

DOI: 10.11931/guihaia.gxzw202209068

陈天翔, 刘慧圆, 韩国霞, 等, 2022. 植物子平台及中国数字植物标本馆(CVH)运维: 做法和经验 [J]. 广西植物, 42(增刊 1): 18–28.

CHEN TX, LIU HY, HAN GX, et al., 2022. Curating and administrating the Plant Specimen Sub-Platform and CVH: activities, experience and lessons learned [J]. *Guihaia*, 42(Suppl. 1): 18–28.



植物子平台及中国数字植物标本馆 (CVH) 运维: 做法和经验

陈天翔¹, 刘慧圆¹, 韩国霞¹, 包伯坚¹, 李敏¹, 许哲平^{2,3,4*}, 覃海宁^{1*}

(1. 中国科学院植物研究所系统与进化植物学国家重点实验室, 北京 100093; 2. 中国科学院文献情报中心智能情报实验室, 北京 100190; 3. 学术期刊新型出版与知识服务重点实验室, 北京 100190; 4. 中国科学院大学经济管理学院, 北京 100190)

摘要: 该文系统介绍了植物子平台所采取的以数字标本质量为导向的数字化技术规范和管理策略, 以及 CVH 网站数据共享规则, 并指出存在的问题及今后努力方向。

关键词: 数字标本, 技术规范, 质量控制, 管理策略, 共享原则

中图分类号: Q94 文献标识码: A 文章编号: 1000-3142(2022)增刊 1-0018-11

Curating and administrating the Plant Specimen Sub-Platform and CVH: activities, experience and lessons learned

CHEN Tianxiang¹, LIU Huiyuan¹, HAN Guoxia¹, BAO Bojian¹,
LI Min¹, XU Zheping^{2,3,4*}, QIN Haining^{1*}

(1. *State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China;*
2. *Key Laboratory of Intelligent Information, National Science Library, Chinese Academy of Sciences, Beijing 100190, China;* 3. *Key Laboratory of New Publishing and Knowledge Services for Scholarly Journals, Beijing 100190, China;* 4. *School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China*)

Abstract: There is a account in this paper on the technical criteria and management policies of digitized specimens used in the Plant Specimen Sub-Platform as well as the sharing principles used in CVH, and the existing problems and challenges are described.

Key words: digitized specimens, technical criteria, quality control, management policies, sharing principles

收稿日期: 2022-09-12

基金项目: 国家植物标本资源库项目 (E0117G1001)。

第一作者: 陈天翔 (1996-), 助理工程师, 主要从事生物多样性数据库建设和共享平台开发与运维, (E-mail) chentx@ibcas.ac.cn。

*通信作者: 覃海宁, 博士, 主要从事植物分类学、生物多样性信息化和生物多样性保护研究, (E-mail) hainingqin@ibcas.ac.cn; 许哲平, 博士, 副研究馆员, 主要从事科学数据、科技情报和开放科学研究, (E-mail) xuzp@mail.las.ac.cn。

1 前言

过去十余年间,植物子平台课题组致力于组织全国科研院所标本馆及部分大学标本馆开展标本数字化工作,最终形成了覆盖百余家植物标本馆近 800 万份标本的数字化标本数据库,并建立中国数字植物标本馆 (Chinese Virtual Herbarium, CVH) 网站,实现数据在线共享,成为国家标本资源共享平台中的骨干子平台和重要门户网站。

作为姐妹篇,前文(刘慧圆等,2022)介绍了植物子平台暨 CVH 作为国家标本资源共享平台项目的主要课题所经历的发展过程和对我国生物信息共享事业发挥的作用,本文则主要介绍植物标本子平台及其共享网站 CVH 运行管理的技术特点,尤其是子平台作为课题操作的管理方式和制订及推广的技术规范等。

高质量的标本数据库是实现信息共享的基础

和开展标本馆现代化管理等活动的重要依据,而标本数字化技术流程及管理策略既是实现高质量数据的路径,也是提高工作效率、实施数字化项目的保障。欧美同行无论是开展单一标本馆的数字化 (Haston et al., 2012; Tulig et al., 2012; van Oever & Gofferjé, 2012) 还是整个国家层面的数字化 (Barkworth & Murrell, 2012; Borsch, 2020) 都十分重视数字化技术流程的设计。我国科学家在 20 世纪数字化早期工作中就已经采用正确合理的技术和方法(凌萍萍和汤傲杉,1995;李鸣光和余萍,2007)。

正如刘慧圆等(2022)指出,21 世纪初国家科技基础条件平台在启动时,标本资源平台项目就在考察分析国内外标本数字化及其共享经验基础上,对国家战略性植物标本资源的收集、保藏及数字化和共享系统进行梳理,形成顶层设计,包括制订国家标本资源共享平台建设框架、主要技术路线(图 1)及一系列技术规范 and 标准。

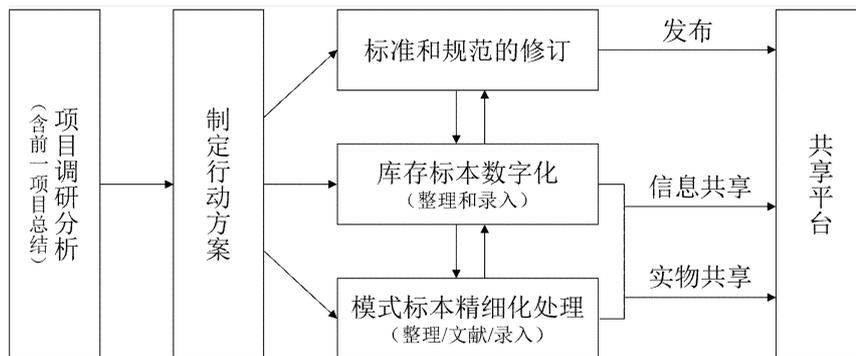


图 1 早期标本数字化项目技术路线

Fig. 1 Technical route of early specimen digitization project

自 2004 年国家自然科技资源共享平台“植物标本描述标准和规范的修订及共享试点”项目启动,2006 年中国数字植物标本馆网站上线,到 2014 年植物子平台(课题)成立直至 2019 年完成历史使命,子平台暨 CVH 在十余年中一直遵循上述主体思想和技术路线进行建设,并根据数字化过程中出现的新情况进行修订、完善和调整,形成良性循环,促进数字化及其共享工作平稳有序推进,从而在项目管理、技术支撑等方面形成了一套

完整的标本数字化平台运作模式。许哲平和赵莉娜(2010)和许哲平等(2012)介绍了在 CVH 早期建设中尝试引入国际主流生物多样性信息学标准以对 CVH 数据挖掘和集成,并尝试与全球生物多样性信息共享建立联系。本文主要介绍植物标本子平台课题的管理模式以及为保障数字标本质量所采取的技术路线和技术服务等内容。

1.1 数字化标准和规范先行

标本资源共享平台项目组基于“资源共享,制

度先行”的原则,在开展大规模标本数字化工作之前,优先着手制订植物标本数字化相关标准和规范。主要是在2004、2005两年间制订了36项标准和规范,内容涉及标本质量评价标准、标本资源信息描述规范、标本数据库建设规范等,并邀请分类学家制订高等植物各个类群的标准名录、物种描述规范以及标本采集、制作及保藏规范等。这些标准和规范的讨论和制订,为后期开展大规模标本数字化提供了工作框架和范本以及技术标准,为实现植物标本信息共享奠定了坚实的基础。

1.2 标本数字化共享试点

国家标本资源共享平台在2004、2005两年间,以“植物标本描述标准和规范的修订及共享试点”为题目开展工作。一方面制订标本数字化规范和标准,另一方面选择代表性标本馆开展试点工作,积累经验,掌握实际的数字化流程。试点单位主要为中国科学院系统标本馆和几所重要大学标本馆,均为历史较长、馆藏量较大和技术力量较强的标本馆,两年数字化共计150多万份植物标本。项目组(课题组)通过前期小规模试点,掌握了组织多家标本馆基于同一技术方案开展标本数字化及其信息共享的基本工作方法,以及可能碰到的困难和解决办法。

1.3 标本数字化建设高峰时期

2006年以后,植物子平台进入标本数字化高峰时期。这一时期分为两个阶段。第一个高峰阶段是2006—2008年,主要完成中国科学院系统标本馆和部分地方科研院所(园、馆)标本馆的数字化,这也是项目原初给植物子平台分配的数字化范围,以及少数几所大学标本馆等,共计28家,完成331万份普通标本和1.7万份模式标本的数字化表达(许哲平和赵莉娜,2010)。第二个高峰为2014—2019年,共完成230万份标本的数字化及共享,参建单位达92家。第二期92家参建单位(标本馆)中,除了涵盖前期建设未及的大部分地方科研院所外,还包括50所高校标本馆。92家参建标本馆中,74家是2014年后首次参加数字化工作,其中59家为馆藏10万份以下的标本馆,包括小型标本馆(3万~10万份)41家、微型标本馆(3万份以下)18家,这些标本馆中不乏标本保藏有特

色的标本馆,如重庆自然博物馆(CQNM,2.9万份)、山东大学(JSPC,1.3万份)、福建省药品检验所(FJIDC,2万份)、四川农业大学小麦研究所(SAUT,0.6万份)等(刘慧圆等,2017;覃海宁等,2019)。可以说,这是近二十年标本集中数字化进程中,涉及标本馆数量最多、影响面最广的时期。

截至2019年,子平台组织全国100余家植物标本馆完成近800万份标本的数字化表达及共享任务(刘慧圆等,2022)。

2 数据质量管理策略

植物子平台按照国家科技基础条件平台统一部署,在实施对全国植物标本数字化的进程中,积极探索,逐步形成了以数字化增量为目标、质量为核心的项目(课题)管理机制和技术方法,主要是通过严格把关/签订任务书、抓中期进展和严格清查结题数据质量的“两头严、中间抓”三步策略来实现,并实施课题启动前的摸底调查和实施过程中的技术培训及保障护航服务。

2.1 调研并掌握本底情况

实体植物标本馆的式微(衰落),如缺乏活力、标本馆信息更新不及时、维护人手及经费不足甚至缺乏等,是世界性的难题(Barkworth & Murrell, 2012; Thiers, 2018; Borsch, 2020),中国也不例外(覃海宁和殷学波,2003a;覃海宁等,2019;葛斌杰等,2020)。中国植物标本馆数量众多,有300多家植物标本馆,馆藏量达到2000多万份,是全球标本馆藏量大国(Thiers, 2018;覃海宁等,2019),但绝大多数标本馆成立时间不超过百年,80%是1950年后成立的,标本馆(对外发布)信息陈旧、缺乏管理员、处于“休眠”关门状态等现象时有发生,有些还相当严重(Qin, 1999;覃海宁和殷学波,2003b;覃海宁等,2019;葛斌杰等,2020)。

显然,对中国植物标本馆资源进行全面系统的调查,了解各家标本馆的馆藏状况、数字化及利用状况和人力资源状况等,对开展全国性标本数字化工作具有重要的意义。早在2000年,我们针对《中国植物标本馆索引》(傅立国,1993)出版后的状况对全国100家标本馆做了调查(覃海宁等,

2002;覃海宁和殷学波,2003a);2003年我们根据科技基础条件平台建设的基本要求,对我国植物标本资源的保存、利用和资源管理等方面的信息展开了较为全面、系统的调查,并建成我国标本资源指南数据库,为启动科技基础条件平台提供前提保障,为标本资源共享平台建设以及信息共享提供决策依据(贾渝等,2005;曹一化等,2006;杜占元等,2007)。

国家标本资源共享平台(NSII)经过10年建设,到2013年时共完成植物标本数字化500余万份并实现共享。此时,平台的发展面临着一个重要的瓶颈时期。基础条件较好的大中型标本馆基本完成大部分标本的数字化工作,没有数字化的标本大多为标本质量参差不齐、缺乏人手的小型微型标本馆中的标本,以及大中型标本馆中的疑难标本。结果导致数字化难度增大,包括标本数字化成本提高、质量下滑、增量放缓以及合格参建单位(标本馆)减少等(刘慧圆等,2017,2022)。

植物子平台在此时启动了新一轮的标本资源调查工作,以突破上述发展瓶颈。此次调查以实地调查为主,由子平台负责人、技术员等人员组队,重点走访未来数字化潜力较大和目前工作存在困难较大的标本馆。调查队每到一地(馆)都详细了解馆藏标本状况,如标本数量、产地及采集时间段、定名情况、保存条件、数字化比例,以及管理技术力量等,并现场教学和解答数字化技术相关问题。2014至2018五年间共出队20次抵达15座城市调查走访50家植物标本馆(表1,封三图片)。

经过调查走访,子平台不仅现场解决了各家标本馆数字化技术上遇到难题和相关疑问,而且掌握了这些标本馆的实时现状,尤其是标本资源状况的第一手材料。我们依据这些最新信息评估各家标本馆的标本重要程度、数字化能力及参与课题的时间点,制订数字化短期(1年)、中长期(3~5年及5年以上)目标和任务。在实际操作中,我们结合对标本馆现况的调查结果,对每年参建单位申报数字化数量进行必要的调整,从源头上确保数字化任务稳步进行,同时为了增加标本的特色和代表性,也给新单位参加数字化建设预留了机会。

2.2 严格签订年度任务书

植物子平台参建单位(又称CVH成员馆)一年一(次)签专题任务书。通常,国家标本资源共享平台下达任务后,植物子平台通过自愿报名和邀请参加两种途径并结合调查摸底,确定年度数字化参加单位,签订年度专题任务书。专题任务书经过多年修订趋于完善。其内容涵盖甲方(子平台)、乙方(参加单位)责任和权利,如标本数字化数量、完成时段及经费额度和拨款时间等,并以考核指标的形式对标本数字化质量作了明确规定,要求选择“三有标本(有花或果、有采集记录签和鉴定签)”进行数字化,并对标本图像尺寸、标本信息录入字段要求和错误率允许值等,均作了明确的规定。子平台工作人员常常为了考核指标落地,反复多次与参建单位联系和说明,确保任务书落实到位。

2.3 抓好中期检查

中期检查是了解和掌握子平台年度参加单位工作进展的最佳办法,通常是在年中举行。各家参建单位按照任务书要求,向子平台提供标本数字化任务量的二分之一(数据,包括图像)接受审查。子平台将审查中发现的问题反馈给参建单位,要求补充、修改和完善,并于年终结题验收时一并提交,子平台同时将中期检查结果存档,作为年终专题验收的参考材料之一。

子平台通过开展当年任务的中期检查工作,全面掌握各个标本馆数字化任务进度以及面临的困难,提出解决问题方案,督促标本馆按计划完成任务。

2.4 严格把控年度验收数据

参建单位(标本馆)在完成年度任务后需要向子平台汇交标本数据和图像,进行验收。子平台需要对标本馆汇交的数据和图像进行审核,确保数据数量、质量符合标准和要求。数据在验收之前一般需要经过2次审核。第1次经子平台审核,抽取一定比例(任务量的5%)的数据和图像,通过计算机软件和人工审核相结合的方式审核数据和图像是否符合数据验收标准。数据和图像审核合格标准为95%,若低于此值,将返回原单位请求返修后重新提交验收。对于抽查数据不合格的,返修

表 1 2014—2018 年子平台 (课题组) 调研标本馆情况
Table 1 Herbaria surveyed by sub-platform research group (2014–2018)

批次 Batch	城市 City	调研标本馆 Surveyed herbaria	时间 Time	调查人员 Researcher
1	贵阳	贵州大学、贵州师范大学	2014.11	覃海宁, 刘慧圆
2	昆明、成都、 重庆、广州	中国科学院昆明植物研究所、云南大学、四川省林业科学研究 院等 11 家标本馆	2015.2	覃海宁, 林祁
3	北京	北京大学、北京中医药大学、北京自然博物馆等 5 家标本馆	2015.2、2015.7	覃海宁, 刘慧圆
4	南宁	广西壮族自治区药用植物园、广西壮族自治区中医药研究所、 广西壮族自治区林业勘测设计院	2015.4	覃海宁, 单章建
5	太原	山西大学、太原师范学院、山西中医学院等 4 家标本馆	2015.6	覃海宁, 单章建
6	南昌、福州	中国科学院庐山植物园、南昌大学、福建农林大学等 6 家标 本馆	2016.11	覃海宁, 谢丹
7	贵阳	贵州大学、贵州师范大学、贵州中医药大学等 6 家标本馆	2017.7	覃海宁, 李奕
8	兰州	兰州大学、西北师范大学	2018.4	覃海宁
9	南京	江苏省中国科学院植物研究所、南京大学、南京师范大学等 7 家标本馆	2018.5	覃海宁, 刘慧圆, 韩国霞, 李奕
10	呼和浩特	内蒙古大学、内蒙古农业大学、中国农业科学院草原研究所	2018.8	覃海宁, 刘慧圆, 李奕

后,第 2 次抽查数据量为原来的 2 倍,第 3 次抽查为第 2 次抽查数量的 2 倍,数据抽查 3 次后仍然达不到 95%的,下一年度将暂停该单位参加数字化分配任务。第 2 次,子平台将审查后的数据和图像上交至 NSII 项目组复审,重复上面的流程,如还有问题则直接返给参建单位进行修改。

植物子平台从两个维度对标本数据进行验收:一是验收数据数量是否满足任务要求,是否有往年重复提交的数据且复份标本的重复率不应超过 5%;二是验收数据质量是否合格,验收项目包括数据的正确性、准确性、完整性、一致性等指标。合格数据在格式和内容上应是统一、规范、准确的,并且数据内容与标本实物上的信息一致,也就是“忠于原文”原则。符合标准规范的数据易于整理、关联、分析、共享,能够保证平台数据管理、用户数据利用的良好体验。标本图像的审核在数量方面,要求每条标本数据记录必须对应至少一幅标本图像;质量方面,则依据发布的图像质量标准,对图片的分辨率、内容、拍摄参数进行审核,检

查图片是否能够还原标本的原貌。近年来,相机、扫描仪、硬盘和互联网等硬件的发展可以支持超高分辨率标本图像的获取、存储和在线共享,项目组也随之提升图像质量标准,以呈现更好的在线共享效果。

标本数据和图像验收的具体流程及内容见附件 3 (<http://www.guihaia-journal.com> 在线发表)。

3 数字化技术服务及技术支持

植物子平台为参建单位提供的技术服务和技术支持,主要围绕获得高质量数字标本数据及图像来开展。一方面是发放“标本数字化流程及技术指南”(附件 1, <http://www.guihaia-journal.com> 在线发表)和标本信息录入系统软件及其使用指南(附件 2, <http://www.guihaia-journal.com> 在线发表)供参建单位使用;另一方面是举行专门的数字化技术培训班,请熟练技术人员和分类学家授课,内容以数字化技术讲解和实际操作演示为主,并

涉及标本管理使用和分类学等相关知识,以全面提升标本馆数字化技术人员的业务水平。在项目实施期间,子平台技术人员会不断通过邮件、微信等联系方式为参建单位排忧解难和提供技术指导及服务。

3.1 发放数字化技术指南及标本信息录入软件

自标本资源共享平台启动建设以来,我们就一直注意建立一套标本数字化技术方法体系,并在实践中不断补充完善。2015年前,数字化技术方法组件,如标本数据录入规范、图像采集规程等文件都是分别发放给参建单位使用的,显得零碎,甚至出现组分之间相互矛盾的现象。2015年起,子平台将数字化技术各个组分联合打包为“标本数字化流程及技术指南”(附件1, <http://www.guihaia-journal.com> 在线发表)和标本信息录入系统指南(附件2, <http://www.guihaia-journal.com> 在线发表),随年度专题任务书发放给平台参建单位,同时在 CVH 网上在线发布。该指南主要包含标本数字化流程、图像采集、数据录入、图像质量评价四部分内容,从数据标准规范出发,对标本整理、数据录入、影像采集3个环节进行了全面的梳理和详细的阐述(刘慧圆等,2017)。其中图像采集部分包含了推荐设备的使用方法、图像质量标准等内容,提供了一套完整的标本图像采集工作流程,有效指导标本馆尤其是初次参建馆在数字化初期阶段对于摄影设备的选择和部署,让平台的新加入单位(标本馆)能够快速建立起符合标准的标本数字化的硬件系统。

早在 21 世纪初,当科技基础条件平台尚在酝酿阶段,中国科学院植物研究所标本馆就安排包伯坚先生开发植物标本信息采集系统(Cathaya 系统),供本单位线上使用。平台项目启动后,我们基于 Cathaya 系统研制单机版植物标本与物种信息系统(图 2),后升级为 Ginkgo 系统供参建单位使用。Ginkgo 系统实现了植物标本信息的录入、物种信息的查询检索和编辑、数据统计、数据备份和导出、查询结果的地理分布图(GIS 图像)等功能,后台数据库均使用平台研制的数据标准。该系统(含各个版本)推广使用后显著提高了标本数字化的效率及植物标本的管理水平。

2015 年,子平台发布了体积更小、安装更简单的标本数据录入软件 Ginkgo-s。Ginkgo-s 软件设计了高效科学的数据录入和管理功能,解决了部分数据录入时的常见困难。在保证数据符合标准规范的同时,降低了录入数据的人员数量、时间等成本,提高了数据录入效率(附件 2, <http://www.guihaia-journal.com> 在线发表)。

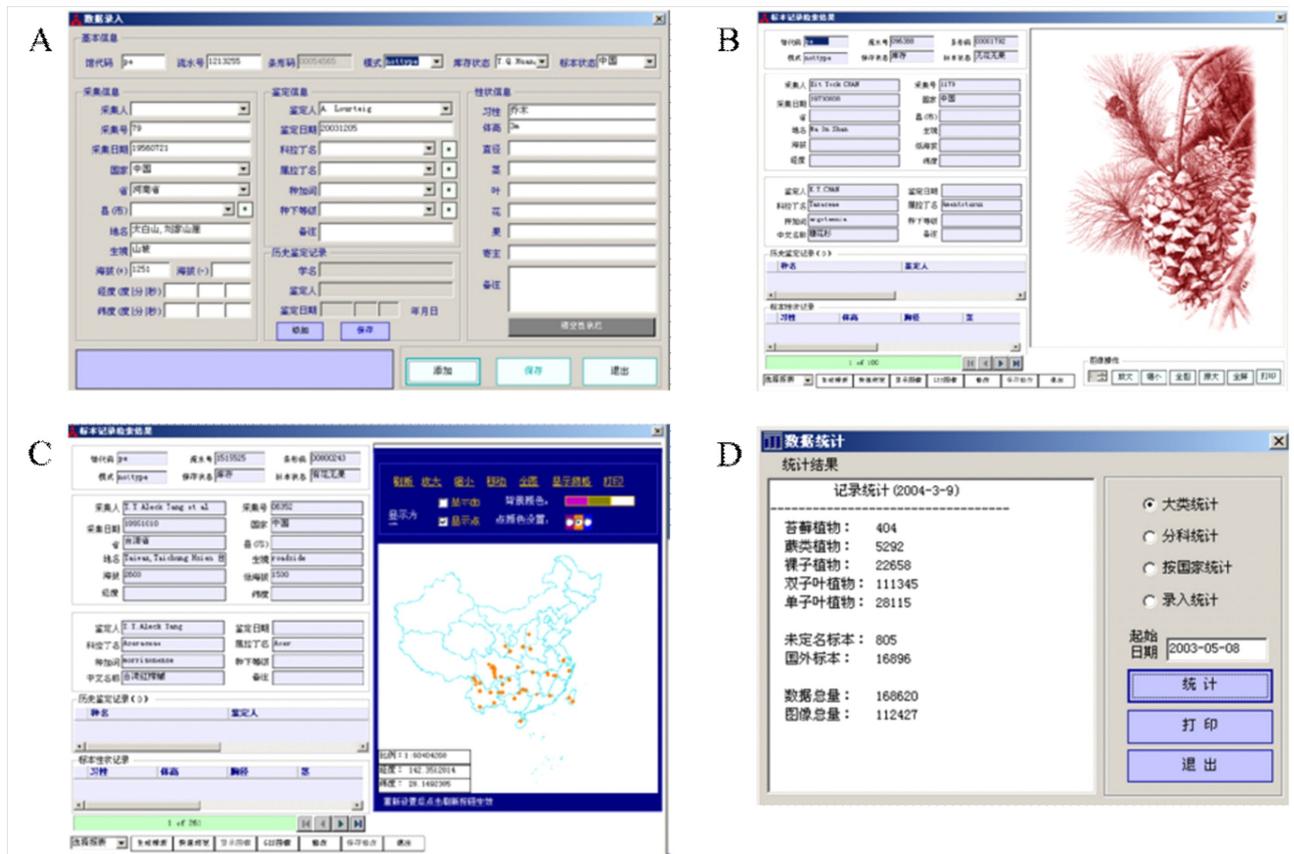
3.2 举办数字化技术培训班

2013—2018 年间,植物子平台共组织开办 4 次全国性的植物标本数字化技术培训班(表 2, 封二图片)。共有来自中国科学院系统、大专院校及地方院所园馆 90 余家单位的 200 多位学员参加,均为平台参建单位的技术骨干人员,培训导师(讲师)则由经验丰富的专家担任。培训班核心内容为植物标本规范化整理、信息录入、图像采集和数据汇交共享等数字化关键环节,此外,还有标本采集和植物分类学原理等学科基本技能和基础理论,以及标本数字化共享国内外动态等(表 3)。

标本数字化技术培训班的开办对统一数字化技术规范、高质量完成数字化任务和培养全国性数字化技术队伍等起到很好的推动作用。培训班为平台参建单位数字化技术人员提供了一次难得的集中式、系统性的学习机会,也是他们与子平台成员当面咨询,与其他同行交流和研讨的良好平台。为了方便没有参与技术培训的用户学习相关内容,子平台还依托培训班录制了《植物腊叶标本数字化技术》视频、编写《标本数字化流程及技术指南》等培训班课件,上传至 CVH 网站在线共享,为用户提供多媒体形式的标本数字化技术学习方式。

4 平台数据共享机制

自国家科技基础条件平台项目启动以来,植物子平台(课题)一方面从制度上建立起一套完整的标本数字化共享体系和技术标准,另一方面从队伍和机构上维持一支稳定的平台管理技术队伍和专家咨询团队,包括植物子平台及 CVH 专家委员会、用户委员会,并持续地按照共享规则实现标



A. 数据录入; B. 标本记录检索结果; C. 标本记录检索结果地理信息界面; D. 数据统计。
A. Data entry; B. Specimen query results; C. Specimen query results of GIS; D. Data statistics.

图 2 植物标本与物种信息系统 3.1 版部分界面

Fig. 2 Some interface of plant specimen and species information system version 3.1

本等信息共享和提升数据质量。

4.1 数字化标本共享原则

早在科技基础条件平台建设初期,我们就对国内外植物标本资源共享政策法规开展研究,并对中国生物(植物)标本资源共享政策法规制订及信息共享机制提出建议(贾渝等,2005)。在2009年内蒙古锡林浩特召开的数字标本馆(CVH)建设研讨会上,全体平台参建单位一起讨论并通过了《中国数字植物标本馆(CVH)数据共享管理暂行条例》(以下简称《条例》)(<https://www.cvh.ac.cn/public/uploaded/files/support/r090826.pdf>)。《条例》对网络管理、数据管理、用户管理、组织机构和职责、考核及罚则进行了规范,并明确了各个参建单位负责人的权责,规范了数据联邦式共享模式等。

这是我国有关生物标本信息共享原则的最早文件之一,它的发布使得 CVH 的数据服务从此步上正轨(许哲平等,2012)。

为了在满足用户批量获取并使用标本数据用于研究的需求的同时,确保数据不被滥用,子平台制定了线下数据共享使用原则,规范用户线下获取数据的流程。用户在阅读并同意遵守数据共享原则后,需要向子平台提交数据申请表(附件4, <http://www.guihaia-journal.com> 在线发表)。子平台会对用户身份、数据用途、外发数据量和数据敏感程度进行审核,决定是否给用户发送其所要求的标本数据。数据共享使用原则除了规范数据发放原则及流程外,还统一了数据引用方式,同时也约束了用户使用数据的权力和义务。据刘慧圆等

表 2 子平台历次标本数字化技术培训班

Table 2 Specimen digitization technology training courses held by the sub-platform

日期 Date	城市 City	参加人数 Number of participants	单位(标本馆) Participant herbaria	讲师人数 Number of lecturers
2013.11.17—2013.11.20	北京	40	24 家	18(含动物及岩矿子平台各 2 人)
2014.11.9—2014.11.10	贵阳	60	28 家	22(含野外考察导师 2 人)
2015.7.19—2015.7.24	北京	67	49 家	15(含野外考察导师 2 人)
2018.8.12—2018.8.15	呼和浩特	79	46 家	16(含壁报交流和分专题讨论)

表 3 第 3 次标本数字化技术培训班课程内容(2015 年 7 月 19—24 日, 北京)

Table 3 Items of the third specimen digitization technology training course (July 19th–24th 2015, Beijing)

项目 Item	培训人 Lecturer
标本鉴定	刘全儒(北京师范大学)
“国家标本资源共享平台(NSII)”网站介绍	许哲平(中国科学院植物研究所)
“中国数字植物标本馆(CVH)”新网站介绍	李敏(中国科学院植物研究所)
标本馆(PE)介绍及参观	杨志荣、傅连中(中国科学院植物研究所)
标本及标本数字化	覃海宁(中国科学院植物研究所)
标本整理准备	林祁(中国科学院植物研究所)
实际操作 I: 标本准备及鉴定	林祁(中国科学院植物研究所)
标本信息录入及管理软件系统	包伯坚(中国科学院植物研究所)
实际操作 II: 标本信息录入	包伯坚(中国科学院植物研究所)
标本摄影技术	上官法智(中国科学院昆明植物研究所)
实际操作 III: 标本摄影技术	上官法智(中国科学院昆明植物研究所)
标本扫描技术	孟世勇(北京大学)
标本图像处理	李敏(中国科学院植物研究所)
实际操作 IV: 标本扫描及图像处理	孟世勇/李敏
标本数据提交方法及注意事项	刘慧圆(中国科学院植物研究所)
标本采集技术	傅连中(中国科学院植物研究所)
信息化标本采集和数据处理技术	陈彬(中国科学院上海辰山植物园)
数字化标本在科研中的应用	孟世勇(北京大学)
野外实习(中国科学院北京森林生态系统定位研究站)	苏宏新(中国科学院植物研究所)

(2022)统计,2007—2020年间,用户基于对从CVH申请获得的标本数据,发表研究论文多达1400余篇,包括中文(含学位论文)1000余篇、英文400余篇,文章涵盖了气候变化、外来入侵物种管理、生物多样性调查、分类学研究、濒危物种评估、植物保护和中医药植物分析等50多个主题。

在CVH网站建设上,则关注于应用共享技术标准。一方面数据库管理中引入国际数据标准软件工具,如引入国际生物多样性标准 Darwin Core,对标本数据做了大量的规范化整理,大大提高了数据质量;另一方面采用生命科学标识 LSID 国际标准对基于物种名的各类数据进行关联,通过为每个物种名赋予一个全球唯一的标识符,然后利用该标识符进行站内物种、标本、文献、图片、植物园等数据的关联,与 EOL、uBio、BHL、IPNI 等国际主流生物多样性信息平台之间的关联,为全球生物多样性信息共享奠定了基础。此外,通过 KML 文件,还可以将地标整理后的标本数据在 Google Map、GIS 等工具上进行展示和分析,大大改善了数据获取途径和分析效果,促进了多源数据的共享和集成(许哲平和赵莉娜,2010)。

4.2 数据共享中的数据质量提升

数据汇交并非标本数据生命周期的终点,子平台后续会持续进行数据修订、数据标准化等数据质量提升相关工作。逐字录入的标本数据有人名、地名拼写不统一,日期、经纬度、海拔、性状测量数据格式和单位不统一等问题,导致数据检索结果容易出现缺漏。子平台定期检查数据质量,修正录入错误,包括行政区划、地名、人名等字段值进行标准化、统一化,对经纬度、海拔、日期等字段值的格式、进制和计量单位进行统一化等数据标准化操作,建立标准化字段。通过使用标准化字段进行检索,符合条件但格式不规范的数据也能展示至检索结果中,从而增加检索结果的完整性,使数据能够更有效地利用和共享。同时,CVH 开通了供用户反馈数据问题的渠道,建立了基于网站用户的数据修订机制:用户通过反馈渠道将发现的数据问题提交至 CVH,子平台定期整理、审核收集的问题,并针对有效的反馈集中修改标本数据。通过建立这种用户贡献的机制,在进一步

提升在线共享标本数据质量的同时,增加网站用户对于科学数据库建设的参与感。

5 存在的一些问题

植物子平台在存续的15年间(刘慧圆等,2022),按照总平台要求,在平台项目领导的指导下,编制大量的标本数字化标准规范,建立了一套较为完备的子平台(课题)及 CVH 共享网站运行机制,形成了颇具成效的数据质量管理体系。这些经验和做法作为国家级的数据集成共享范式,对于其他类似项目的运行管理有良好的指导和借鉴意义。同时,子平台及 CVH 在运行和共享工作中仍存在一些问题未能得到有效的解决,包括:(1)存量数字标本中不少照片可用率低——早期工作受当时数码相机和扫描仪性能所限,所制作标本图像分辨率偏低,细节模糊,这些标本又多是大型标本馆的馆藏,重新翻拍难度很大;(2)存留数字标本数据清理进展缓慢——如鉴定名称修订、采集地名新旧对比,以及信息录入错误的纠正等;(3)在 CVH 平台上搭建分类学辅助研究系统——未能成功;(4)CVH 共享平台与参加单位/实体馆的关联未能打通——如相互间未能及时有效地传递标本信息增加和更改的信息,成员馆子站系统久建未成等。这些问题都由来已久,并且子平台(课题)都尝试努力,但“久攻未下”,半途而废或效果不佳。此外,许哲平等(2012)还指出植物子平台及 CVH 缺乏有效的评估机制和评价指标,以及技术人才队伍的稳定性等问题;刘慧圆等(2017)则从顶层设计的角度分析人才队伍建设所存在的问题等。希望上述问题和我们过往尝试解决过程中积累的经验教训能在“国家植物标本资源库”建设中予以借鉴、思考和解决。

致谢 本文所介绍的标本数字化技术规范是百余家植物子平台参建单位暨 CVH 成员馆千余人长期实践的经验和总结。感谢他们的支持和帮助!感谢严令斌同学、王加国同学(贵州大学)、孟世勇博士(北京大学)和上官法智工程师(中国科

学院昆明植物所)帮助撰写标本影像制作技术指南!感谢国家标本资源共享平台项目负责人马克平老师和陈铁梅老师的指导和帮助!感谢原研究组李奕、何强、王利松、赵莉娜、单章建、谢丹等同事同学的支持和帮助。感谢贾渝、陈建平、林祁等老师的支持、指导和帮助!

参考文献:

- BERENDSOHN WG, GÜNTSCH A, 2012. OpenUp! Creating a cross-domain pipeline for natural history data [J]. *ZooKeys* 209: 47-54.
- BORSCH T, STEVENS AD, HÄFFNER E, et al., 2020. A complete digitization of German herbaria is possible, sensible and should be started now [J]. *Res Ideas Outcomes*, 6: e50675.
- 曹一化, 刘旭, 许增泰, 等, 2006. 自然科技资源共性描述规范 [M]. 北京: 中国科学技术出版社: 46-53.
- CHEN T, CHEN D, WU DL, 1995. An introduction to the data system for computer-aided label preparation of specimen collection [J]. *J Trop Subtrop Bot*, 3(2): 90-92. [陈涛, 陈都, 吴德邻, 1995. 植物标本采集标签计算机印制数据系统 [J]. *热带亚热带植物学报*, 3(2): 90-92.]
- 杜占元, 刘旭, 郭志伟, 等, 2007. 自然科技资源共享平台建设的理论与实践 [M]. 北京: 科学出版社: 523-532.
- FU LG, ZHANG XC, QIN HN, et al., 1993. *Index Herbariorum Sinicorum* [M]. Beijing: China Science and Technology Press. [傅立国, 张宪春, 覃海宁, 等, 1993. 中国植物标本馆索引 [M]. 北京: 中国科学技术出版社.]
- GE BJ, YAN J, DU C, et al., 2020. A brief introduction to world and Chinese herbaria [J]. *Plant Sci J*, 38(2): 288-292. [葛斌杰, 严靖, 杜诚, 等, 2020. 世界与中国植物标本馆概况 [J]. *植物科学学报*, 38(2): 288-292.]
- HASTON E, CUBEY R, PULLAN M, et al., 2012. Developing integrated workflows for the digitisation of herbarium specimens using a modular and scalable approach [J]. *Zookeys*, 209: 93-102.
- 贾渝, 马克平, 覃海宁, 2005. 生物标本资源 [M]//王东阳. 自然科技资源共享政策法规研究. 北京: 科学出版社: 232-263.
- KNIGHT-DAVIS S, BRUNS T, TUCKER GC, 2015. Big things have small beginnings: Curating a large natural history collection-processes and lessons learned [J]. *J Libr Scholarly Commun*, 3(2). DOI:/0.7710/2162-3309.1240.

林祁, 杨志荣, 包伯坚, 等, 2017. 植物模式标本的考证与数字化: 以中国国家植物标本馆为例 [J]. *科研信息化技术与应用*, 8(4): 63-76.

LI M, XUAN J, ZHAO MY, et al., 2018. Birth and development prospect of iHerbarium [J]. *Front Data Comput*, 9(5): 36-40. [李敏, 宣晶, 赵明月, 等, 2018. 标本馆伴侣的诞生与发展前景 [J]. *科研信息化技术与应用*, 9(5): 36-40.]

LI MG, XU ZR, GUAN DF, et al., 1995. Conceptual issues in the development of English - Chinese bilingual botanical databases [J]. *Acta Sci Nat Univ Sunyatseni*, 34(4): 76-81. [李鸣光, XU Zhaoran, 关朵霏, 等, 1995. 植物标本汉英双语数据库管理系统的概念与实践 [J]. *中山大学学报(自然科学版)*, 34(4): 76-81.]

李鸣光, 余萍, 2006. 原始数据与标准化数据必需在植物标本数据库中并存 [C]//国际生物多样性计划中国委员会 (Chinese National Committee for DIVERSITAS), 中国科学院生物多样性委员会 (Biodiversity Committee, the Chinese Academy of Sciences), 国家环境保护总局自然生态保护司 (Department of Ecological and Natural Conservation, State Environment Protection Administration). 中国生物多样性保护与研究进展Ⅶ——第七届全国生物多样性保护与持续利用研讨会论文集. 北京: 气象出版社, 4: 220-223.

凌萍萍, 汤徽杉, 1987. 江苏省植物研究所微型计算机标本管理系统 [C]. 南京中山植物园研究论文集: 39-42.

LIU HY, QIN HN, LI M, 2017. Plant specimen resource sharing platform and plant specimen digital capability construction [J]. *e-Sci Technol Appl*, 8(4): 13-23. [刘慧圆, 覃海宁, 李敏, 2017. 植物标本资源共享平台与标本数字化能力建设 [J]. *科研信息化技术与应用*, 8(4): 13-23.]

LIU HY, QIN HN, BAO BJ, et al., 2022. An analysis of digital specimens of higher plants in China [J]. *Guihaia*, 42(Suppl. 1): 48-64. [刘慧圆, 覃海宁, 包伯坚, 等, 2022. 中国高等植物数字化标本分析 [J]. *广西植物*, 42(增刊1): 48-64.]

QIN HN, YIN XB, 2003a. Recent Advances in the Conservation of Biological Collections [C]//QIN HN. *Proceedings of the Third International Conference on the Preservation of Botanical Collections, Herbarium Techniques*, Beijing, September, 23-26, 2001. Beijing: China Science and Technology Press: 41-51. [覃海宁, 殷学波, 2003a. 国外生物标本保藏工作新进展 [C]//覃海宁. 第三届国际生物标本保藏会议(2001年9月23-26, 北京)论文集. 北京: 中国科学技术出版社: 41-51.]

QIN HN, YIN XB, 2003b. Preliminary results and analysis of a survey of China herbaria [C]//QIN HN. *Proceedings of the*

- Third International Conference on the Preservation of Botanical Collections, Herbarium Techniques, Beijing, September, 23 - 26, 2001. Beijing: China Science and Technology Press: 124-130. [覃海宁, 殷学波, 2003b. 全国植物标本馆现状调查初步分析报告 [C]// 覃海宁. 第三届国际生物标本保藏会议(2001年9月23-26, 北京)论文集. 北京: 中国科学技术出版社: 328-341.]
- QIN HN, 1999. Herbaria in China: past, present and future [J]. *Korea J Plant Taxon*, 29(4): 363-382.
- QIN HN, LIU HY, HE Q, et al., 2019. Index herbariorum sinicorum [M]. 2nd ed. Beijing: Science Press. [覃海宁, 刘慧圆, 何强, 等, 2019. 中国植物标本馆索引 [M]. 2版. 北京: 科学出版社.]
- THIERS B, 2018. Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium [J]. [2017-07-16]. <http://sweetgum.nybg.org/science/ih/>.
- TULIG M, TARNOWSKY N, BEVANS M, et al., 2012. Increasing the efficiency of digitization workflows for herbarium specimens [J]. *ZooKeys*, 209: 103-113.
- VAN OEVER JP, GOFFERJÉ M, 2012. 'From Pilot to production': Large Scale Digitisation project at Naturalis Biodiversity Center [J]. *ZooKeys*, 209: 87.
- WANG YH, WU X, ZHAO W, 2009. Research on users and characteristics of shared services for natural scientific and technological resources [J]. *Sci Technol Manage Res*, 29(3): 310-312. [王运红, 吴霞, 赵伟, 2009. 自然资源共享服务用户及共享服务的特点研究 [J]. 科技管理研究, 29(3): 310-312.]
- WANG YH, ZHANG G, SHEN XY, 2008. Research and practice on national infrastructure of natural resources for science and technology of China [J]. *Chin Sci Technol Resour Rev*, 4: 16-19. [王运红, 张莞, 沈欣媛, 2008. 国家自然资源 e-平台建设实践 [J]. 中国科技资源导刊, 4: 16-19.]
- 伍玉明, 张春光, 覃海宁, 等, 2010. 生物标本的采集、制作、保存与管理 [M]. 北京: 科学出版社: 297-384.]
- XU ZF, 2017. Kingdonia project: A herbarium based citizen science practice [J]. *e-Sci Technol Appl*, 8(4): 97-105. [徐洲锋, 2017. 结合公民科学的 Kingdonia 协同工作平台的构建与应用 [J]. 科研信息化技术与应用, 8(4): 97-105.]
- XU ZP, QIN HN, MA KP, et al., 2012. Research on management, sharing and application of natural science and technology resources: Taking Chinese Virtual Herbarium (CVH) for an example [J]. *Chin Sci Technol Resour Rev*, 44(1): 27-33. [许哲平, 覃海宁, 马克平, 等, 2012. 自然资源的管理、共享和应用研究——以中国数字植物标本馆为例 [J]. 中国科技资源导刊, 44(1): 27-33.]
- XU ZP, ZHAO LN, 2010. Chinese Virtual Herbarium (CVH) platform [J]. *Sci Data Commun*, 3: 33-36. [许哲平, 赵莉娜, 2010. 中国数字植物标本馆平台(CVH) [J]. 科学数据通讯, 3: 33-36.]

(责任编辑 蒋巧媛 邓斯丽)

本文附录请到本刊网站(http://guihaia-journal.com/ajax/publisher/download_pdf.aspx?psu=2CCD15A72AA65E14BBCDF14DB5AA9DCA898549C99465EE9B&journal_id=gxzw&file_no=220004&year_id=2022&issue=Z1) 下载。