

doi:10.3969/j.issn. 1672-5166.2016.04.07

# HL7开发框架在国家卫生数据字典研制中的应用

杨 喆<sup>①</sup> 刘丹红<sup>①△</sup> 娄苗苗<sup>②</sup> 郝惠娟<sup>①</sup> 张黎黎<sup>③</sup> 汤学军<sup>③</sup>

文章编号: 1672-5166 (2016)04-0364-06 中图分类号: R-101; R-37 文献标志码: A

**摘 要** **目的** 应用HL7标准开发框架(HDF)的方法学,构建国家卫生数据字典。**方法** 采取自底向上的方法构建国家卫生信息概念数据模型;在模型的基础上,采用自顶向下的方法定义类的属性,赋予属性的数据类型和值域,形成规范的数据元及其定义。**结果** 国家卫生信息概念数据模型包含6个大类和若干子类;每个类有若干描述属性,属性的HL7数据类型及所包含的元素形成规范数据元,其语义和表示来源于模型及属性的值域。**结论** 应用HL7模型驱动的标准开发框架研制数据字典,数据元描述更加清晰,可为多来源数据提供统一的元数据参照。通过特化数据类型制定必需的词汇表等工作,数据字典将在数据共享和系统互操作方面发挥重要作用。

**关键词** 数据字典 数据模型 数据元 数据类型 HL7

## Application of HL7 Development Framework to National Health Data Dictionary Development

Yang Zhe, Liu Danhong, Lou Miaomiao, Hao Huijuan, Zhang Lili, Tang Xuejun

Institute for Health Informatics, Fourth Military Medical University, Xi'an 710032, Shannxi, China

**Abstract** **Objective** To build a national health data dictionary based on the HL7 development framework. **Methods** The health conceptual data model was built by bottom-up approach, while the attributes and data types for the classes of the model were defined by top-down approach. Vocabularies were set and finally standardized data elements established. **Results** Conceptual data model contains six classes and several sub-classes; each has a few descriptive attributes. The components of each data type assigned to each attribute are the fundamental of data elements. The semantics and format of data elements come from the model and domain values. **Conclusions** Using HL7 model-driven development framework in developing data dictionary, data element definitions is traceable and able to provide a common metadata reference for multi-source data. Data dictionary will play an important role in data sharing and system interoperability through specifying data types and developing necessary vocabulary domain values.

**Key words** Data dictionary, Data model, Data element, Data type, HL7

基金项目: 国家卫生计生委统计信息中心 卫生信息标准开发与维护项目(项目编号: 20150301)

① 第四军医大学卫生信息研究所, 西安市, 710032

② 第四军医大学西京医院数字中心, 西安市, 710032

③ 国家卫生计生委统计信息中心, 北京市, 100044

作者简介: 杨喆(1983), 女, 硕士学位, 讲师; 研究方向: 卫生信息标准化; E-mail: tanger\_003@163.com

通讯作者: 刘丹红(1964), 女, 博士学位, 教授; 研究方向: 卫生信息标准与信息建模; E-mail: liudanh@fmmu.edu.cn

△ 通讯作者

## 1 背景

为了从多方面应对互操作性挑战, HL7 制定了一系列标准和规范, 包括信息模型、数据类型和词汇; 消息; 临床文档; 语境管理 (context management) 标准; 应用技术、框架和符合性规范。尽管 HL7 标准多种多样, 但贯穿开发过程的主线是模型驱动的方法 (model-driven methodology), 即从一组通用参考模型中导出标准、规范及其他中间产品。模型驱动的标准开发方法学规范 (HDF methodology specification) 是 HL7 开发框架 (HL7 development framework, HDF) 项目的重要产出之一, 用于 HL7 消息及其它所有技术规范的开发<sup>[1, 2]</sup>。在整个 HL7 开发框架中, 参考信息模型 (reference information model, RIM) 发挥关键作用<sup>[3]</sup>。RIM 定义与医疗活动有关的一系列类和子类、类的属性、属性的数据类型和值域, 为描述对象提供明确、通用的语境和语义, 从而使基于 RIM 开发的所有标准和规范具备协调一致性。

国家卫生数据字典的主要作用是为多来源的卫生数据提供共同的标准参照, 促进数据之间的一致性和可比性<sup>[4]</sup>。目前卫生行业各专业领域定义和使用的数据元数量庞大, 相互之间缺乏对照和协同, 数据元的含义和允许值存在较大变异<sup>[5]</sup>。因此, 借鉴 HDF 的理念和方法学, 通过数据模型表达数据的语境和关联关系<sup>[6]</sup>, 并在模型的框架下规范数据的语义和形式, 是值得探索的

国家卫生数据字典开发途径。

## 2 方法

(1) 采用自底向上的方法构建卫生信息概念数据模型, 并与 HL7 RIM 的类及类的属性进行对照。

(2) 属性的表达采用 HL7 数据类型, 并对其进行特化 (specify)。

(3) 以属性的数据类型为基础, 按照数据类型包含的语义成分展开, 每个语义成分对应一个数据元。对所有属性及其数据类型进行归纳整理, 形成规范的数据元。

(4) 对直接来源于 HL7 RIM 的属性, 采用了 HL7 RIM 词汇域定义属性值 (值域); 从我国电子健康记录数据元抽象而来的属性, 采用我国已有的代码表定义其值域。

通过上述过程, 规范数据元的语义和表示类型, 构成国家卫生数据字典的基本内容。

## 3 结果

### 3.1 概念数据模型

以国家卫生信息标准专业委员会制定的与医疗活动有关的 59 个数据集为研究对象<sup>[7]</sup>, 采用自底向上的方法建立卫生数据概念模型。模型共包括 6 个大类和若干子类, 每个类具有一组描述属性见图 1。

对象类及其属性组成数据元概念, 是构建标准化数据元的基础。

将上述概念数据模型与 HL7 RIM 的类和属性进行对照。概念数据模型的类及子类均可与 RIM 对应, 但是类的属性设置稍有不同,

例如“Entity”的属性行政区划代码, RIM 无此属性, 同样 RIM “Entity”的属性 RiskCode 本模型不涉及。“实体”及其属性与 RIM 对照不一致结果见表 1。

### 3.2 数据类型

HL7 数据类型包括三大类: 基础数据类型 (foundation), 基本数据类型 (basic) 和数量数据类型 (quantities)<sup>[8, 9]</sup>。本模型全部采用 HL7 数据类型, 注重数据元的语义表达, 按照抽象层次就高不就低的原则选择使用, 目的是便于定义数据元的形式表达。选用的数据类型有 II、ED、BL、INT、PQ、Real、MO、URL、ST、TS、AD、EN、CS、CV、CE 和 CD, 共 16 类。将 HL7 数据类型包含的元素展开, 以此为基础构建数据元。每个 HL7 数据类型包含的元素比较多, 根据我国目前卫生业务领域数据标准化的实际需求, 只选取了其中部分主要元素来定义数据元。数据元的表示类型与 HL7 数据类型之间存在承接关系, 即表示类型是数据类型中某个组成成分或其具体形式。目前按照 ISO/IEC 11179 中指定的表示类型划分为 5 大类<sup>[10]</sup>: 文本、符号、数值、日期时间和代码。表 2 是部分数据类型包含元素的展开结果及对应的表示类型。

对象类、属性和数据类型是数据元的基本组成成分。

### 3.3 规范化数据元

依据概念数据模型的对象类、

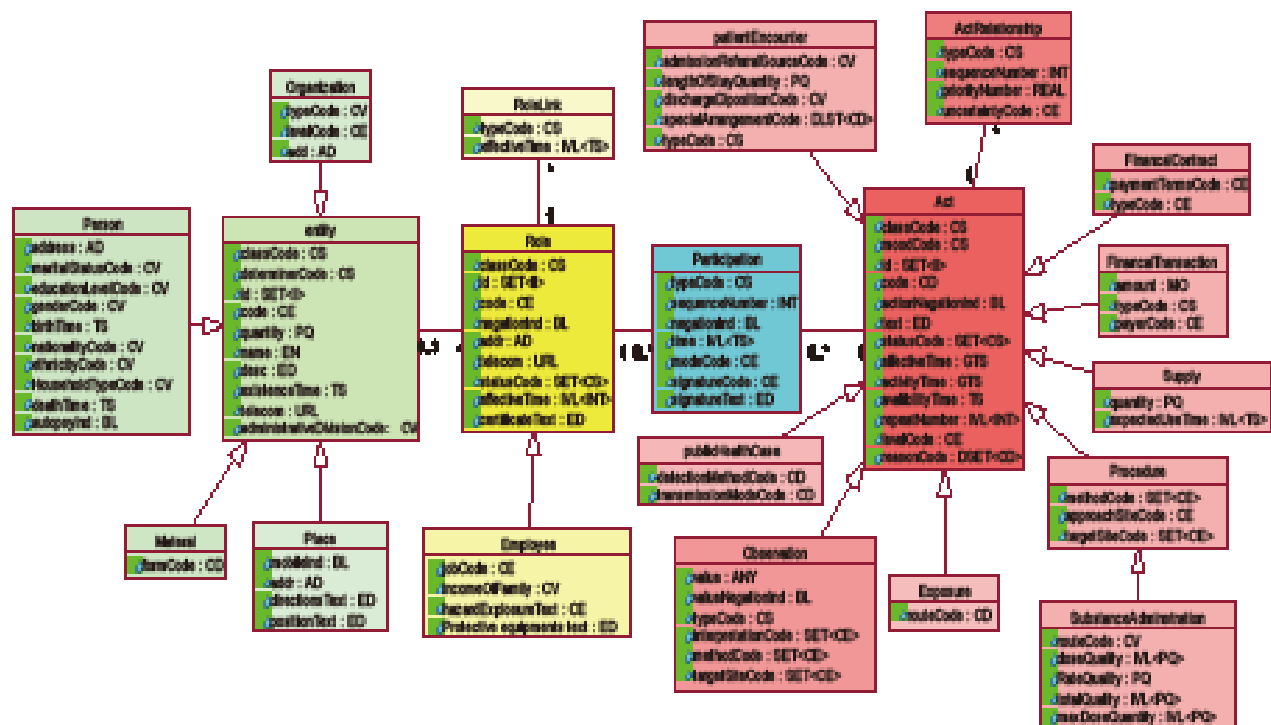


图1 卫生数据概念模型

表1 概念数据模型与 RIM 对照（实体类）

| 概念数据模型        |                            | RIM            |               |
|---------------|----------------------------|----------------|---------------|
| 类             | 属性                         | 属性             | 类             |
| Entity        | administrativeDivisionCode | —              | Entity        |
| Entity        | —                          | StatusCode     | Entity        |
| Entity        | —                          | RiskCode       | Entity        |
| Entity        | —                          | handlingCode   | Entity        |
| Entity—Place  | —                          | gpsText        | Entity—Place  |
| Entity—Person | nationalityCode            | —              | Entity—Person |
| Entity—Person | ethnicityCode              | raceCode       | Entity—Person |
| Entity—Person | householdTypeCode          | —              | Entity—Person |
| Entity—Person | deathTime                  | deceasedInd    | Entity—Person |
| Entity—Person | autopsyInd                 | organDonorInd  | Entity—Person |
| Entity—Person | —                          | disabilityCode | Entity—Person |

## General Health Information Management 综合卫生信息管理

续表

| 概念数据模型              |           | RIM                       |                     |
|---------------------|-----------|---------------------------|---------------------|
| 类                   | 属性        | 属性                        | 类                   |
| Entity—Person       | —         | mivingArrangementCode     | Entity—Person       |
| Entity—Person       | —         | ehtnicGroupCode           | Entity—Person       |
| Entity—Person       | —         | religiousAffiliationCode  | Entity—Person       |
| Entity—organization | typeCode  | standardIndustryClassCode | Entity—organization |
| Entity—organization | levelCode | —                         | Entity—organization |

表2 HL7 数据类型包含元素

| 数据类型 | 语义成分           | 表示类型         | 数据类型 | 语义成分              | 表示类型 |
|------|----------------|--------------|------|-------------------|------|
| EN   | Formatted      | 文本           | CS   | code              | 代码   |
|      | Use Code       | 代码           | CV   | code              | 代码   |
| AD   | Use Code       | 代码           |      | codeSystemUID/OID | 符号   |
|      | UseablePeriod  | 时间段          |      | codeSystemName    | 文本   |
|      | Formatted      | 文本           | CE   | Code              | 代码   |
| URL  | Address        | 文本           |      | displayName       | 文本   |
|      | Scheme         | 代码           |      | CodeSystemUID/OID | 符号   |
| II   | root UID/OID   | 符号           |      | CodeSystemName    | 文本   |
|      | Extension      | 文本           |      | Translation       | 代码   |
|      | identifierName | 文本           |      | Value             | 数值   |
| ED   | mediaType      | 代码           | PQ   |                   |      |
|      | thumbnail      | 文本、PDF、图片或影像 |      | Unit              | 代码   |

对象类的属性、属性对应的数据类型及其对包含元素形成了规范数据元。HL7 代码型属性均有对应的词表,本模型也定义了代码型属性的取值范围,对应 HL7 词汇域 (vocabulary domain)。这些

属性的数据类型包括:CS (模型自定义)、CD (概念描述符)、CE (可选的代码体系及代码)。对直接来源于 HL7 RIM 的属性,采用了 HL7 RIM 对应的词汇域定义属性值 (值域)。从我国电子健康记录数据

元抽象而来的属性,采用了我国已有的代码表定义其值域。这些代码表主要来源于 WS 364-2011<sup>[11]</sup>。

表3是对象类“实体”形成的数据元。例如实体类代码,对应的属性 classCode,数据类型为CS,



表3 实体类的规范化数据元

| 数据元         | 数据类型     | 表示类型     | 定义  |
|-------------|----------|----------|---|
| 实体类代码       | CS       | 代码       | 某个实体的实例所属的实体类别, 参照 HL7 vocabulary domain values   |
| 实体限定码       | CS       | 代码       | 规定实体对象指宽泛的(类)还是特定的(实例)。多数情况下, 一个实体可能表示有关某个特定实例的信息, 有时也可能表示一般意义上的一大类实体信息                                     |
| 实体标识符根      | Set< II> | 符号       | 任何实体均可用具备一定内部规则的标识符进行标识, 以实现该实体在一定范围内的唯一可标识性。为了保证标识符对实体标识的准确性, 必须为每个产生标识符的系统设定根标识, 即标识符根。例如, 居民的身份证号码系统标识符等 |
| 实体标识符       | Set< II> | 符号       | 实体特有的唯一标识符。实体实例的标识符是一一对应的, 不表示类别, 即实体标识符针对实体的每个实例, 而不是某类实体。例如, 居民身份证号 101110200001010011                    |
| 实体标识符名称     | Set< II> | 文本       | 除了标识符号(机器识别)外, 有时还需要为标识符赋予名称, 便于人类阅读和识别。例如中华人民共和国居民身份证号码等   |
| 实体代码        | CE       | 代码       | 实体的某个实例所属的特定类别。对于所有实体, 其代码来源于若干个代码系统中的一个代码系统, 该系统由实体类代码 Entity.classCode 提出, 比如生命体(动植物等)、化学物质、机构等           |
| 实体数量值       | PQ       | 数值       | 规定某个实体对象所代表的客观事物的物理数量, 可以是一组事物的计数, 也可以是其他物理数量   |
| 实体数量的测量单位   | PQ       | 数值       | 用于计量实体数量的单位。计量单位的类型可用代码表示   |
| 实体的名称       | COLL<EN> | 文本       | 实体的非唯一的文本标识符或名称。例如, 人的姓名、医疗机构名称等  |
| 实体的名称类型代码   | COLL<EN> | 文本       | 实体的某个名称所属的类别在 HL7 实体名称使用 (EntityNameUse) 代码表中的编码  |
| 实体描述        | ED       | 文本或其他多媒体 | 采用文本或者多媒体形式对实体的描述   |
| 实体存在时间      | IVL<TS>  | 时间段      | 实体以物理形式存在的时间区间。说明实体的有效存在时间对正确理解该实体的相关信息有意义  |
| 实体通信地址      | URL      | 文本       | 由 Internet standard RFC 2396 规定的实体的电信地址 URL, 即通用资源定位。常用的 URL 地址类型有电话号码、传真号码、电子邮件地址、超文本参照、FTP 等              |
| 实体通信地址类别代码  | URL      | 代码       | 表示实体的通信方式的类别代码。该代码由 Concept Domain URLScheme 定义, 其中包括 ftp、http、tel 等  |
| 实体通信地址可用时间段 | URL      | 时间段      | 规定一个时间段, 在此期间该通信地址可以使用。   |
| 实体行政区划代码    | CV       | 代码       | 中华人民共和国县级及县级以上行政区划在特定编码体系中的代码, 见 GB/T 2260-2007   |

## General Health Information Management 综合卫生信息管理

表示类型为代码,通过定义来指向对应的词汇表(HL7 vocabulary domain values.EntityClass)。对于代码型数据类型,提炼出了两个通用的数据元,即代码系统标识符和代码系统名称。

#### 4 结论与讨论

本研究应用HL7模型驱动的标准开发框架,基于现有卫生领域数据集中的数据元,采取自底向上的方法构建了国家卫生信息概念数据模型,包含6个大类和若干子类;在模型的基础上,采用自顶向下的方法定义类的属性,属性的HL7数据类型及所包含的元素形成规范数据元,数据元语义和表示类型来源于模型及属性的值域。采用此种方法研制我国卫生数据字典形成的数据元具有清晰的定义和语境,可为多来源数据提供统一的元数据参照,使数据元定义具有可溯源性。随着数据类型特化、词汇表的制定等工作的开展,数据字典将在数据共享和系统互操作方面发挥重要作用。

通过数据类型的特化,可将字典中的规范数据元特化为特定信息系统使用的数据元。数据类型的特化主要针对基本数据类型ANY,代码类数据元CV、CE及测量类数据元PQ、MO等。例如身高,其数

据类型为PQ。在领域应用时,可以根据业务需要,规定具体的测量单位,例如厘米,其表示形式进一步明确,数据元就具体化为一个具有一定取值范围的数值。除了ANY数据类型,其他比较通用和抽象的数据类型,如CD、ED等有时也需要特化,例如将CD特化为代码,将ED特化为文本或其他多媒体形式等。另外,数据元在应用时还可通过添加本领域、本业务场景中特有的约束或限定来实现数据元的重用。约束或限定要遵循统一的规则,促进信息交换和共享时语义互操作的实现。规范的约束主要通过制定和采用数据模型中属性对应的词汇域完成。例如,实体类通过类代码和代码而特化为一个个具体的客观事物的实例,如人、机构、组织器官、建筑物等。■

#### 参考文献

- [1] Health Level seven.HL7 Development Framework[EB/OL][2015-06-20]. <http://www.hl7.org/search/index.cfm?&criteria=HDF&startRow=11>.
- [2] 刘丹红,徐勇勇. HL7 互操作框架与语义标准需求分析[J]. 中国卫生信息管理杂志, 2014, 11(4): 376-380.
- [3] HL7 Reference Information Model. Health Level seven[EB/OL][2015-06-20]. <http://www.hl7.org/implement/standards/rim.cfm?ref=common>.

- [4] 姜苗苗,杨喆,刘丹红,等. 卫生数据标准化方法研究[J]. 中国卫生信息管理杂志, 2013, 10(5): 440-443.
- [5] 中华人民共和国卫生部.WS 363-2011 卫生信息数据元目录[S]. <http://www.moh.gov.cn/publicfiles/business/htmlfiles/zwgkzt/pwsbz/index.htm>. 2015-06-20.
- [6] 姜苗苗,杨喆,刘丹红,等. 基于领域信息的卫生信息概念数据模型构建方法[J]. 中国数字医学, 2015, 10(1): 74-77.
- [7] 国家卫生计生委.WS 375.10-2012《疾病控制基本数据集》第10部分: 传染病报告[S]. <http://www.nhfp.gov.cn/zwgkzt/s9497/201207/55525.shtml>. 2015-06-20.
- [8] Health Level seven. HL7 V3 .Data Types-Abstract Specification[EB/OL][2015-05-20]. <http://www.hl7.org/v3ballot/html/infrastructure/datatypes/datatypes.html>.
- [9] 杨喆,刘丹红,徐勇勇. 卫生信息标准化中的数据类型[J]. 中国数字医学, 2012, 7(10): 6-8,16.
- [10] International Organization for Standardization.ISO/IEC 11179, Information Technology -- Metadata registries (MDR) [S].<http://metadata-stds.org/11179/2015/06/20>.
- [11] 中华人民共和国卫生部.WS 364 卫生信息数据元值域代码[S]. <http://www.moh.gov.cn/zwgkzt/s9497/201108/52771.shtml>.2015-06-20.

[收稿日期: 2016-05-17]

修回日期: 2016-07-15]