

# 动物标本规范化整理、录入与审核工作细则

(中国科学院动物研究所 北京 100080)

动物标本是动物物种信息的基本载体。保护动物标本的内在和外延信息并使其与各种服务体系进一步结合，共同造福于社会是我们工作的重要目标。因此，明确动物标本信息数字化的工作内容，严格按照要求进行数字化处理是必要的。

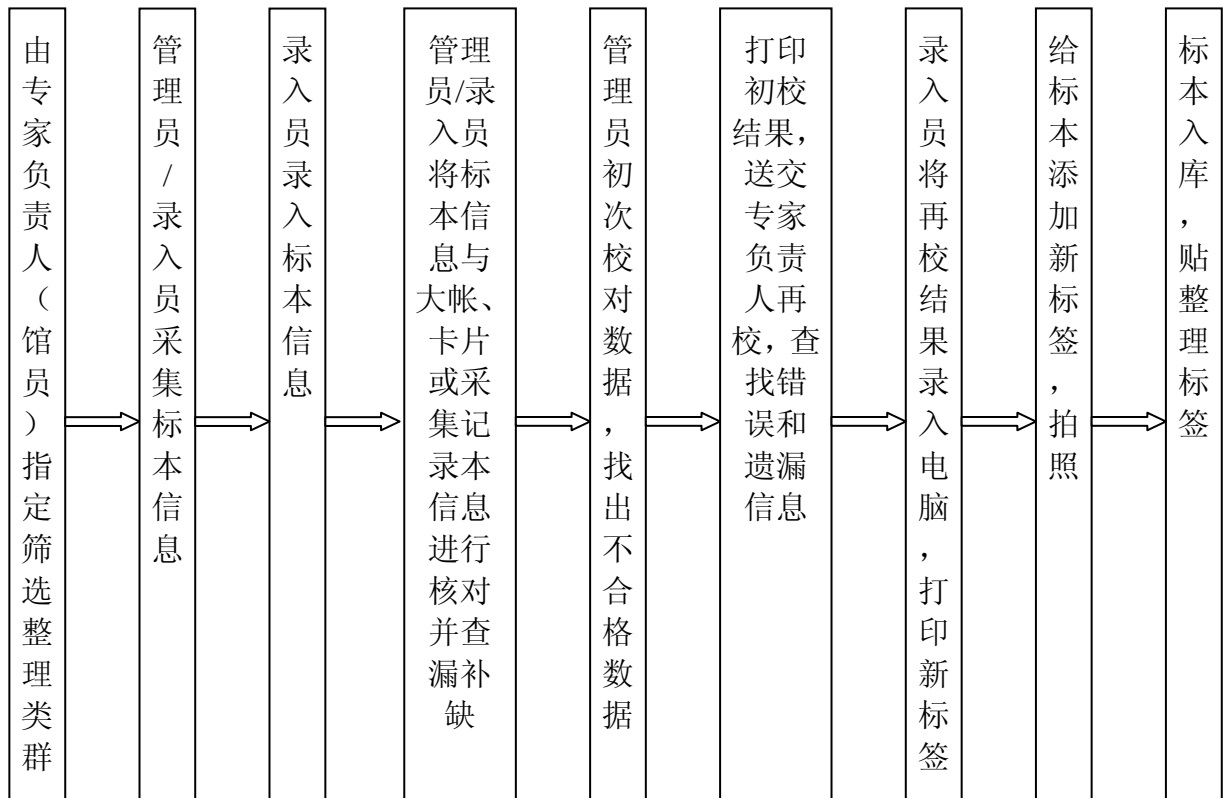
动物标本信息主要包括文字信息和多媒体信息，前者包括标本信息和文献信息，后者则一般包括图像信息和声像信息。它们的数字化处理过程主要有三个步骤：标本信息的整理、录入及审核。标本信息的整理是指对馆藏标本进行普查，收集所有的原始信息并记录标本的保藏状态、保藏方式、保藏位置等内容；标本信息的录入是指将整理时收集和记录到的标本信息录入到相应数据库中；标本信息的审核可分为录入人员的自我核对、专家的审核及数据管理人员的审核。前者主要核对由于录入者失误所造成的信息失真、缺失及偏差，后二者则主要检查录入的信息是否符合科学要求、是否符合项目要求、是否按既定进度进行等。

标本数据信息的整理、录入及审核可能会因类群的不同而有一定差别，但其基本原则是统一的，即全面而真实地反映标本的实际情况。根据以往的工作经验和教训，我们认为规范化的整理、录入及审核将有益于各项工作的顺利进行，因而将主要的工作内容、要求及工作细则详述如下（以动物所工作流程为例，供大家参考）：

## 1 动物标本规范化整理、录入及审核的类型与工作流程

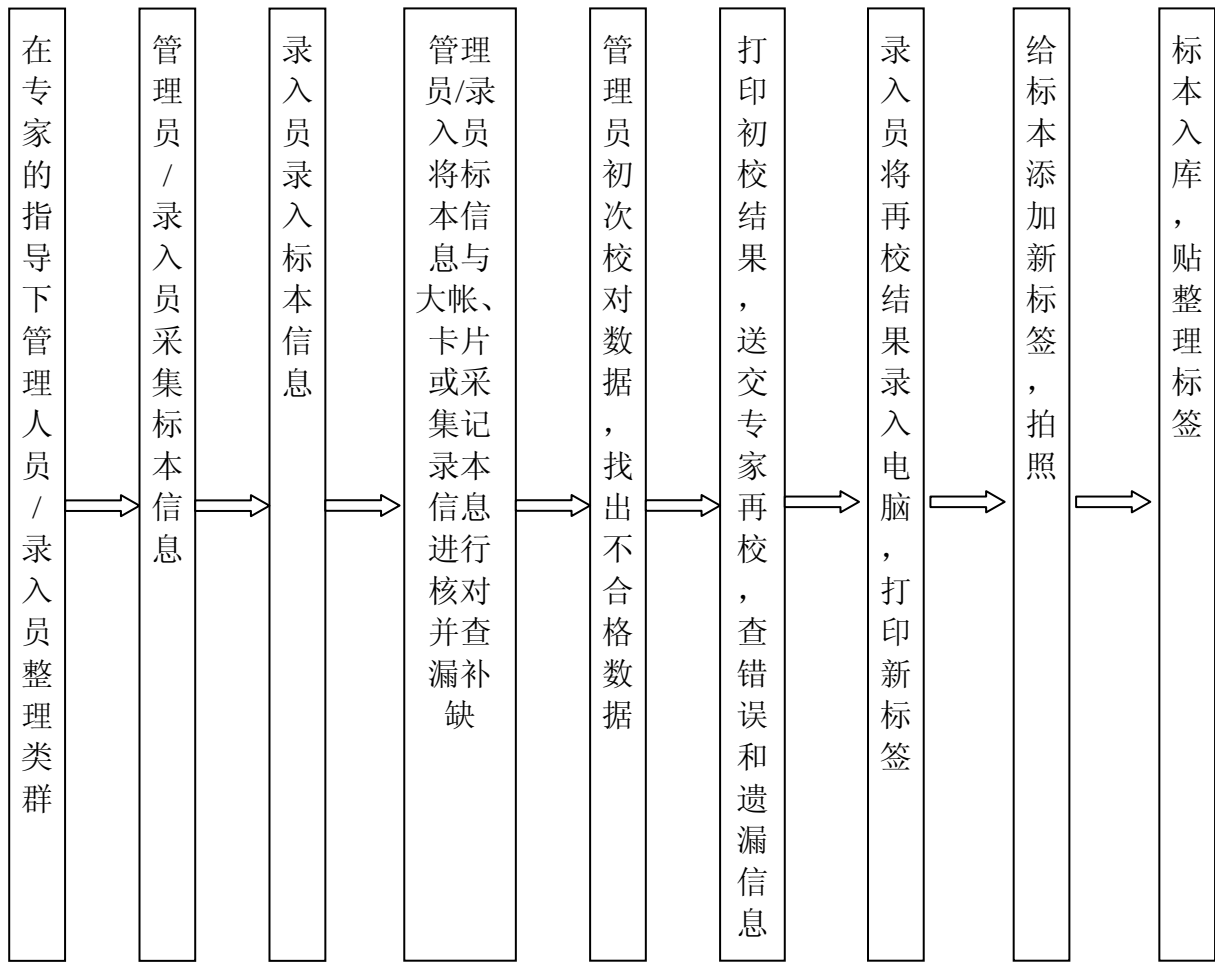
我们根据专家参与的程度，将标本信息的规范化整理、录入和审核分为三种方式，并对管理人员如何协助专家更好地完成工作做了较详细的工作流程设计：

### 1.1 无专家参与



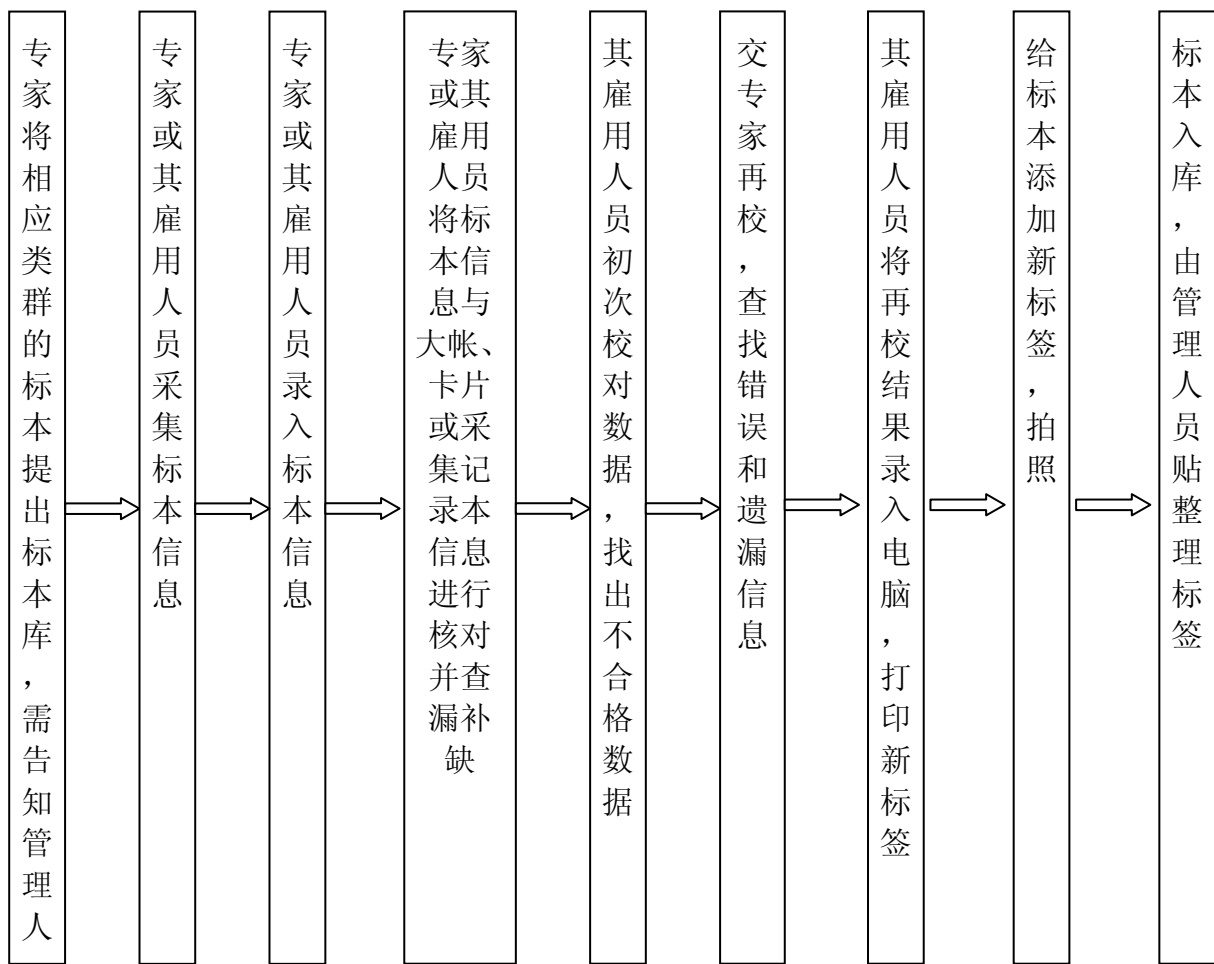
## 1.2 专家部分参与

此流程仅供各成员单位的专家参考。专家根据自己类群的特点及工作习惯对流程进行变更。



### 1.3 专家独自完成

此流程仅供各成员单位的专家参考。专家可根据自己类群的特点对流程进行采纳和变更，甚至舍弃。



## 2 动物标本信息的整理

标本及其信息的整理是指对馆藏标本进行普查，收集所有的原始信息并记录标本的保藏状态、保藏方式、保藏位置等内容，同时对一些受损标本进行适当的保护性复原。标本信息中，标本的原始文字信息主要体现在采集记录本、标本的标签以及入馆记录本（俗称“大帐本”）中。整理过程中，应全面收集并整理以上材料，并通过查阅各种文献资料或与专家的交流对一些字迹模糊、地名不规范、时间不明确等信息进行更正或补充。在采集原始信息时要遵循全面、准确的原则，尤其在对时间和地点等信息进行记录时需更加仔细。目前动物所采用的主要工作流程是，先将标签信息及标本的保藏状态、保藏方式和保藏位置等抄录在“标本信息记录表”中，再通过大帐本和采集记录本等补充其他信息，最后将这个信息表送交专家进行核对。在专家的指导下，对于一些异常的原始信息，包括物种名、标本号、人名、地名、寄主名称、采集地信息等进行校正。“标本信息记录表”需要永久保存，作为数据查找、备份和校错的原始依据之一。

在进行标本整理时还须注意以下内容：

- 各个类群按所制定的标准对标本分级，至少要区分为模式标本和普通标本；并可进一步按标本质量、标本信息完整性、保藏状况等再分级。
- 脊椎动物整理主要完成信息标准化、对可修补的残损标本的修补、对不可修补标本的标注、核对卡片、目录（大账）、采集本与标本的一致性、核对标本分离部分（如兽类骨骼和皮张分别保藏的情况）的一致性。
- 鱼类标本目前一般仅在标本上有一个记录标本号的竹签，因此核对卡片、目录（大账）、采集本与标本的一致性显得更为重要。另外，鱼类标本有时在一瓶中有不同物种的情况，因而需要分瓶。最后，应对过去的一些劣质标本瓶（瓶口不严等）进行换瓶，每个标本瓶中的每号标本尽可能加一个耐液浸的标签，并在标签上标注全部采集信息。
- 整理后的标本必须按相应类群的学科规范制作（如针插、整姿等）；
- 玻片（液浸）标本若每片（瓶）有多个个体，应保证每片（瓶）只包含一个物种，且所有个体应属同一采集时间和地点，并具准确的个体数；这样的一片（瓶）为一号。
- 每号标本要加一个计算机打印的标本号签；
- 每号标本均须有采集标签，并以适当的方式与标本保存在一起；对于不规范的旧标签，应保留原标签，并添加相应的新标签，新标签应对旧标签的信息进行标准化处理。
- 所有动物类群的模式标本标签均须要有明确的标识，按国际惯例正模用红色标签，标注“Holotype”，配模用蓝色标签，标注“Allotype”，副模用黄色标签，标注“Paratype”；其它类型模式标本待进一步统一。
- 整理的标本应鉴定到属或种，对于部分标本暂时不能鉴定到属的，必须鉴定到科，并说明情况和研究现状，供研究人员和共享使用人员参考。
- 标本鉴定信息应与标本一起完整保存，对不规范、手写难以辨认的鉴定标签应添加计算机打印标签，该标签制定不同类群的标准格式，并加注“鉴定人”项。
- 尽可能对不规范、手写难以辨认的记录，包括地名、人名、有疑问种名，进行收集记录，以便汇总成册
- 所有新制标签均应不改变原有数据的内容和性质，只是补充、完善和标注新旧名称等。
- 以计算机打印条附在标本上为准（脊椎动物）或计算机打印标本号签（昆虫和无脊椎动物）。

### 3 动物标本信息的录入(含图像采集)

目前我们采用的录入系统是《动物信息系统 2.2》，主要有两大部分，即标本信息和物种信息，共 10 个录入界面。标本信息为主要的录入界面。对于物种信息，我们目前

仅要求填写基本物种信息，模式标本的物种还须填写原始文献名称及出处。该系统的具体使用方法可参见《动物信息系统 2.2》使用手册或软件中的帮助。

为了保证标本基本信息的完整性及实用性，我们希望标本的所有信息均能体现在录入系统中，尤其是一些重要的基本信息不能为空白。它们是馆藏信息（物种编码、标本号、标本属性、保藏方式、保藏位置、保藏状况、标本馆名称、标本来源、鉴定人、鉴定日期）、采集信息（采集国家、采集省区、采集县市、采集日期、采集人）和物种信息（拉丁动物科名、拉丁动物属名、拉丁动物种名、原始文献信息、物种定名人和时间、科以上分类信息）。其中馆藏信息和采集信息是在录入系统的“标本信息”界面中录入，大部分物种信息在录入系统的“物种信息”界面中录入，另外科以上分类信息需要在“科名库”界面中录入。需要注意的是，所有录入的同类型数据必须统一录入格式，例如时间的录入格式等。

### 3.1 文字信息

#### 3.1.1 标本信息

标本信息主要是指所有跟随标本的原始文字记录，根据不同的信息记录状态可分为大帐信息、采集本信息、卡片信息以及标签信息等。它们是数字化标本馆的基础支撑信息。标本信息录入就是对这些写在纸上的原始记录进行收集并转变为数字化信息。以下是录入时的必填数据项：

- 标本号
- 拉丁动物科名
- 拉丁动物属名
- 拉丁动物种名（如未定到种可写 sp.）
- 采集国家
- 采集省区
- 采集县市（或山脉、河流、地点）
- 采集日期
- 采集人（1980 年以前标本可以使用“不详”，地方来源标本可以使用“native coll.”）
- 标本属性（是否模式标本等）
- 保藏方式
- 保藏位置

- 保藏状况
- 模式标本必须包括原始文献信息

### 3.1.2 文献信息

文献信息主要是指与馆藏标本相关的文献，其中模式标本的原始文献为主要收集对象。

文献名称及出处的录入是在程序中各类群物种信息库界面下的物种信息选项中填写。具体操作是：进入物种信息编辑的卡片状态后，再进入“分类系统与引证”子选项。在此选项下将收集到的文献名称及出处录入在“原始文献”或“重要参考文献”的编辑栏中。录入格式以“动物志”的要求为准。

收集到的文献无论是什么形式，最终都要整理成以文献题目为名的 PDF 文件。在动物志上发表的文献，它的 PDF 文件可以只包括相关内容的篇章和所在动物志的封面，其文件名为该动物志的卷册名。所有的 PDF 文档单独保存在独立于目前使用的数据库软件之外的文件夹中。

## 3.2 多媒体信息

目前多媒体信息主要是指图像信息，多为标本的照片。为了更好地将图像信息与文字信息链接起来，我们对图像信息作了统一的规定，如下：

- 图像以“组”为计量单位，每组的张数由专家根据各类群特点确定，原则上要求每组图像应该具有分类学上的参考意义和价值。
- 当遇到标本与标签不能在同一画面中同时清晰体现时，应该单独为标签拍摄一张照片。要注意标签中应包含定名签。
- 图像以标本号为文件名，缀以“-1、-2、-3...”区别同一号标本的不同图像。
- 照片保存格式为“JPEG”，像素至少为 2560×1920。
- 每张图像都要有详细的记录，至少包括 ID（物种代码）、中文种名、拉丁种名、标本号、科名、保藏位置以及备注等。
- 建立“标本图像信息记录表”，该表必须包括上交数据库中已录入的所有物种，若某物种因特殊原因无法完成图像采集，则该物种的图像采集记录为空，并在备注中说明无法采集的原因。备注中也可说明标本的属性等信息（建议在 excel 中进行记录，见样表）。
- 普通标本每个物种提供一组图像。

- 模式标本每号标本提供一组图像。
- 若以标本扫描的方式获得标本图像，则扫描精度不得低于 600dpi。
- 拍摄主体应不小于图像总画面的 1/2。

标本拍摄属于静物拍摄，因此拍摄时的光线、背景以及构图是影响照片质量的主要因素。标本拍摄时的光线以白光最佳，要避免出现阴影，标本体色与背景颜色的反差要大，在图片中严格避免出现无关内容。如果遇到较小的标本（体长小于 0.5cm），建议采用显微拍摄。图像的保存是独立于《动物信息系统 2.2》的。保存图像信息的文件夹应以类群名命名。昆虫建议以科名命名文件夹，文件夹名称为“拉丁科名+中文科名”。

#### 4. 动物标本信息的审核

标本信息的审核根据人员可分为录入人员的自我核对、专家的审核及数据管理人员的审核，也可根据工作深度分为一般性审核和详细审核。

录入人员在录入时应不定期进行自我核查，避免由于工作失误，例如笔误，造成的信息失真、缺失及偏差。数据管理人员必须周期性的备份数据并进行数据审核，避免录入数据中出现不符合项目要求的内容。建议，审核周期每个月进行一次一般性数据审核，每个季度进行一次详细审核。图像信息以季度为周期进行审核。在整个整理、录入及审核过程中，专家要不定期对数据进行普查或抽查，避免录入数据中出现专业性知识的录入错误。

**需要特别注意的是，每次上交数据前必须进行详细审核。**为了使审核工作切实起到查漏补缺的作用，动物所针对不同类型的数据规范了基本的审核内容和方法。

#### 4.1 文字信息

##### 4.1.1 标本信息

标本信息的审核主要分 5 个步骤：前期准备、筛选数据、查错并记录、总结。以下各个步骤主要针对数据审核人员：

前期准备

步骤	操作方式
提取需审核的数据文件	拷贝《动物信息系统 2.2》中 works 文件夹下的 a01、a02 相关文件（不同后缀代表不同类群：c 昆虫、n 鸟、s 兽、w 无、y 鱼、z 两爬）
清除无用信息	启动 Foxpro 程序，用“pack”命令彻底删除已删除记录



数据格式转换*	将上述文件在不丢失信息的前提下转化为审核人员熟知的文件格式，建议转换为 access 文件类型
---------	---

注：“\*”视审核人员习惯而定的操作

#### 筛选数据

步骤	操作方式
选择指示数据项	从必填数据项（见前述）中选择可以区分数据是否整理的选项作为指示数据项，如保藏位置有记录即表示该标本是整理过标本，应该对其进行审核
提取审核数据	根据指示数据项筛选出将要审核的数据

#### 查错并记录

步骤	操作方式
选择查错对象	根据必填选项（如前述）选择将要进行审核的项目
查看对象中的错误	对查错对象进行排序*，筛选出空白记录、不规则记录以及不正常记录等其他不符合要求的记录
记录错误	记录上一步骤中筛选出的错误记录
分析其他选项	重复前三步

注：“\*”由于录入形式的不规则会使排序无法达到预期目标，因此在录入时应严格按统一的录入原则进行录入

#### 总结

步骤	操作方式
分析出错原因	在记录错误的文件中说明这些纪录被筛选出的原因，并指出更正方案
更正错误	与管理人员或录入人员协商解决录入中出现的问题；若不属于录入错误，与专家讨论解决方法

#### 4.1.2 文献信息

文献信息的审核主要包括两部分。首先，文献的保存文件格式是否为 PDF，文件名是否为文献题目，文献内容是否完整，字迹是否清晰；其次，在《动物信息系统 2.2》的物种编辑表中可以看到完整的符合要求的文献名的记录。

## 4.2 多媒体信息

此处只针对图像信息的审核作说明。图像信息的备份必须含有“图像信息记录表”的电子版和文本材料。图像信息的审核从三方面进行。

### 4.2.1 图像的质量

这一层面的审核需要相关专家的参与。当拍摄人员按要求完成拍摄任务后，专家需察看照片的质量和效果。对不符合要求的照片，专家应指导拍摄人员重新拍照。当照片的质量和效果达到预期目的后，专家应在文本材料的“图像信息记录表”上签名。

### 4.2.2 图像的格式

这一层面的审核由数据审核员进行。主要审核照片的命名是否以标本号为标题，且后缀为“-1、-2、-3...”以及照片保存格式是否为“JPEG”，像素至少为 2560×1920；

### 4.2.3 图像信息与标本及物种信息的匹配

这一层面的审核可主要由数据审核员进行。其步骤和方法参见下表：

步骤	操作方法
提取每个图像的名称	将图像的名称全部提取出来（有小型程序可简化操作）
审核“标本图像采集记录表”的准确性	将图像的名称（前一步提取的结果）与“标本图像采集记录表”中记录的标本号进行比对，查找“标本图像采集记录表”中遗漏的或多余的图像的名称。 方法：可将两个记录表单都转变为 access 文件，利用 access 查询选项中的对比功能进行筛选；或在 excel 下用宏命令查找。
确定有图像的标本记录	通过上一步骤，将原“标本图像采集记录表”中多余记录或遗漏图像名称，总结成表并在日后返还给管理人员。
审核图像采集标本记录与物种信息的匹配	将录入数据中以核对的物种代码与经过上一步骤的“标本图像采集记录表”中的物种代码进行比对，查找“标本图像采集记录表”中遗漏的或多余的已核对物种。
确定有误的图像信息	通过上一步骤，将原“标本图像采集记录表”中多余的或遗漏的物种代码总结成表并在日后返还给管理人员，日后进行补充拍照或补充核对标本信息。

以上动物标本规范化整理、录入与审核的主要内容、要求及工作细则，将伴随着数据库工作的不断深入而做出进一步的调整和完善。