

医学数据资源共享中数据元著录的标准化方法研究

胡凯¹, 刘丽华^{2a}, 徐勇勇¹, 尹岭^{2b}

[摘要] 目的 研究对数据元进行著录的标准化方法。方法 首先建立数据元结构术语条目库,然后选择结构术语条目组成完整的数据元,最后进行数据元属性著录。结果 根据数据元的定义、构成以及国家标准设计了一种数据元标准化著录方法。结论 对数据元进行标准化著录是医学数据资源共享的重要工作,需要按照相关标准选择合理的方法进行。

[关键词] 数据元;标准化;著录

Standardization of Data Element in Medical Data Resource Share HU Kai, LIU Li-hua, XU Yong-yong, et al. The Department of Health Statistics, the Fourth Military Medical University, Xi'an 710032, Shaanxi, China

[Abstract] **Objective** To study a standardization to establish data element in medical data resource share. **Methods** The library of data elements structural glossary items was established first, then whole data element was established by selecting structural glossary items, finally, the input of data element's character was executed. **Results** A standardization method to establish data element was designed based on definition, structure and national standard of data element. **Conclusion** It's a important work to share medical data resource that establish data element by standardization, and it must be followed correlative standard and select logical method.

[Key words] data element; standardization; share

中图分类号: R195.1 文献标识码: A 文章编号: 1006-9771(2006)04-0365-03

[本文著录格式] 胡凯,刘丽华,徐勇勇,等. 医学数据资源共享中数据元著录的标准化方法研究[J]. 中国康复理论与实践, 2006, 12(4): 365—367.

要对医学领域的数据资源进行共享与交流,就需要保证数

据交流与共享用户对事物与概念表示的一致性与准确性理解,这就需要对共享服务体系所涵盖数据内容的数据元(即数据的基本单位)进行标准化。数据元标准化是为共享服务范围内每一个数据的基本单元及其相关特性进行规范化说明,研究合理的方法进行数据元标准化著录是其中一项重要的工作。

1 数据元的定义与构成

1.1 数据元的定义 用一组属性描述其定义、标识、表示和允

作者单位: 1. 解放军第四军医大学卫生统计学教研室, 陕西西安市 710032; 2. 解放军总医院 a. 医疗统计科; b. 神经信息中心, 北京市 100853。作者简介: 胡凯(1976-), 男, 河南洛阳市人, 主治医师, 硕士, 主要研究方向: 医院信息资源分析利用、医学信息标准化。通讯作者: 刘丽华。

许值的数据单元,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语义的信息单元^[1]。

数据元由数据元概念和表示两部分组成。当一个表示被联合到一个数据元概念时就能够产生一个数据元。按图 1 所示,数据元和数据元概念之间存在多对一的关系,也就是一个数据元必须要对应一个数据元概念,而一个数据元概念可以对应多个数据元。换言之,多个数据元可以共享一个数据元概念。

数据元需要并且只能有一个表示,当数据元的概念模型相同而表示不同时就是两个不同的数据元。数据元中的表示是描述数据元中数据元概念的特性的。也就是说,在数据元中,特性都具有且只有一个表示。

在数据元概念中,对象类和特性之间是一对一的关系,一个对象类需要且只需要一个特性(或者特性类),一个特性只描述一个对象类,当一个特性和一个对象类建立关联时就产生了一个数据元概念^[1]。

1.2 数据元的构成 数据元的组成结构如图 1 所示,由对象类词术语、特性术语、表示术语 3 个部分组成。要形成数据元,就要首先对这些术语进行录入。

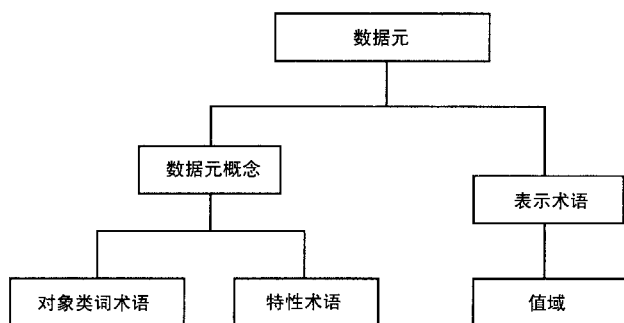


图 1 数据元的结构术语构成

2 数据元标准化著录的步骤

数据元标准化著录分为以下步骤:①数据元结构术语条目的著录:著录的内容将成为数据元构建形成的备选术语条目;②数据元的构建形成:通过选择结构术语条目形成一个完整数据元;③数据元属性著录:对已经形成的完整数据元进行数据元各个属性的著录。

3 数据元结构术语条目著录

分为对象类词术语条目著录、特性术语条目著录以及表示术语条目著录。

3.1 对象类词术语条目的著录 著录的内容作为数据元构成时对象类词术语的备选项。

3.1.1 对象类词分类的层级构建 根据本体论的方法对数据元的对象类词术语条目建立一个由不同的层级构成的分类体系。所有的对象类词术语条目根据各自的层级与分类构成一个分类体系的树状结构,其中的一个节点就代表一个对象类词术语,它可以有下一级的对象类词术语(见图 2)。

3.1.2 对象类词术语条目的著录与编码 确定了对象类词的层级与分类位置之后,进行对象类词术语条目的著录。要求著录的对象类词术语条目应尽可能参照相关标准。

在用户著录对象类词术语条目的同时,需要对所著录对象类词术语条目进行编码,编码规则可以自行定义,但要求以该术语条目的分类与层级进行编码。

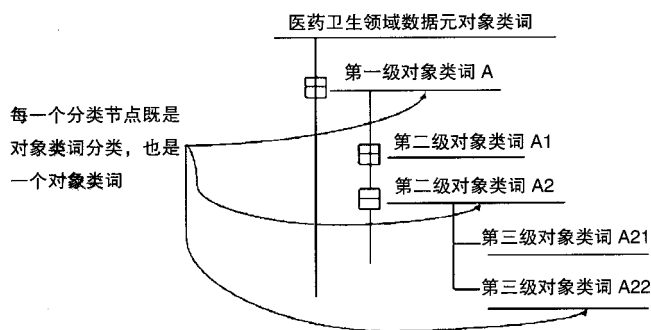


图 2 对象类词术语条目分类体系树状结构图

3.2 特性术语条目的著录 数据元特性术语条目的著录针对每一个对象类词术语条目进行。为了减少数据元特性术语条目著录的工作量以及减少数据冗余,设计了数据元特性术语条目按照对象类词分类层级自动继承的方式进行。在这种著录方式中,每一个对象类词的术语条目由以下两部分共同构成:①上级对象类词的特性术语继承:该对象类词所归属的上一级对象类词所具有的特性术语,自动继承作为该对象类词的特性术语,用户不必再进行录入;②对象类词的自身特性术语:该对象类词不同于上级对象类词,也不同于相同层级、相同分类其他对象类词的自身特性。

如图 3 所示,用户对对象类词进行特性术语著录时,自动获得上两级对象类词术语的特性术语条目,只需再追加录入 A22 本身具有的特殊特性术语。需要说明的是,如果用户对某个对象类词著录的特性术语条目与所继承的特性术语相同,则不允许进行著录。另外,如果相同层级、相同分类的所有对象类词具有一个相同的特性术语条目,那么这个特性术语条目应当作为这些对象类词上级对象类词的特性术语条目。

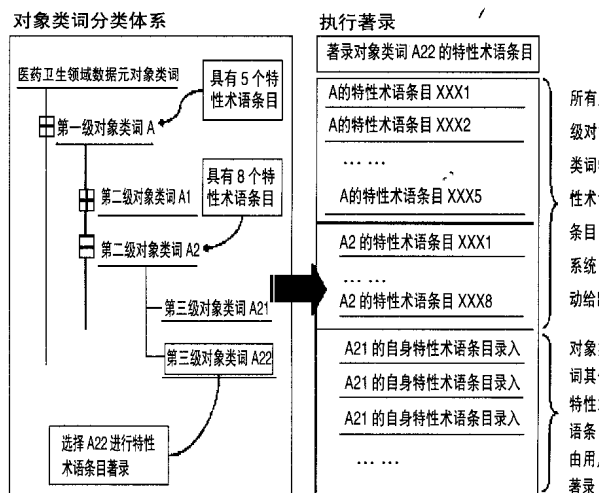


图 3 特性术语条目著录

3.3 表示术语条目著录 表示术语条目著录后,作为数据元构建形成时表示术语的备选项语。

表示术语的著录不针对具体对象类词和特性,由用户根据需要进行表示术语字典的设置与管理。目前国际上认可的表示术语如表 1 所示,著录时可以根据医学数据资源共享与交流的需要进行添加。

表 1 国际范围内认可的表示词的列表

表示词	含义
金额	货币单位的数量,通常与货币类型有关。
日期	特定的年、月、日,格式参照 GB/ T 7408。
日期时间	特定的年、月、日中的特定时间点,格式参照 GB/ T 7408。
代码	表示一组值中的一个值的字符串(字母、数字、符号)。
描述	表示一个人、客体、地点、事件或概念的一系列句子,既可用于定义(通常用一两个句子),也可用于较长文本,在数据元的中文名称中通常使用“说明”、“备注”、“意见”等词。
名称	为一个人、客体、地点、事件或概念指定的一个词或短语,该词或短语是该人、客体、地点、事件或概念的称谓。
号码	一个特定的值的数字表示,通常暗示顺序或一系列中的一个。
百分比	具有相同计量单位的两个值之间的百分数形式的比率。
量	非货币单位数量,通常与计量单位有关。
比率	一个计量的量或金额与另一个计量的量或金额的比。
指示符	两个且只有两个表明条件的值,如 On/ Off、True/ False,又称标志。

4 数据元构建形成

数据元的构建包括数据元概念的形成与表示术语的选择,需要通过选择数据元术语条目构成一个数据元^[2]。

4.1 对象类词选择 在数据元对象类词分类体系中选择对象类词术语条目。

4.2 数据元概念形成与选定 对象类词术语一旦选定,则该对象类词术语与它的所有特性术语条目组合形成的数据元概念将自动出现,用户可以选择某条数据元概念准备后续的表示术语条目著录。

4.3 表示术语选择 针对用户选定的数据元概念,从表示术语条目中选择。注意:表示术语如果与特性术语重复,则无须录

入。

5 数据元属性著录

对已经形成的数据元执行属性著录,属性包括基本属性和自定义属性。

5.1 数据元基本属性著录 数据元基本属性按照国标 GB/ T 18391 规定,如表 2 所示,约束条件为 M 的表示“必选”;约束条件为 C 的表示“条件必选”;约束条件为 O 的表示“可选”^[3]。

5.2 数据元自定义属性添加与著录 如果数据元的基本属性不能满足用户要求,则需要添加数据元的自定义属性。数据元的自定义属性添加要求按照国标 GB/ T 18391 的规定,进行属性描述符的著录,属性描述符如表 3 所示。数据元的自定义属性添加后,进行数据元属性著录时也与基本属性一同进行著录。

表 2 数据元属性输入表格

属性种类	数据元属性名称	约束	属性种类	数据元属性名称	约束
标识类	名称	M	表示类	表示类别	M
	标识符	C		表示形式	M
	版本	C		数据元值的数据类型	M
	注册机构	O		数据元值的最大长度	M
	同义名称	C		数据元值的最小长度	M
	相关环境	C		表示格式	C
定义类	定义	M	管理类	数据元允许值	M
关系类	分类模式	O		责任机构	O
	关键字(词)	O		注册状态	C
	相关数据参照	O		提交机构	O
	关系类型	C		备注	O

表 3 数据元属性描述符

属性描述符	约束	属性描述符	约束
名称(中文)	M	数据类型	M
名称(英文)	M	最大长度	O
定义	M	字符集	C
约束	M	语言	C
条件	C	备注	O
最多实例数	O		

[参考文献]

[1] 中华人民共和国国家质量监督检验检疫总局. GB/ T 18391 .1-2002/ ISO/ IEC 11179-1 :1999, 信息技术 数据元的规范与标准化 第 1 部分:数据元的规范与标准化框架[S]. 北京:中国标准出版社,2002 .

[2] 中华人民共和国国家质量监督检验检疫总局. GB/ T 18391 .5-2001/ ISO/ IEC 11179-5 :1995, 信息技术 数据元的规范与标准化 第 5 部分:数据元命名和标识原则[S]. 北京:中国标准出版社,2002 .

[3] 中华人民共和国国家质量监督检验检疫总局. GB/ T 18391 .3-2001/ ISO/ IEC 11179-3 :1994, 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[S]. 北京:中国标准出版社,2002 .