

附件 1 标本数字化流程及技术指南

植物标本子平台课题组

(2015 年 5 月)

前言

经过多年实践,植物标本子平台及 CVH 参建单位在标本数字化技术的各个环节上积累了丰富的经验。为总结经验和促进今后工作,子平台在各位同行的支持及配合下,撰写本技术指南,内容包括标本数字化工作流程、标本影像制作技术指南及标本影像验收标准等内容。指南中对标本影像部分作了重点描述。我们认为,符合指南要求的标本影像既是数字化工作成果也是用户共享的最重要信息,高质量的标本影像还可以弥补信息录入的不足,例如信息录入误漏等,因而,对长期保存数字化标本信息也具有重要的意义。对于那些存在不足和未能全面达标的标本照片,若再予以重照补拍,将带来标本再次遭受损耗和人力资源浪费的后果。由于时间匆忙,此版指南可能会有错漏,希望各位同仁不吝批评指正。

1 标本数字化工作流程

标本的数字化流程主要包括标本整理、信息录入、影像制作和信息及图像上传四个步骤(图 1):

1. 标本整理

在进行标本信息录入和照相之前,需要对馆藏标本进行整理及清理,目的是选出合格标本。合格标本的主要标准是有花果或其中之一、有采集记录签和鉴定签——即“三有标本”(图 2)。采集记录至少包含采集人/队(和采集号)、采集时间和采集地等基本信息;定名签或鉴定名称要求到种,至少到属(拉丁名);标本本身即使缺少花果,至少也有相当的茎叶枝条,否则不予数字化和提交。

标本整理的难点在于邀请专家鉴定标本,还有所缺标签的增补、褪色标签的更换,以及加固标本和清扫灰尘等等。

2. 信息录入

将采集记录签、定名签和条形码等信息录入到下发的录入系统中(附件 2,图 3),要求“忠于原著”——即将标签信息原原本本地录入到相应的字段中!注意辨别标签上的字体,尤其是人名和地名,一旦错误录入将难于发现和纠正。标本信息录入好后应及时校对,保证录入正确率符合任务书的要求;

3. 标本影像制作——拍照或扫描

对整理好的标本进行影像制作,不外乎是通过拍照或扫描的方式来实现(图 4)。无论是哪种方式,都应该贴好条形码并在合适位置放置标尺及色卡(图 6)。

条形码和标尺的位置应选在台纸中心附近及靠右侧处,靠左的位置常常被属/种夹盖住而不便查看,同一标本馆的位置应固定,且摆放正,不要倾斜。

原则上条形码应贴在台纸下方边缘中部或靠外侧,也可以贴在台纸右上角,避免贴在里侧或左上角。条形码由子平台课题统一制作,如果自行制作要求打印清晰、粘贴牢固、不易脱落、不反光;编码必须同馆不断号、不重复。

原则上标尺及标准色卡位置摆放也应相对固定。标尺通常是放在台纸右侧,摆正,以不遮盖标本为宜;

4. 信息及图像上传提交

照片拍摄之后,进行批量裁边及调光,然后以条形码号命名好,妥善保存,等待上传至平台。

2 标本影像验收标准(与任务书上的描述一致!)

拍摄标本必须使用 1200 万像素及以上的数码相机。照片的尺寸为 3000 像素×4000 像素以上,图像 1200 万像素以上。每张至少 3 兆字节以



图 1 标本数字化流程图

上(像素推荐:2400万像素。建议使用 Nikon D3200 或者更高档次的数码相机拍摄)。

相片的质量要求成像清晰(包括所照的植物体和标本上的标签、标尺、色卡、条码),白平衡准确,色彩自然,画面光亮统一适度,大小要求覆盖整份标本。

模式标本:所使用的扫描仪规格为 A3 型,分辨率为 600 线,照片大小是: $12 \times 600 = 7200$ (宽); $16 \times 600 = 9600$ (长); $9600 \times 7200 = 6912$ 万像素;原始文献扫描使用:300 线。

照片拍摄后,需要批量进行裁边及调光,使画面被整个台纸铺满。

照片处理后,以条形码命名图像文件(图 5)。

3 标本影像制作技术指南

1. 基本概念

1.1 条形码 参加本项目数字化的标本必须有条形码。条形码是数据库中每份标本的特异编码,同一标本馆条码号应该是连续(不间断)和不重复的。项目组统一规定,条形码构成为:标本馆代码+7 位标本编码,如 PEY0066773 代表北京大学生命科学院植物标本馆第 0066773 号标本。

建议条形码应贴在台纸下方边缘中部或右(外)侧,避免贴在里侧或左上角,因台纸里侧被

插入属种夹子而不易查看,另外,通常台纸左上角通常贴采集签、上方边缘粘/别有散落材料包、右下角有鉴定签而不适。注意条形码不可覆盖原来的标本馆代码或图标,也不要被标本覆盖。

1.2 标尺和色卡 标尺和色卡同样是数字化标本所必备的。前者便于研究者丈量标本各个部分大小,后者(色卡)主要是用于纠正颜色的偏差,从而精确区分花瓣、果实等结构的颜色。项目组(课题组)提供统一的标尺和色卡,各单位也可以自己制作,只要质量合格即可。

建议标尺置于台纸右侧或下方,与边缘平行为好,色卡位置则稍灵活,以不压住标本和其他签(鉴定签、采集签、号牌等)为准(图 6)。

1.3 号牌 多数标本都挂有号牌,为硬纸质,方形或圆形,统一印制或手写的,也有用纸张做成的号牌。号牌在标本鉴定、研究及引证中具有重要的意义。拍照或扫描标本时,须将号牌有记录的一面朝上。

1.4 像素 像素是组成位图的最小单位。像素多少决定了位图图像细节的精细程度。通常情况下,图像的分辨率越高,所包含的像素就越多,图像就越清晰,印刷的质量也就越好。同时,它也会增加文件占用的存储空间。

1.5 分辨率 分辨率(亦称为解析度)是指单位英寸中所包含的像素点数,即像素密度。往往用于

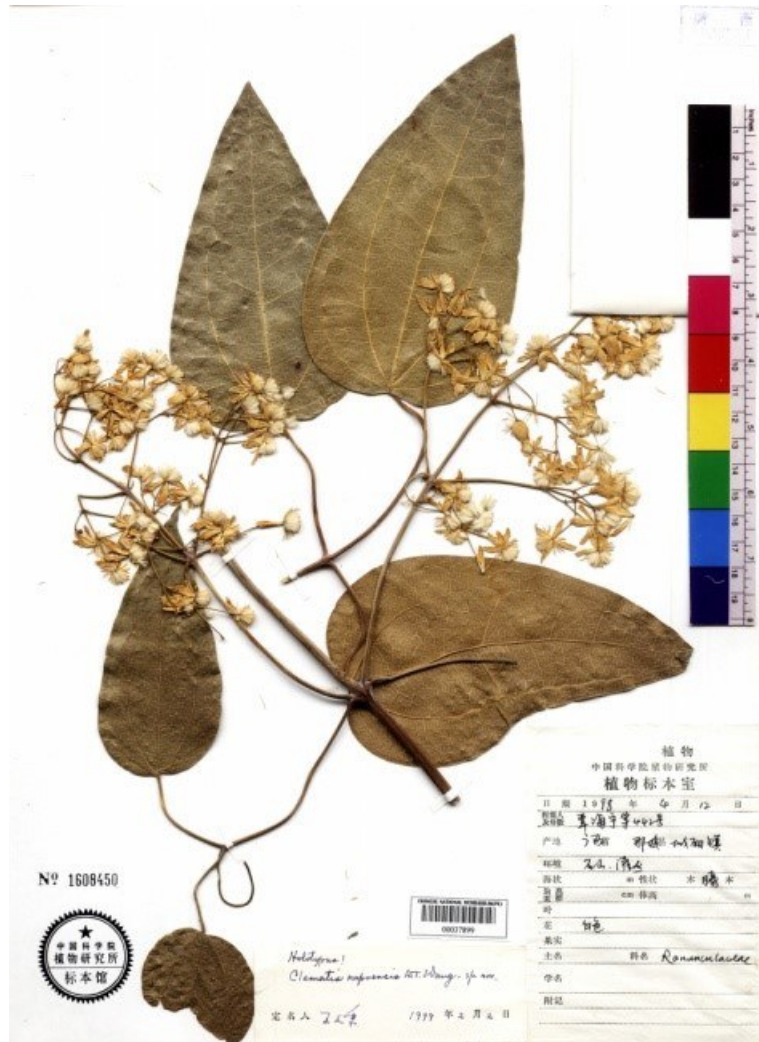


图2 “三有标本”——有花果、采集签、鉴定签(PE: 数字化以后)

扫描、打印设置或者显示设备,通常表示成每英寸像素(Pixel per inch, ppi)和每英寸点(Dot per inch, dpi)单位为PPI或者DPI。一般显示器的显示分辨率为72 ppi、视网膜屏144 ppi,打印分辨率一般为300 dpi;一般打印机分辨率为300 dpi。

注意:实际使用中“分辨率”常与“像素”混用。分辨率被表示成图像或设备每一个方向上的像素数量(比如4K屏幕分辨率其实是指长边像素为4000像素左右,16:9的显示器一般为4096×2160)或总像素(尼康D800相机的最高分辨率为3600万像素)。

1.6 像素与分辨率的关系 边长像素=尺寸×分辨率,总像素=长边像素×宽边像素

如:标本尺寸29×42cm(12×17inch),推荐的

扫描分辨率为300 dpi,则短边12 inch×300 dpi=3600pix 总像素为:3600×5100=18360000,即一千八百万像素。

2. 标本影像制作途径

标本影像主要通过2条途径获得,一是扫描仪扫描,二是相机拍摄。

扫描仪扫描 扫描数字化途径通常使用平板扫描仪扫描,其优点是分辨率可调,结构简单,操作容易。但普通扫描仪有两个直观的缺点,1.景深(清晰范围)不可调控,限于扫描仪结构,材料比较厚的标本,例如果实比较大的树木标本,靠近扫描一端清晰,压不到扫描平板的地方则模糊;2.颜色表达不准确,可以调整白平衡,但是调整过程比较麻烦,扫描仪使用时间久了之后,冷阴极灯管



图 3 标本数据录入软件 (GINKGO-s 3.4)



图 4 仪器设备 (电脑、扫描仪、标本照相系统等)

色偏严重,后期需要使用软件根据标本上的颜色标尺校准颜色后方可达到项目合同要求。

欧美发达国家目前普遍采用专用扫描仪扫描植物标本,较好地克服上述的缺点。我们因条件所限未用普遍推广该办法。

相机拍摄 使用相机拍摄优点在于:景深易于控制,对于藏量较小标本馆,不拍摄标本时相机可作它用,增加实用性;缺点在于:需要操作人员具有一定的摄影知识和一定的时间的操练才能灵活掌握。

使用相机快门线和软件控制相机拍摄 一般拍摄时将相机安装在翻拍架(推荐)或者三脚架上,手动按下相机快门拍摄。此方法比较费时,且拍摄完成后,需要将存储卡取下,导入计算机中,费时

费力;大量拍摄使用存储卡,易达到使用寿命,可能会有照片损坏,此外,长期插拔存储卡,会磨损插卡槽。

建议有条件标本馆,尤其是大中型标本馆采用相机快门线和软件操控相机法。该法除了解决上述问题,还能将所拍摄照片直接存储到计算机,根据一定规则对照片进行命名,实时查看照片拍摄结果,等等。但此法的应用需要操作人员具有较高的技术水平和耐心,一旦熟练使用则大大提升拍摄速度和工作效率。

选择什么途径 以上途径中,选择相机+三脚架/翻拍架+软件控制拍摄标本(图4)是比较经济划算、可行,易于实现的方案。具体方式的选择由各单位根据主客观条件来定。

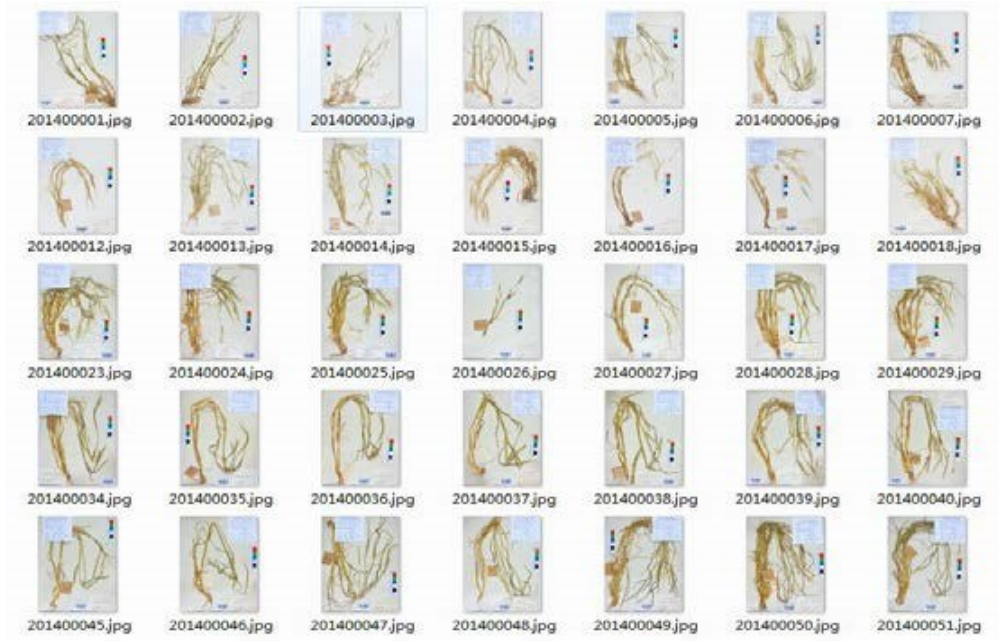
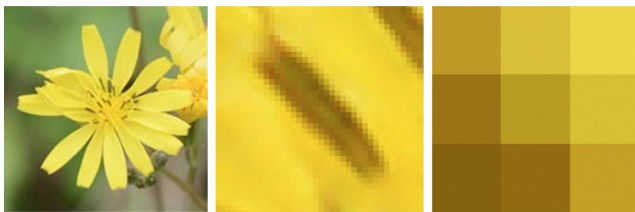


图 5 上传的合格标本图像(四川农大小麦研究所标本馆)



图 6 合格数字标本(注意标签、标尺及条码的位置)



400×400 pixel

40×40 pixel

3×3 pixel

无论哪种标本影像制作途径，对影像的质量要求都是一致的。本项目对标本影像质量有 4 点基本要求：清晰、色准、覆盖标本、像素大于 1200 万。覆盖(整张)标本、1200 万像素主要是对硬件配置的要求；清晰和色准是人为操控校准上的要求，也是难度较大、最需下功夫的技术要点，包括事先设置和事后处理。

附件 2 Ginkgo-s 3.8 用户手册(2015 年 4 月修订)

植物子平台课题组
(2015 年 5 月)

一、安装

1. 程序下载

Ginkgo-s 3.8 安装包可以从中国科学院植物研究所植物标本馆网站免费下载, 下载网址: <http://pe.ibcas.ac.cn>, 在菜单中, 进入数据中心后点击程序下载标签, 在下载页面中选择 Ginkgo-s 软件, 下载到电脑或直接安装。

2. 程序安装

点击安装文件开始安装, 一路点 next 即可。在安装过程中如果遇到防火墙出现安全提示并询问是否允许本次操作, 选择允许本次所有操作。杀毒软件可能会提示软件有危险, 是因为它未识别的软件都认为有危险。

3. 程序升级

点击主菜单中的升级项可将系统升级到最新版本, 升级时需要电脑连接互联网。

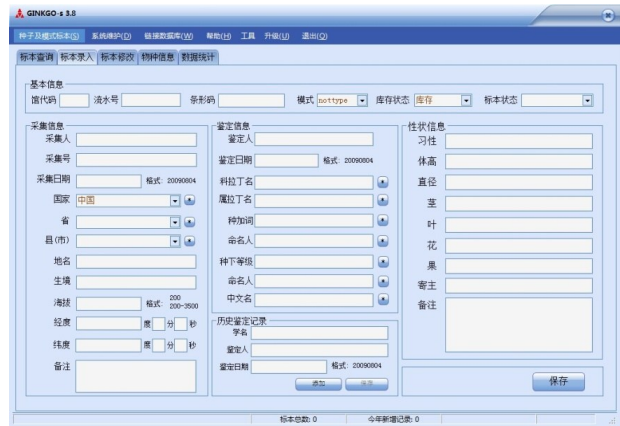
如果采用重新安装方式升级需要注意的是要将现有数据库备份, 因为安装后数据库将被替换为新数据库, 原有数据将全部丢失。

4. 新功能

Ginkgo-s 3.8 主要更新包括以下几个方面: 预装了物种学名库, 可以直接选择录入; 录入学名时直接弹出科、中文名、命名人等信息; 查询界面增加多个自动提示功能; 数据合并工具中增加了错误数据导出功能。

二、数据录入

启动程序, 点击标本录入标签打开录入界面(见下图)。



1. 通则

馆代码和条形码是必录项, 这两项不能空。输入某项内容后按回车键可以自动进入下一项。

所有数据录入完成后点击保存按钮, 如果中途点保存则只能到标本修改界面添加还未录入的信息。

2. 采集地录入

采自中国的标本要录入省份和县的行政区名称, 省级只能选择, 县一级最好选择, 如果选择列表中没有要录入的名称可点击右侧带 * 号按钮添加新名称。县以下小地名录入到地名栏中。

国外标本除国名外其余各级别名称均录入到地名中。

3. 学名录入

本软件已将所有国产物种名称预装到数据库中。学名共七项内容: 科拉丁名、属拉丁名、种加词、种级命名人、种下等级、种下等级命名人和中文学名。拉丁名在录入时会根据已输入字符列出与之匹配的名称, 如果没有所需要输入的信息可点击右侧带 * 号按钮添加新名称。

学名录入技巧: 直接在学名处理开始录入, 如果弹出所需名称选中之后其他信息会自动填入, 包括科名。

学名信息可以在物种信息界面中录入或修改。如图。



4. 历史鉴定记录录入

历史鉴定记录是指标本台纸上有多个鉴定签时鉴定时间较早的若干个标签内容。录入应该是在输入该份标本条形码后进行。先点击此栏目内的添加按钮后再录入相关内容,录入完成之后要按保存按钮。如果有多个历史鉴定记录重复上述操作。



三、数据修改

有两种方式进入修改界面。一是点击标本修改标签进入,此外,也可以由查询界面直接进入修改界面。

1. 点击标本修改标签进入修改界面



进入修改界面后输入查询条件后点击开始检索按钮,待数据返回后即可修改相关信息。

2. 在查询结果中点击图标进入修改界面

如果需要在查询中修改当前记录,可点击图标直接进入修改界面进行修改。



3. 注意事项

修改数据的要求与录入基本一致,主要注意学名、行政区名称及经纬度格式等。

4. 录入新鉴定标签

点击转至历史鉴定按钮即可将现有鉴定信息保存到历史鉴定中,如保存成功学名信息会清空,此时可录入新的鉴定信息,完成后点击保存按钮保存新的鉴定信息。

5. 其他

条形码和馆代码是不能修改的,如发现条形码输入错误可删除该条记录重新录入,或者打开数据直接删除。

四、数据查询

程序启动后首先看到的的就是查询界面。查询界面分四个区域,分别是查询条件设置、结果显示、照片显示、分布图显示。

1. 查询条件设置



根据需求填写查询条件,查询内容如上图所示,各条件之间是和的关系,也就是说输入的条件越多返回的记录数越少。设置好检索条件后点击开始检索按钮开始检索。如要重新设置检索条件可点击清空选项按钮清空全部已设内容。

有些项目是要在列表中选择,如科、属、种等。涉及日期的项目可以限定范围。选择不同符号(<、\、=)可设置该项目的时段。

2. 数据浏览

移动记录有两种方式,一是点击左侧的条形码列表,另外就是点击列表下面的导航按钮。如下图。

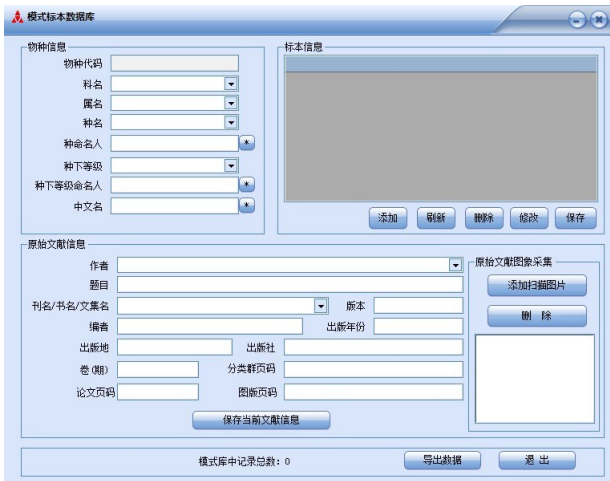


点击列表显示按钮可以表格形式显示数据。

3. 照片浏览

照片需要放到安装目录下的 image\specim\目录下, 文件为 JPG 格式, 文件名为该标本的条形码号。

照片可进行放大镜、原大和保存等操作。



4. 报表打印

如果需要打印检索出来的数据可选择报表类型后点击显示报表按钮, 这样会生成所选报表格式, 预览没问题可打印, 也可另存为 Word 或 Excel 等文件格式。

也可输出为采集记录签。

5. 数据导出

检索出来的数据可以直接导出为 EXCEL 文件, 点击"导出"按钮系统会将数据存为 EXCEL 文件。

6. 修改数据

点击"修改"按钮可进入修改界面修改当前数据。

7. 删除数据

点击"删除"按钮可直接删除当前记录。

五、模式标本

模式标本录入与普通标本相同, 所不同的是模式文献录入。

在菜单中选择种子及模式标本-模式标本录入与查询, 弹出模式标本数据库窗口, 如下图。



在此窗口操作之前应已经完成模式标本基本信息的录入, 在这里只是将相关的物种信息与其绑定。首先要做的是选择物种学名, 之后添加模式标本并保存。

选择物种学名之后还可以添加原始文献信息, 添加后保存即可。

点击导出数据按钮可以导出数据库中所有模式标本信息。

六、系统维护

1. 数据备份

在菜单中选择系统维护-数据备份, 弹出文件管理器窗口, 在此选择一个路径用于保存数据库即可。

2. 数据恢复

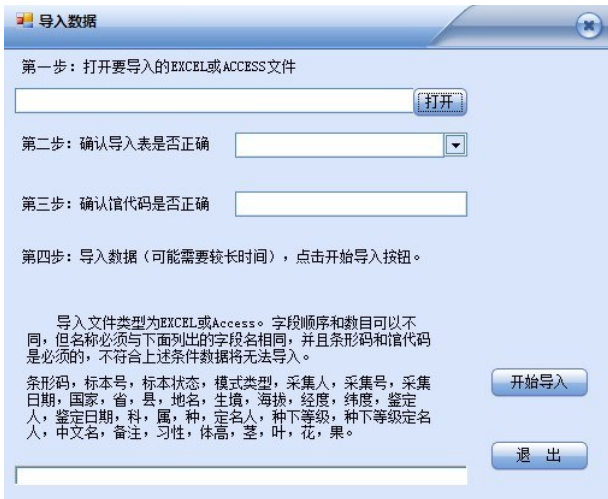
在菜单中选择系统维护-数据恢复, 弹出文件管理器窗口, 在此选择一个备份的数据库, 此时将替换现有数据库, 这也是和数据合并的区别。

3. 数据合并

数据合并与数据导入的区别在于数据合并是将两个分别录入的数据库

taxon.mdb 的数据合并, 而数据导入是指将其他方法录入的 EXCEL 表导入数据库。

在菜单中选择系统维护-数据合并, 弹出合并数据窗口(见下图)

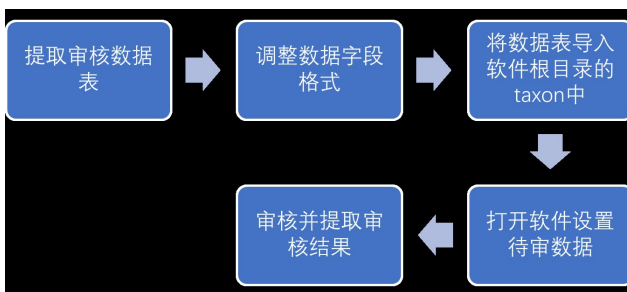


在上图中第一步是点击打开文件按钮选择要导入的数据库文件,选择好后点击开始导入按钮直到出现成功提示,在导入过程中不要终止此项操作。导入的数据库必须是 GINKGO2 及以上版本,因之前版本数据库在结构上与 2.0 之后是有区别的,无法正常导入。如果遇到数据无法导入系统会将未导入的数据列表于当前窗口,用户可以将其导出至 EXCEL 表中。

4. 数据导入

数据导入主要是指将 EXCEL 格式数据导入到数据库中,也可是与本系统不同的 Access 数据库,但对数据表的字段有一定要求。如果是 Ginkgo2 的数据库可采用合并的方式更为简单。

在菜单中选择系统维护-数据导入,弹出导入数据窗口(见下图)



导入数据分为四步,首先是打开导入的文件,第二步选择一个要导入的数据表,第三步确认馆代码是否正确,如不正确可修改,最后点出开始导入按钮开始数据导入。

导入的数据必须包含所提示的所有字段,字段内容除条形码外均可以空,字段顺序可不同。

在导入过程中不要终止操作。

5. 数据库操作

本系统采用 Access 数据库,文件名是 taxon.mdb,在安装目录下可以找到。如果需要直接操作数据库可以将其打开编辑数据。标本数据保存在 spcim 表中,物种信息保存在 taxon 表中,行政区名称在 geog 表中,历史鉴定记录保存在 spcim_his_iden 表中,性状保存在 spcim_plus 表中。spcim 是系统主表,与其他表关联是依赖于条形码号或相关表 ID,因此,要确保这些关联字段的完整性。需要注意的是子表 taxon 和 geog 中不应有重复数据,如有重复数据会导致查询时出问题。

七、照片编辑

在系统中提供一个简单的照片编辑工具。可以对照片进行旋转、剪裁和条形码识别。

八、技术支持

在使用过程中如发现有任何问题或疑问请与课题组(子平台)联系。

附件 3 标本数据审核的一般流程及审核内容

(2017 年 5 月)

一、植物子平台数据审核软件(CVH_Data_Check)操作流程及注意事项

1. 提取审核数据表

将同一单位不同时间上交的数据合并到一个数据表中。(注意:上交的全部数据表格都要备份)

2. 调整数据字段格式及顺序

将该单位的数据表中的字段顺序调整成数据审核软件(CVH Data Check)中自带的“待审数据”表中的字段顺序,以便下步审核使用。

3. 将数据表导入软件根目录的 taxon 中

可以复制、粘贴实现此步,也可以通过数据库自带的“导入外部数据”来实现。

4. 打开“CVH_Data_Check.exe”进行设置

指定照片位置,选择该单位的馆代码,随机选择待审记录(一般选择 10%),审核后转入总汇。(注意:在第 3 步后,应将“CVH”数据库关掉,然后启动 CVH_Data_Check.exe)

5. 审核并提取审核结果

进入审核界面,左边是录入的标本信息,右侧

是标本照片。将标本信息与照片上的采集签、鉴定签、吊牌以及条形码进行比对,并查看照片的亮度、清晰度以及标本的完整性,若都合格,就点击“合格”,进行下一份标本;若不合格,写清楚不合格的原因,点击不合格进行下一份标本。当所有标本都审核完毕,返回“设置”页面,点击“导出抽样不合格记录”便可得到不合格的 excel 表格。(注意:在写不合格原因时,最好越详细越好,并且同一不合格原因用同一语句表达,便于最后统计。)

二、审核内容

1. 数据审核要求:

录入信息与台纸上的标本及标签信息一致;重点检查必录字段,即包括采集人-号码-采集时间、采集地点(至少到省和县级)和植物名称(科名和学名)——如果原标本上没有这些信息,需要补全之后再数字化,否则为不合格数据!

需要输录的字段要求及注意事项

字段名	要求	备注	
条形码	馆代码+数字编码。数据信息与扫描的台纸信息完全匹配。		
名称	科名	1.名称体系每个部分单独一个字段; 2.不能出现其他无关的符号,如“?”、汉字等; 3.拉丁名要求跟台纸上的名称完全一致,而不是接受名或异名; 4.如果台纸上有多个拉丁名,录入鉴定时间最新的那个。	
	属名		
	拉丁名		属名+种加词+种下等级
	命名人		
	种下等级		
	种下命名人		
	中文名	如果有多个中文名称,需要录入前三个信息。	
采集人	按采集签录入。如果有多个采集人,全部罗列出来,不同的采集人名字之间用逗号隔开。	1.不能出现一些不确定的数字或?号; 2.如果大批量缺失采集人信息需要核实资料补充后录入。	

采集号	按采集签录入。如果有多份标本，同一个采集号只算一次。	不能出现一些无关的数字格式或?号等字符。
采集日期	格式要求：YYYYMMDD(年月日)	格式必须全部统一。如：19380913, 20051201, 19000000。
采集地	国家	按采集签录入。 不能出现无关字符，如数字、?号等。 采集地信息至少补充到县级。
	省份	
	区县	
	小地点	
鉴定人	要求同采集人。	
鉴定时间	格式同采集时间。	
经纬度	全部统一成小数点格式。	eg: 37°15'54"=37.265 如果原标签非小数点格式，建议： 1 保留原始格式； 2 增加一个字段，为小数点格式。
海拔	格式要统一。对于区间海拔，录入高的那个数字，比如1000-1200米，则录入1200米。	负海拔指的是低于海平面的海拔，而不是区间中的低的那个海拔数值，比如1000-1200米，负海拔不是1000米。
标本性状描述： 花、果、植物高度、颜色等	按照台纸上的内容逐一录入。	
生境	按照台纸上的内容录入。	

2. 照片审核要求

审核项目	要求
方向	全部直立
裁边	全部裁边
清晰度	清晰(包括所照的植物体和标本上的标签)，大小要求覆盖整份标本，图片大小短边不小于4800像素或长边不小于6400像素，局部放大仍然清楚。
色彩	曝光适度

三、审核合格率

要求合格率不能低于95%。以标签信息上的必录字段(采集人号时间,地点,植物名称)及照片为计。

四、生成该馆标本核查记录表

xx 标本馆数据核查记录表(模板)

核查时间:

核查人:

核查结果: 有关该批标本数量及质量情况的描述。

结论(简要): 合格与否,存在的主要问题及建议。

附件 4 “中国数字植物标本馆(CVH)”数据申请表

申请人姓名		
** 身份	1 研究生: 硕士 <input type="checkbox"/> 博士 <input type="checkbox"/> ; 2 职工及职称: 初级 <input type="checkbox"/> 中级 <input type="checkbox"/> 高级 <input type="checkbox"/> ; 3 其他 <input type="checkbox"/>	
单位及部门		
通讯地址		
申请人邮箱及电话		
导师/组长/主任姓名及邮箱		
*** 项目/课题名称	国家重点基础研究发展计划(973 项目)	
	国家高技术研究发展计划(863 项目)	
	国家科技支撑计划项目	
	重大专项	
	国家自然科学基金: <input type="checkbox"/> 面上; <input type="checkbox"/> 青年; <input type="checkbox"/> 地区; <input type="checkbox"/> 重点; <input type="checkbox"/> 杰青	
	中国科学院项目	
	其他项目(来源部门及名称)	
	研究生论文(题目)	
申请数据描述		
数据应用说明		
发送数据描述 [这部分由 CVH 联系人填写, 包括总份数, 及馆藏、省份和类群等细节]		
数据使用承诺	<p>本人郑重承诺, 严格遵守 CVH 有关数据共享使用的规定: 1) 所得数据需妥善保存, 未经许可, 不得整体发给第三者, 也不直接放置网络上; 2) 在论文中注明标本数据来自“中国数字植物标本馆”网站, 并把论文 pdf 文档发给 CVH 联系人存档。</p> <p style="text-align: right;">申请人签名: _____ 日期: _____</p>	

* 申请流程: 申请人从 CVH 网上下载表格填写→CVH 审核→申请人签名→CVH 发放数据。

** 身份: 如果申请人是研究生或非高级职称者, 需填写导师或课题组长(室主任)信息;

*** 项目/课题名称: 需填写标本所用于基金(或其他项目/课题)的名称。