

DOI: 10.11931/guihaia.gxzw202209065

刘启新, 褚晓芳, 董晓宇, 等, 2022. 中国植物标本馆数字化发展的缩影——江苏省中国科学院植物研究所标本馆 (NAS) [J]. 广西植物, 42(增刊 1): 71–86.

LIU QX, CHU XF, DONG XY, et al., 2022. Development epitome of digitization of plant specimens of herbaria in China: Herbarium of Institute of Botany, Jiangsu Province and Chinese Academy of Sciences (NAS) [J]. *Guihaia*, 42(Suppl. 1): 71–86.



中国植物标本馆数字化发展的缩影 ——江苏省中国科学院植物研究所标本馆 (NAS)

刘启新, 褚晓芳, 董晓宇, 惠红, 周义峰*

(江苏省中国科学院植物研究所(南京中山植物园)标本馆, 南京 210014)

摘要: 江苏省中国科学院植物研究所标本馆 (NAS) 是中国最早的植物标本馆之一, 也是国内最早开展植物标本数字化的标本馆, 其标本数字化发展经历了 4 个阶段: 20 世纪 80 年代后期尝试的标本文字信息数字化的起步阶段; 20 世纪 90 年代末的标本图像数字化和文字信息数字化规范阶段; 2004 年以后的标本批量数字化与信息网络共享快速发展阶段; 2018 年后的标本数字化信息维护与优化阶段。这一过程集中代表和反映了中国植物标本数字化的发展历程。此外, 近年来开始了发掘和利用江苏植物标本的数字化信息工作, 包括建设江苏省级数字植物标本馆、开发江苏省维管植物标本时空分布可视化系统、开展标本采集-入库过程数字化等。今后, 将不断深化标本数字化的工作, 以期形成有 NAS 特色的数字化植物标本馆。

关键词: 江苏省, 植物标本, 标本图像, 标本信息, 数字化, 数据库

中图分类号: Q94 文献标识码: A 文章编号: 1000-3142(2022)增刊 1-0071-16

Development epitome of digitization of plant specimens of herbaria in China: Herbarium of Institute of Botany, Jiangsu Province and Chinese Academy of Sciences (NAS)

LIU Qixin, CHU Xiaofang, DONG Xiaoyu, HUI Hong, ZHOU Yifeng*

(*Herbarium of Institute of Botany, Jiangsu Province and Chinese Academy of Sciences (Nanjing Botanical Garden Mem. Sun Yat-Sen), Nanjing 210014, China*)

Abstract: The Herbarium of Institute of Botany, Jiangsu Province and Chinese Academy of Sciences (NAS) is one of the earliest herbarium in China, and it is also the earliest herbarium to carry out the digitization of plant specimens in

收稿日期: 2022-08-12

基金项目: 国家标本资源共享平台项目(2005DKA21400); 国家科技资源共享服务平台项目(E2117G1001); 江苏省省属公益类科研院所植株科研课题(JSPKL202058)。

第一作者: 刘启新(1958-), 硕士, 研究员, 主要研究方向为植物分类学, (E-mail) naslx@ qq.com。

*通信作者: 周义峰, 硕士, 副研究员, 研究方向为植物资源学, (E-mail) njgzhou@ 163.com。

China. The digitization development process has gone through four stages: the initial stage of trying to digitize the text information of specimens in the late 1980s, the standardization stage of digitization of specimen image and text information in the late 1990s, the stage of batch digitization of specimens and sharing specimens information on network after 2004, and the stage of maintenance and optimization to digitization information of specimen after 2018. This process concentratedly represents and reflects the development of digitization of specimens in China's herbaria. In addition, in recent years, digital information of plant specimens from Jiangsu Province has been excavated and utilized, including the construction of Jiangsu Virtual Herbarium, development of a visualization system for the temporal and spatial distribution of vascular plant specimens in Jiangsu Province, and digital management in the whole process from specimen collection to specimen warehousing. In the future, the work of specimen digitization will continue to deepen, with a view to forming a digital herbarium with NAS's characteristics.

Key words: Jiangsu Province, plant specimens, image of specimens, information of specimens, digitization, database

植物标本是开展植物学研究的一类重要实物材料,不仅承载了植物的形态特征和遗传物质信息,而且记录或携带着植物的产地、采集和资源信息,甚或载有植物种类的鉴定信息、标本研究信息以及鉴定与研究的历史信息,在植物学研究领域,诸如植物分类、植物地理、植物区系等学科的研究中备受关注,特别是模式标本,是植物种类及其名称的唯一实物凭证,具有不可替代性,其地位尤其特殊和重要。因此,业界极为重视对于植物标本的收集、保藏、研究与应用。

标本馆的重要任务之一,就是长期妥善保藏植物标本,并为学术研究和社会利用提供支撑和服务。但是,植物标本(尤其是腊叶标本)具有随着时间的延长而逐渐老化或极易破损的局限性,加之保存条件的不适或过多翻阅易造成标本实体损坏,甚或天灾及人为失误可造成标本损毁或丢失,因此在实际工作中标本的保护与利用往往产生矛盾,特别是对于保存时间久且已“老化”的标本,矛盾愈显突出。这一现象在很长一段时间内困扰着标本馆管理人员,也是标本馆工作的难点和必须解决的问题。

20世纪中后期,随着计算机的普及与应用,利用计算机技术将标本信息数字化,为标本馆解决管理和应用方面的矛盾和困难提供了十分有效的方法和途径。数字化的植物标本信息以数据库的数字形式与标本实体分离,不仅在一定程度上减少直接查阅植物标本实物才能获取相关信息的依赖性 or 局限性,而且可以减少人为活动造成标本

损失,实现多人同时、多次、异地获得完全相同的标本信息。这是传统方式管理植物标本所不能及的,更能有效地延续实体标本寿命。无疑,标本的数字化为标本的管理、检索、利用和保护提供了崭新的途径。正因如此,计算机一经引入标本馆管理,国内外的主要标本馆都陆续启用并迅速推进馆藏标本的数字化工作(王利松等, 2010; Sobeeon, 1999; 李鸣光等, 1995; 马波等, 2001)。

江苏省中国科学院植物研究所标本馆(NAS, 原名为中国科学院南京中山植物园标本馆;以下简称“本馆”)源于1923年的我国第一个生物研究机构——中国科学社生物研究所和建立于1934年的原动植物研究所的标本室。与国内其他植物标本馆相比,本馆应是我国最早的植物标本馆,至今已近百年,是国内四大植物标本馆之一,也是华东地区最大的植物标本馆,保存有被子植物、裸子植物、蕨类植物及苔藓植物的腊叶标本80余万份,模式标本5000余份。本馆最大的保存特色和优势是有众多模式标本、大批20世纪20年代至40年代我国早期植物学家采集的标本以及来自世界其他标本馆早年的标本。中国植物学研究最早发源于南京和上海,我国早期植物学家在开展植物学研究过程中采集有许多植物标本,如秦仁昌、方文培、耿以礼、蒋英、王启无、蔡希陶、陈焕镛、贺贤育等在我国植物分类学发展中做出杰出贡献的专家和学者,他们在西南(滇川贵)、华东、华南等地采集的大量标本,多数保存于本馆。此外,还有一些外国植物学家(如P. J. M. Delavay、C. C.

Courtois、H. Migo、J. F. Rock、Y. Yabe 等人)在中国各地采集的许多标本和震旦博物院(Musee Heude)的馆藏标本等。这些都是新中国成立中国科学院时收藏的主要植物标本,本所作为中国科学院专设的高等植物分类研究所将这部分标本予以全盘接收并贮存于本馆中。这些植物标本经历了抗日战争期间从南京西迁至重庆、战后又迁回的颠沛,以及“文革”时期的易地搬迁,保存至今,非常不易(薛攀皋, 1992)。上述标本,许多仅保存于本馆,在我国近代植物学研究中发挥了重要作用,具有特殊的地位和极高的价值,十分珍贵。如何妥善保存这些标本是本馆工作的重中之重,也成为本馆积极开展标本数字化工作最直接、最迫切的原动力之一。

除此之外,随着学科的发展和信息化的深入,植物科学工作者,尤其是植物分类学家,除了到馆实地查看植物标本细节,更希望通过网络平台获取或关注标本的相关信息,加之为了更好地全面保护馆藏标本,也为了让人们更方便、有效地利用这些标本的信息资源,在较大的范围内做到信息共享,本馆不仅充分认识到植物标本数字化的重要性,而且及早开展标本数字化工作,并在 1984 年就率先在国内开始建设植物标本数据库,从此拉开了我国植物标本馆开展标本信息化建设的序幕。

1 本馆标本数字化的经历与发展阶段

纵观中国植物标本馆将计算机技术应用于植物标本的管理并进行标本信息数字化的过程,本馆最有代表性——开始于 20 世纪 80 年代,率先在国内建立了植物标本数据库,历经了各个发展阶段,成为我国植物标本数字化平台的重要参与单位,是国内开展标本数字化最早、历程最长的标本馆。回顾本馆植物标本数字化过程以及开展的工作,即可管窥我国植物标本馆数字化的历程,概为缩影。就本馆开展的标本数字化的工作与经历,大体可以分为 4 个阶段。

1.1 第一阶段:率先在国内尝试将计算机应用于标本管理——建立植物模式标本管理系统(MSBS)

20 世纪 60 年代开始,国外就将计算机技术应用于植物标本的管理,主要用于简单的植物类群检索系统和列表;20 世纪 70 年代中期,一些有代表性的植物研究单位的标本馆根据各自的特点和优势独立研发,扩大标本信息的贮存和检索功能,如密苏里植物园的植物标本数据库(1970s)、爱丁堡皇家植物园植物标本馆的 PANDORA 项目(1986)、史密森研究院的植物标本数据库(1970s)(Shetler, 1973; Pankhurst, 1993; 李鸣光, 2000)。在此国际发展大趋势的背景下,伴随着密苏里植物园与我国签署编写 *Flora of China* 的合作协议的逐步实施,本标本馆于 1984 年至 1986 年与南京工学院(现东南大学)生物医学工程系联合研发了植物标本计算机管理系统,并利用该系统对 125 科 290 属约 2 000 份模式标本进行了信息数字化(凌萍萍和汤傲杉, 1988)。这是我国标本馆第一次尝试开发的植物标本信息系统,并开始标本数字化管理。

该系统实现了植物标本信息的存贮、排序、查询、检索、增加和删除等功能,显著提高了植物标本(特别是模式标本)的管理水平。一方面避免了人工查阅标本造成的磨损,能更有效地保护标本,有利于标本的长期保存;另一方面又可准确、多次显示植物标本的信息,避免因人工观察和记录造成所获信息的误差,尤其是在进行大量标本查阅和涉及的分属群较多时,人工查询往往会产生各种差错,而且工作量大。此外,实现了标本信息的异地查阅,可以大幅度节省研究费用,缩短查阅时长。

由于技术的限制,当年所有标本信息首先采用人工登记(图 1),再通过人工将这些信息录入计算机(微机)系统中。整个标本管理系统软件是在 IBM-PC/XT 计算机和 Lotus 1-2-3 集成软件基础上开发的(图 2),资料数据的输入工作以 Lotus 1-2-3 的编辑方式进行,输入区、输出区和标准查询区的范围用资料命令功能块进行制作,而应用程序则用宏汇编进行编制。用户可以通过显示屏窗口查看数据库的内容,包括标本信息和操作程

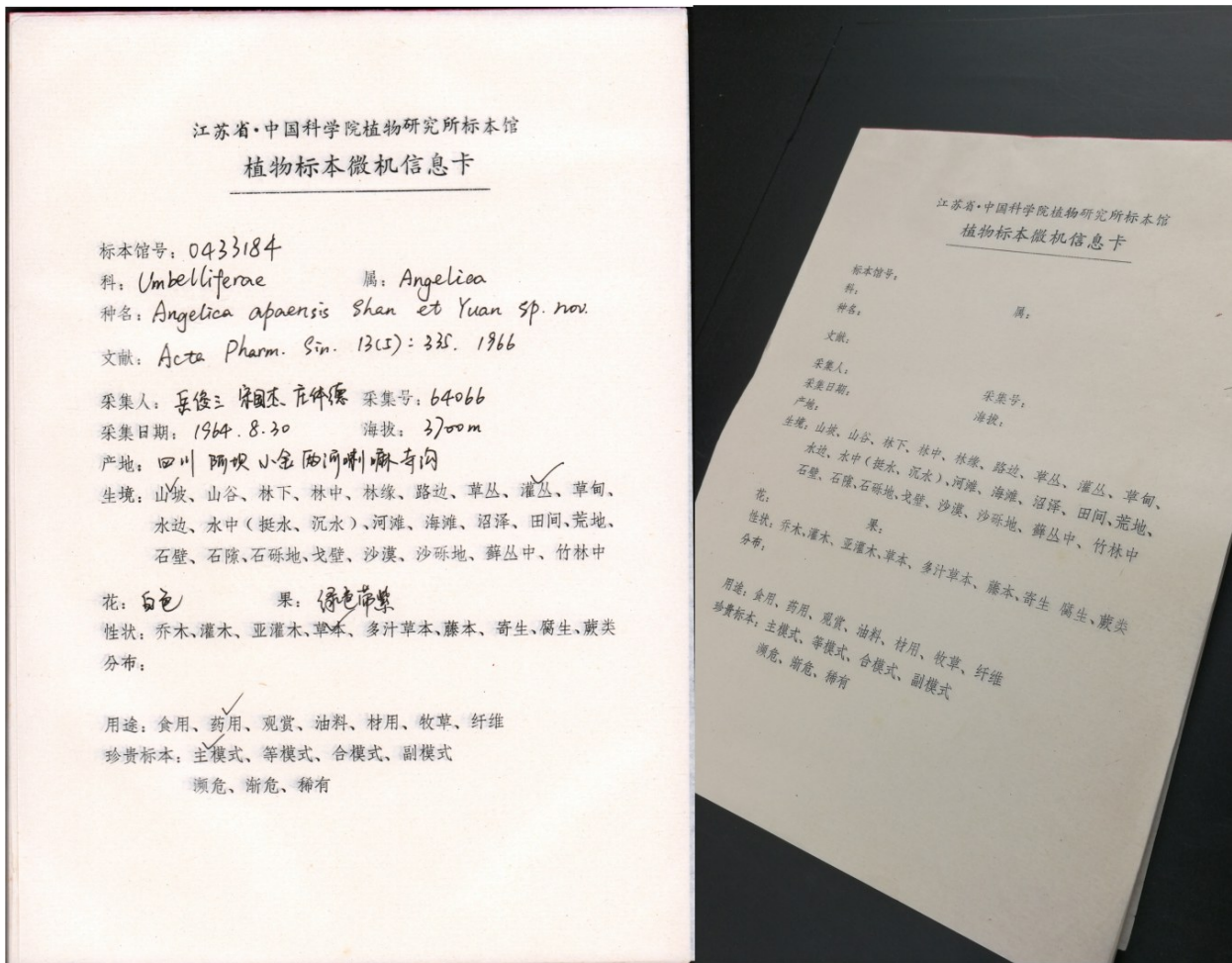


图 1 20 世纪 80 年代 NAS 馆藏植物标本的微机信息登记卡

Fig. 1 Microcomputer information record card of plant specimens collected by NAS in 1980s

序,使得植物标本管理实现了计算机化。

本馆标本数字化最早使用的是美国国际商用机器公司(IBM)于1983年发布的个人电脑产品IBM PC/XT台式计算机,这是当时比较先进的硬件设备,采取XT设计,使用了Intel的8088芯片,引入了标准的16位ISA总线以及采用了当时最新的英特尔80286处理器,硬盘为10MB,预装了DOS 2.0系统,支持“文件”的概念,并以“目录树”形式存储文件。这是硬盘第一次成为PC的标准配置。标本数据贮存器是5.25 in的软磁盘(1.2M),后期改用了3.5 in软磁盘(图3)。

该系统将所有标本分别存放于11个信息库中,每份标本信息包括(标本馆)标本流水号,标本

植物隶属的科号、科名、属名和种名,种名的异名或俗名,种名发表的原始文献,标本采集的采集人、采集号、采集日期、产地、生境和海拔,植物的习性、叶、花、果、雄蕊、雌蕊、用途、分布,标本的模式类型等等,共30项信息。

用户可以调用任何一个库中的标本信息,也可以随时增加、删除和修改这些信息,还可以根据需要对存放的标本信息进行检索,不仅能对每份标本进行检索,而且能从10个角度或方面查询标本。例如:①植物种名发表的原始文献;②植物标本隶属的科、属、种的名称;③标本的采集人;④标本的产地;⑤标本的生境;⑥标本采集地的海拔高度;⑦植物的性状;⑧花色;⑨模式类型;⑩植物用

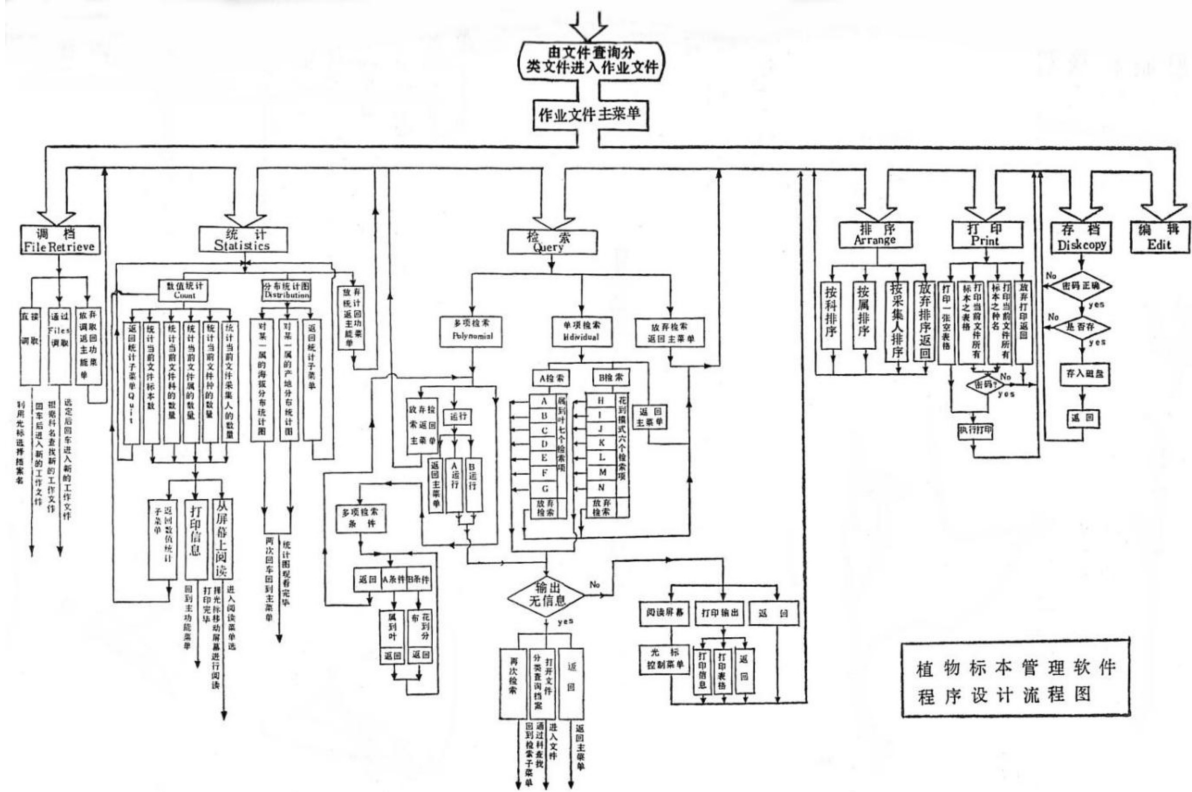


图 2 20 世纪 80 年代 NAS 植物标本管理系统程序设计图 [引自凌萍萍和汤徽杉 (1988)]

Fig. 2 Programming flow chart of management system of plant specimens collected by NAS in 1980s

途。也可以采取多种或复合条件进行查询,如查询某人在某地、某生境下采集的某科草本植物标本,还可以对植物标本的科、属、种的数目和标本份数进行统计。

系统操作以“菜单”的选择方式进行。先显示出文件索引(调档)、统计、检索、排序、打印、贮存、编辑等主菜单;选择了主菜单的某菜单项后,可点击它,以显示出下级的子菜单;再根据需要,在子菜单中进行选项操作;如此逐层往复,直至完成一次检索,系统会返回到主菜单。系统易于掌握,使用方便。

根据标本管理的要求,将植物标本管理软件系统划分为资料数据库、表格图形区和应用程序区 3 个主要部分:(1)资料数据库可分为输入、标准查询和输出 3 个区,其中输入区供用户手工输入每份标本的 30 项信息;标准查询区供用户建立栏位的查询标准,以满足用户不同的查询要求;输

出区供用户显示各种检索结果,可在显示器上视阅,也可打印输出。(2)表格图形区可将输出的任意植物标本中每一项信息填入表格中,得到标本信息的表格形式,或者可将输出的某些数据加以统计,获得直观的图形显示,这些表格或图形也可打印输出。(3)应用程序区供存放应用程序,应用程序将资料数据库与表格图形区联系在一起,组成层状叠套的“菜单”工作方式,使系统能按一定方式自动进行(图 2)。

但是,限于当时的技术水平,该系统也有不足之处。一是系统使用的是英文;二是系统采用的 Lotus 集成软件具有一定局限性,植物标本信息需分别存放于若干个库中,使用前要根据所查询标本所在的科或属进行调库,操作基本上按库进行;三是打印信息只能按已编好的格式进行,而且用户自行进行组织打印格式比较复杂;四是受限于当时的计算机容量等硬件条件,该系统只有标本

的文字信息,缺乏标本的图像信息,不能直观地看到标本图像。

考虑到电子产品的性能具有一定不稳定性和标本数据库软件系统的安全性,本馆在进行标本数字化时首先将标本信息进行文字纸质记录,以作为数字化信息的一种传统备份形式。

本馆首次开展的标本数字化仅涉及模式标本。之所以如此,一是模式标本是一类比普通标本更加重要和特殊的标本,二是本馆的模式标本是单库另存,便于试点和操作。但是,如同所有标本馆一样,本馆保存最多的标本是普通标本,对标本馆的管理和标本查询而言,更需要对普通植物标本信息进行数字化。然而,本馆这一时期开发的标本数据库未涉及普通标本。同时,此数据库仅有标本的文本信息,而且数据库开发尚未采用 Internet 技术——可使世界各地的用户通过网络直接查询数据库信息,能在更大范围、更有效地实现标本信息资源共享。

尽管如此,这项工作在我国具有开先河性,为本馆以及我国植物标本馆开展植物标本数字化建设奠定了基础,而且本馆植物模式标本数据库(名为 MSBS)也被美国 IOPI 网站转载。这是当时我国唯一一个被国外知名网站转载的植物标本信息管理系统,说明该项工作跟上了国外标本馆现代化管理的发展潮流。此后,我国其他部分植物标本馆闻息或来访交流后,逐渐开始了植物标本数字化软件的开发工作。可见本馆在我国馆藏标本数字化启动阶段中起到了示范或促进作用。

1.2 第二阶段:开始普通标本图像数字化和标本数字化规范建设

1.2.1 参与国家科技部首次启动的植物标本数字化项目——馆藏植物标本图文信息系统建设

20 世纪 90 年代,伴随着数码相机和扫描仪的兴起,标本图像数字化成为标本数字化的重要内容。此时,国外标本馆愈加重视标本的数字化,不仅有了标本图像信息,而且标本数字化的规模也越来越大。如阿拉巴马大学于 20 世纪 80 年代末牵头建设了美国东南部地区各小型标本馆的植物标本信息系统 SERFIS;1991 年宾夕法尼亚州 Carnegie 博物馆开始建设植物标本数据库;20 世纪 90 年代初

加利福尼亚大学 Berkeley 分校牵头的美国西部地区加利福尼亚植物标本管理系统 SMASCH 等。在此期间,本馆在原有数据库基础上,开始普通标本的数字化工作,同期国内其他植物标本馆也纷纷开始建立自己的标本数字化软件系统,如 1994 年中山大学植物标本馆与美国阿拉巴马大学合作研制植物标本数据库,1995 年中国科学院华南植物研究所标本馆开始尝试建立植物标本馆数据库,1997 年在国家自然科学基金的资助下,中山大学标本馆与广西植物研究所标本馆联合研制共享标本数据库。与此同时,中国科学院植物研究所标本馆也开始研发植物标本管理系统。

20 世纪 90 年代末,科技部首次启动标本数字化建设项目。1999 年科技部基础司同时资助了与植物标本馆的标本数据库有关的 4 个基础性工作项目(由华东地区的江苏省中国科学院植物研究所、华南地区的中山大学、西南地区的四川大学和东北地区的东北师范大学等单位承担),并加以合并,形成由 4 个单位共同参与建设的“植物标本信息系统”项目,本馆为参与单位之一。同年 12 月,科技部在中山大学主持召开了由广东省科委、4 个参建单位以及特邀的中国科学院植物研究所和华南植物研究所等单位的领导和科研人员(共 36 人)参加的专题讨论会,以期统一数据库的规范,为今后数据共享打下基础,建成的信息系统可以在全国通用。会上同时演示了中山大学基于 Access 系统软件和中国科学院植物研究所基于 Foxpro 系统软件各自开发的植物标本信息系统数据库,2 个软件平台各有利弊,但都能适应或满足当时的发展需求,会议最后确定各单位可以根据自身需要任选其中之一。这次会议标志着植物标本数字化以及建立信息系统平台引起了国家的重视,并有意向标准化、共享和普及推进,成为后来国家科技部启动植物标本标准化整理、整合及共享平台项目的基础。

随后,项目组以中山大学生命科学学院标本馆为主,联合研制了植物标本信息共享数据库,并拟定了植物标本信息录入内容的试行标准,以及可选项及其录入内容。

在该项目支持下,本馆以此系统为基础开展



图 3 20 世纪 80 年代用于 NAS 植物标本管理系统的 IBM PC/XT 计算机以及 2 种规格的存储介质(软磁盘)

Fig. 3 IBM PC/XT computer and two size storage media (soft disk) used for management system of plant specimens collected by NAS in 1980s

了馆内普通植物标本的数字化——标本文本和图像数字化,其中植物标本数据库的录入系统采用的是“植物标本信息系统 4.4”(李鸣光, 2000),共完成了 4 万份普通标本信息的录入,包括标本的文字信息和图像信息,并采用光盘和硬盘保存。为了保证数据的安全,所有数据分别采用了系统、光盘和硬盘的双备份(图 4)与之前本馆开发的植物模式标本数据库相比,本数据库具有如下的特点和改进:(1)为了适应我国各标本馆仍以个人电脑为主的现状,该数据库基于个人电脑,并以微软公司的 Access 软件为平台开发。(2)除了进行标本文字信息数字化,还对标本图像进行数字化(数码图像),但受限于当时数码相机的成像技术,图片像素只有 300 万。(3)标本信息中增加了标本鉴定的历史记录,即将标本上不同时期的所有鉴定标签信息(包括学名、鉴定人和鉴定时间)均全部录入数据库。(4)数据库服务器软硬件配置有了明显的提升。此时已有了 Microsoft windows 2000 server version 网络服务器操作系统、Microsoft SQL Server 数据库管理系统、双硬盘的服务器等,同时数据保存条件也有了极大改进,存储性能和容量有明显提高(如光盘和移动硬盘)。(5)新开发的标本数据库是教学和科研两大系统的多个单位合作形成的,兼顾多方需求和特点,适应性强,便于推广使用。

1.2.2 参与中国科学院的生物标本信息系统建设项目——江苏省中国科学院植物研究所植物标本数据库建设 2000 年,中国科学院生物局开展了生物标本馆网络建设项目。本馆作为参与建设单位承担了本所标本馆网络信息系统建设任务,其中植物标本数据库建设作为独立的部分,重点是进行标本数字化。由于该项目恰巧与科技部的“标本馆基础设施维修改造及信息网络建设”项目相衔接或相近,因此实际上这 2 个项目有部分套合,但又有所改进。

在该项目的录入工作中,标本文字信息或数据项共有 20 个:标本馆代码、标本流水号(台纸号)、拉丁科名、拉丁属名、拉丁种名、种下等级拉丁名、定名人、模式标本类型、采集人、采集号、采集时间、采集地[包括国家、省、县、具体采集地(如乡镇、水系、山体、景区等)]、生境、海拔、经度、纬度、鉴定人、鉴定日期、标本照片、备注等。此外,数据库中的标本信息还包括图像信息,其中普通标本每个物种(包括种下等级)须包括 1 张标本整体实物照片,模式标本原则上须附 1 张标本整体实物照片。

在该项目的支持下,本馆共完成了 8 万份标本的数字化,包括馆藏模式标本、蕨类植物和裸子植物两大类群的全部标本以及被子植物部分科的部分标本。每份标本附图片 1 张,并且每份标本附



图 4 “植物标本信息系统 4.4”的安装程序、标本数据界面和数据存储介质

Fig. 4 Installation program of “Plant specimen information system 4.4”, specimen data interface and data storage medium

有唯一的条形码,而标本图片与文字数据通过标本条形码一一对应。通过该项目,本馆初步建设了具有一定规模的植物标本信息数据库,并且该数据库可通过网络进行标本信息查询。

此次开展的标本数字化项目,与前述的两次数据库建设相比,在如下几方面有不同或改进。

(1)在标本信息内容上有别于其他标本馆:①将馆代码编入标本条形码中,即馆代码+8位数字编码,如NAS00003450,以便与其他标本馆的标本区别或进行网络平台检索时能区分出本馆的标本;②增加了本馆的标本流水号(即台纸编号),因为本馆的标本台纸流水号在馆内是唯一的,而且流水号的先后次序表明每一份标本装订时的先后时间和次序,可以从一个侧面反映标本入馆的时期和馆藏的历史阶段;③增加了备注,用以记录标本上采集信息和鉴定信息以外的其他文字说明,以及录入标本信息时需要对本标进行必要的补充说明。

(2)标本信息全部提供给中国科学院的标本信息网络平台,供网络查询。

(3)标本实物采用数码相机进行拍摄记录,图片的像素达到500万。但根据网络平台要求,限于平台的容量,真正上网的图片经过压缩,仅有200 K。

(4)第一次采取了馆内与馆外、在职与非在职人员相结合的方式共同开展和完成相关工作。

(5)本馆标本的保存采用的是具轨集装箱式密集柜方式,并且标本库内空间有限,标本数字化不宜在库内随取随录,需在库外进行。因此,为了最大限度地保护标本,减少标本移动次数,在标本的鉴定基础上,采取标本整理与标本图像拍摄合一、标本文字信息直接录自标本图片、文字信息校对直接参照标本图像信息的做法。只有当录入信息有误时,才按条形码回库查找标本。

(6)结合标本数字化,对馆内标本加强了精细化管理。在此次标本数字化的过程中,除了对尚

无种名的已装订标本集中进行物种鉴定,还对馆内已有种类鉴定信息的标本进一步精细化管理,重点开展了4个方面的工作:①对标本上原有信息进行审核认定,查阅有关资料,弄清每个信息的来源和正误,包括标本上文字的书写辨析和确认;②结合标本鉴定和标本图像采集,对馆内标本进行全面的次序整理,对于错误鉴定的标本,根据新定种名进行必要的位置调整;③对于破损标本进行补救,尤其是有一些老标本使用频率高,加上早期保存条件有限,急需补损补缺;④对于早期标本的采集信息,由于记载简约、地名更替、记录丢失等不同历史原因,重点进行了查考。

(7)由于标本信息数据量大,这些信息的保存开始采用了DVD光盘和大容量移动硬盘。

1.3 第三阶段:植物标本批量数字化与信息共享平台的快速发展时期

1.3.1 参与科技部的“植物标本标准化整理、整合及共享平台建设”项目(I期) 2004至2008年,本馆参加了国家科技部项目“植物标本标准化整理、整合及共享平台建设”(I期)。对于本馆而言,这是标本数字化的第三阶段。

经过近10年的发展,我国各个标本馆的标本数字化软件和硬件条件都得到了改善,植物标本信息系统开发也逐渐成熟。在此期间,科技部更是积极推进这项事业。在1999年启动了“植物标本信息管理系统”项目之后,2000年又在四川大学主持召开了“实物数据共享研讨会”,在原有参与单位的基础上增加了中国科学院动物研究所、中国科学院细胞研究所、中国医学科学院基础医学研究所和中国地质大学4个单位。本馆参与了该会议。此会为科技部提出“植物标本标准化整理、整合及共享平台建设”项目做了充分的准备。

本馆作为重要的参与单位之一参加了该项目。在此期间,根据项目的要求,标本数字化采用的是中国科学院植物研究所生态中心包伯坚先生等人研发的输录软件系统——植物标本数据管理软件(网络版)Davidia 2.0及Davidia 3.0(图5),将植物标本的采集信息、鉴定信息以及标本的图像信息进行了数字化。在此期间,本馆共完成了30万份普通标本的数字化。所有标本的数据(文

字和图像)均提交至项目管理和共享平台中心,其中本馆的图像数据均采用移动硬盘保存。

1.3.2 继续参加“植物标本标准化整理、整合及共享平台建设”项目(II期) 从2009年至今,本馆继续参与了科技部的标本数字化项目(II期),并成为国家数字植物标本馆(CVH)平台建设的主要标本馆之一。通过该平台,本馆完成的数字化标本信息向社会公开,并借以此平台与教学平台和保护区平台合并后构建的国家标本资源共享平台NSII,进一步扩大了这些标本的共享范围和价值。

从2009至2012年,本馆标本数字化录入采用的是中国科学院植物研究所包伯坚等人研发的植物标本录入查询软件Ginkgo 3(单机版)以及后期的在线录入信息系统,继续对普通标本进行数字化。其中本馆于2010年后开始在以往开展标本数字化的基础上,对原有数据进行补充和修订,以及数据库的维护和升级。2013年开始,重点开展了下列多类标本的图文数字化,包括全部单子叶类群的标本、西藏植物专项采集的标本、新编《江苏植物志》连续多年普查的标本、邬文祥先生家人捐赠的标本、蔷薇科和菊科新鉴定的标本,以及近年新采集的标本。在此期间共完成12.5万份标本的数字化。此时标本图像的拍摄设备有了很大改善(图6),图片的像素和质量都有了进一步的提升。

1.4 第四阶段:植物标本数字化信息的维护与优化

时至今日,本馆已数字化的标本共约50万份,其中包括蕨类植物39科2.06万份、裸子植物10科6951份、被子植物230科47.24万份,其中模式标本4000余份。但是,本馆的标本数字化工作远没有结束,其中重要的工作之一就是对于已经数字化的标本信息进行维护和优化,尤其对于早期数字化的标本,或作信息补充、纠正,或进行图像更新等。这是本馆植物标本数字化建设的最重要阶段,也是今后相当长一段时间内的重点工作。

1.4.1 馆藏植物模式标本高精度数字化 20世纪80年代,本馆的植物标本数字化工作始于模式标本的数字化。但彼时植物标本数字化工作属于起步阶段,有许多不完善:一是仅记录了标本采集和鉴定的文字信息;二是尚无标本图像信息;三是仅对少量标本进行了数字化信息采集。

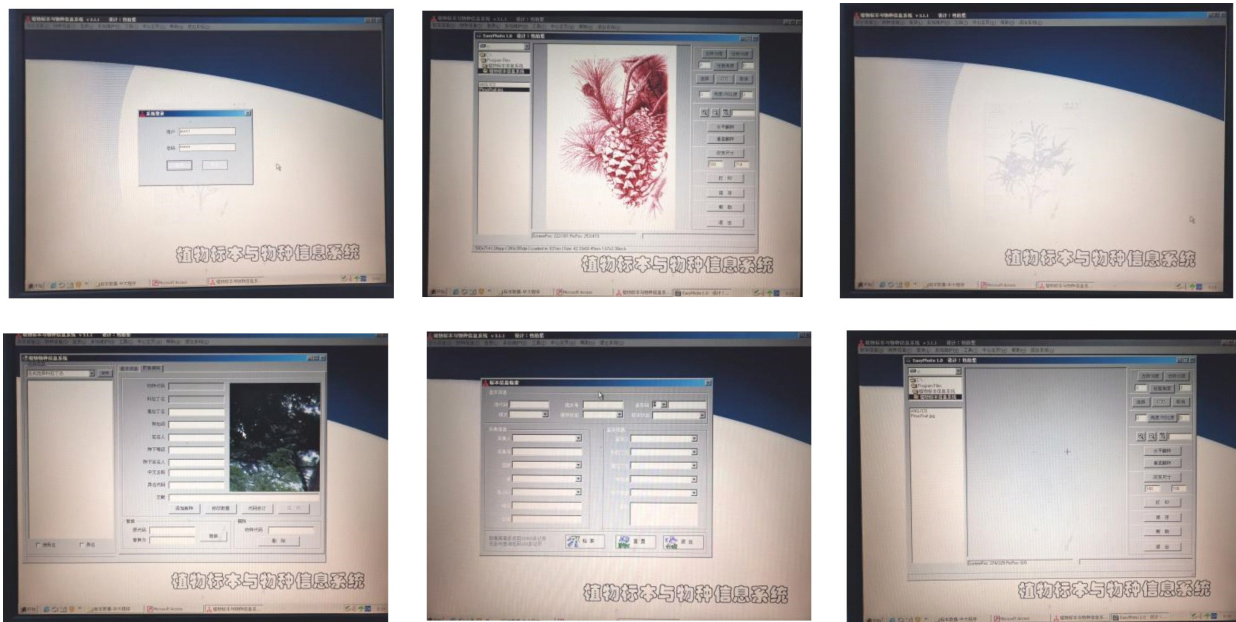


图 5 植物标本数据管理软件 Davidia (包括标本信息的录入、复查、修改和图像处理等)
Fig. 5 Management software of plant specimen data “Davidia” (including entry, recheck and amendment to specimen information, and processing of specimen image)



图 6 2013 至 2019 年用于标本信息数字化的标本图像获取和录入设备
Fig. 6 Equipments of obtain and entry to specimen image used for digitization information of specimen from 2013 to 2019

随着本馆标本数字化的不断深入和硬件设备的改进,在开展“植物标本标准化整理、整合及共享平台建设”项目的同时,另行专项重新对馆藏植物模式标本进行了数字化——标本图像高精度数

字化。此次数字化工作,采用了共享平台提供的、量身定制的标本扫描仪采集图像信息(图 7),像素可达到 6 000 万。在该项目的支持下,共完成了 4 000 份模式标本的高精度数字化。此次模式标本



图 7 用于模式标本图像采集的高像素植物标本扫描仪

Fig. 7 Plant specimen scanner with high pixel used for image collecting of type specimen

图像数字化是在第一次模式标本数字化基础上的升级版,无论在模式标本的数量上还是在图像质量上都有了全面的提升,极大提高了模式标本的利用效率和价值。

与第一次模式标本数字化工作相比,有 2 个重要进展:一是所有的标本都进行了高像素的扫描;二是所有标本都按种类附上新种发表原始文献的扫描版(图 8)。同时还有两个特点:一是采用了扫描仪获取图像信息,像素精度有了极大提高;二是每份标本既有高像素的标本图片,也有新种发表原始文献的图片,完全不同于普通标本的数字化。

1.4.2 馆藏标本早期数字化信息与图像的更新优化 本馆标本数字化起步比较早,开展图像数字化的时间跨度也比较长,馆内保存的不同时期的数字化标本图像在质量上存在很大差异。相比之下,早期拍摄的图片质量普遍不佳,像素较小、分辨率低,主要是当时的数码相机分辨率不高,同时与标本拍摄经验和拍摄条件有关,但这部分照片的数量有限。就数码相机而言,从当初的富士、柯

达,到现在使用的 Canon 5DSr,其中经历了多代更新,像素分辨率从原来的 100 万到目前的 5 000 万,相差 50 倍。此外,早期数码相机拍摄的标本照片,虽然原图清晰,但放大后部分细微结构的清晰度已远不如当今高清晰的数码相机。由于现在拍摄条件有了很大的改善,对于早期拍摄的少量标本进行必要的复拍。

与上相似,本馆的标本数字化经历了不同发展时期、采用不同数据库及其标准,一些条目和规定有一定的差异,加之本馆保存的许多早期采集的标本的信息甄别和录入比较困难,诸如手写文字不易辨认(主要源于手写字体不规整、不规范或潦草,还因标本年代久远而字痕变淡),非英文的外文采集信息难以辨识和录入,纪年随时代和个人习惯变更而不同,地名随时代变更而变得难以考证等等,导致有些标本信息需要考证或纠正。这些数据错误中,除少量采集人、鉴定人名错误能够通过批量处理来修改以外,其他均需人工校对,投入更多时间,提高早期数据的质量。

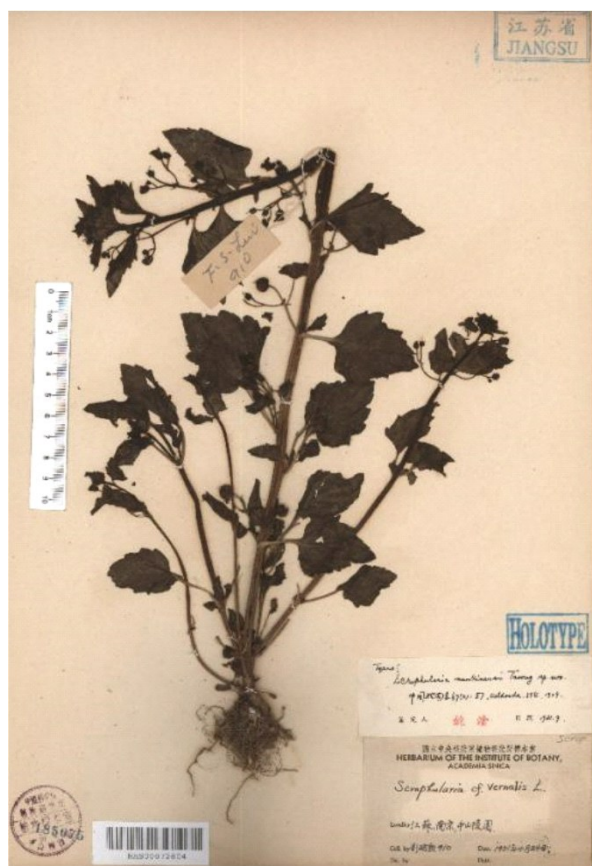


图 8 NAS 模式标本及其原始文献扫描图像

Fig. 8 Scanning image of type specimen and its original literature in NAS

1.4.3 开展全国江苏标本信息的专家核查 中国国家标本网络平台的标本数据来自国内各个研究单位和教学单位的标本馆/室,参与人员广泛,标本来源复杂,标本信息质量难免参差不齐,需要提高准确率。作为我国国家标本网络平台标本信息数据的主要参建单位之一,理应在这方面有所贡献;身为华东及江苏的标本馆,更应该加强江苏标本的核查。近年来,本馆除了继续推进库存标本数字化工作,还开始对包括本馆早期录入在内的、网络平台上已有(含国内其他标本馆录入)的江苏标本信息进行清理与核查,例如,2021年起我馆开展了 NSII 平台中江苏地区全部标本数据的专家核查工作(不含花果、树皮等非主要检索性状),对其进行清理校对。一方面用以支持区域植物学、生物多样性等方面的研究工作,另一方面可指导本馆后续的标本采集与数字化的方向。

2 本馆标本数字化的深化与利用

截止到 2018 年,本馆标本数字化工作仅局限于将标本图片数字化,以及采集和鉴定的文本信息数字化。但对于植物研究者来说,这些工作及其信息有一定的局限性;对于标本馆的数字化管理而言,这些工作和信息有待完善和提高。有鉴于此,根据馆内标本收藏特点和数字化现状,本馆从 2018 年开始,在对余下的标本继续进行常规标准的图文数字化的同时,加强了标本数字化的信息深化工作,重点对已有数字化的标本信息进行发掘与利用。

2.1 开展江苏省标本数字化网络平台建设——省级数字标本馆

2018 至 2019 年,本馆参加了 NSII 平台的省

级数字植物标本馆联盟,籍于此创建了省级数字植物标本馆 Web 平台(简称 PVH)(图 9),为开展本省植物研究的研究者提供植物名录的编辑与维护,以及名录参考的标本库、图像库、文献库的在线维护与管理等服务,形成开放式的区域植物学研究平台。其要点:一是根据新版《江苏植物志》(2013—2015)与文献,整理并上传江苏植物基础名录(每种均有出处),包括蕨类植物、裸子植物和被子植物共计 235 科 1 240 属 3 494 种;二是对所有植物种类的凭证标本进行了审核,共计 14.9 万份,其中对 2 037 种指定了 2 804 份凭证标本;三是开展了江苏植物的原色图像库建设,并上传了 1 318 种的彩色图片共计 4 182 张。

2.2 开发江苏植物时空分布可视化软件

为了利用标本馆数字化建设的优势,为植物研究者提供更加丰富的地方性植物分布信息,为地方建设发挥应有的作用,在上述工作的基础上,本馆开展了江苏地区植物时空分布规律与变化的分析,同时与软件公司合作,开发了标本时空分布的可视化软件——维管植物标本时空分布可视化系统(图 10)。目前,该软件的制作已经完成。将校准、“清洗”过的标本数据导入软件后,可以实现:(1)在指定时间段内,基于标本采集数量和物种数,估算出实际物种数量,并投射在本省的网格化地图上;(2)可对比指定的两个时间段内各网格化区域中物种数量的增减;(3)可展示各时间段中各科的占比。

3 开展标本采集—入库全过程数字化

为了加强馆内工作暨标本的信息化管理,从 2019 年开始,本馆开始采用时任中国科学院昆明植物研究所标本馆馆员的徐洲锋工程师开发的“Biotracks”手机 App、“标本馆助手”和“Kingdonia”系列软件,实现野外标本采集至入库全过程的标本信息化管理(图 11)。

本馆的做法大体如下:(1)在野外采集标本的同时,应用“Biotracks”手机 App 同步记录采集信息,包括采集的时间、地点和地理坐标,以及采集

人、采集号和植物性状信息,还有采集轨迹和同步拍摄的植物数码照片(每份标本的实物及其生境);压制标本时,可补充野外记录以及相关标注(如留存 DNA 样本)。这些信息可同步到服务器,同时整理所有数码照片至按标本采集号命名的文件夹。(2)标本鉴定后,自“Biotracks”下载当次野外采集全部记录,生成 Excel 文本,录入相应鉴定信息,并使用“标本馆助手”程序,将其统一为 DarwinCore 格式的表格,并转换为打印格式,打印一份留作采集记录存档,另外根据采集份数打印带条码格式的采集-鉴定合并记录签。(3)标本经装订和再次冷冻杀虫后,即可入中转库,进行数码照片的拍摄。标本采集信息通过“标本馆助手”转化,生成 Kingdonia 可读取格式后,上传至 Kingdonia 标本管理平台的缓冲区,即可与上传至 Kingdonia 的标本照片自动识别及匹配。标本数据经人工审核与校错后,便可以释放至公共数据库,也可以其他格式导出保存,提交或共享至其他平台。目前,由于本馆原采集数据关键检索词的分割方式与 Kingdonia 平台不同,且旧数码相片的像素无法满足平台自动识别与关联的要求,需要大量人工处理的工作,目前这部分工作尚在进行中。

Kingdonia 平台除了能让标本采集至入库的流程更系统化以外,也可添加标本对应的野外照片,并自动关联“Biotracks”中野外记录标本信息时拍摄的照片。这就为标本查阅者提供了该植物更翔实丰富的外部形态信息,也可作为文本信息中“生境”信息的补充。同时,Kingdonia 与“标本馆助手”的配套使用也为标本提供了库区内定位的功能,能够通过在庫内扫描网格二维码及标本条形码,来标注标本在庫区内的位置及出入库情况,避免了过去清理和借还标本时因疏漏而导致标本无法查找乃至遗失的问题。这项工作的开展,极大提高了标本管理的效率。由于本馆标本数量较大,因此目前该工作正在起步阶段。

4 结论

本馆是我国开展植物标本数字化最早的标本馆,也是我国最有代表性的植物标本馆,标本数字



图 9 “江苏数字植物标本馆”网页
Fig. 9 Web page of “Jiangsu Virtual Herbarium”

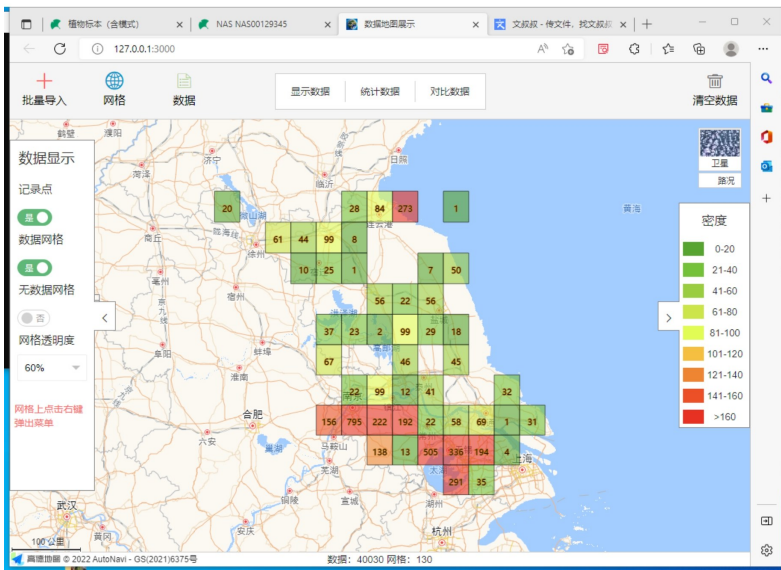


图 10 江苏植物时空分布可视化示意图及软件证书
Fig. 10 Visualization diagram of spatial and temporal distribution of plant in Jiangsu and its software certification

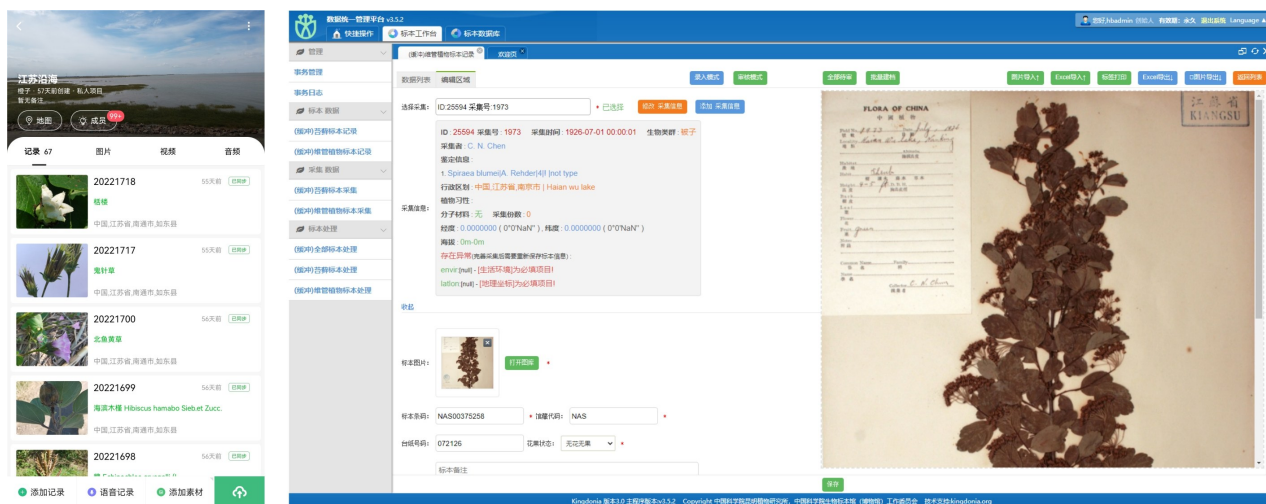


图 11 NAS 标本采集至入库过程数字化界面图

Fig. 11 Digital interface diagram of whole process from specimen collection to put in storage in NAS

化工作前后经历了四个发展阶段,包括 20 世纪 80 年代末尝试标本文字信息数字化的起步阶段,20 世纪 90 年代末的标本图像数字化和文字信息数字化规范阶段,2004 年后的标本批量信息化以及信息网络共享的快速发展阶段,以及 2018 年后的已有标本数字化信息维护与优化阶段,其中第一个阶段对我馆的数字化建设有着重要的意义,是国内其他标本馆未经历过的。作为我国植物标本数字化网络平台(CVH 和 NSII)建设的重要单位之一,本馆进行的数字化标本信息已经共享于平台,而且也开始数字化标本信息的深化工作。但是,截至目前,数字化工作主要局限于将标本图片和文本信息的数字化,国内其他标本馆也几乎类似。从标本馆的数字化管理和开展的研究看,仍存在一些不足和问题,例如:(1)标本文本信息化过程中标准化程度不高,导致关键词检索匹配度较低,甚至出现错误;(2)标本采集制作过程中,更多有价值的信息未录入数据库,包括采集时拍摄的大量生境、伴生植物和植物分类特征细节等的数码照片,无法展现给查阅者;(3)目前的数字化工作并未备注其在实体馆中的位置,无法实现馆内库区的信息化管理;(4)可提供数据统计分析的平台尚未完善,不足以支撑更多的应用需求。由此可见,时至今日,虽然本馆的馆藏标本数字化工

作从馆藏量上看已近完成,但是标本数字化的工作仍任重道远,不仅尚有部分标本(与之前已数字化的标本相比在信息完善上往往要花费更多的时间和精力)未及时进行数字化,而且标本数字化信息的维护、优化、发掘、利用则刚刚开始。

致谢 本馆植物标本数字化工作得到标本馆和信息中心许多同仁的参与和帮助,如刘心恬高级实验师作为早中期阶段的主要成员负责了数据的整理与核查;邓懋彬研究员为该项工作鉴定了大量的标本,解决了许多信息录入过程中的书写疑问;姚淦工程师和徐增莱研究员在后期阶段的标本信息录入过程中,在疑难标本的鉴定以及文字识别方面给予了很多帮助;陈岳高级工程师随时解决软硬件出现的问题;熊豫宁和马振秀参与标本的整理,做了大量工作。正因为上述各位的大力协助,使得本馆的标本数字化得以顺利开展,在此一并表示诚挚的感谢。

参考文献:

- LI MG, 2000. Plant specimen information system 4.4: system management [J]. Acta Sci Nat Univ Sunyatseni, 2: 29. [李鸣光, 2000. 植物标本信息系统 4.4: 系统管理 [J]. 中山大学学报(自然科学版), 2: 29.]

- LI MG, MIAO RH, XU ZR, et al., 2000. Plant specimen database and biodiversity research [M]// Biodiversity Community, Chinese Academy of Sciences. Biodiversity conservation in China facing the 21st century. Beijing: China Forestry Press: 450-454. [李鸣光, 缪汝槐, Zhaoran XU, 2000. 植物标本数据库与生物多样性研究 [M]//中国科学院生物多样性委员会. 面向 21 世纪的中国生物多样性保护. 北京: 中国林业出版社: 450-454.]
- 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, Zhaoran XU, 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.]
- LING PP, TANG JS, 1988. The microcomputer management system of specimens in Jiangsu Institute of Botany [M]// *Bulletin of Nanjing Botanical Garden Mem. Sun Yat-Sen* 1987. Nanjing: Jiangsu Technology and Science Press: 39-42. [凌萍萍, 汤傲杉, 1988. 江苏省植物研究所微型计算机植物标本管理系统 [M]//南京中山植物园研究论文集 1987. 南京: 江苏科学技术出版社: 39-42.]
- MA B, LI ML, LIU JK, et al., 2001. The construction of Sichuan University botanical specimen database [J]. *J Sichuan Univ (Nat Sci Ed)*, 38(6): 839-843. [马波, 李梦龙, 刘冀昆, 等, 2001. 四川大学植物标本馆馆藏标本网络数据库的构建 [J]. *四川大学学报(自然科学版)*, 38(6): 839-843.]
- PANKHURST R, 1993. Taxonomic databases: the PANDORA System [M]//FORTUNER R. *Advances in computer methods for systematic biology*. Baltimore: The Johns Hopkins University Press: 229-590.
- SHETLER SG, 1973. An introduction to the botanical type specimen register [M]. Washington: Smithsonian Institution Press: 1-25.
- SOBEEON J, 1999. Linking biodiversity information sources [J]. *Trends Ecol Evol*, 14: 291.
- WANG LS, CHEN B, JI LQ, et al., 2010. Progress in biodiversity informatics [J]. *Biodivers Sci*, 18(5): 429-443. [王利松, 陈彬, 纪力强, 等, 2010. 生物多样性信息学研究进展 [J]. *生物多样性*, 18(5): 429-443.]
- XU PG, 1992. The first biological research institute in China [J]. *Chin Mater Sci Technol*, 13(2): 47-57. [薛攀皋, 1992. 中国科学院生物研究所——中国最早的生物研究机构 [J]. *中国科技史料*, 13(2): 47-57.]

(责任编辑 蒋巧媛 邓斯丽)