理记录	HDSD00. B6
11. 8	中医临床护理信息基本数据集：第 8 部分：手术室护理记录	HDSD00. B7
11. 9	中医临床护理信息基本数据集：第 9 部分：消毒供应中心记录	HDSD00. B8
12	中医医院伦理审查信息基本数据集	HDSD00. B9
13	基层医疗卫生机构中医诊疗区（中医馆）健康信息平台电子病历信息基本数据集	HDSD00. C1
14	基层医疗卫生机构中医诊疗区（中医馆）健康信息平台知识库信息基本数据集	HDSD00. C2
15	基层医疗卫生机构中医诊疗区（中医馆）健康信息平台基本医疗信息系统基本数据集	HDSD00. C3
16	基层医疗卫生机构中医诊疗区（中医馆）健康信息平台治未病信息基本数据集	HDSD00. C4
17	基层医疗卫生机构中医诊疗区（中医馆）健康信息平台远程会诊信息基本数据集	HDSD00. C5

4.2 编码数据集内部标识符

每个数据集标准中内部标识符自成体系，同一个数据元在同一个数据集标准中的内部标识符相同

且唯一^[5]，见图 5。数据集分类编码后 3 位数字的数据元顺序号代表数据元在某特定数据集的序号，每个数据集集中的数据元均从 001 开始顺序编码。

（下转第 48 页）

结和凝练,信息化人员常常忙于医院信息系统和硬件维护、网络安全保障、数据处理分析等日常工作,读文献、做科研的时间少,缺乏科研热情,医院对信息化人员科研能力重视不够,仅作为技术支撑部门,保障医院信息系统正常运行即可。对此三级中医医院要发挥在中医药领域的领头军作用,重视培养信息化人员科研创新能力,关注“互联网+医疗健康”、“互联网+中医药健康服务”研究动态,加强医院信息化建设与发展研究。医院科研部门要带动信息管理部门人员积极投身科研,参与管理、临床等课题研究,支持申报各级各类信息化科研课题,及时转化医院信息化建设成功案例和成果,鼓励撰写和发表高质量论文。

5 结语

三级中医医院是中医医疗服务的龙头单位,具备信息化建设较好资源和优势。随着中医药传承创新发展的推进、信息技术的日益发展,中医医院信息化建设也将面临更多困难和挑战。信息管理部门

作为医院信息化建设的执行者,需明确自身定位和职责,强化信息化人才队伍建设,注重知识培训,提升专业技能,营造环境促进科研创新,为智慧中医医院建设发挥重要的基础支撑和人才作用。

参考文献

- 1 任浩, 马兰, 刘燕燕. “互联网+医疗”背景下枢纽型信息部门的建设 [J]. 医学信息学杂志, 2019, 40 (12): 37-40, 57.
- 2 国家卫生健康委统计信息中心. 全民健康信息化调查报告 [M]. 北京: 人民卫生出版社, 2019.
- 3 张爱超, 陈庆锬, 雷行云, 等. 我国三级医院卫生信息人力资源建设情况研究 [J]. 中国数字医学, 2019, 14 (4): 15-18, 25.
- 4 国家中医药管理局. 中医药信息化发展“十三五”规划 [EB/OL]. [2019-12-02]. <http://www.satcm.gov.cn/guicaisi/gongzuodongtai/2018-03-24/2144.html>.
- 5 肖勇, 沈绍武, 田双桂, 等. 我国基层中医医院信息化人才队伍建设的思考 [J]. 中国数字医学, 2013, 8 (12): 50-53.

(上接第 42 页)

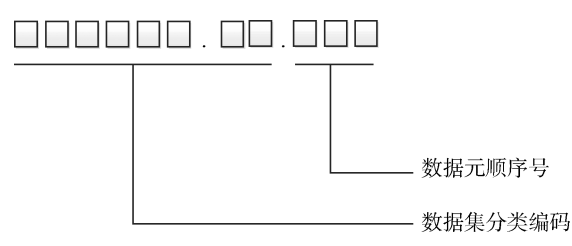


图 5 数据集内部标识符

5 结语

中医药数据元、数据集标准规范化是一个严谨的过程,数据量大、易出错,每个步骤都很关键,但同时此类标准具有很强的规律性,EXCEL 2010 作为常用办公软件,功能强大,在数据处理上操作便捷,将 EXCEL 应用于数据元、数据集标准规范化工作中能够提高工作效率,降低出错率,为后续

开展类似研究提供参考。

参考文献

- 1 舒亚玲, 沈绍武, 肖勇, 等. 我国中医药信息标准化建设现状及其思考 [J]. 医学信息学杂志, 2018, 39 (7): 46-49, 65.
- 2 张艺然, 朱佳卿. 中医药信息标准制定的组织管理研究 [D]. 北京: 中国中医科学院, 2018.
- 3 张盼, 毛树松, 邓文萍. 中医药信息标准术语规范化研究 [J]. 医学信息学杂志, 2018, 39 (5): 12-15.
- 4 徐倩, 熊振芳, 毛树松. 中医药领域信息数据元标准化与规范化工作研究 [J]. 医学信息学杂志, 2017, 38 (9): 56-60.
- 5 舒亚玲, 沈绍武, 赵移珍, 等. 省级全民健康信息平台数据标准研究 [J]. 医学信息学杂志, 2018, 39 (10): 55-59.