

登革热监测系统概念数据模型构建^{*}

陈 强 王松旺 李言飞 戚晓鹏

(中国疾病预防控制中心公共卫生监测与信息服务中心 北京 102206)

[摘要] 介绍概念数据模型的内涵以及应用概念数据模型的优势, 应用场景分析方法对我国登革热监测各项活动进行梳理, 构建由实体类以及实体类关联关系组成的登革热监测概念数据模型, 为疾病监测相关信息系统建立统一的业务概念。

[关键词] 登革热; 概念数据模型; 数据建模; 数据标准; 场景分析; 类

[中图分类号] R-056 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2018.06.010

Building of the Conceptual Data Model of Dengue Fever Monitoring System CHEN Qiang, WANG Song-wang, LI Yan-fei, QI Xiao-peng, National Center for Public Health Surveillance and Information Services, Chinese Center for Disease Control and Prevention, Beijing 102206, China

[Abstract] The paper introduces the connotation of conceptual data model and the advantage of using it, and arranges Dengue Ferver (DF) monitoring activities in China by applying scene analysis method. To build the DF monitoring conceptual data model that is comprised of entities and correlated entities so as to establish the unified service concept for disease monitoring related information system.

[Keywords] Dengue Fever (DF); Conceptual data model; Data modeling; Data standard; Scenario analysis; Class

1 引言

登革热 (Dengue Fever, DF) 是由登革病毒引起的急性传染病, 主要通过埃及伊蚊和白纹伊蚊传播, 广泛流行于全球热带和亚热带地区, 是分布最广、发病最多、危害较大的一种虫媒病毒性疾病。登革病毒分为 1-4 型, 均可致病, 不同型病毒可重复感染并

增加发生重症风险。人群对登革病毒普遍易感染, 重症登革热病死率可达 1% 以上。近几十年全球登革热发病率大幅度增长, 全球约有一半的人口处于登革热发生的风险区。近年来登革热在国内疫情高发, 针对登革热的传染病防控, 国家建有全国登革热监测系统、全国病媒生物监测系统、全国重要病媒生物抗药性监测系统和口岸的监测信息系统。各信息系统之间缺乏有效的衔接, 也很难连接, 数据不能共享使用。系统间的数据定义、编码标准等都不同, 形成信息孤岛, 不利于疾病的综合防控。

2 概念数据模型及优势

2.1 概念数据模型

数据模型 (Data Model) 是数据特征的抽象,

[收稿日期] 2018-03-14

[作者简介] 陈强, 助理研究员; 通讯作者: 戚晓鹏, 副研究员。

[基金项目] 国家重点研发计划“国家生物安全监测网络系统集成技术研究、生物安全检测标准化与数据转换技术研究”(项目编号: 2016YFC1200801)。

组织数据元素并标准化它们之间的相互关系以及现实世界实体的属性。是指在特定应用程序领域中找到对象和关系的抽象形式化（如在公共卫生中找到的易感人群、蚊虫和诊疗），定义这种形式化的一组概念（如实体、属性和关系）。数据模型显示确定数据的结构，在企业模型的背景下，通过提供数据的定义和格式来支持信息系统的开发。根据美国国家标准协会（American National Standards Institute, ANSI）的定义，数据模型可以分为3种实例：概念数据模型（Conceptual Data Model, CDM），逻辑数据模型（Logical Data Model, LDM），物理数据模型（Physical Data Model, PDM）。CDM用来描述系统模型所涉及的领域、定义的范围，由重要事物的以及之间的关联关系组成，模型指定使用数据表达的事实种类，定义和限制模型语言所允许的表达式。CDM的目标是统一业务概念，是业务人员和技术人员沟通的桥梁，能确定不同实体之间最高层次的关系。LDM是用户从数据库所看到的模型，是对CDM进一步的分解和细化，反映系统设计人员对数据存储的观点，主要用于数据库管理系统（Data Base Management System, DBMS）的实现^[1]。PDM是在LDM的基础上考虑使用的具体技术，进行数据库体系结构设计，真正实现数据在数据库中的存放。概念模型是桥梁和纽带，模型重用关系，见图1。对于具体信息的开发过程而言概念模型是抽象和概括的，不涉及过多的细节，独立于具体的信息系统；对整个领域的信息化建设都具有指导意义，是信息标准化的基础。本文不是从软件工程角度阐述疾病监测系统的建立方法，而是从数据库概念模型来重新梳理国家登革热监测系统。

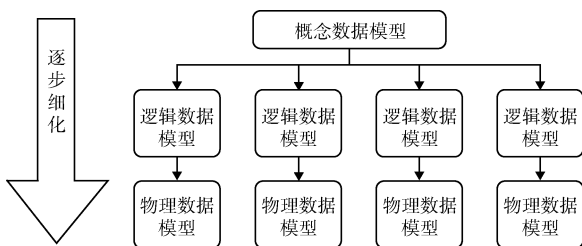


图1 模型重用关系

2.2 应用概念数据模型的优势

概念数据模型是对信息需求的高级描述，包括主要概念及其之间的主要关系，是识别不同实体之间的最高级关系。构建疾病监测的概念模型就是构造易于理解的疾病信息的整体框架，它是数据表示和代码值定义的标准，为建立公共卫生数据字典、组织数据标准提供基础框架；此外可以为全国性的疾病监测制定单位间的协作标准，为公共卫生机构和医疗单位提供数据交换定义；减少公共卫生疾病监测信息系统的开发时间，国家和地区疾控中心在开发相关信息系统时使用统一的概念模型可以大幅减少在采集、数据需求分析和数据库设计上花费的精力，减少开发时间和费用。

3 我国登革热监测基本情况

我国登革热监测的业务目标是：早期发现，及时采取控制措施，防止疫情扩散；了解登革热疾病负担和流行特征；掌握登革热媒介伊蚊种类构成、密度、分布及季节变化和长期趋势；动态监测疫点、疫区媒介伊蚊密度，评估疫情传播风险和伊蚊控制效果。监测对象包括：登革热、登革出血热（Dengue Haemorrhagic Fever, DHF）和登革休克综合征（Dengue Shock Syndrome, DSS）的疑似、临床诊断和实验室诊断病例。监测内容包括：病例报告、个案调查、密接调查、筛查入户调查、突发事件报告、实验室检测、双层叠账法调查和诱蚊诱卵指数调查等。通过梳理我国政府印发的登革热监测、诊疗方案和指南等，整理登革热监测流程，见图2，涉及多个、多级单位或部门。通过归纳和汇总，整理出5个主要流程：病例监测、媒介伊蚊监测、实验室监测、病媒生物抗药性监测和干预（分媒介和人群）。5个流程涉及多个现有系统，根据流程的业务范畴，尽可能将每个流程所涉及的具体业务进行场景分析，提取类和类关联，建立数据模型。

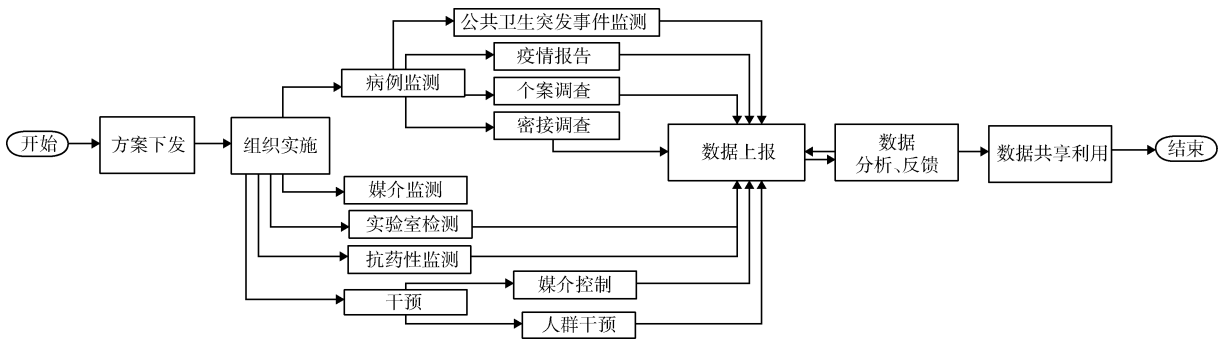


图 2 登革热监测流程

4 登革热概念数据模型设计

基于美国卫生信息传输标准 (Health Level Seven, HL7) 相关信息模型 (RIM), 参考美国公共卫生概念数据模型 (Public Health Conceptual Data Model, PHCDM)^[2] 和其他与健康相关的数据模型来开发。概念数据模型由类类关联以及关联的次數构成。类是具有相同属性的同类事物、概念的集合; 而泛化和特化是形成类和子类的方法。泛化是指对所涉及的类的同一特征综合为同一类; 所谓特化是对上位类按照一定的原则进行细分。类之间通过类属性来区分、描述所需要的特性。类属性是信息系统的基本数据元。本研究中类属性参考《公共卫生信息分类与基本数据集》和 HL7 RJMV3.0^[3]。类与类之间的关系称为类关联, 具体包括: 父子关联, 体现类之间的继承关系; 自关联, 同一类实体之间的关联; 一般关联, 起源不同类之间的关系。关联的基数即类关联彼此间发生的关联次数, 包括一对

一、一对多、多对多 3 种。多对多关联时, 引入关联类, 形成两组关联。本研究应用场景分析方法^[4], 结合我国登革热监测的各技术指南、规范, 对每一类登革热监测事件分别进行场景描述, 然后邀请疾病预防控制部门熟悉登革热业务的专家对每个场景发生或可能发生的活动进行描述和论证, 务必体现事件的触发、发展、结局全过程, 最后对这些场景进行归纳提取类、类关联、类属性。如涉及患者、临床医生、疾控人员的类可以归纳为参与者; 食物、蚊帐、消杀药水等可以归纳为材料; 家、医院、疾控中心等可以归为位置; 就诊、治疗、诊断、隔离等可以归为活动^[5]。按照这种方式, 根据所有业务流程所涉及的场景, 利用泛化和特化的方法建立模型。登革热监测概念数据模型, 见图 3。概念模型共分为参与者、位置、材料、活动 4 大主题域, 不同的主题域有相应的子类、子子类, 通过类与类关联结合形成一个整体。活动是整个登革热监测的核心, 包括监测/调查、观察结果、干预、报告、评价、实验室检测 6 大子类。

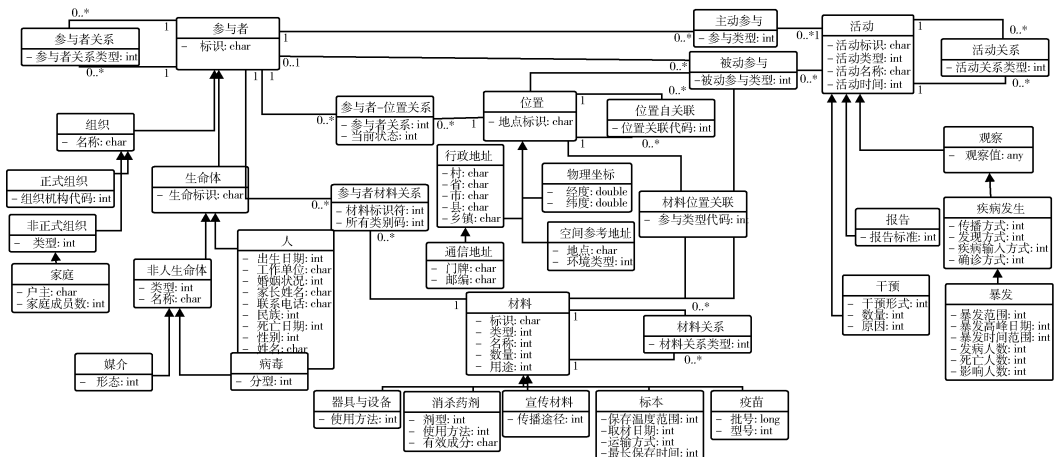


图 3 登革热监测概念数据模型

5 业务场景应用举例

建立模型，通过具体的业务场景检验模型的合理性。此处简单描述在登革热疑似病例实验室检测场景中模型是否合理。登革热病例实验室检测流程，见图 4。根据国家相关指南，登革热实验室检测包括临床标本采集、标本的保存运输、实验室检

测过程、复核检测以及结果撰写与报告等流程，通过分析业务场景，提取相关类，使用模型来覆盖处理。实验室检测应用场景模型，见图 5。可知实验室检测的场景活动及其关联活动都可以用概念模型来进行标注，从而一定程度上验证模型的合理性。本研究还验证个案调查、密切接触者调查、病历报告、入户筛查登记、诱蚊诱卵指数监测等业务场景，不再赘述。

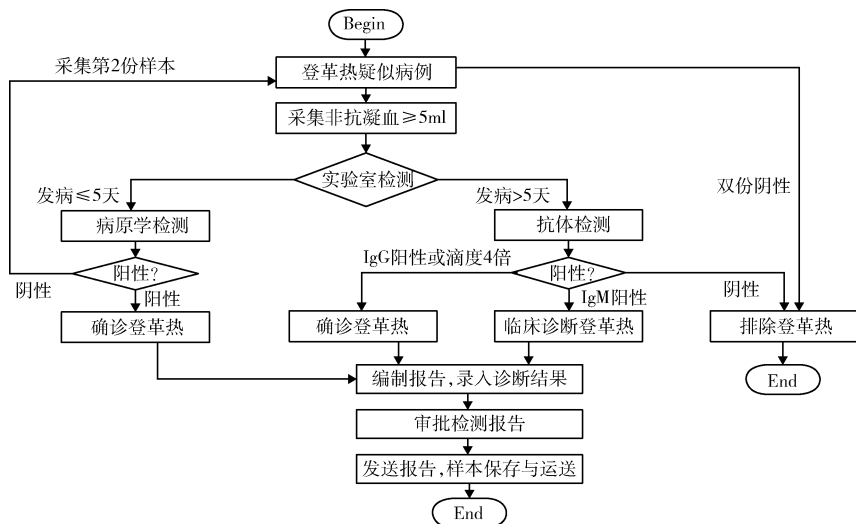


图 4 登革热病例实验室检测流程

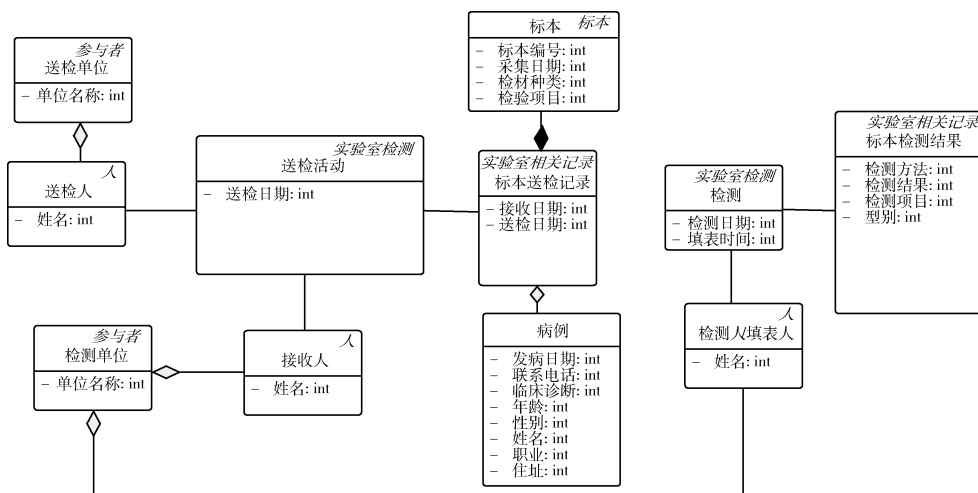


图 5 实验室检测应用场景模型

6 结语

于美国各级疾控中心及其所在的州和当地公共卫生部门，其建立和使用产生巨大的收益。本研究参考 PHCDM，从登革热监测着手，希望建立统一的顶层

PHCDM 是以美国疾控中心为主而建立，服务

(下转第 60 页)

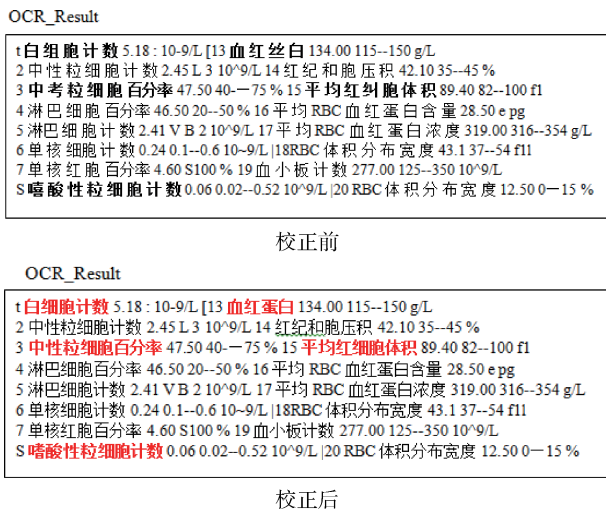


图 6 医学检验单字符识别结果校正前后对比

5 结语

应用字符识别校正算法，建立专业的医学词库，实现对医学文本图像的字符识别校正，提高医学字符识别准确率。该方法可以直接应用于有文本数据录入需求的医疗信息系统，用于辅助用户自动、高效、准确地将医学数据输入到信息系统中，降低人力

(上接第 46 页)

结构，解决相同疾病监测在不同部门、单位之间“信息孤岛”的问题，使新建系统不仅限于满足本单位、部门的需要^[6]。本研究中的概念模型是基础也是关键，最终目标是增强数据利用，促进数据共享。使用该概念模型，必然在数据库设计中重复使用相应构成，这种重复使用会为独立开发的各个软件系统在数据上展现一致性。数据的一致性是非常重要的特征，其提高将使疾病监测应用在不同信息系统之间共享数据更加便捷，使复杂数据映射和转换过程变得简单。数据的一致性还将允许数据跨越多重系统进行比较和连接，同时有利于分析、发现趋势，促进公共卫生大数据平台的建立。提供可重复使用的数据分析及数据库设计，最终开发出通用平台，将大幅减少开发时间和费用。本研究模型还不够完善，能否代表登革热疾病防控的所有业务细节尚有待验证，检验模型最好的方法是通过将模型不断应用在系统建设中来发现不足，后续将通过逻辑、物理建模来分析数据模型的优劣，进一步完善模型。

成本。将医学文本图像识别技术与校正方法结合，开发相应的医学数据拍照自动录入模块嵌入到医疗信息系统中，可以有效地提高医学检验数据的录入效率和准确率，具有一定的实际应用价值。

参考文献

- 1 向明华，向国华. 基于 OCR 技术的医疗档案管理系统研究与构建 [J]. 中国医疗设备, 2015, 30 (10): 106 - 107.
- 2 郭世雄. 医疗仪器中的数字识别技术研究 [D]. 西安: 西安建筑科技大学, 2009.
- 3 Shaohua Yang, Hai Zhao, Xiaolin Wang, et al. Spell Checking for Chinese [C]. Istanbul: Proceedings of the Eight Interational Conference on Language Resources Evaluation (LERC'12), 2012; 730 - 736.
- 4 王宸敏. 基于 OCR 技术的化验单识别方法研究 [D]. 杭州: 浙江大学, 2016.
- 5 Opencv Dev Team Histogram Comparison [EB/OL]. [2017 - 02 - 14]. https://docs.opencv.org/2.4/doc/tutorials/imgproc/histograms/histogram_comparison/histogram_comparison.html.
- 6 Ray Smith, Daria Antonova, Dar - Shyang Lee. Adapting the Tesseract Open Source OCR Engine for Multilingual ORC [C]. Barcelona: Proceedings of the Interational Workshop on Multiligual OCR 2009, 2009.

参考文献

- 1 孟凤霞, 王义冠, 冯磊, 等. 我国登革热疫情防控与媒介伊蚊的综合治理 [J]. 中国媒介生物学及控制杂志, 2015, (1): 4 - 10.
- 2 胡大权. 数据库概念模型的分析与应用 [J]. 计算机工程与应用, 2002, (22): 211 - 214.
- 3 刘丹红, 徐勇勇, 王霞, 等. 卫生信息概念数据模型与数据元标准研究 [J]. 中国卫生质量管理, 2005, (6): 1 - 3.
- 4 张先波, 金水高, 刘丽华. 公共卫生实验室检测活动类的提取与泛化 [J]. 中国医疗器械杂志, 2007, (4): 248 - 252.
- 5 郭赞, 金水高, 刘丽华. 场景分析在公共卫生信息概念模型构建中的应用研究 [J]. 中国医院, 2007, (7): 32 - 34.
- 6 Boudreaux E D, Cydulka R, Bock B, et al. Conceptual Models of Health Behavior; Research in the Emergency Care Settings [J]. Academic emergency medicine?: official journal of the Society for Academic Emergency Medicine, 2009, 16 (11): 1120 - 1123.