

国家重要监测数据整合平台的建设

岳玉娟, 吴海霞, 任东升, 王君, 刘起勇

中国疾病预防控制中心传染病预防控制所, 传染病预防控制国家重点实验室, 北京 102206

摘要: **目的** 开发国家重要监测数据整合平台。**方法** 基于数据的时空特征和生物学特征, 采用 J2EE (Java2 平台企业版) 体系的 B/S (浏览器/服务器) 架构设计和 JAVA 编程技术开发平台。**结果** 实现了多源、多尺度的重要病媒生物、重要病媒传染病、入侵动植物、病原微生物、生物安全组学等国家重要监测数据的查询与可视化平台开发, 平台主要功能包括平台介绍、分类数据库查询可视化、数据整合查询可视化、登革热预测预警等功能。**结论** 国家重要监测数据整合平台的建设, 有利于解决数据标准不一、管理分散、质量参差不齐等问题, 有利于多部门、多单位之间数据交换和共享。

关键词: 时空特征; 生物学特征; 数据库; 数据整合; 平台

中图分类号: R181 文献标志码: A 文章编号: 1003-8280 (2019) 04-0387-04

DOI: 10.11853/j.issn.1003.8280.2019.04.007

Construction of a national important surveillance data integration platform

YUE Yu-juan, WU Hai-xia, REN Dong-sheng, WANG Jun, LIU Qi-yong

State Key Laboratory of Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China

Corresponding author: LIU Qi-yong, Email: liuqiyoung@icdc.cn

Supported by the National Key Research and Development Program of China (No. 2016YFC1200802) and Project of State Key Laboratory of Infectious Disease Prevention and Control (No. 2018SKLID101)

Abstract: Objective To construct a national important surveillance data integration platform. **Methods** Based on the spatio-temporal characteristics and biological characteristics of national important surveillance data, the B/S (browser/server) architecture design and JAVA programming technology development platform of J2EE (Java 2 Platform Enterprise Edition) system were adopted to construct this platform. **Results** A query and visualization platform of national important surveillance data such as multi-source and multi-scale important vector data, important vector-borne diseases, invasive animals and plants, pathogenic microorganisms, and biosafety profile was developed. The main functions of the platform included platform introduction, sub-database query and visualization, data integration query and visualization, and dengue fever prediction and early warning. **Conclusion** The construction of a national important surveillance data integration platform helps to solve problems such as different data standards, decentralized management, and uneven quality. It is conducive to better data exchange and sharing between multiple departments and multiple units.

Key words: Spatio-temporal characteristics; Biological characteristics; Database; Data integration; Platform

互联网技术的飞速发展使企业和科研工作迈向了数据化,但同时也带来了数据分散、数据孤立、数据整合困难等问题^[1]。国内数据整合研究侧重于基础和技术研究,应用上侧重商务智能、面向服务应用等商务方面^[1],从万方数据库以“数据整合”为关键字搜索到的文献大多是关于房地产、地籍、海洋沉积物等方面的信息整合^[2-5]。而国外数据整合研究更多关注地理信息系统(GIS)、遥感(RS)、生物信息学、基因表达等公共事业方面^[5-14]。此外,以往相关

领域信息整合平台构建,为本次国家重要监测数据整合平台的实现提供了参考^[15-16]。

由于现有的国家重要监测的数据量庞大且复杂、时空跨度大、数据标准不一、数据格式多样、管理分散、质量参差不齐等,造成了“数据孤岛”和“信息孤岛”,不利于数据交换和共享^[17]。随着国家重要监测数据信息化的发展和应用,基于统一标准和格式的国家重要监测数据库,建立国家重要监测数据查询与可视化平台,数据整合技术是基础和关键。本课题组从全局角度出

基金项目:国家重点研发计划(2016YFC1200802); 传染病预防控制国家重点实验室项目(2018SKLID101)

作者简介:岳玉娟,女,博士,副研究员,从事“3S”技术与公共卫生研究,Email: yueyujuan@icdc.cn

通信作者:刘起勇,Email: liuqiyoung@icdc.cn

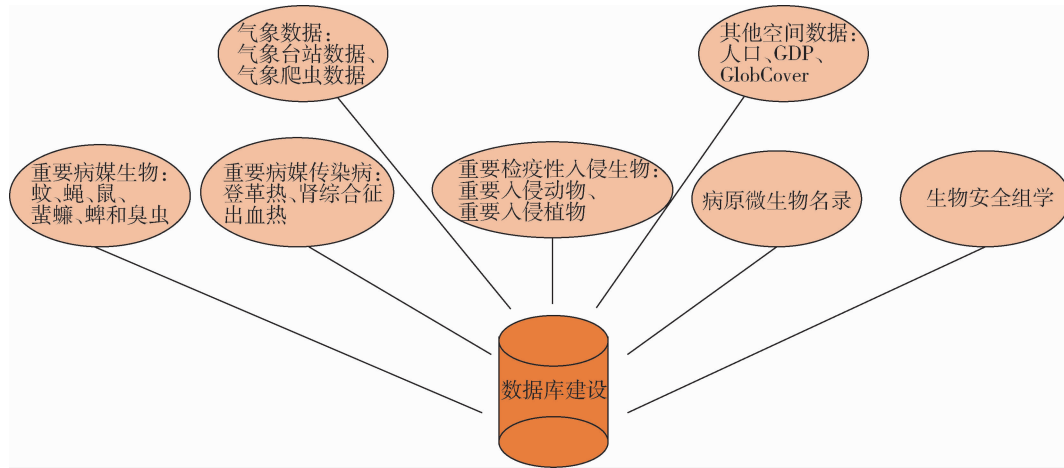
网络出版时间:2019-06-20 19:57 网络出版地址: <http://navi.cnki.net/knavi/JournalDetail?pcode=CJFD&pykm=ZMSK>

发,厘清数据的准确性、一致性、完整性、可用性,重点解决数据整合平台构建过程中的数据过剩、数据冗余问题。

1 材料与方法

1.1 材料 本研究依托国家重点研发计划,其中国

家重要监测数据包括重要病媒生物、重要病媒传染病、重要检疫性入侵生物、病原微生物名录和生物安全组学数据,也包括气象和空间地理环境数据,数据时间跨度为2005年至今,空间跨度为全国(图1)。



注:GDP. 国内生产总值; GlobCover. 全球覆盖计划

图1 国家重要监测数据整合平台数据库建设

Figure 1 Construction of the national important monitoring data integration platform database

1.2 方法 参照黄华南^[2]提出的以数据为核心,基于数据自身特点的数据整合方案。根据时间属性、空间属性和生物学属性,对国家重要监测数据开展数据整合。其中,重要病媒生物、重要病媒传染病、重要检疫性入侵生物和气象数据时间属性为“年/月”,空间属性为“省/市/区(县)”,从而该类数据之间可通过时间和空间属性开展数据整合。病原微生物名录中“宿主”字段,存储了诸如“蚊、蝇、鼠”等宿主。生物安全组学数据中“名称”字段,存储了诸如“蚊、蝇、鼠”等名称,因此,病原微生物名录和生物安全组学数据可通过生物学属性与病媒、病媒传染病和检疫性入侵生物等开展数据整合。

本平台采用基于J2EE(Java2平台企业版)体系的B/S(浏览器/服务器)架构进行设计,采用JAVA编程技术,按3层的设计思路进行扩展,在用户终端和最终数据端之间建立接入中间层,方便系统的部署和维护,增强系统的可扩展性和灵活性。提供数据集成、应用集成、流程集成、界面集成等技术手段,实现各功能的总集成。

以数据整合为核心,搭建国家重要监测数据整合平台(图2),平台功能包括:①平台介绍,整体介绍平台功能以及数据情况;②分类数据库,提供原数据的表格、统计图和地理信息系统(GIS)地图的查看,以及数据下载;③数据整合,根据数据整合规则,对多类数据进行交叉查询,提供查询结果以表格和

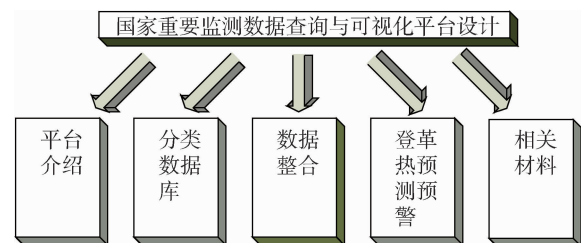


图2 国家重要监测数据整合平台设计

Figure 2 Design of the national important monitoring data integration platform

空间地图方式查看以及结果下载,并提供数据变化监测GIS地图的可视化;④登革热预测预警,根据模型方法,调用平台数据,展现登革热预测结果;⑤相关材料,提供材料的上传下载。

2 结果

2.1 平台介绍 国家重要监测数据平台涵盖内容,所拥有的数据和模块见图3。

2.2 分类数据库 就重要病媒生物、重要病媒传染病、重要检疫性入侵生物和气象数据,根据时间、空间范围的设定,输出查询结果表,并针对输出结果,制作统计图和GIS地图(图4)。就病原微生物名录和生物安全组学数据,采用关键字段,如病毒名进行查询,输出查询结果表。就空间环境数据,如人口和土地利用分类,输出栅格值阈分级图。

2.3 数据整合 数据整合功能是平台的核心功能,



图3 国家重要监测数据整合平台介绍

Figure 3 Introduction of the national important monitoring data integration platform



注:GIS. 地理信息系统

图4 国家重要监测数据库输出

Figure 4 Export of the national important monitoring database

数据整合实现过程见图5。在数据查询窗口选定时间范围、空间范围和数据类型(图5A),得出数据联合查询结果表(图5B)。根据结果表数据,实现该类数据在该时空范围内变化监测可视化(图5C)。根据查询结果表数据、生物学属性,查询病原微生物信

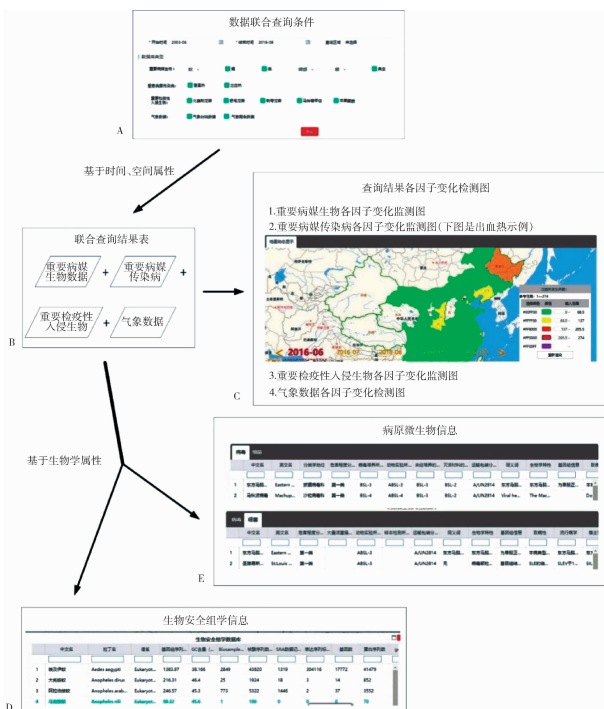


图5 国家重要监测数据整合可视化

Figure 5 Visualization of national important monitoring data integration

息(图5D)和生物安全组学信息(图5E)。

2.4 登革热预测预警 登革热预测预警模块功能的实现贯穿了数据调用、方法选择和模型实现,见图6。可调用前1个月的蚊虫数据、登革热发病数据、气象数据,预测当月的登革热病例数。图中实例选择2010年1月至2015年12月的数据,根据零膨胀与广义加性模型构建登革热预测模型,并预测2016年1月登革热发病数。



$$D_t = \lambda \{ [b_0 + g_1 \ln(D_{t-1} + 1)] + g_2 (V_{t-1}) + g_3 (T_{t-1}) + g_4 (P_{t-1}) \} \times \exp \{ b_1 + g_5 \ln(D_{t-1} + 1) + g_6 (V_{t-1}) + g_7 (T_{t-1}) + g_8 (P_{t-1}) + \delta_t \}$$

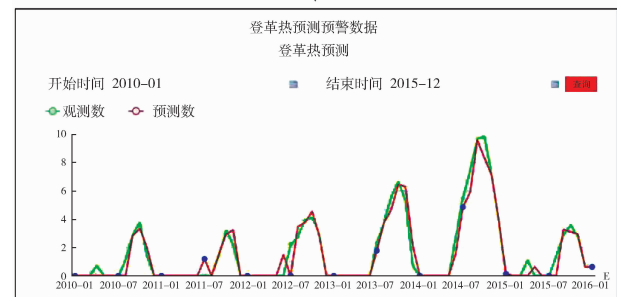


图6 登革热预测预警可视化

Figure 6 Visualization of dengue fever prediction and early warning

3 讨论

根据国家重要监测数据的时间属性、空间属性和生物学属性,建立数据整合规则,为国家重要监测数据的集成、共享、交换等奠定了技术基础;将多源、多尺度、多平台、动态化的国家重要监测数据进行抽取、清洗以及转换成覆盖面广、结构统一、标准统一、内容完整的“数据中心”,为国家重要监测数据库建设提供强有力的数据基础。以数据整合功能为核心,创建国家重要监测数据整合平台,为病媒及病媒传染病等科学研究提供可靠的数据支持与保障。GIS技术的发展将给国家重要监测数据整合平台提供更多、更新、更优的技术手段。

参考文献

[1] 李伟华,郑彦宁,刘志辉. 国内外数据整合研究进展分析[J].

- 数字图书馆论坛, 2015(6):54-61. DOI: 10.3772/j.issn.1673-2286.2015.06.009.
- [2] 黄华南. 基于GIS的国土房产数据整合方案设计[J]. 测绘通报, 2007(10):64-66. DOI: 10.3969/j.issn.0494-0911.2007.10.020.
- [3] 张李军, 刘寿福, 黄东海, 等. 南京市城乡地籍数据整合若干问题探讨[J]. 安徽农业科学, 2015, 43(1):295-297. DOI: 10.3969/j.issn.0517-6611.2015.01.101.
- [4] 荣芳. 基于数据整合的地籍信息系统设计与实现[J]. 地理空间信息, 2015, 13(6):36-37, 43. DOI: 10.3969/j.issn.1672-4623.2015.06.012.
- [5] 刘志杰, 金秉福, 张瑞端, 等. 海底沉积物碎屑矿物数据整合技术研究[J]. 海洋通报, 2015, 34(6):657-662. DOI: 10.11840/j.issn.1001-6392.2015.06.008.
- [6] Ma Z, Hart MM, Redmond RL. Mapping vegetation across large geographic areas: integration of remote sensing and GIS to classify multisource data [J]. Photogramm Eng Remote Sens, 2001, 67(3):295-307.
- [7] 李勇, 包世泰, 周品, 等. ECDIS与GIS数据集成研究[J]. 测绘科学, 2007, 32(4):135-137. DOI: 10.3771/j.issn.1009-2307.2007.04.050.
- [8] Karnatak HC, Shukla R, Sharma VK, et al. Spatial mashup technology and real time data integration in geo-web application using open source GIS-a case study for disaster management [J]. Geocarto Int, 2012, 27(6):499-514. DOI: 10.1080/10106049.2011.650651.
- [9] Salleh SA, Hamid JRA, Ariffin IM. Investigation of potential integration of spectroradiometer data with GIS technology: the spectro-GIS tools [J]. IOP Conf Ser Earth Environ Sci, 2014, 18(1):012053. DOI: 10.1088/1755-1315/18/1/012053.
- [10] Goble C, Stevens R. State of the nation in data integration for bioinformatics [J]. J Biomed Inform, 2008, 41(5):687-693. DOI: 10.1016/j.jbi.2008.01.008.
- [11] Seoane JA, Aguiar-Pulido V, Munteanu CR, et al. Biomedical data integration in computational drug design and bioinformatics [J]. Curr Comput Aided Drug Des, 2013, 9(1):108-117. DOI: 10.2174/15734099112089990011.
- [12] Fagan A, Culhane AC, Higgins DG. A multivariate analysis approach to the integration of proteomic and gene expression data [J]. Proteomics, 2007, 7(13):2162-2171. DOI: 10.1002/pmic.200600898.
- [13] Rezola A, Pey J, Tobalina L, et al. Advances in network-based metabolic pathway analysis and gene expression data integration [J]. Brief Bioinform, 2015, 16(2):265-279. DOI: 10.1093/bib/bbu009.
- [14] de la Prieta F, Rodríguez S, Bajo J, et al. Data integration in cloud computing environment [C]//Proceedings of the 4th International Workshop on Knowledge Discovery, Knowledge Management and Decision Support. Mazatlan, Mexico: Atlantis Press, 2013:180-185.
- [15] 邹晓妮, 吴菲, 李静静, 等. 医院内法定传染病信息推送平台开发与实现[J]. 中国卫生信息管理, 2014, 11(6):596-601.
- [16] 李梅, 苗润莲, 张岸, 等. 基于GIS的京津冀科技资源数字地图服务平台构建[J]. 现代情报, 2017, 37(6):172-177. DOI: 10.3969/j.issn.1008-0821.2017.06.028.
- [17] 唐健, 沈陈华, 周国峰, 等. 国土资源数据整合方案设计及其实现研究[J]. 中国土地科学, 2009, 23(9):72-78. DOI: 10.3969/j.issn.1001-8158.2009.09.014.

收稿日期: 2019-02-26 (编辑: 陈秀丽)