植物标本信息共享与整合——以广西植物标本馆为例

沈晓琳，赵志国，刘 演*

（广西壮族自治区 中国科学院 广西植物研究所，广西 桂林 541006）

摘 要：以植物标本资源整合共享为主线，寻求和探索“信息孤岛”现象的原因，进一步体现共享整合的价值。通过开展深入调研，找出存在的问题和原因，并采取灵活多样的整合方式和共享模式，来实现植物标本信息资源的高效利用。

关键词：整合；共享；植物标本；资源

中图分类号：G262 **文献标识码**：A **文章编号**：1000-3142(2010)06-0899-04

Information sharing and integration of plant specimens——a case in IBK

SHEN Xiao-Lin, ZHAO Zhi-Guo, LIU Yan*

(*Guangxi Institute of Botany, Guangxi Zhuang Autonomous Region and the Chinese Academy of Sciences*, Guilin 541006, China)

Abstract: This article took plant specimen resources sharing and the conformity as a master line, seeked and explored the reason of“the information isolated island”phenomenon, further manifested the value of sharing conformity. Through thorough investigation and research, the authors discovered the existing questions and the causes, realized the highly effective use of plant specimen information resources by adopting the nimble diverse conformity way and the sharing pattern.

Key words: integration; sharing; herbaria specimens; resources

植物标本是植物学家长期从事科研活动的积累和人类自然遗产的永久记录之一，是研究物种的分布及其历史、现状、系统演化的证据。广西植物资源丰富，维管束植物就有 8 354 多种，物种数量仅次于云南、四川而位居我国第三位，是发展经济、建设生态广西得天独厚的重要战略资源。随着社会的发展、年代的积累，馆藏的标本数量日益增多(梁畴芬等，1985)，广西植物标本馆馆藏标本已达 40 余万份；而全国其他植物标本馆也相应地收藏了大量植物标本，如何能从海量信息中，快速、准确、便捷地从一个查询界面上能获取全部信息成为一个急需解决的难点问题。为了解决信息孤岛问题，把分散的、单一的标本数据库通过网络技术和相应的技术规范标准；由中科院植物标本馆牵头，各地方标本馆配合，构建数字植物标本馆是为了更好地整合植物资源信息共享，也是相关科研人员必不可少的，是科技创新体系的重要组成部分，对提升产业创新能力、促进全社会科技进步具有重要意义。

1 数据平台构建的基础

1.1 广西植物标本数据库构建情况

广西植物标本馆(IBK)创建于 1935 年，是我国建立较早的植物标本馆之一，在全国 318 个植物标本馆收藏量的排名中居第七位，为全国十大标本馆之一(傅立国，1993)，总建筑面积为 2 000 m^2。广西

收稿日期：2009-09-15 **修回日期**：2010-12-05

基金项目：广西科技攻关项目(桂科能 0815011-6-2)；广西植物研究所基本业务费项目(桂植业 09002)[Supported by the Key Technologies Research and Development Program of Guangxi(0815011-6-2); Fundamental Research fund of Guangxi Institute of Botany(09002)]

作者简介：沈晓琳(1975-)，女(黎)，广西全州人，硕士，工程师，主要从事植物资源信息管理工作，(E-mail)sxl@gxib.cn。

* 通讯作者(Author for correspondence, E-mail: liuyan@gxib.cn)

植物标本馆的标本来源以广西各地为主，经过几代植物学家的艰苦奋斗，广西植物标本馆馆藏维管束植物标本已达40万份(沈晓琳等，2005)，共收录国内外植物物种约12 000种，涵盖了广西植物8 000多种，其中模式标本4 000余份，涉及150科1 100余个分类群(刘演等，2000)，馆藏标本年代最早的采集于1889年。其中尤以石灰岩地区的植物标本最为齐全，占本馆馆藏标本的30%，是全国馆藏广西石灰岩石山地区植物标本最多的植物标本馆(林春蕊等，2008)。在石灰岩石山地区植物研究领域中，具有不可替代的地位。此外，标本馆还广泛收集华南、西南和东北等其他各省区的标本，并同时积极与周边国家和其它国家的标本馆建立标本交流关系，收藏了美国、英国、日本、印尼、新西兰、越南等国的部分标本。目前已完成了20万份维管植物标本信息数据库的建立，正逐步实现标本数字化、网络化。把网络共享整合数据库技术引入植物标本的管理与应用，有助于提高对植物资源的理解与利用，实现资源共享。这使得标本馆从传统的以标本借阅为主的单功能服务向以信息和知识收集、传播发布、检索为主的多功能服务转化，由此从“被动式”服务转向“主动式”服务，使植物标本馆充分发挥了效益。

1.2 中国数字标本馆数据库构建情况

“中国数字植物标本馆(Chinese Virtual Herbarium，简称CVH”网站(www.cvh.org.cn)是在科技部“国家科技基础条件平台”项目资助下建立的，其宗旨是为用户提供一个方便快捷获取中国植物标本及相关植物学信息的电子网络平台。CVH建设的目的包括：(1)提供中国植物标本及相关植物学的全面和最新的信息，供专家及一般用户上网查询；(2) 为国内同行间交流与合作提供平台，并实现与国际接轨；(3)提供政府及民间对植物多样性保护和可持续利用的参考资料；(4)促进参与标本馆的现代化管理建设进程。最终目标是把CVH建设成为中国植物标本信息及植物学科的国家型门户网站。

网站提供国内主要标本馆数字化标本信息，包括一般标本及模式标本，每份标本信息包括标签信息及图像信息，前者包括标本采集人、采集日期，地点、生境与海拔以及科名、种名和鉴定信息及标本存放地点(标本馆)等信息。模式标本还附有原始文献pdf文件及较高画质影像。网站还提供了与分类学研究及相关领域的数据库，包括标本采集地名库、模式标本文献库、植物名称及分布、植物名称作者及其论著目录以及《中国植物志》和《中国高等植物图鉴》的电子图书等。目前CVH网站包含数据库20余个，数据量达3.3 TB。参与建设单位(共建单位/成员单位)达20余家，包括中国科学院和地方科学院及一些大学标本馆，基本上包含了我国主要和重要的标本馆。

2 数据共享的技术方法

2.1 数据共享的技术方法

全国各标本馆都有自己各自的平台，所使用的系统和数据库各不相同，但是各成员馆必须使用统一的标准，其共享数据标准采用Darwin Core V1.4，共享系统推荐采用TapirLink，或者支持导出为Darwin Core格式的系统；数据交换采用通用的XML语言作为标准；数据资源共享采用资源目录以文件或数据库形式提供资源共享；整个平台使用webservice技术、ASP.NET技术、异步调用技术来解决各标本馆彼此之间的差异，使得标本数据可以无平台差异、无数据库差异的完美组合。在信息交换与共享的技术实现方面，采用了J2EE体系结构，利用JSP、资源目录、数据交换等技术实现(王卫玲，2007)。所有“CVH”系统中的数据，其开放程度由各成员馆决定，但至少要提供Darwin Core中规定的7个必需字段和省份字段。

2.2 数据共享的网络结构

共享整合数据库是由一组数据库组成，这些数据库物理地分布在计算机网络的不同结点或者场地上，而在逻辑上这些数据库组成一个整体、一个系统，即构成一个共享整合数据库。共享整合数据库使用计算机网络将地理位置分散而管理和控制又需要不同程度集中的多个逻辑单位(通常是集中式数据库系统)连接起来，共同组成一个统一的数据库系统。

该系统由一台主服务器和各地标本馆客户端及相应的Web服务网站构成。由主服务器调用各节点服务器上的数据，调用是异步进行的，不会因为某一节点不通而影响整个系统的正常运行。各节点数据在主服务器上汇总，将结果返回给WEB用户。为了保证整个系统运行的可靠性和安全性，降低网络拥塞可能带来的系统迟滞和不稳定性，提高客户端处理的效率，该系统数据库分别放在各自的局域网内部，彼此相互独立，通过Web Service穿越防火墙与外网交换信息。不能通过Internet网络直接访

问(沈晓琳等,2008),网络结构如图 1 所示。

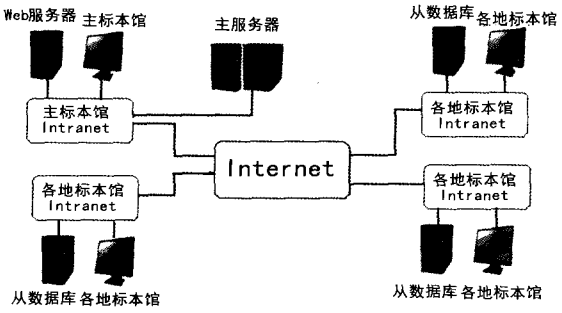

图 1　数据共享网络结构

Fig. 1　Network structure of data sharing

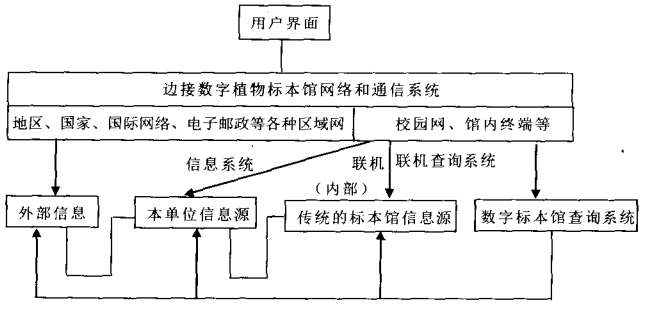

图 2　数字植物标本馆的模式

Fig. 2　The model of digital Herbarium

2.3 植物标本数据共享模式

植物标本数据共享模式是一个开放式的硬件和软件集成平台,通过数码摄象机和数码像机把植物的各种特性数字化,并通过计算机技术和网络技术把植物标本的各种特性全方位地集成在网上服务。在这个工程中需要较大的软件工程、网络工程、计算机工程、信息组织工程、面向市场的用户运营等部分的有机组合。按照"整合、集成、共享、提升"的基本思路,构建具有基础性、公益性、开放性等特点的标本信息库(陈三茂等,2003)。将有效地推动数字植物标本馆的应用,为研究生物多样性及植物现时的、历史的特性提供一个快捷查询平台,其共享模式如图 2 所示。

3　实现共享结果与分析

3.1 共享系统分析

各标本馆数据库服务器作为该标本馆局域网上的一个网络结点,在服务器上安装相应的数据库程序(如广西植物标本馆安装的是 SQL server 2000 数据库),该数据库中存放着该标本馆的本地数据,不同标本馆之间的访问是访问放置于各标本馆数据库服务器上,通过分布式事务或复制来保持数据的一致性;另外,各标本馆数据库服务器在主标本馆数据库服务器上必须注册,才能实现各标本馆对总部数据库服务器的访问。

采用这种方案,一方面,减少网络上本地数据库服务器的负载,降低了不同标本馆之间互访的复杂性;另一方面,即使整个系统中有某个标本馆出现故障,系统仍能运行,提高了系统的可靠性,由于各标本馆数据库在主标本馆数据库服务器上都有备份,所以一旦某站点出现故障数据损坏,通过 SQL Server 的复制很容易实现数据的迅速恢复(张庆莉等,2007)

3.2 共享运行分析

整合植物标本数据库目前参与测试并采用这种方式接入的单位有 20 余个,可检索总数据量约 250 多万条记录(http://pe.ibcas.ac.cn/sptest/syninvok.aspx),这 20 余家单位的植物标本数据已经实现共享整合,各单位使用的操作系统和数据库不完全相同,各成员馆使用的数据库类型有:SQL Server2000、MS Access 和 MY SQL。但各单位的整合数据库运行情况良好,从调试至今,日均独立 IP 稳定在 2 万以上;年均访问量超过千万;访问前五名域名是:成都生物所、福建省福州市、兰州大学、北京植物所、中科院西北高原研究所;查询前三名的属分别为 *Blumea*(166)菊科、*Clematis*(19)毛茛科、*Schisandra*(16)木兰科;查询前三名的省:四川(15487),甘肃(1437),西藏(1133);访问人数最多的前 10 份标本如表 1 所示;目前系统运行稳定,查询速度快而准,可扩展性强,易维护。

3.3 实现共享的意义

植物标本蕴含的信息是研究植物系统分类、区系、进化、种群、群落等的基本资料,从中可提取物种现时及历史上的分布特征、濒危植物的历史和现状,而现代生物多样性保护的正确决策在很大程度上也有赖于对已知标本信息的综合概括能力(包秀艳,2007);植物标本信息资料越全,统计出来的信息就越精确,对鉴定物种、分析物种多样性、研究物种分布、系统发育、演化等提供第一手数据源,共享整合之前查询的标本只能是一个单一的数据库中的数据,不能调用其他标本馆的相应的数据,如在广西植物标本馆网站(http://www.gxib.cn/WebSetup/VHAdvQuery.aspx)上查询的植物标本只有 20 多万份标本资源,且只是本馆的标本数据,不能同时调

用其他标本馆的标本数据，而共享之后在中国数字标本馆（http://pe.ibcas.ac.cn/sptest/syninvok.aspx）上查询，可以互访、互通、协调统一；目前已有250多万份标本资源共享上网查询，共享整合后的数据库使植物标本信息远远大于单一标本馆的信息，且查询速度快，资料全，不会因某一网站瘫痪而影响查询资料；通过数据整合，可以快速找到各自所需的资料，不受地域和时间限制，随时随地都可以查询全国各地标本馆的植物标本资源。这样可以减少人力物力的浪费，还可以有效地保护植物标本的损坏；同时可以提高本馆的知名度。所以通过对植物标本信息共享整合，将为从事植物方面研究及相关领域研究的人员提供一个简便、快捷、全面的查阅植物标本的技术平台，同时可以充分有效地开发和利用植物标本信息资源，能有效地保护馆藏的实物标本。

表 1 访问人数最多的前 10 份标本

Table 1 The top 10 specimens with the most visit records

访问数 Visit number	馆代码 Hall code	条形码 Barcode	采集人 collecter	采集号 Collect code	拉丁名 Latin name	采集年份 Collect year
779	IBK	IBK00059649	梁向日	68787	*Leucas zylanica*	1936
520	IBK	IBK00185083	弄岗综考队	20263	*Leucas zylanica*	1979
335	IBK	IBK00059650	梁向日	68787	*Leucas zylanica*	1936
332	IBK	IBK00059651	海南东路队	295	*Leucas zylanica*	1954
313	IBSC	0173005	M. Sousa	8390	*Amicia zygomeris*	1977
304	IBSC	0354568		7658	*Ficus formosana*	1952
265	IBSC	0517259	A. Kolakovsky	3245	*Anthemis zygia*	1935
258	PE	00098709	Y. Y. Pai	46	*Oxytropis glabra*	1933
252	IBK	IBK00059655		32998	*Leucas zylanica*	
243	IBK	IBK00059662	黄志	35474	*Leucas zylanica*	1933

4 讨论

随着计算机技术在生物科学中的渗透和应用，基于在植物研究的大量原始材料，以及该领域几代人努力的丰硕成果，整合各类生物资源信息的共享平台是科技创新体系的重要组成部分，是服务于科技创新与产业发展的物质基础，对提升产业创新能力、促进全社会科技进步具有重要意义，数字植物标本馆是一个新生事物，它不可避免地存在一些缺陷，例如保存问题、网络安全问题等；目前CVH的标本信息是通过集中式实现共享查询的，其主要缺陷是更新周期长，存储压力大。“中国数字植物标本馆”仍在不断建设和完善之中。现阶段主要任务是在增加数据库记录数和提高数据质量。通过对平台、网络、数据库、应用以及终端等一系列的整合，提高信息化的效益，以共享机制建设为核心，按照“整合、集成、共享、提升”的基本思路，构建具有基础性、公益性、开放性特点的科技基础条件平台，整合而成的植物资源信息共享平台已基本完成，由于该平台还处在调试阶段，后台程序还没完全定格，还需进一步完善，功能还需在使用中不断提升。

参考文献：

包秀艳. 2007. 试论保护生物多样性的意义及措施[J]. 赤峰学院学报(自然科学版)，**23**(5)：33－34 张庆莉，周红雷. 2007. 现代校园一卡通系统的设计[J]. 河南科技学院学报，**35**(3)：94－96

傅立国，等. 1993. 中国植物标本馆索引[M]. 北京：中国科学技术出版社，366－371

Chen SM(陈三茂)，Tian YL(田晔林). 2003. Digital herbarium, the trend of herbarium development in the 21st century(21世纪植物标本馆的发展方向——数字植物标本馆)[J]. *J Beijing Agric Coll*(北京农学院学报)，**18**(3)：208－210

Li CF(梁畴芬)，Huang GB(黄广宾)，Lu YX(陆益新). 1985. Guangxi's punt resources being exploited(在开发利用中的广西植物资源)[J]. *Guihaia*(广西植物)，**5**(3)：227－243

Lin CR(林春蕊)，Liu Y(刘演)，He CX(何成新)，*et al*. 2008. Statistics and analysis of digital information in the herbarium of Guangxi Institute of Botany(广西植物标本馆标本数字化信息统计与分析)[J]. *Guihaia*(广西植物)，**28**(2)：278－284

Liu Y(刘演)，Wen HQ(文和群)，*et al*. 2000. A multimedia information system for the maineconomic plants in Guangxi 广西主要经济植物的多媒体信息系统[J]. *GuiHaia*(广西植物)，**20**(1)：94－96

Shen XL(沈晓琳)，Yu X(于翔)，Guo LF(郭伦发). 2005. Design and implement of network information system of IBK(广西植物标本馆网络信息系统的设计与实现)[J]. *J Guangxi Acad Sci*(广西科学院学报)，**21**(2)：110－112

Shen XL(沈晓琳)，Zhang XL(张向利)，*et al*. 2008. Distributional plant specimen database design and realization(分布式植物标本数据库的设计与实现)[J]. *Computer & Telecommunication*(电脑与电信)，**150**(8)：20－21

Wang WL(王卫玲). 2007. Research on web services integration basesd on SOA(基于SOA的Web Services集成技术研究)[D]. 广西大学硕士论文，**28**：22