

中华预防医学会预防医学情报专业委员会

第十六届学术交流会

论文集

中华预防医学会

预防医学情报专业委员会

2005年9月 辽宁大连

中华预防医学会预防医学情报专业委员会

第十六届学术交流会

论文集

中华预防医学会
预防医学情报专业委员会
2005年9月 辽宁大连

日 程 表

2005年9月 辽宁大连

9月23日

全 天 报 到

9月24日

上午8:30—12:00

一. 开幕式、全体与会人员合影 (8:30—9:20)

二. 专家报告 (9:20—10:40)

1) 中国公共卫生数据字典研究进展 (9:20—10:00)

(中国疾病预防控制中心 金水高 研究员)

2) 医学图书馆文献资源整合的探讨 (10:00—10:40)

(解放军医学图书馆 雷春炳 研究员)

3) 《预防医学情报杂志》介绍 (10:40—11:00)

(《预防医学情报杂志》编辑部 胡亚君 主任)

三. 专题讲座 (11:00—12:00) (按公司名称字母顺序排名)

1) 竞争情报与疾控 (11:00—11:30)

(百度在线网络技术(北京)有限公司 王啸)

2) 信息资源挖掘开发及利用的解决方案、网站建设整体思路及应用开发管理 (11:30—12:00)

(北京航天四创软件技术有限公司 姚卫东)

下午2:00—5:30

一. 专题讲座 (2:00—4:30) (按公司名称字母顺序排名)

1) 《中国专业图书馆网》基于WEB的集群图书馆管理系统 (2:00—2:30)

(北京华夏网信科技有限公司 王洪力)

2) 医学科研与公共卫生之电子信息运用 (2:30—3:00)

(Ovid公司 萧金生)

3) 《中国疾病预防控制中心仓库》V5.0版介绍 (3:00—3:30)

(清华同方光盘股份有限公司 王芬)

4) 外文生物医学期刊全文服务系统 (3:30—4:00)

(深圳市康健查查网络科技有限公司 李玉)

5) 万方数据公共卫生信息系统介绍 (4:00—4:30)

(万方数据股份有限公司 林云水)

二. 委员会会议 (4:30—5:30)

9月25日

上午(8:30—12:00)

一. 大会学术论文交流

(8:30—11:00 每人15分钟,大会发言的名单根据到会人员的情况将做适当调整)

1. 军队医药卫生类期刊现状分析研究

(解放军医学图书馆 杜云祥)

2. 提升能力,更好地应对突发公共卫生事件

(解放军第三〇二医院 《传染病信息》杂志社 李军)

3. 医疗卫生机构数字图书馆建设的探索与实践

(广州市疾病预防控制中心 杨智聪)

4. 中国疾控中心门户网站建设思路

(中国疾病预防控制中心 李海燕)

5. 天津市公共卫生信息化现状与发展

(天津市卫生防病中心卫生信息中心 霍飞)

6. 网络资源的应用

(福建省疾病预防控制中心 朱凯)

7. 网络环境下医学图书馆的知识管理

(山东省疾病预防控制中心时念玲)

8. 情报信息服务在疾病预防控制中的应用

(上海市疾病预防控制中心 喻海兰)

9. 试论重大食物中毒事件的信息工作

(宁波市卫生监督所 邱维丰)

10. 云南省地方病防治所档案信息资源的开发与利用

(云南省地方病防治所 马荣)

二. 评选优秀会议论文(11:00—11:15)

三. 闭幕式(11:15—11:30)

下午

参观学习

9月26日

参观学习

9月27日

撤离

目 录

《中国公共卫生数据字典》研究内容与进展.....	1
天津市公共卫生信息化现状与发展.....	17
在中华预防医学会第四次全国会员代表大会开幕式上的致辞.....	25
文献资源整合与西文期刊文献数据库建设.....	26
中国疾控中心新版门户网站建设思路及疾控机构门户网站建设相关问题探讨.....	32
突发公共卫生事件的情报咨询研究.....	36
情报信息服务在疾病预防控制中的应用.....	39
网络环境下医学图书馆的知识管理.....	42
医疗卫生机构数字图书馆建设的探索与实践.....	46
国内泡型包虫病文献分布.....	49
网络资源的应用.....	52
网络环境下疾控文献信息服务的实践.....	56
中文医学核心期刊简介.....	59
谈互联网上生物医学信息资源的获取.....	63
军队医药卫生类期刊现状分析研究.....	66
国内 5 大权威中文生物医学期刊数据库文献收录比较研究.....	70
提升能力,更好地提供预防传染病突发公共卫生事件的信息服务.....	74
预防医学信息管理员的素质要求.....	76
网络环境下医学图书馆的参考咨询工作.....	77
论专业图书馆信息服务工作模式的转变与对策.....	79
医学参考文献的查询与应用.....	81
我国卫生信息化建设浅析.....	85
突发公共卫生事件应急机制建设的探索与思考.....	88
培养基层疾病预防控制工作者信息服务能力探讨.....	92
清远市疾病预防控制中心网站建设介绍.....	93
豫南某县献血浆人群 HIV、HCV、HBV 感染流行病学分析.....	95
福安市一起农村麻疹暴发流行的调查分析.....	98
公卫信息快速发展应注重人才继续教育.....	100
突发公共卫生事件的构成要素和主要特征探讨.....	102
西双版纳州法定传染病疫情信息网的组建及效用.....	105
专业图书馆知识信息特色化.....	108
疾病给社会造成的损失.....	110
立足本职,服务大众.....	111
味精中漂白剂监测结果分析.....	113
云南省地方病防治所档案信息资源的开发与利用.....	114
试论重大食物中毒事件的信息工作.....	116
探索知识服务模式 提升创新服务水平.....	118
浅议现代图书馆自动化网络建设.....	120
网络环境下翻译实践与分析.....	121
中、小学传染病防治效果考核评价.....	123
规范档案管理 服务疾控事业.....	126
网络环境对疾控系统图书馆的影响.....	128
网络环境下疾控系统图书馆的新模式.....	131
大连市 2004 年餐饮具卫生状况调查.....	133
2004 年《浙江预防医学杂志》论文引文分析.....	135
谈医学图书信息促卫生监督发展之三要素.....	137
我国疾病预防控制领域图书馆的现状与发展对策.....	139

《中国公共卫生数据字典》 研究内容与进展

中国疾病预防控制中心 金水高
2005年9月27日 大连

主要内容

- 一、项目概述
- 二、数据建模 (PHCDM、NHIM、HL7RIM)
- 三、数据元标准 (Data elements)
- 四、元数据标准 (metadata)
- 五、国家最小数据集 (NMDS)
- 六、国家卫生数据字典 (NHDD、NHS)

一、项目概述

《公共卫生信息基本数据集标准研究》关注的是数据内容的表达标准与规范，研究的最终产品是**国家公共卫生数据字典**，具体研究内容如下：

- 公共卫生信息概念模型
- 公共卫生最小数据集标准
- 公共卫生数据元与元数据标准

一、项目概述

标准化有记载的历史可以追溯到秦始皇时代。公元前二世纪秦始皇统一中国后，就推行了“书同文，车同轨，统一弛道，统一货币，统一兵器”等，今天我们可以将其称之为标准方面的重大政策。

随着时代的发展，标准越来越为全人类各行各业所高度重视。国际和许多发达国家已建立了较完善的信息标准发展协调促进组织。

一、项目概述

我国公共卫生信息系统建设是国家卫生信息化建设的重要组成部分。多年来已经建立并运行了覆盖公共卫生各个领域，数十个以传染病疫情与突发公共卫生事件监测报告为主体、包含不同专病监测报告、管理的公共卫生信息系统，初步形成了从基层到中央的疾病监测报告信息网络，积累了大量的原始数据。

一、项目概述

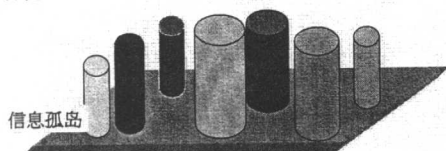
但是正如我国科学家在分析我国近年来各个行业信息系统建设与信息资源利用间存在的普遍问题时所说的“中国是一个数据大国，但不是一个数据强国”。

一方面，我国在不同的地区和领域运行着众多的信息系统，存储了大量的数据资源；

另一方面，我国各个领域又普遍存在着严重的数据“荒”。

一、项目概述

重要原因是信息系统的规划与建设缺乏统一的标准,表现为存在着众多的“难以交换、无法共享”的信息孤岛。相对于“书同文,车同轨”时代,我们可将其称为信息系统建设的“春秋战国”时代,小作坊的“烟囱”式开发。



一、项目概述

卫生领域的信息孤岛问题受到国家卫生部的的高度重视,于2003年始先后启动了四个标准化研究组重点课题:

- 医疗卫生信息标准体系框架 (HISAF)
- 医院基本数据集标准 (HIBDSS)
- 公共卫生基本数据集标准 (PHBDSS)
- 社区医疗基本数据集标准 (CHBDSS)

一、项目概述

国家卫生信息标准的系列研究是为消除我国卫生信息孤岛,促进系统间、部门间卫生信息的互联互通、协调运作、广泛共享并满足政府职能部门对卫生信息综合利用的基础性工作。

《公共卫生基本数据集标准》研究是国家卫生部2004年启动,委托中国CDC 承担的国家卫生信息标准化研究体系的重要组成部分。

一、项目概述

公共卫生基本数据集标准研究目的:

- 规范公共卫生领域信息系统的开发;
- 实现公共卫生数据有效交换和广泛共享;
- 发挥公共卫生信息在管理评价中的支持作用。

主要内容

- 一、项目概述
- 二、数据建模 (PHCDM、NHIM、HL7RIM)
- 三、数据元标准 (Data elements)
- 四、元数据标准 (metadata)
- 五、国家最小数据集 (NMDS)
- 六、国家卫生数据字典 (NHDD、NHS)

二、数据建模

什么是数据建模?

数据建模是分析和描述组织必须理解的事物(也就是类)的过程;一个数据模型描述了那些类的事实(也就是属性)以及与之相关的关联关系和事物规则。数据建模需要业务人员基于框架进行反复研讨与斟酌,直至达到具体类及相关属性定义上的一致性。

数据建模的目的是面向用户信息需求和事务处理,开发的一个精确模型或者图示表达。一个成功的数据模型是使数据库和应用开发具有良好稳定性和可扩展性的设计图。数据模型在新系统开发或增强应用中扮演一个框架角色。

二、数据建模

《公共卫生基本数据集标准》课题组在翻译和研读 PHCDM、NHIM 和 HL7RIM 基础上, 将研究建立中国公共卫生信息概念数据模型。

二、数据建模

关于美国公共卫生概念模型 (PHCDM)

Public Health Conceptual Data Model

U.S. Department of Health and Human Services

Public Health Service
Centers for Disease Control and Prevention (CDC)
Atlanta, Georgia 30333

July 2000

二、数据建模

PHCDM的基本功能

1. 建立公共卫生数据标准, 包括数据定义、组成结构 (如复杂的数据类型)、代码表值和数据应用。
2. 与国家卫生信息标准框架主体协同, 为公共卫生机构与健康提供者间的信息交换定义标准。
3. 遵照已确定的数据交换标准, 构建计算机信息系统, 并应用于公共卫生相关的数据处理。

二、数据建模

PHCDM的目标:

1. 为公共卫生组织数据标准和指南提供一个框架;
2. 降低公共卫生信息系统的开发强度;
3. 通过一致性增强数据共享;
4. 满足在国家标准框架中描述公共卫生数据的需求
5. 促进CDC与政府和地方合作者在公共卫生领域的协作。

二、数据建模

PHCDM使用了一个合理的高层面抽象的方式来记录公共卫生概念, 避开了通常在创建逻辑和物理数据模型中经常出现的许多细节。PHCDM采用统一建模语言 (UML) 的建模规则, 用到的UML概念:

主题域

类和关联

属性和数据类型

二、数据建模

(一)主题域

主题域是模型中属性相同部分类的集合。一个主题域是在允许范围内用子集的方式把一个模型划分为若干更容易理解的区块。PHCDM中有四个主题域: 卫生相关活动、位置、客体和参与者。

卫生相关活动	位置	客体	参与者

二、数据建模

1. 卫生相关活动主题域

卫生相关活动主题域包含卫生服务、卫生环境和影响健康行为的信息。一个卫生相关活动可以是一个观察、一个干预、一个治疗安排或者一个信息发布。

二、数据建模

2. 位置(场所)主题域

包含与参与者、卫生相关活动或者客体相关联的地点。位置信息可以是一个邮政地址、一个通讯地址或者一个物理地址。

二、数据建模

3. 客体主题域

包含与卫生相关活动和参与者有关的物资、设备和产品。如食物、药品以及与公共健康相关的物理实体和其他有形内容。客体的典型例子包括食品、标本、药物、医疗设备和医疗供给;与公共健康相关的物理实体可包括湖、池塘、船或者飞机等健康的危险因素暴露的潜在发源地。

二、数据建模

4. 参与者主题域

包含卫生相关活动的参与对象信息。一个参与者可以是一个独立的个人和非人生命有机体,或者一个正式或非正式组织。参与者的典型例子包括患者、医师、流行病学家、医院和实验室。参与者也包括像抽烟者、五岁以下儿童这样的具有公共特征的群体。

二、数据建模

(二)类与关联

PHCDM有29个类。类就是一些被采集的信息,它们可以是人、地点、物品、概念或事件。

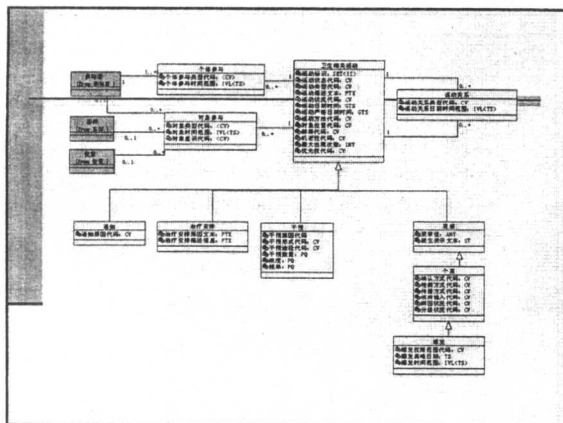
PHCDM有4个核心类,对应于4个主题域,即卫生相关活动、位置、客体和参与者。每个核心类包含若干个类。

二、数据建模

PHCDM中的类:

卫生相关活动主题域含10个类:

- | | |
|----------|---------|
| ① 卫生相关活动 | ② 行动者参与 |
| ③ 对象参与 | ④ 活动关系 |
| ⑤ 通知 | ⑥ 治疗安排 |
| ⑦ 干预 | ⑧ 观察 |
| ⑨ 个案 | ⑩ 爆发 |

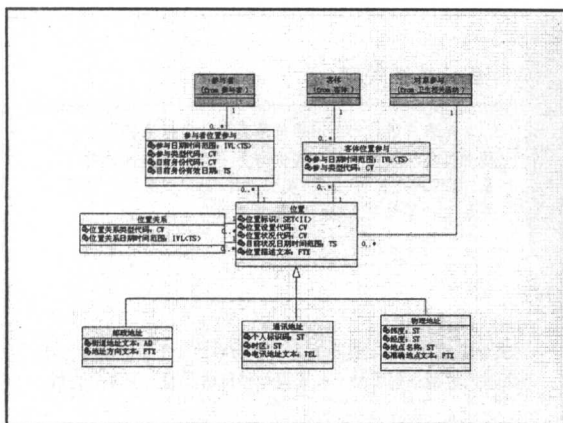


PHEDSS

二、数据建模

位置主题域含7个类:

- ① 位置类
- ② 参与者位置参与
- ③ 客体位置参与
- ④ 位置关系
- ⑤ 邮政地址
- ⑥ 物理位置
- ⑦ 通信地址

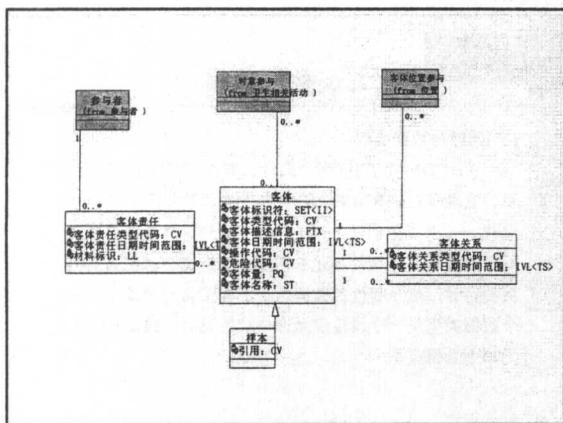


PHEDSS

二、数据建模

客体主题域含4个类:

- ① 客体类
- ② 客体关系类
- ③ 客体责任类
- ④ 样本类



PHEDSS

二、数据建模

参与者主题域含8个类:

- ① 参与者类
- ② 参与者关系类
- ③ 组织类
- ④ 正式组织类
- ⑤ 非正式组织类
- ⑥ 个体类
- ⑦ 人类
- ⑧ 非人生命个体类

二、数据建模

属性与数据类型举例

卫生相关活动	
活动标识: SET(II)	
活动状态代码: CV	
活动类型代码: CV	
活动描述文本: FTX	
活动状态代码: CV	
活动日期时间: GTS	
活动方法代码: CV	
对象位置代码: CV	
解释代码: CV	
机理性代码: CV	
最大出现次数: INT	
优先级代码: CV	

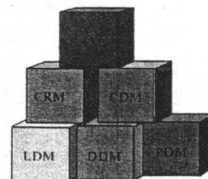
属性名称: 活动状况代码
属性描述: 表示活动状态的代码 (例如, 仅是意向、已命令、进行中、已完成), 这个属性并非用来描述一个类的分类地位
数据类型: CV-代码值数据类型

属性名称: 活动日期时间
属性描述: 已发生的、安排或计划发生的、或条件成熟将发生的行为时间。时间规范可以是一个时间点、一个活动发生的时间范围或者只是假设发生的时间。
数据类型: GTS-通用时间规格数据类型

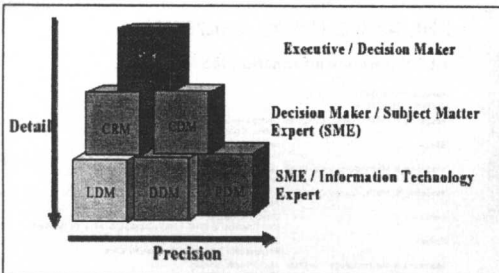
数据模型类别层次

Hierarchy of Data Model Types

- Subject Area
- Class Relationship
- Conceptual Data
- Logical Data
- Database Design
- Physical Database

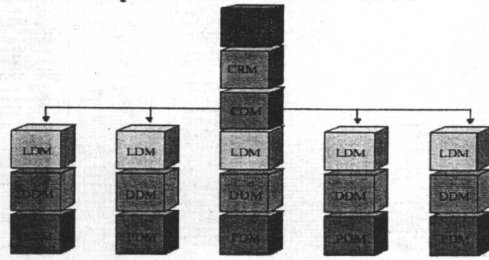


数据模型层次关系



概念数据模型与物理数据模型关系

Multiple PDMs from one CDM



二、数据建模

关于澳大利亚国家卫生信息模型 (NHIM)

National Health Information Model

Version 2.0

Prepared by the National Data Standards Unit
Australian Institute of Health and Welfare
GPO Box 570, Canberra ACT Australia 2601
Phone: (02) 6244 1000 Fax: (02) 6244 1166

二、数据建模

NHIM提供了如何在卫生系统内部对信息进行组织和结构化的方法。其第二版的发展标志着从第一版中的E-R模型提升到一个高层面、自由关联和多事务处理框架模型形成的一个变革。这个变革是由于认识到NHIM作为一个高层面框架而应具有通用性效果的同时还应考虑建模过程中的多层面的需要。亦即, 这个变革既反映了在国家层面上实体一致性标识的重要性, 也考虑了在较低层面上关联或事务规则的重要性。

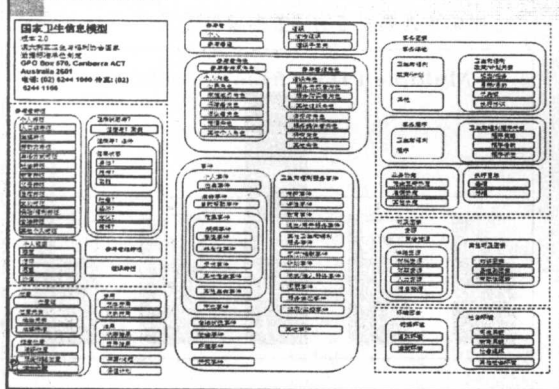
二、数据建模

NHIM是一个信息模型，是独立于过程的。即，它并不关心事情是如何发生的，关心的是潜在于发生这些事情的构成和相应的国家卫生服务政策。通过理解卫生信息资源构成，就能更好地开发这些资源所包含的信息。NHIM是一个概念模型，目标在于建立一个高层面上具有一致性结构，能够标识和描述一个广泛的实体，同时为更加详细的低层面于模型的开发提供一个框架。

二、数据建模

NHIM有12个主要超级实体，可以概要地组织成四个逻辑范畴：参与者和状态，状态变化事件，环境因素和分类系统。每个实体具体覆盖的内容和重要性可以通过实体中所含数据元的总体来评估。一些实体模型中有很少甚至没有数据元，这可能是为今后进一步的开发或对NHIM的结构再评估提供的有用的标识区域。

数据实体模型：



二、数据建模

关于HL7参考信息模型 (HL7RIM)

HL7 Reference Information Model

Version: V 02-04 (7/28/2004)

ModelID: RIM_0204

Modeling & Methodology Co-Chair George Beeher, Jr., PhD.

Editor James Case, DVM, PhD.

Modeling & Methodology Co-Chair Jane Curry

Modeling & Methodology Co-Chair Ann Huber

Editor Lloyd McKenzie

Editor Gunther Schadow, M.D., PhD.

Modeling & Methodology Co-Chair Abdul-Malik Shahr

Modeling & Methodology Co-Chair Shahr Consulting LLC

HL7® Version 3 Standard, © 2004 Health Level Seven® Inc. All Rights Reserved.

二、数据建模

HL7RIM是在HL7标准开发活动中，概览医疗和卫生服务信息的一个静态模型，是集HL7工作组和国际会员多数人的观点所达成的意见，是HL7V3.0所有协议规范标准提取其相关信息内容最根本的消息来源。

RIM是卫生保健信息需求的一个模型。RIM的抽象手法以及通过词汇规范的扩展能力，使得RIM可以适用于任何可以想象的卫生保健系统的信息交换场合。RIM对于任何信息域包括实体扮演角色和活动的参与都是概念性地应用。

二、数据建模

RIM采用了非常抽象的建模方法。RIM的主干类和它们的结构属性是它的核心。RIM由六个‘主干’类组成：

1. **Act行为**：描述卫生管理和服务事务中执行的并必须被记录的活动。
2. **Participation参与**：是用诸如执行者、执行对象以及执行地点等一类术语来表达一个行为的相关环境。
3. **Entity实体**：描述参与卫生服务活动并对其有影响的客观事物和有生命体；

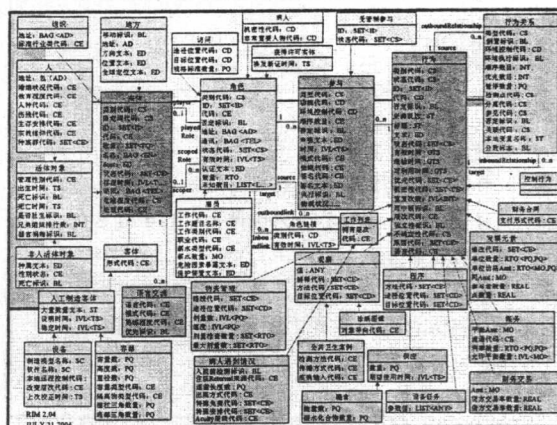
二、数据建模

4.Role 角色: 确立实体在参与卫生服务行为中担负的角色;

5. Act Relationship 行为关系: 指的是行为之间的相互关系, 例如一个发生的观察事件与对该观察事件作出的指令之间的关系;

6.Role Link 角色链接: 描述个体角色间的相互关系。

六个主干类图如下:



主要内容

- 一、项目概述
- 二、数据建模 (PHCDM、NHIM、HL7RIM)
- 三、数据元标准 (Data elements)
- 四、元数据标准 (metadata)
- 五、国家最小数据集 (NMDS、DSS)
- 六、国家卫生数据字典 (NHDD、NHS)

三、数据元

定义：用一组属性描述其定义、标识、表示和允许值的数据单元。

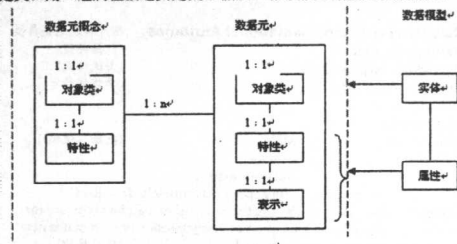


图 4-1 数据元的结构模型

三、数据元

数据元的基本属性 (GB)

- ①标识类属性(中文名称、英文名称、中文全拼、标识符、版本、注册机构、同义名称、相关环境)
- ②定义类属性(定义、对象术语、特性术语)
- ③关系类属性(分类模式、分类模式值、关系)
- ④表示类属性(表示词、数据类型、数据格式、值域)
- ⑤管理类属性(状态、提交机构)
- ⑥附加类属性(备注)

三、数据元

数据元的分类

- ### 1. 数据元分类模式
- 关键字、主题词表、分类法和本体论
- ### 2. 数据元的分类属性
- 数据元的若干成分导致了对分类的需求。主要的成分包括对象类、特性、表示、值域、数据元概念以及数据元本身。分类属性的使用使各分类模式与数据元选定的成分联系起来。

三、数据元

数据元值域及其表达模式

数据元在具体应用中表现为值，通常是一个允许值的集合，被称之为值域，它们在信息交换等功能中得到共享。

值得注意的是数据元从不表示为一个单个数值，因为它是一个类（数据值完整的集合），不是一个单个事例。

三、数据元

数据元属性规范性描述（以澳大利亚Sex数据元为例）

Sex		标识和定义类属性
Identifying and Definitional Attributes		(知识库ID)
Knowledgebase ID:	000149	(版本号)
Version:	No: 3	(元数据类型)
Metadata type:	Data Element	(当前管理状态)
Admin. status:	Current	
	01/07/03	
Definition:	The sex of the person.	(定义)
Context:	Required for analyses of service utilization, needs for services and epidemiological studies. (进一步的说明)	

三、数据元

数据元属性规范性描述（以澳大利亚Sex数据元为例）

Sex		关系和表达类属性
Relational and Representational Attributes		(数据类型)
Datatype:	Numeric	(表达格式)
Representational form:	Code	(字符串格式)
Representational layout:	N	
Minimum size:	1	
Maximum size:	1	
Data domain:	1 Male	(数据取值域)
	2 Female	
	3 Indeterminate	
	9 Not stated/inadequately described	
Guide for use:	An indeterminate sex category may be necessary for situations such as the classification of perinatal statistics when it is not possible for the sex to be determined. (使用规则)	

三、数据元

数据元属性规范性描述（以澳大利亚Sex数据元为例）

Sex		验证和收集方法
Verification rules:		Code 3 Indeterminate should be queried for people aged 90 days (3 months) or greater. (值的合法性)
Collection methods:		Code 9 is not to be an allowable option when data is being collected ie it is not to be a tick box on any collection forms or computer screens. Systems are to take account of any null values that may occur on the primary collection form. (数据收集方法)
What is your (the person's) sex?		Male_Female
The term 'sex' refers to the biological differences between males and females, while the term 'gender' refers to the socially expected/perceived dimensions of behaviour associated with males and females - masculinity and femininity.		
The Australian Bureau of Statistics advises that the correct terminology for this data element is sex.		
To avoid problems with edits, transsexuals undergoing a sex change operation should have their sex at time of hospital admission recorded.		
Related metadata: is used in the derivation of Diagnosis related group vers 1 supersedes previous data element Sex ver 2 (相关元数据)		

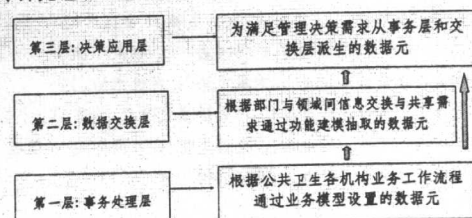
三、数据元

数据元属性规范性描述（以澳大利亚Sex数据元为例）

Sex		管理类属性
Administrative Attributes		(源文档)
Source document:	National Health Data Committee	
Source organisation:	NHIM Demographic characteristic (信息模型链接)	
Information model link:	NHIM Demographic characteristic	
(数据说明):	开始和结束日期	
Data Set Specifications:	Start date	End date
NMDS - Admitted patient care	01/07/2003	
NMDS - Admitted patient mental health care	01/07/2003	
NMDS - Non-admitted patient emergency department care	01/07/2003	
DSS - Cardiovascular disease (clinical)	01/01/2003	
DSS - Diabetes (clinical)	01/01/2003	
DSS - Health care client identification	01/01/2003	
Comments:	(注释)	

三、数据元

《公共卫生基本数据集标准》中关于数据元标准的研究包括以下三个部分:



三、数据元

目前课题组完成了公共卫生数据项的定义以及著录工具设计:
数据项定义内容如下:

编号: 标识每个数据项的唯一序列号
组号: 公共卫生基本数据集讨论组号
子系统编号: 各讨论组内部子系统编号
子系统名称: 各讨论组内部子系统名称
中文名称: 数据元素的中文名称
中文简称: 数据元素的中文简称
英文名称: 数据元素的英文全称
英文缩写: 数据元素的英文名称缩写

三、数据元

数据项定义内容(续):

描述: 数据项的用途、意义、范围等信息的详细描述。

数据类型: 数据项的数据类型

数据长度: 数据项可选值域数据的长度

值域数据类型: 数据项值域的数据类型(引用HL7定义的数据类型)

值域字典名称: 分类代码标准表的名称, 如: 职业代码表

值域字典依据: 来源于什么标准, 或制定该数据元素的参考依据。
如: GB/18392

交叉引用: 如果该数据项是引用的数据项, 说明原数据项的编号

主要内容

- 一、项目概述
- 二、数据建模 (PHCDM、NHIM、HL7RIM)
- 三、数据元标准 (Data elements)
- 四、元数据标准 (metadata)
- 五、国家最小数据集 (NMDS)
- 六、国家卫生数据字典 (NHDD、NHS)

四、元数据

• 元数据的定义

元数据最本质、最抽象的定义为 data about data, 在许多领域有其具体的定义和应用。

• 元数据的用途:

一是提供基于用户的信息, 帮助用户使用数据;

二是支持系统对数据的管理和维护, 使系统以最方便最有效的方式展现数据。

四、元数据

• 元数据概念理解

国际电器和电子工程师协会 (IEEE) 在1994年2月的白皮书中指出: 元数据是关于数据的信息, 它使得有效地管理和访问数据成为可能。元数据的主要目的是解决数据管理和共享问题。该协会认为至少有两种元数据, 即系统级元数据和应用级元数据。前者是用于管理数据的信息, 包括数据访问时间、数据量、当前的存储位置等; 后者是帮助用户的信息, 包括找到、评价、访问和使用数据的信息。

四、元数据

• 对元数据的认识

元数据并不是一个新的概念。实际上传统的图书馆卡片、出版图书的版权说明、磁盘的标签等都是元数据。当资源转换为数据形式时, 资源的管理和应用均产生了新问题, 例如: 数据生产者需要归纳和整理好本部门的业务数据, 数据服务者需要管理和维护好本行业不同部门提供的海量数据; 数据使用者需要通过方便简洁的途径远程访问数据。通过元数据辅助, 可帮助数据生产者和用户解决这些问题。

四、元数据

元数据的应用: 都柏林核心元数据集 (DC: 对图书文献数据的描述)

DC元数据是在充分吸纳了图书情报界所具有编目、分类、文摘等经验, 同时在利用计算机、网络的自动搜索、编目、索引、检索等研究成果的基础上发展起来的。它是描述、支持、发现、管理和检索网络资源的信息组织方式, 其最大特点是数据结构简单, 信息提供者可直接编码。

四、元数据

元数据的应用: 医药卫生科学数据共享工程元数据 (对数据集的描述)

该元数据标准主要是说明医药卫生领域共享数据集的。目前已经建立了若干个数据集的元数据库, 并提供查询工具。用户填写查询界面上的选项, 如目录名称、关键字、查询条件、组合关系、查询检索区域范围等, 即可获取所需数据的元数据信息。

四、元数据

医药卫生科学数据共享元数据标准中包括子集、实体和元素的概念:

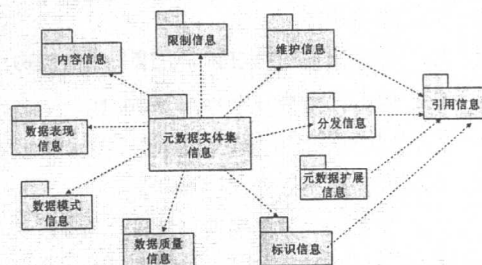
元素: 是元数据最基本的单元。本标准有22个核心元素, 132个可选元素。一个元数据元素说明一个数据集某一方面的特征。

实体: 是由一个或若干个元数据元素组成。复合实体: 则由元数据实体、元数据元素和/或其他复合实体构成。每个元数据元素、实体或复合实体均需说明其名称、定义、类型、值域、简称等特征信息。

子集: 是由若干元素、简单的或复合的元数据实体组成的集合。

四、元数据

医药卫生科学数据共享元数据标准中的9个子集



四、元数据

从UML图可看出医药卫生科学数据共享领域描述数据集元数据的子集由下列内容聚集:

标识信息、内容信息、分发信息、数据质量信息、数据表现信息、扩展信息、数据模式信息、限制信息、维护信息。

四、元数据

标识信息子集由下列实体和元素组成:

标识信息:

名称/角色名称 (中文)	名称/角色名称 (英文)	描述	定义	数据类型	是否必填	是否唯一	是否主键	是否外键	备注
标识信息子集	Identification Subset	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集
标识信息子集	Identification Subset	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集
标识信息子集	Identification Subset	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集
标识信息子集	Identification Subset	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集
标识信息子集	Identification Subset	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集
标识信息子集	Identification Subset	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集
标识信息子集	Identification Subset	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集
标识信息子集	Identification Subset	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集
标识信息子集	Identification Subset	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集	标识信息子集