ICS 11.020
C07
备案号:25593—2009

WS

中华人民共和国卫生行业标准

WS/T 304—2009

卫生信息数据模式描述指南

Guidlines for data schema description of health information

2009-01-22 发布　　2009-08-01 实施

中华人民共和国卫生部 发布

目　　次

前　言

本标准参考 ISO/IEC 19501《信息技术　开放式分布处理　通用建模语言(UML)(版本 1.4.2)》(2005 年英文版)的表述,结合卫生信息数据内容与特征进行编写。

本标准附录 A、附录 C、附录 D 是资料性附录,附录 B 是规范性附录。

本标准由卫生部卫生信息标准专业委员会提出。

本标准由中华人民共和国卫生部批准。

本标准负责起草单位:中国人民解放军总医院。

本标准主要起草人:刘丽华、王才有、李包罗、胡建平、胡凯、王骏、张黎黎、饶克勤。

卫生信息数据模式描述指南

1 范围

本标准规定了卫生信息主题域模式、类关系模式、数据集模式的描述规则。

本标准适用于医药卫生领域信息资源的组织与规划、卫生信息系统设计与开发以及具体数据资源描述中的数据模式描述。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件，其随后所有的修改单(不包括勘误的内容)或修订版均不适用于本标准。然而，鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件，其最新版本适用于本标准。

WS/T 305—2009 卫生信息数据集元数据规范

WS/T 306—2009 卫生信息数据集分类与编码规则

3 术语和缩略语

下列术语和缩略语适用于本标准。

3.1 术语

3.1.1 数据模式 data schema

数据的概念、组成、结构、相互关系的总称。

3.1.2 主题域 subject area

根据主题对某一个卫生信息对象进行分析、归纳，最终划分为若干个具有相关内容主题、更加容易理解的下一级主题内容区域。

3.1.3 数据集 dataset

具有一定主题，可以标识并能够被计算机化处理的数据集合。

3.2 缩略语

HL7 (health level 7)美国卫生信息传输标准

UML (unified modeling language)统一建模语言

XML (extensible markup language)可扩展标记语言

SQL (structured query language)结构化查询语言

4 数据模式描述

4.1 描述指南的编写方法

根据不同使用需求和适用环境，对医药卫生领域的数据模式进行分类，然后针对不同种类的数据模式，分别制定相应的描述规则和方法，作为该类数据模式的描述指南。针对每种数据模式，描述指南内容如下：

a) 表达样式：说明某类数据模式描述所应当包含的描述内容，以及描述结果的表达方式。

b) 描述规则：说明描述过程中具体的描述方法、步骤与约束，阐述表达式样中各种表达形式所表示的含义，以及实现方法。

c) 描述参照：提供数据模式描述结果的例子供用户作为使用参照。

4.2 数据模式种类划分

根据不同使用需求和适用环境，将医药卫生领域的数据模式划分为以下三类：

a) 主题域模式：以主题为驱动，对数据的内容、构成和关系进行描述。针对一个卫生信息对象，可以通过分析、归纳与区别，对其进行不同主题的划分；针对划分得到的主题，可以继续进行子主题、子子主题的划分，同时对于不同级别的主题内容，进行关系与结构的描述。

b) 类关系模式：以应用为驱动，通过对业务活动中所涉及的数据对象与数据内容的分析、归纳和概念性抽象，用类、类之间的关系以及类的属性等，对数据的内容与关系进行描述。

c) 数据集模式：数据集是具有一定主题，可以标识并可以被计算机处理的数据集合。通过对数据集内容的分析、归纳，用属性来表示数据元，并把一组具有共同描述主题的属性组合在一起用实体来表示，并且描述实体之间的关系。

这三类数据模式的用途、主要内容与适用情况说明见表1：

表1 三类数据模式的用途与主要内容

数据模式种类	用途	主要内容	适用情况举例
主题域模式	卫生信息资源的规划、组织、收集与管理	按主题对数据进行相同属性归并，不同属性进行区分，并用通用方式进行表达	公共卫生信息资源规划、组织、主题数据集设计与发布等
类关系模式	卫生信息系统的设计与开发	按活动对数据的类、类关系、属性、属性数据类型、值域等进行设计，并用通用方式进行表达	医院信息系统的设计与开发
数据集模式	数据集内容构成的描述	把数据集的内容划分为实体、属性，并用通用方式进行表达	医药卫生科学数据共享数据集的规范化描述与收集

以上对数据模式进行种类划分的目的，是让用户根据自己的情况选择适用的数据模式种类，然后根据自己的选择，参照该类别数据模式的描述规则与方法(分别见第5章、第6章和第7章)进行描述。本标准的种类划分，以及对三类数据模式的命名，是以保证本标准的可操作性为原则，仅适用于本标准范围内。

5 主题域模式描述

针对主题域模式的具体应用，规定了三种允许的表示方法，见表2：

表2 主题域模式允许的三种表示方法

表示方法	适用情况	描述内容
主题划分表示法	适合于表达一个卫生信息资源顶层主题域划分	描述卫生信息资源整体的顶层框架性构成
主题包含表示法	针对需要对资源进行逐级主题划分的描述需求	描述主题的逐级划分与包含关系，还可以描述同级别主题域之间的关系
主题关系表示法	针对需要表示主题之间框架性关系的描述需求	主题域的划分，以及主题域之间的框架性关系

5.1 主题划分表示法

5.1.1 表达样式

适用于表达一个信息资源整体顶层构成的描述需求，可以直观地描述资源整体由几大部分构成，但是不能进行逐级划分。对于一个卫生信息资源的整体概念域，按照主题的抽象、概括与归纳，进行主题内容划分，表达式样示例见图1：

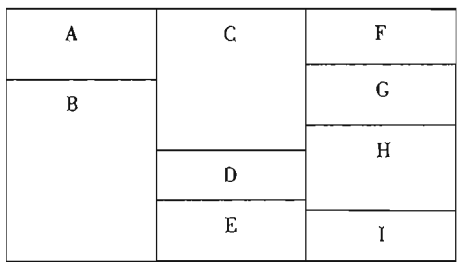

图1 主题域模式中主题划分表示法的表达样式示例

5.1.2 描述规则

主题划分表示法的描述规则如下：

a) 以一个矩形方框表示信息资源对象的整体；

b) 在矩形内部使用直线分隔的方法，得到若干个矩形区域表示构成信息资源整体的主题域；

c) 在每个区域内，标明该主题域的名称。

5.1.3 描述参照

以《2007年中国卫生统计年鉴》的内容划分为例，按照主题的划分，可以把它划分为13个大主题域，见图2：

图2 主题域模式中主题划分表示法的描述实例

5.2 主题包含表示法

5.2.1 表达样式

适用于需要对卫生信息整体进行多级主题划分的需求。可以进行主题的逐级划分，并可以根据需要描述同级别主题域之间的关系，表达样式示例见图3：

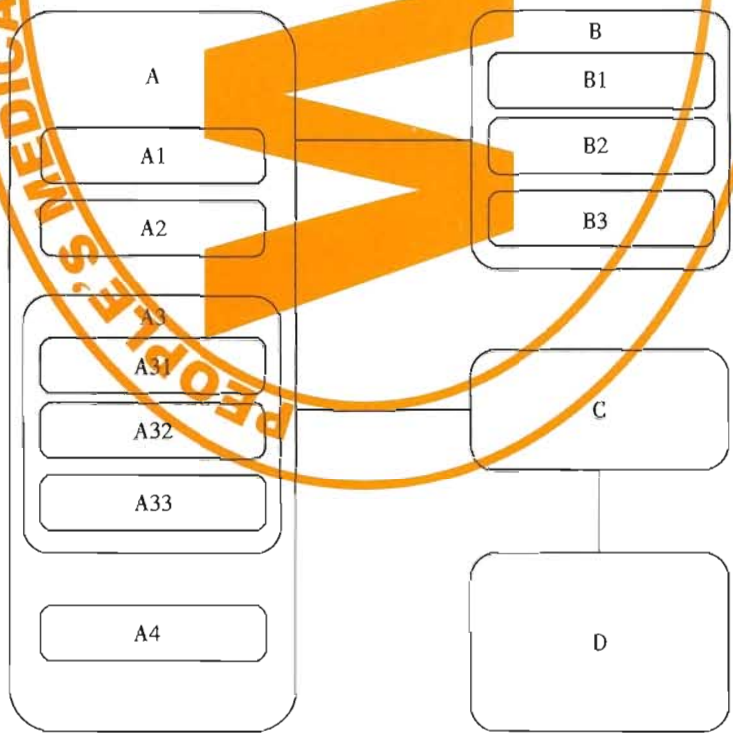

图3 主题域模式的主题包含表示法的表达样式示例

5.2.2 描述规则

主题包含表示法的描述规则如下：

a) 使用圆角方框表示按照主题划分所得到的主题域；

b) 对任何一个主题域，可以根据需要和实际的内容主题，在它的内部划分出新的主题域，外围不再有圆角方框的主题域(如图 4 中的 A、B、C、D)，是整个资源整体最顶层的主题域；

c) 每个主题域的名称应当写在所标识的主题域内部，而且不得写在下一级主题域的线框内，以确保不会被歧义理解其命名的对象；

d) 采用直线描述两个同级别主题域之间发生的相互关系，这种主题域之间的关系描述可以根据需要而定。

5.2.3 描述参照

主题包含表示法的实例：以澳大利亚国家卫生数据模型为例。该描述从整体层面描述了主题域的划分，并且分别以不同层面的包含关系描述了各个主题域内部不同等级内容之间的包含关系，参见附录 A 中的图 A.1。

5.3 主题关系表示法

5.3.1 表达样式

适用于需要表达信息资源的主要构成及其相互之间关系的需求。描述的内容包括顶层的主题域，以及它们之间关系，表达样式示例见图 4：

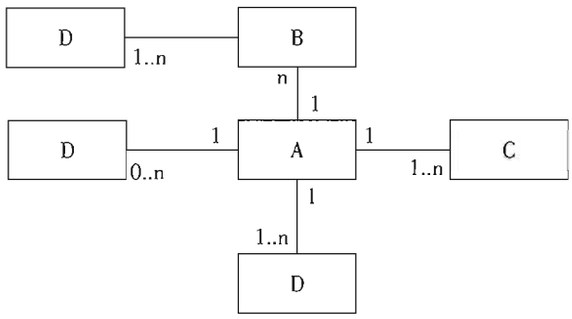

图 4 主题域模式的主题关系表示法的表达样式示例

5.3.2 描述规则

主题关系表示法的描述规则如下：

a) 使用方框表示构成信息资源的主题域，并在方框内标记主题域的名称；

b) 标记主题域之间的关系，在发生相互关系的主题域之间，使用直线连接；

c) 在表示主题域之间关系的连线两端，可以根据发生相互关系的主题域之间的数量对应关系，进行标记，标记的规则如下：

0..1	表示没有，也可以有并最多有一个
0..n	表示有，也可以没有，最多有无穷多个
1	表示必须有，且只能有一个
1..n	表示必须有，最多可以有无穷多个
n	表示必须有指定数量的

5.3.3 描述参照

主题关系表示法：以出院病人相关信息数据内容为例，说明构成这一统计信息资源收集需求的主题域模式，见图 5：

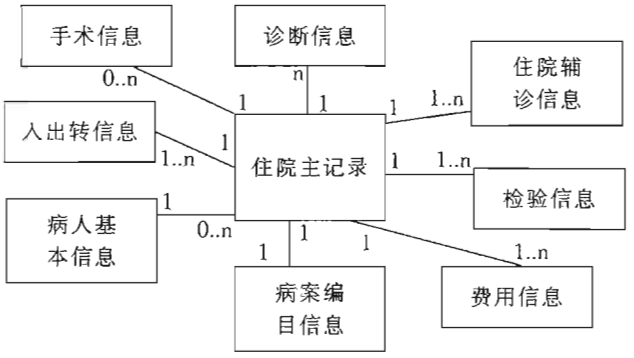

图 5　主题域模式中主题关系表示法描述实例

6　类关系模式描述指南

6.1　表达样式

类关系模式的描述方式使用 UML 类图的描述方式，结果的表达样式见图 6：

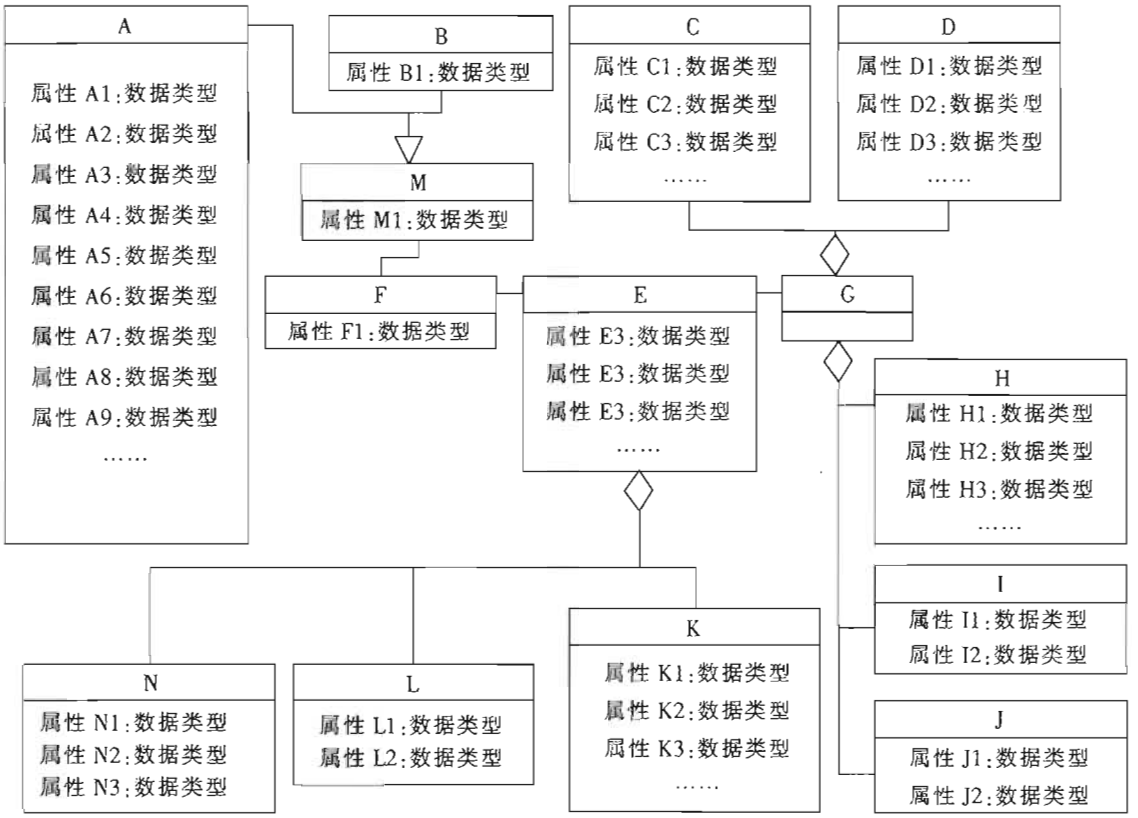

图 6　类关系模式表达式样示例

6.2　描述规则

描述规则按照 UML 类图的规则，参照 UML 类图的描述规则，主要使用到的描述规则见附录 B。

6.3　描述参照

数据关系模式描述的例子包括用于指导卫生信息应用系统设计与规划，以及具体描述系统的设计与开发的 UML 类图模型。

例如 HL7 RIM 模型，描述了实体、角色、参与、活动四个核心类及其子类的属性、关联关系，以及属

性的数据类型等信息，具体参见附录 C 中图 C.1。

例如美国公共卫生概念数据模型，描述了参与者、位置、客体、卫生相关活动四个核心类及其子类的属性、关联关系，以及属性的数据类型等信息，具体参见附录 C 中图 C.2。

7 数据集模式描述

7.1 表达样式

对数据集概念的详细解释及卫生信息数据集的阐述见 WS/T 306—2009 的相关内容。

数据集模式的描述方式包括由表示数据集整体数据模式的 UML 类图，以及分别用来描述数据集、实体与属性摘要信息的数据字典，表达样式见图 7：

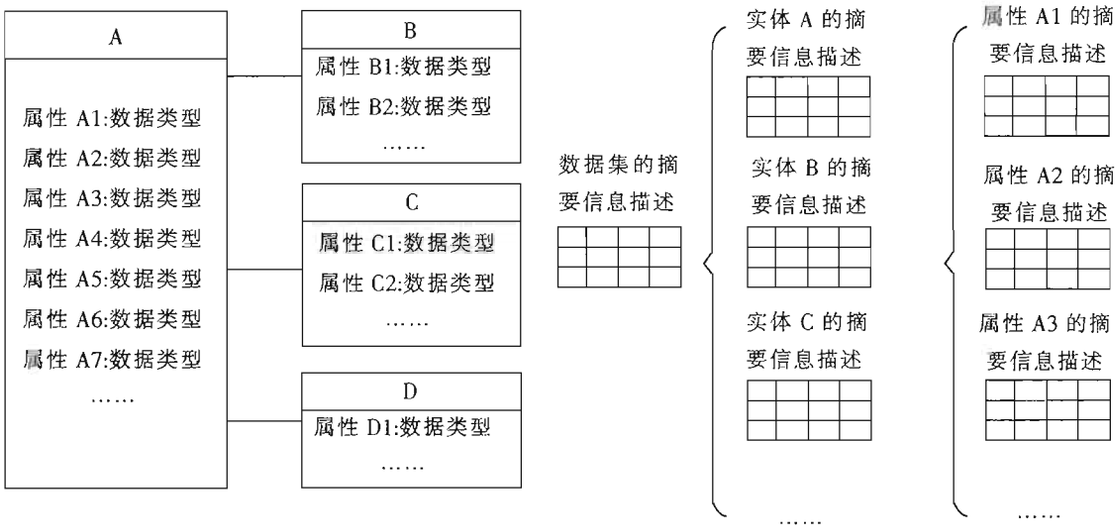

图 7 数据集模式的表达样式示例

7.2 描述规则

数据集模式具体步骤和规则如下：

a) 参考 WS/T 305—2009，选择其中关于数据集本身属性的内容描述数据集的属性信息。具体按照表 3 所列的属性来表示数据集的摘要信息。

表 3 对数据集进行摘要信息描述

描述项目		约束	属性定义
数据集名称		必选	能够简要描述卫生信息数据集主题与内容的标题
数据集标识符		必选	卫生信息数据集的唯一标识符
数据集摘要		必选	数据集内容的简单说明
数据集提交或发布方	负责单位名称	必选	提交或发布数据集，并对数据集的真实性、正确性负责的单位或部门的单位名称
	联系人姓名	可选	联系人姓名
	联系电话	可选	可以与负责人或负责单位联系的电话号码
	通讯地址	可选	能够进行邮政联系的详细地址
	邮政编码	可选	进行邮政联系的邮政编码
	电子邮件地址	可选	联系人或负责单位的电子邮件地址
关键词说明	关键词	必选	用于描述数据集主题的通用词、形式化词或短语
	词典名称	可选	正式注册的词典名，或类似的权威关键词资料名称

续表

描述项目		约束	属性定义
数据集语种		必选	数据集采用的语言
数据集特征数据元		可选	卫生信息数据集中，能够表达数据集核心内容与特征资源的数据元列举
数据集发布日期		可选	卫生信息数据集进行提交或发布的日期
数据集发布格式	发布格式名称	可选	数据集分发格式名称
	版本	可选	数据集分发格式所对应的软件版本（日期、版本号等）
在线访问地址		可选	可以对数据集进行在线访问或获取的信息
数据集分类	类别名称	可选	对应于所使用的某种分类方法所得到的具体类目名称
	类别编码	可选	类别名称对应的编码
	分类标准	可选	所依据的分类标准名称
相关环境说明		可选	说明数据集生产的处理环境，包括软件、计算机操作系统、文件名和数据量等

b) 数据集整体数据模式的 UML 类图描述：用 UML 图的方式表现数据集的实体、属性、属性的数据类型等，UML 类图描述规则见附录 B。在 UML 类图中属性的数据类型选用 HL7 RIM 的部分数据类型来表示（参见附录 D）。

c) 用数据字典描述实体：数据字典来进行实体的摘要信息描述见表 4。

表 4　对实体进行摘要描述的数据字典内容

描述项目	约束	属 性 定 义
中文名称	必选	实体的标识，一般使用名词表达，通常名称都能反映出实体的属性和特征
中文别名	可选	实体的别名，一般使用名词表达
英文名称	必选	实体的英文全称
英文短名	必选	实体的英文名称缩写
定义	必选	实体定义的详细描述
注释	可选	和实体相关的其他信息
版本标识	必选	用于实体的配置管理和控制
状态	必选	0：讨论版本；1：正式版本
来源	可选	说明实体定义的来源，来源包括已有的数据模式标准、已有的信息系统以及其他来源
安全说明	必选	说明该属性的安全限制信息，包括访问和使用限制等

d) 用数据字典分别描述属性：数据字典进行属性的摘要信息描述见表 5。属性的 RIM 数据类型选用 HL7 RIM 的部分数据类型来表示（参见附录 D）。

表 5　对属性进行摘要描述的数据字典内容

描述项目	约束	属 性 定 义
中文名称	必选	属性的标识，一般使用名词表达，通常名称都能反映出属性的属性和特征
中文别名	可选	属性的别名，一般使用名词表达
英文名称	必选	属性的英文全称
英文短名	必选	属性的英文名称缩写
定义	必选	属性定义的详细描述
RIM 数据类型	必选	属性的概念数据类型，选择 HL7 RIM 数据类型
SQL 数据类型	可选	该属性在关系型数据库中的数据类型，按照结构化查询语言的数据类型表达方式进行描述，例如 char(100)，代表可变长字符串，最大长度单位 100 个字符

续表

描述项目	约束	属 性 定 义
键	必选	是否为主键、外键等。如果是，则详细说明
注释	可选	和属性相关的其他信息
版本标识	必选	用于属性的配置管理和控制
状态	必选	0：讨论版本；1：正式版本
来源	可选	说明属性定义的来源，来源包括已有的数据模式标准、已有的信息系统、数据元以及其他来源
值域	必选	属性的取值范围
安全说明	必选	说明该属性的安全限制信息，包括访问和使用限制等

7.3 描述参照

下面以 1992 年全国肝炎流行病学数据集作为实例说明数据集模式的描述规则和步骤。

a) 数据资料示例的概况和内容

数据集的相关信息如下：

——数据集名称：1992 年中国病毒性肝炎血清流行病学调查；

——数据资源获取方式：全国 31 个省 1 岁～59 岁人群抽样调查获取；

——组织调查单位：中国 CDC；

——数据管理部门：中国 CDC 信息中心；

——数据存储形式：SAS V604 数据库；

——总例数：67 214 条记录；

——变量数：42 个；

——数据建立时间：Monday，February 13，1995 03：01：00 PM。

——数据集数据结构见表 6：

表 6 1992 年中国病毒性肝炎血清流行病学调查 SAS V604 数据集结构

序号	字段中文名称	英文名称	字段类型	字段含义	值域	主/外键
1	调查日期	Survey_Date	DATE	调查日期	具体的日期时间	
2	调查地	Survey_Place	CHAR	调查地	使用省国家地区编码	
3	监测点号	Surveillance_Place_ID	NUM	监测点号		
4	村号	Village_ID	NUM	村庄编号		
5	家庭号	Household_ID	NUM	家庭编号		
6	个人号	Individual_ID	NUM	被调查者个人号		
7	血清号	Serum_No	NUM	每个血清样本的唯一编号		主键
8	家庭应调查人口数	Household_Size	NUM	被调查者家庭应调查人口数	>0 整数	
9	家庭被调查人口数	Household _ Number _Surveyed	NUM	被调查者家庭被调查人口数	>0 整数	
10	性别	Sex	CHAR	被调查者性别	GB 性别编码	
11	出生年月	Birth_of_Date	TS	被调查者出生年月		
12	户主性质	Household_Nature	CHAR	被调查者户主性质	1,2	

续表

序号	字段中文名称	英文名称	字段类型	字段含义	值域	主/外键
13	与户主关系	Rel_to_HH_head	CHAR	被调查者与户主关系	1-6	
14	职业	Occupation	CHAR	被调查者职业	1-5	
15	民族	Nationality	CHAR	被调查者民族	1-8	
16	文化程度	Education	CHAR	被调查者文化程度	1-4	
17	既往病毒性肝炎病史患病年龄	Past_Hepatitis_Age	CHAR	被调查者既往病毒性肝炎病史患病年龄	≥0 整数	
18	现症病毒性肝炎病期长短	Current_Hepatitis_Duration	CHAR	现症病毒性肝炎病期长短	0,1	
19	流行病学调查项目	Survey_Item	CHAR	流行病学调查项目	1-9	
20	流行病学调查结果	Survey_Result	CHAR	流行病学调查结果	1-3	
21	乙肝疫苗接种史	HEB_Vaccination_History	CHAR	乙肝疫苗接种史	1-3	
22	第一针乙肝疫苗接种时间	first_HEB_Vaccination_Date	DATE	第一针乙肝疫苗接种时间	__年__月	
23	乙肝疫苗接种针数	Times_Hepatitis B_Vaccination	NUM	乙肝疫苗接种针数	1-4	
24	甲肝疫苗接种史	History_Hepatitis A_Vaccination	CHAR	甲肝疫苗接种史	1-3	
25	甲肝疫苗接种日期	Date_Hepatitis A_Inoculation	DATE	甲肝疫苗接种日期		
26	检验项目名称	Lab_test _Item	CHAR	检验项目名称	1-7	
27	检验结果	Lab_test _Result	BL	检验结果	0,1	

b) 数据集的描述

——数据集的字典描述：1992 年全国病毒性肝炎流调数据集数据字典描述见表 7：

表 7 数据集的摘要信息描述实例

描述项目		属 性 值
数据集名称		1992 年全国病毒性肝炎流调数据集
数据集标识符		中国疾病预防控制中心
数据集摘要		1992 年全国肝炎流行病学调查数据集，包括调查个体的基本信息、接种信息、被调查信息、预防接种信息和化验信息等内容
数据集提交或发布方	负责单位名称	中国疾病预防控制中心
	联系人姓名	
	联系电话	
	通讯地址	北京市宣武区南纬路 27 号
	邮政编码	100050
	电子邮件地址	
关键词说明	关键词	肝炎，流调
	词典名称	
数据集语种		
特征数据元		调查日期，调查地，血清号，流行病学调查项目

续表

描述项目		属 性 值
数据集发布日期		
数据集发布格式	发布格式名称	SAS
	版本	V60
在线访问地址		
数据集分类	类别名称	
	类别编码	
	分类标准	
相关环境说明		以血清为标本，全国31个省1岁～59岁人群抽样调查获取67 214条记录

——数据集的UML类图描述：1992年全国病毒性肝炎流调数据集的UML类图描述如图8所示：

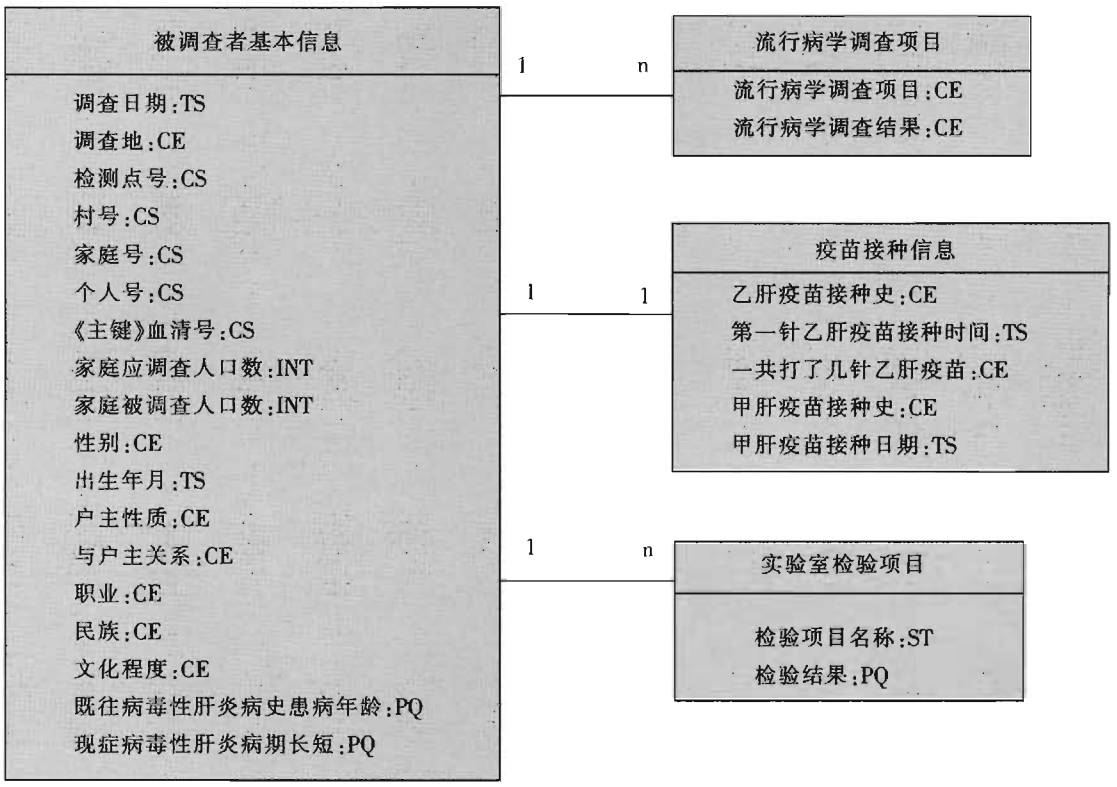

图8 数据集的UML类图描述实例

c) 实体的描述：实体使用数据字典进行摘要信息描述见表8，以“被调查者基本信息”实体为例说明。

表8 被调查者基本信息实体的摘要描述

描述项目	填 写 信 息
实体名称	被调查者基本信息实体
别名	被调查者基本信息实体
英文名称	Investigated_Subject_Info
短名	Investigated_Subject_Info
定义	被调查者个人基本信息以及与调查主题有关的属性信息

续表

描述项目	填 写 信 息
备注	包含调查时间空间信息
版本标识	0.1
状态	0
实体来源	流行病学调查

注：其他实体依次按照这种 UML 图和数据字典相结合的方法进行描述。

d）属性的描述：属性的描述采用数据字典进行描述，以“被调查者基本信息”实体的“调查日期”属性为例说明。针对第一个属性“调查日期”进行填写见表 9：

表 9　调查日期的摘要信息描述

描述项目	填写信息	描述项目	填写信息
属性名称	调查日期	状态	0
别名	调查日期	属性来源	流行病学调查
英文名称	Survey_Date	RIM 数据类型	TS
短名	Survey_Date	SQL 数据类型	DateTime
定义	调查实施的具体日期	值域	具体的日期
备注		安全说明	专业内部使用
版本标识	0.1		

注：其他属性依次按照这种方法进行描述。

附 录 A
（资料性附录）
主题域模式表示实例

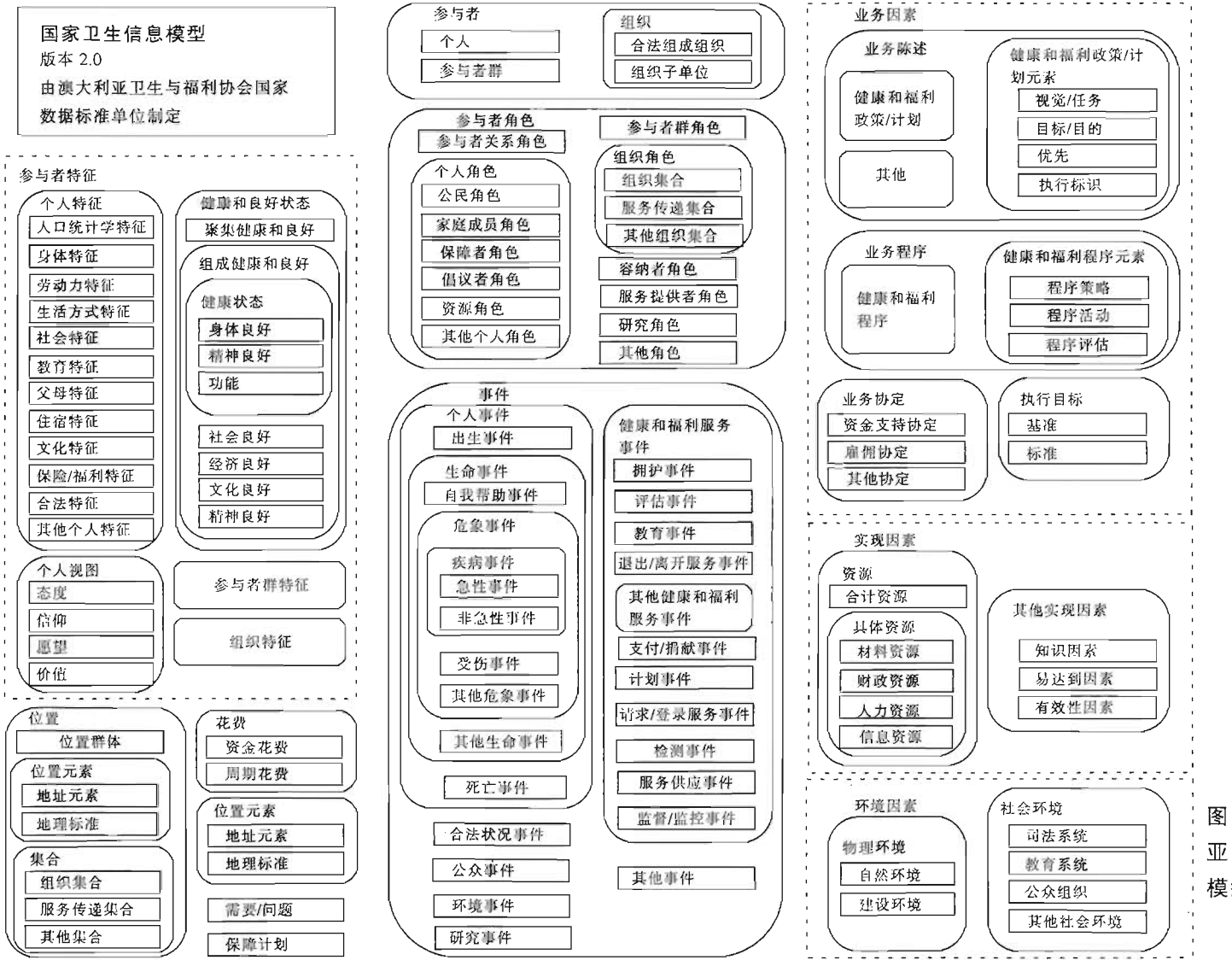

图 A.1 澳大利亚国家卫生信息模型

附 录 B
（规范性附录）
本标准用到的 UML 类图描述方法

B.1 实体

B.1.1 定义

一个实体代表一组真实的或抽象的事物（参与者、位置、客体、事件、多个事件的结合体等），它们拥有共同的属性和特征。

该组中的一个叫做该实体的一个实例。

一个实体对应于现实世界中的一个物体，或者是一个抽象的概念。

例如：

参与者：患者、医生、调查者……

位置：医院、调查地点……

客体：检查设备、实验室、药物……

事件：诊断、观察……

一个实体可以是独立确定的，也可以是存在依赖关系的。对立的实体是指可以独立确定而不需要决定与其他实体关系的实体；而依赖性的实体需要确定与其他实体间的关系。

B.1.2 图形符号

实体使用 UML 中的类（class）来表示，同时设定该类的原型为“实体”。类的图形符号是一个矩形框，其中标注出实体的名称。

在类表示符号的方框中，由两条横线将类符号划分为三个部分，三个部分要填写的内容如下：

a) 第一条横线的上方写实体的名称（如图 B.1 中的“医生”）；

b) 第一条横线与第二条横线之间的位置注明类的静态属性（属性的定义参见 B.2.1）；

c) 第二条横线下方部分要求在技术实现时注明类的操作属性，本标准不涉及具体这方面的内容，不进行标注。

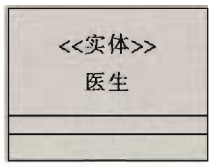

图 B.1 实体

B.2 属性

B.2.1 定义

属性代表实体的一个特征，或一个与真实或抽象事物关联的特性。通常一个实体包含多个属性，一个属性实例包括特征类型及其值。比如“调查信息”这个实体包含有“调查时间”、“调查地点”、“执行单位”等多个属性。

实体实例的相互不同是通过各个属性的取值而不同的。

属性是实体到值域的映射。在一个实体中，属性必须有唯一的名字，同样的属性名字代表同样的意义。反过来，同样含义就需要相同的属性名字，使用别名时例外。一个实体可以拥有任意数量的属性。一个属性由属性名称或其别名来标注。

B.2.2 图形符号

属性使用的属性(attribute)来表示，见图 B.2。

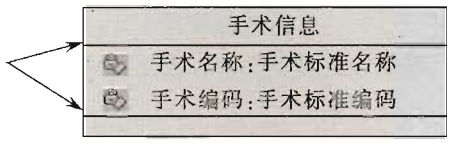

图 B.2 属性

示例中，实体“手术”包含有两个属性：手术名称、手术编码。

B.3 值域

B.3.1 定义

一个值域代表一组命名并定义好的值，属性(attribute)从中取值。值域的单独定义，为的是可以重用，即多个属性可以使用同一个值域。

值域被视为一个拥有确定数量的无限实例的类。例如：“状态码”可以被认为是一个值域，任何允许的值都必须满足它的定义，“状态码”包含三个枚举型的值，那么该值域的取值只有三种可能。另一个例子，“姓名”可以有无穷的实例，但必须由汉字和字母 a-z，A-Z 组成。

定义一个值域通常有两种方式：定义值列表和给出取值范围规则。定义值列表方式会给出所有允许取的值，一个属性取的值只有在列表中出现才是有效的。如：“性别”。给出取值范围规则方式通常会给出取值的下边界和(或)上边界。如：方位角取值必须在−360°和+360°之间。

B.3.2 图形符号

值域没有对应的 UML 图形符号。而通过设定属性的类型来实现。

B.4 主键和外键

B.4.1 定义

主键是一种特殊的属性，对该实体属性取值作出了唯一性的限制。即，所有实体实例的该属性取值不会出现重复。通过该键值可以唯一确定一个实体。

实体实例的相互不同是通过各个属性的取值而不同的。为了通过某个属性而唯一地标识出各个实体，这是采用主键。

例如：把“病种信息”实体的 ICD-10 编码定义为主键，这样，每一个病种的 ICD-10 编码互不相同，不会出现重复现象。

外键也是实体一种属性，它是与该实体相关联的另一个实体的主键。例如：在学校的图书管理系统中存在“患病人群”实体和“病种信息”两个互相关联的实体，“患病人群”实体中包含属性“人群编号”、“病种编码”等属性，“病种编码”的取值应从“病种信息”的实例中取得。“病种信息”中以“病种编码”为主键，那么“病种编码”就是“患病人群”实体的外键。

B.4.2 图形符号

在 UML 中通过设定属性的原型(stereo type)为“主键”和“外键”见图 B.3。

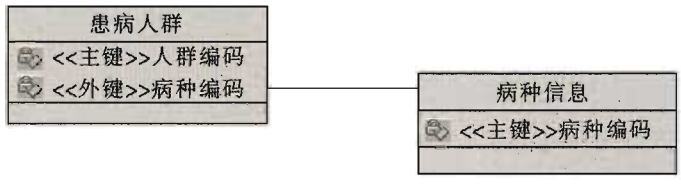

图 B.3 主键

B.5 实体间关系

B.5.1 继承关系

a) 定义:继承关系,表示若干数据实体继承自某个数据实体,并因此具有该实体的部分属性。

b) 图形符号见图 B.4:

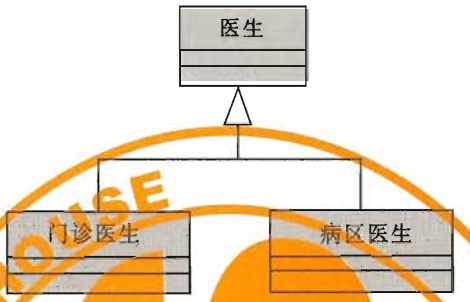

图 B.4 继承关系

B.5.2 包含关系

a) 定义

——包含关系,表示实体(数据单元)包含其他的实体(数据单元),被包含的实体(数据单元)是前者的一个组成部分。例如:一个人包括头、躯干、四肢。

——指定包含关系同时要指定包含的数量。

——包含数量通常包括:

0..1	表示没有,也可以有并最多有一个
0..n	表示有,也可以没有,最多有无穷多个
1	表示必须有,且只能有一个
1..n	表示必须有,最多可以有无穷多个
n	表示必须有指定数量的

——包含关系有两种:一种是聚集,另一种是强势聚集。聚集表示一个实体由一组实体所组成。强势聚集表示每一组成部分是不可分割的,组成部分与整体是"终身关系",同时建立和清除。

b) 图形符号见图 B.5,图 B.6。

——菱形符号出现在整体一侧。

——包含数量要在关联线靠近组成部分的位置上标识。

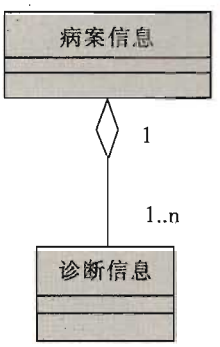

图 B.5 聚集关系

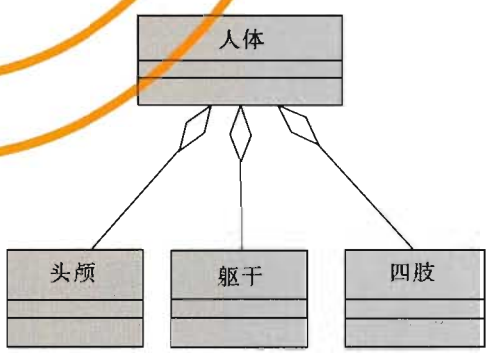

图 B.6 强势聚集关系

B.5.3 依赖关系

a) 定义:依赖关系,表示对实体(数据单元)的理解、使用等依赖其他的实体(数据单元)。

b) 图形符号见图 B.7。

——依赖关系使用带箭头的虚线表示。箭头的方向表示依赖方向。

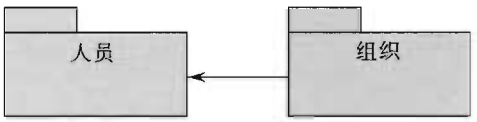

图 B.7 依赖关系

——在示例中,对数据单元"组织"的理解,依赖于数据单元"人员",在"组织"数据单元中涉及了人员数据单元中的实体,因此存在依赖关系。

B.5.4 关联关系

a) 定义:用于表示除泛化、包含、依赖关系以外的数据实体之间的关系。关联两边的数字表示两个实体之间在关联中的数量对应关系,具体表示规则与 B.5.2 中的包含数量相同。

b) 图形符号见图 B.8。

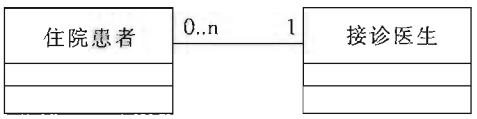

图 B.8 关联关系

B.5.5 注释

a) 定义:注释是自然的文档,对模型的有机组成部分作出限制。注释通常出现在数据单元中。

b) 图形符号见图 B.9。

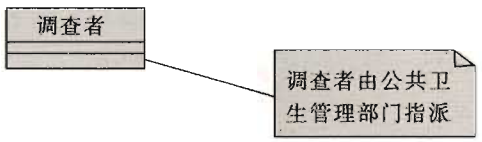

图 B.9 注释关系

B.6 实体的组织

B.6.1 定义

实体的组织是出于某种目的,由若干个实体和指定的域(属性)组装而成的集合。

一个数据模型通常包含一至多个数据单元。数据单元之间的表现同时也暗示了数据单元之间的关系。如:数据单元"课程"包含在数据单元"学校"中,暗示了"课程"是"学校"的逻辑上一部分。

数据单元之间的关系包括:依赖关系和包含关系。依赖关系通过 B.5.3 中定义的 UML 图形符号表示。包含关系通过 UML 模型本身的层次关系确定。

B.6.2 图形符号

使用 UML 中的包(Package)表示见图 B.10。

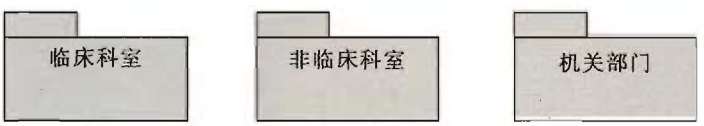

图 B.10 实体的组织

附 录 C

（资料性附录）

类关系模式描述实例

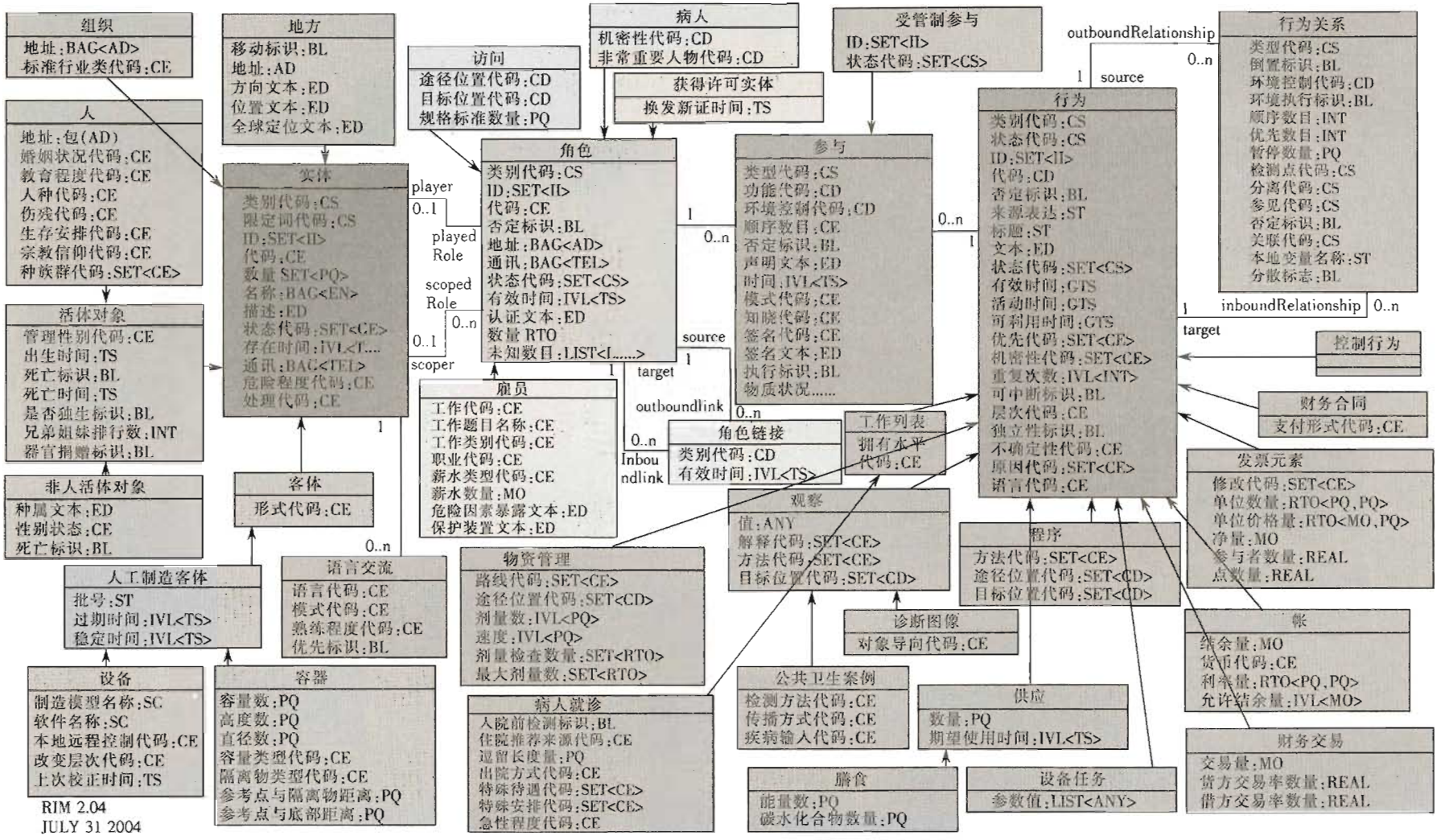

图 C.1 HL7 RIM

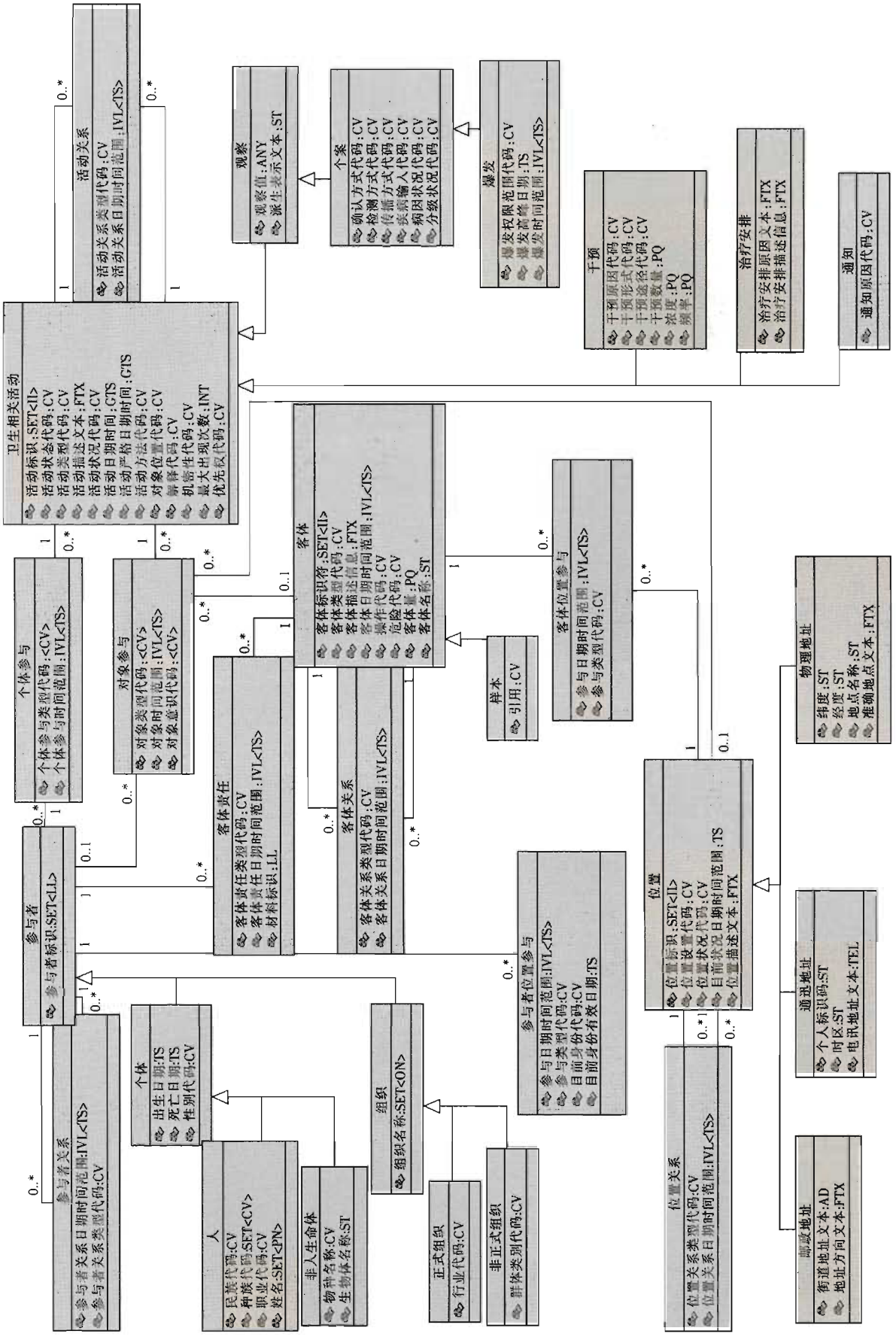

图 C.2 美国公共卫生概念数据模型

附 录 D
（资料性附录）
HL7 RIM 部分常用数据类型概述

表 D.1 HL7 RIM 部分常用数据类型概述

名称	符号	描 述
数值型	ANY	定义每个数据值基本的特性。这是一个抽象类型，意味着没有值能够只是一个数值而不属于任何具体的类型。每一个具体类型都是通用抽象数据值类型的特化。
布尔型	BL	布尔数值型，表示两值的逻辑值。一个布尔数值要么是真，要么是假，或者，另一个值不存在。
压缩数据型	ED	主要用于 HL7 之外的人类翻译或者进一步机器处理的数据。包括非格式化或格式化的书写语言、多媒体数据或者被不同标准定义（如，XML-签名）的结构化信息。除了数据本身，ED 可以包括仅有的参考（见 TEL）。应该注明字符串数据类型是当 ED 媒介类型是文本/普通文字时 ED 数据类型的详细解释。
字符串数据类型	ST	字符串数据类型表示文本数据，主要为了机器处理（如分类、查询、检索等）。用作名称、符号和正式表达。
概念描述符型	CD	一个通过给定的编码来表达任意种类概念的描述符，这个给定编码是在一个编码系统中已经被定义过的。一个概念描述符能够包含初始文本或字段，字段用来服务于基本编码和一个或更多个翻译成不同的编码系统。概念字码也包含描述的修饰词，如，“左脚”的概念作为结合后词语项，建立自主要编码“脚”和修饰词“左”。在额外情况下，概念字码需要不止包含一个编码，只要描述概念的初始文本。
被编码的简单值型	CS	用最简单形式编码的数据，在此只有编码和展现的名称没有被预先决定。编码系统和编码系统版本取决于 CS 型数据出现的语境。CS 型被用于被编码属性，这些属性具有唯一的、由 HL7 定义的值集合。
等价编码型	CE	由一个编码值和一个来自于表示相同概念的其他编码体系的编码值（可选）组成的编码数据。当两选一编码可以存在时被使用。
编码字符串型	SC	一个可以有编码（可选）与之绑定的字符串。如果编码存在，则文本通常必须存在，编码往往是范畴内专用编码。
实例标识符型	II	对一个事务或对象进行唯一性标识的标识符。例如 HL7 RIM 对象的对象标识符，诊疗记录号、次序标识符、服务种类项目标识符、车号（VIN）等。实例标识符基于 ISO 对象标识符被定义。
通讯地址型	TEL	电话号码（声音/传真）、邮箱地址或其他依靠远程通讯设备的资源定位。地址被详细表示为具有时间规范许可的 URL（统一资源定位符），并且通过编码来决定对于给定时间和目的的情况下，应当使用哪个地址。
邮政地址型	AD	邮递和家庭或办公地址。一个地址组件序列，如，街道、邮寄箱、城市、邮编、国家等。
实体名称型	EN	人、组织、地点或事务的名称。一个名称组件序列，如名或姓、前缀、后缀等。实体名称值的举例，有“Jim Bob Walton, Jr.”，“Health Level Seven, Inc.”，“Lake Tahoe”等。实体名称可能同字符一样简单或者由几个实体名称部分组成，如“Jim”、“Bob”、“Walton”和“Jr.”，“Health Level Seven”和“Inc.”，“Lake”和“Tahoe”。
通俗名称型	TN	用于表示事情和地点的简单名称，被用作有效简单字符实体名称的限制。
人名型	PN	人的姓名。一个姓名组件序列，如，名或姓、前缀、后缀等。
组织名称数据类型	ON	组织的名称。一个名称组件序列。

续表

名称	符号	描　述
整数型	INT	整数(−1,0,1,2,100,3 398 129,等)是准确的数字,是可计数和可数的结果。整数是不连续的,整数集是无限的,但也是可数的。没有专门的限制施加于整数系列。两个“为空”的特性被定义为正负无穷大。
实数型	REAL	分数,典型用于被测量、被评价或被计算的计量数,源自其他实数。当有效小数位数被精确知道的情况下,典型的表达是小数。
率型	RTO	分子计量数除以分母计量数的商。分子和分母的共同点是不被自动抵消。RTO 型支持滴定率(如,1:128)和其他实验产生的、真实表达“率”的数量。率不是简单的“结构化数字”,如血压测量(如“120/60”)不是率。很多情况中用实数来代替率。
物理量数据类型	PQ	表达测量结果的维度化数量。
货币量数据类型	MO	货币量是以某种现金币种为表示形式的资金数量,现金的币种是在不同经济区域中用以表达资金数量的单位。虽然资金量是一种单一的物理量,但是不同货币单位之间发生交换时的比率是变化的。这是物理量和资金量之间的主要区别,也就是货币单位不能算是物理单位的原因。
时间点型	TS	表示自然时间轴上一个点的量,时间中的一个点经常用日历表示。
集合型	SET	包含其他特定值的值,无特别次序。
序列型	LIST	包含其他离散值的值,具有特定的顺序。
包型	BAG	值的无序集合,每个值都可以在包中多次出现。
间隔型	IVL	基本数据类型中有序相邻值的集合。
历史型	HIST	一组数据值的集合,这些数值遵循历史条目(HXIT)类型(如存在一个有效时间属性)。历史信息不受过去限制,未来期望值也可以出现。
未确定值概率型	UVP	一般数据类型的扩展,用于详细说明给定值所能够表达信息生产者观点信息的概率。
参数概率分布型	PPD	一般数据类型的扩展,详细说明使用分发功能和分发参数的量化数据的未确定性。如果用来接收的应用无法与一个特定的概率分布相协同,那么除了具体的分发参数,均数(预期值)和标准差用来协助维持一个互用性的最小层次。
通用时间规格型	GTS	时间点的一个集合,详细说明事件和行为的时间,某类信息中存在的循环有效模式,如,电话号码(早上、白天),地址(冬天居住在南方,夏天居住在北方的候鸟)和办公时间。

参 考 文 献

[1] Canberra：National Health Data Committee. National Health Data Dictionary，Version 12 [M]. 2003.

[2] Canadian Institute for Health Information. Conceptual Health Data Model V2. 3[M]. Ottawa：2001.

[3] U. S. Department of Health and Human Services Public Health Service Centers for Disease Control and Prevention (CDC). Public Health Conceptual Data Model，Premiere Edition [M]. Atlanta，Georgia：2000.

[4] Health Level Seven，Inc. Reg. U. S. Pat & TM Off. HL7 Reference Information Model [M]. Version：V 02-04 (7/28/2004)：2004.

[5] National Electronic Disease Surveillance System Working Group. NEDSS Logical Data Model Overview and Users' Guide，Version 1. 0 [M]. 2002.

[6] 中华人民共和国卫生部. 2007 年中国卫生统计年鉴[EB/OL]. http://www. moh. gov. cn/newshtml/20754. htm.

[7] 中华人民共和国卫生部. 2007 年中国卫生统计提要[EB/OL]. http://www. moh. gov. cn/newshtml/19165. htm.

[8] 闪四清，陈茵，程雁. 数据挖掘——概念、模型、方法和算法[M]. 北京：清华大学出版社，2002.

[9] 苏新宁. 数据仓库和数据挖掘[M]. 北京：清华大学出版社，2006.

[10] Ken Lunn. UML 软件开发[M]. 北京：电子工业出版社，2005.

[11] 徐宝文，周毓明，卢红敏. UML 与软件建模[M]. 北京：清华大学出版社，2006.

[12] 中华人民共和国国家质量监督检验检疫总局，中国国家标准化管理委员会. GB/T 7714—2005 文后参考文献著录规则 [S]. 北京：中国标准出版社，2005.