

科学标本资源汇交整编规范与实践

谢 园 杨 眉 何明跃 吴志远
(中国地质大学(北京), 北京 100083)

摘要: 科学标本资源是科技资源的重要组成部分, 科学标本资源分为八大类: 植物种质资源、动物种质资源、微生物菌种资源、人类遗传资源、生物标本资源、岩矿化石标本资源、实验材料资源及标准物质。本文在国家科技资源共享平台建设和科技基础性工作项目数据汇交过程管理实践经验的基础上, 研究科学标本的汇交要素、基础性资料元数据标准, 以及标本资源在名称标识、归类编码、原产地、来源、特征特性、图像、保存收藏信息、共享信息等8个方面的共性描述规范。进一步以“中国古人类遗址”为例, 通过数据表设计、原始数据分析与导入、元数据编写和数据文档整理对标本数据进行整编, 并探讨了标本数据库的应用前景, 对国家科技计划项目执行过程中产生科学标本资源的汇交、数据集成与共享具有重要意义。

关键词: 科学标本; 描述规范; 数据规范; 标本数据库; 数据共享; 整编

中图分类号: G312

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2017.05.003

Aggregation Standard and Experience of Scientific Specimen Resources

XIE Yuan, YANG Mei, HE Mingyue, WU Zhiyuan
(China University of Geosciences, Beijing 100083)

Abstract: The scientific specimen resource is important constituent parts of scientific and technical sources. The scientific specimen resources can be divided into eight categories: plant germplasm resources, animal germplasm resources, microbial strains resources, human genetic resources, biological specimen resources, rock and fossil specimen resources, experimental material resources and standard materials. They are important scientific and technological resources. Based on the experiences of the construction of national scientific specimen resources sharing platform and the aggregation of basic scientific projects data, this article discuss aggregation factors of scientific specimen resources, basic data metadata standards, and how to describe normatively the 8 kinds of scientific specimen resources including the name identifications, code classifications, sources areas, origins, characteristics, images, storage information, shared information. By taking ancient Chinese ruins as an example, this article study the application and development of specimen resources prospect and discuss such subject with the methods of design of data sheets, analysis and import original data, profiling metadata, collating data documents and reorganization of specimen resources, which will be of great importance in data collection, integration and sharing of scientific specimen resources in the implementation of the national scientific and technical programs.

Keywords: scientific specimen, description specification, data specification, specimen database, data sharing, reorganize

作者简介: 谢园 (1994—), 女, 中国地质大学(北京)硕士研究生, 研究方向: 材料工程; 杨眉 (1980—), 女, 中国地质大学(北京)助理研究员, 研究方向: 标本资源数字化与共享(通讯作者); 何明跃 (1963—), 男, 中国地质大学(北京)教授, 研究方向: 矿物学、岩石学、矿床学; 吴志远 (1979—), 男, 中国地质大学(北京)助理研究员, 研究方向: 数据库技术。

基金项目: 科技基础性工作专项重点项目“科技基础性工作数据资料集成与规范化整编”(2013FY110900); 国家科技基础条件平台项目“国家岩矿化石标本资源共享平台”。

收稿时间: 2017年7月14日。

1 引言

科技资源包括科技财力资源(经费)、科技人力资源(研究和技术人员)、科技物力资源(仪器设备和工具)、科技信息资源(各类数据库或信息资源)等4个方面^[1]。自然科技资源则是可供科学研究用的、具有一定学术价值的自然资源,主要包括植物、动物、微生物、人种遗传、岩石化石标本、生物标本、实验材料和标准物质等^[2]。标本资源作为典型的自然科技资源是指科研人员在诸如生态考察、地质调查、地震监测、环境监测、气象观测、科学实验等科技工作中,收集、研制或培育的标本或材料^[3]。它凝聚了科技进步与创新的成果,并且极具代表性,能够代表某一类的一种物质,是科学研究的重要物质基础。

在过去的20多年中,多个学术机构及地区制定了各类标本资源描述规范,比如矿物标本资源描述规范^[4]、人类遗传资源信息共性描述规范^[5]、林木种质资源共性描述规范^[6]等,并建立了相关的数据库,为政府决策、科技进步、人才培养和社会科学普及提供支撑服务。然而,各项目制定的标本资源描述标准侧重点互不相同、计量单位互不统一、字段类型互不一致、存储格式要求互不相近,给标本资源数据共享与交流带来了阻碍,这也对八大类科学标本资源共性描述规范的制定提出了更高的要求。

科学标本作为科技资源的重要部分,其特殊性表现在以下几个方面。

(1) 科学标本以实体的形式存在,其自然属性多种多样,有天然产出的,如岩矿化石标本、生物标本、种质资源,也有人工制造的,如标准物质;有可再生的,如种质资源,也有不可再生的,如岩矿化石标本;有宏观的,也有微观的。

(2) 科学标本的保存需要具有专业设备的库房,需要特殊的保存条件和专业人员的维护。因此,标本资源的实体无法汇交,需要汇交的是标本的信息。应面向全社会公布科学标本资源的目录、共性信息、共享信息,以避免重复建设。

(3) 科学标本资源的共享有别于科学数据共享,科学数据是可复制的,共享利用不损耗数据实体。科学标本作为一种物质,其数据可无限共享,实体需要有条件的共享,还要保存数量有限的标本,不损耗破坏标本实体。

我国投入大量科研经费支撑了大量的科技创新活动,产生了一大批重要的科技创新成果^[7],形成了包括科学标本在内的一大批重要科技资源。早期产生的科学标本保存在项目承担单位或项目负责人手里,信息不公开,数据较为分散,并且有的数据库缺乏长期经费支持难以为继,制约了科学标本资源的利用,非常不利于资源的管理、保护,也无法发挥其对科技的支撑作用。

随着信息时代的发展,标本资源保存机构、产出标本资源的科技计划项目开始研究制定描述标准、数据标准,逐渐建立标本数据库实现信息化、网络化管理。但由于标本资源数据在采集和加工过程中,各项目采用的方法、执行标准各不相同,尤其是科研项目产生的科学标本数据,难以整合利用。由国家财政支持创造的科技成果属于国有资产,应向全社会公开(涉密数据除外)。因此,国家高度重视科技计划资源汇交长效机制的建立。目前,国家重点研发计划对科技资源管理单位的权力、义务、责任等作出了明确规定^[7]。

为了规范科技基础性工作专项项目科学标本数据汇交,建立数据共享平台,同时与国家平台接轨,在研究国家科技基础条件平台制定的自然科技资源共性描述规范的基础上,本文档提出“元数据—共性描述规范—标本原始文档”三位一体的科学标本汇交规范,依据规范开展了科学标本资源的规范化整编与建库实践。

2 科学标本汇交要素分析

科学标本资源包括实物标本及与之相关的数据。科学标本资源的数据汇交要素为元数据、与元数据对应的标本共性描述信息表、标本图像文件、标本原始数据表。因此,在科学标本汇交工作中,需要对标本数据实体进行规范化整理,编撰元数据,生成共性信息表,整理标本图像及原

始数据表。而实物标本资源保存于各资源单位中，不作为汇交的要素，如图 1 所示。

科学标本资源元数据即内容、质量、采集方法和其他特性的信息描述。与元数据对应的共性信息表为结合科学性、实用性、可操作性以及建立标本资源数据库的需求，建立的八大类（植物种质资源、动物种质资源、微生物菌种资源、人类遗传资源、生物标本资源、岩矿化石标本资源、实验材料资源及标准物质）科学标本的共性描述信息表。共性信息包括标本资源的名称标识、归类编码、原产地、来源、特征特性、图像、保存收藏信息、共享信息 8 个部分。标本图像文件即经标准化命名的标本资源的数码照片文件，每一号标本资源至少采集一张数码照片。标本资源原始数据表即根据各类标本的特征信息所制定的标本原始数据表，原始数据表可作为附件数据包下载。

3 科技基础性资料元数据标准

元数据是关于数据的数据^[8]，是对数据的内容、质量、采集方法和其他特性的信息描述。作为科学标本汇交的主要要素，科学标本元数据规范化的编写和描述对数据整编共享起着至关重要的作用。

科学标本资源的元数据是对一类标本资源共

性特征（基本内容、保存管理、共享服务等）的简要描述。用于数据的分类导航、查询检索与共享访问。一条元数据对应一个标本共性信息描述数据集。按照《科技基础性工作专项项目数据汇交元数据标准》，元数据的数据项内容包括：资源标识、资源学科分类、中文名称、英文名称、资源描述摘要、关键词、资源类型（此处均为“标本资源”）、资源格式、资源时间、资源地点、最新修订时间、共享方式、资源质量描述、在线链接地址、缩略图、来源项目、资源负责方、资源管理费、元数据管理信息。

4 科学标本的共性描述规范

根据国家科技基础条件平台制定的自然科技资源共性描述规范^[1]，将科学标本资源分为八大类：植物种质资源、动物种质资源、微生物菌种资源、人类遗传资源、生物标本资源、岩矿化石标本资源、实验材料资源及标准物质。结合科学性、实用性、可操作性以及建立标本资源数据库的需求，科学标本的共性描述信息应包含标本资源的名称标识、归类编码、原产地、来源、特征特性、图像、保存收藏信息、共享信息 8 个部分（表 1）。下面对每个部分数据的内容、描述方法、计量单位等进行详细说明。

（1）名称标识

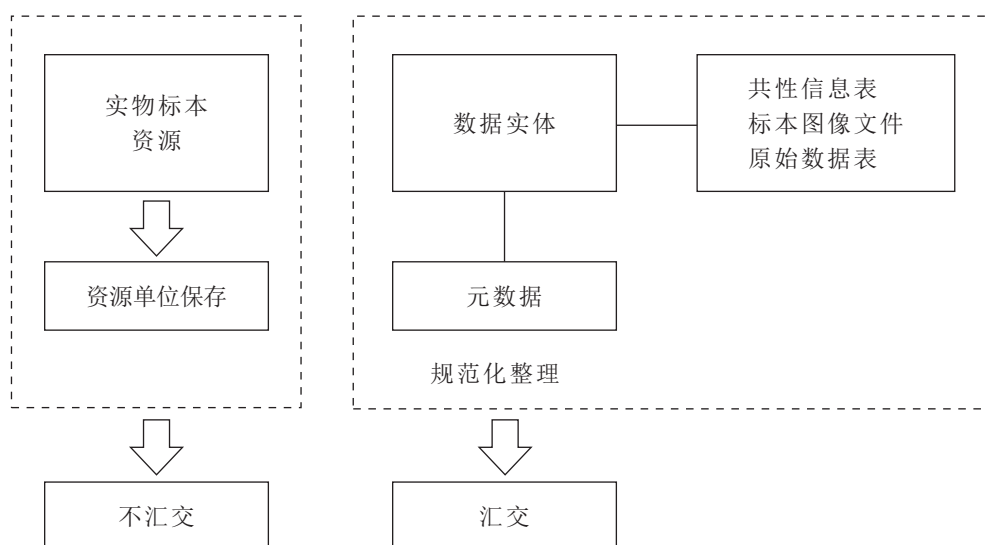


图 1 科学标本资源的实体汇交要素图

表1 科学标本资源描述规范表

	植物种质资源	动物种质资源	微生物种质资源	人类遗传资源	生物标本资源	岩化石资源	实验材料资源	标准物质资源
名称标识	中文名, 外文名, 科名, 属名, 种名	资源名称, 资源别名, 资源外文名, 科名, 属名, 种名或亚种名	中文名称, 属名, 种名加词, 种下分类	资源编号, 样本类型, 性别, 健康状况, 资源分类, 民族, 籍贯, 出生年月	中文名称, 属名, 种名, 本名/种加词, 种下名称, 纲名称, 目名称, 科名称	中文名称, 外文名称	中文名称, 外文名称, 研制培育机构, 研制培育年份, 来源	中文名称, 英文名称
归类编码	资源归类编码							
来源			原产国, 来源历史, 原始编号				单位, 时间, 研制培育机构, 研制培育年份	研制单位名称, 标准物质分类, 研制负责人
产地	国家, 省, 县, 乡, 村, 经度, 纬度, 海拔	国家, 省, 县, 乡, 村, 经度, 纬度, 海拔			国家, 省, 采集地点, 海拔, 经度, 纬度	采集地点, 省, 国家, 经度, 纬度, 海拔, 资源形成时代, 地质产状或层位		
特征特性	资源类型, 主要特征, 主要用途, 气候带, 生长习性, 生育周期, 观测地点, 系谱, 选育单位, 育成年份, 土壤类型, 生态类型, 年均降雨量	资源类型, 功能性, 主要用途, 气候带, 年均温度, 极端平均高温, 极端平均低温, 年均湿度, 年均降水量, 生活生态习性, 繁殖周期, 性成熟期, 生命周期, 形态特征, 具体用途, 生态系统类型	模式菌株, 主要用途, 特征特性, 生物危害程度, 寄主名称, 致病对象, 致病名称, 传播途径, 分离基物, 采集地, 培养基编号, 培养基温度, 基因元件	用途, 血型资料, 流行病学资料, 诊断资料, 治疗资料, 随访资料, 样本采集时间, 样本保存期限, 样本保存条件, 家系标本, 生物安全级别, 其他补充说明, 疾病别名	特征描述, 生境, 寄主	繁殖代数, 主要特性, 主要用途, 微生物质控, 遗传特征, 组织器官来源, 理化指标, 特征特性	主要用途, 特征描述	特征形态, 标准值, 基体, 不确定度, 规格, 保存条件, 使用注意事项
图像	图像记录相对地址							
保存收藏信息	保存单位, 单位编号, 库编号, 圃编号, 引种号, 采集号, 保存资源类型, 保存方式, 实物状态, 记录编号, 记录地址	保存单位, 单位编号, 保存资源类型, 保存方式, 实物状态, 记录编号, 记录地址	保存单位, 资源保存类型, 保存方法, 实物状态, 菌株保藏编号, 其他保藏中心编号, 保藏时间, 记录编号, 记录地址	保存单位, 保存地, 成果类型, 实物状态, 记录编号, 记录地址	保存单位, 采集人, 采集时间, 采集号, 标本号, 鉴定人, 鉴定时间, 标本属性, 保藏方式, 实物状态, 记录编号, 记录地址	保存单位, 资源提供者, 资源提供时间, 采集号, 标本编号, 库存位置号, 保存资源类型, 实物状态, 记录编号, 记录地址	保存单位, 单位编号, 库编号, 引种号, 保存资源类型, 保存方式, 保存条件, 实物状态, 记录编号, 记录地址	保存单位, 库编号, 标准物质中文证书封面, 标准物质英文证书封面, 记录编号, 记录地址
共享信息	共享方式, 获取条件, 联系方式							
		提供形式					运输条件, 获取途径	

包含科学标本的中文名称、外文名称。植物种质资源、动物种质资源和生物标本资源还包括科名、属名、种名，其外文名称为拉丁文名称，其他类资源的外文名称为英文名称。人类遗传资源的标记信息包括资源编号、样本类型、性别、健康状况、资源分类、民族、籍贯和出生年月。

（2）归类编码

分类信息须填写资源归类编码^[1]，该编码按照学科进行分类，分三级，为字符型，共 11 位。详见国家自然科技资源平台资源分级归类与编码。资源分类是根据生命科学研究并结合现实保存样本类型进行分类编码^[1]。

（3）来源

实验材料资源的来源信息包括该资源最近一次引进的单位和时间^[1]以及研制培育机构、研制培育年份。研制培育机构是每种实验材料的原研制或培育机构（人）^[1]。对于不同物种有不同的研制培育年份。对实验动物而言，为品种、品系培育成功年份；对于实验细胞而言，为细胞建株成功年份；对于微生物培养基而言，为研制成功年份。研制培育年份的数据类型是日期型。微生物菌种资源的来源信息包括原产国、来源历史、原始编号。标准物质资源的来源信息包括研制单位名称、标准物质分类以及研制负责人。

（4）原产地

指科学标本的原产地，须记录原产国家、省、县、乡、村名称，还包括经度、纬度、海拔。省是国内种质资源原产省份名称或国外引进种质原产国家一级行政区的名称，国家是种质资源原产国家名称、地区名称或国际组织名称，均为字符型。海拔经度纬度为数值型，海拔以米为单位，保留小数点后两位，经度和纬度以十进制的形式填写，保留小数点后 4 位。

（5）特征特性

各领域标本资源的特征特性不同，其内容也有所差异。

植物种质资源的特征特性信息包括：资源类型、主要特征、主要用途、气候带、生长习性、生育周期、观测地点、系谱、选育单位、育成年

份、土壤类型、生态系统类型、年均温度和年均降水量。植物种质资源的资源类型分为野生资源（群体）、野生资源（家系）、野生资源（个体）、地方品种、选育品种、品系、遗传材料和其他等。主要特性分为高产、优质、抗病、抗虫、抗逆、高效和其他等。主要用途分为食用、纤维、嗜好、药用、生态、观赏、材用和其他等。

动物种质资源的特征特性信息包括：资源类型、功能特性、主要用途、气候带、年均温度、极端平均高温、极端平均低温、年均湿度、年均降水量、生活生态习性、繁殖周期、性成熟期、生命周期、形态特征、具体用途和生态系统类型。动物种质资源的资源类型分为野生种质资源、地方种质资源、培育种质资源、引进种质资源和其他等。主要功能特性：如高繁殖力、高生产力、优质、抗病虫、抗逆、耐粗饲、耐高温高湿、耐高寒、耐极端干旱、耐药、致病、其他等。主要用途包括肉、蛋、奶、蜜、纤维、役用、药用、保健、生物防治、研究教学、观赏、竞技、娱乐、毛皮、其他等。

微生物菌种资源的特征特性信息包括：模式菌株、主要用途、特征特性、生物危害程度、寄主名称、致病对象、致病名称、传播途径、分离基物、采集地、培养基编号、培养温度和基因元器件。主要用途包括分类、研究、教学、分析检测、生产和其他等。生物危害程度包括一类、二类、三类、四类和不清楚。致病对象包括人、动物、人畜共患、植物、微生物、无和不清楚。传播途径主要包括接触传播、空气传播、食物传播、水传播、血液传播和其他等。其中，微生物菌种资源的寄主名称为菌种资源寄生宿主的中文或拉丁名称^[1]。

人类遗传资源的特征特性信息包括：用途、血型资料、流行病学资料、诊断资料、治疗资料、随访资料、样品采集时间、样品保存期限、样品保存条件、家系标本、生物安全级别、其他补充说明和疾病别名。人类遗传资源的用途按照资源的采集目的划分为科学研究、临床治疗和其他等。诊断资料是根据研究是否可以提供诊断相

关资料进行编码^[1]。例如诊断时间,根据病史资料、检验项目及所对应的检测值对本遗传资源对应的生理病理状况进行编码。治疗资料是根据研究是否可以提供治疗相关资料进行编码^[1],如治疗方法、治疗药物名称、剂量、疗程等。生物安全级别是依据样本的潜在传染性进行划分^[1]。其他补充说明需注明相关文字信息,即提供资源特性的部分信息,如样本处理方法、ATCC号等^[1]。疾病别名需给出该疾病的其他名称、叫法^[1]。

生物标本资源的特征特性信息包括:描述、生境和寄主。描述指生物标本的鉴定特征描述,植物标本需描述根、茎、叶、花、果、种子等的特征,动物标本需要根据纲目不同细分,如哺乳纲需描述年龄、颅长、体重、体长、前臂长、后足长、耳长、肩高、臀高、性别等。生境指生物标本采集的生态环境。寄主可分为菌物、无脊椎动物、昆虫和植物的寄主以及植物的附生和腐生。

岩矿化石标本资源的特征特性信息包括:主要用途和特征描述。岩矿化石标本资源的主要用途包括工业原料、药用、建材、科学研究、观赏及其他。特征描述需尽可能全面地描述标本的特征:矿物标本需描述其物理性质、结晶学特征、化学组分等包括颜色、晶体习性、条痕、透明度、光泽、摩氏硬度、密度、解理等特征;岩石和矿石标本需描述其颜色、结构、构造,矿物组成等特征^[9];化石标本需描述其外形、生物特征、分类特征等。

实验材料资源的特征特性信息包括:代数、主要特性、主要用途、微生物质控、遗传特征、组织器官来源、理化指标和特征特性。实验材料资源的代数是指实验动物或实验细胞的繁殖代数^[1]。主要特征针对于实验动物来说,包括常用动物、模型动物、基因突变动物、遗传修饰动物和其他等;针对实验细胞来说,包括贴壁生长和悬浮生长;针对微生物培养基来说,包括液体、流体、半固体和固体。主要用途针对实验动物来说,包括研究、生产、检定和教学;针对实验细胞来说,包括研究和生产;针对微生物培养基来

说包括研究、生产、检测和教学。微生物质控针对实验动物来说分为普通级、清洁级、SPF级和无菌级;针对实验细胞来说为无外源微生物污染(具体描述质控微生物);针对微生物培养基来说是针对每一种培养基确定质控微生物的菌株及指标。遗传特征针对实验动物来说分为近交系、远交群、基因突变系、遗传修饰系和其他等;针对实验细胞来说分为原代培养细胞、有限细胞系和连续细胞系。实验细胞的组织器官来源分为正常组织、肿瘤组织和其他。微生物培养基的理化指标包括性状、pH值、凝胶强度、澄清度、色泽、干燥失重和其他。实验动物的特征特性包括体型、体重(成年动物的正常体重)、毛色和其他等,实验细胞的特征特性包括遗传标志、免疫标志和生化特性。

标准物质资源的特征特性信息包括:特征形态、标准值、基体、不确定度、规格、保存条件和使用注意事项。标准物质的特征形态是指标准物质的物理形态,包括气态、液态和固态。其规格是指标准物质最小包装单元重量。

(6) 图像

每一号标本资源至少采集一张数码照片,图像精度不低于300万像素,保存图片格式可以为.jpg、.png等。在数据表中应记录种质资源的图像记录相对地址名称,即文件夹路径+图像名,图像名为资源编号,如有多张图像应命名为“资源编号-1”“资源编号-2”,依此类推,数据汇交时,需要提供对应的图像文件。

(7) 保存收藏信息

指资源的保存单位、单位编号、保存资源类型、保存方式、实物状态和记录地址。此外,植物种质资源还包括库编号、引种号、圃编号和采集号。微生物菌种资源还包括菌株保藏编号、其他保藏中心编号以及保藏时间。生物标本资源还包括采集人、采集时间、采集号、鉴定人、鉴定时间和标本属性。岩矿化石标本资源还包括资源的提供者、提供时间以及采集号、标本编号和库存位置号。实验材料资源还包括库编号和引种号。标准物质资源还包括中、英文证书封面,即

中、英文证书封面图像记录相对地址名称^[1]，数据汇交时，需要提供对应的图像。记录地址是数据生产方已建设的可对外访问的科学标本资源详细信息的网址或数据库记录链接。

(8) 共享信息

指资源的共享方式、获取条件和联系方式。其中，共享方式分为：完全开放共享、协议共享、暂不共享。联系方式包括联系人、单位、地址、邮编、电话和电子邮箱。实验材料资源还包括运输条件和获取途径。

5 科学标本数据资料规范化整编实践

科学标本数据资料规范化整编是按照统一的描述标准和技术要求，对标本资源的原始数据资料进行审核、转换处理、集成，整理形成系统化、规范化的数据。以科技基础性工作项目“中国古人类遗址”为例，对科学标本数据资料规范化整编过程的4个步骤（数据表设计、原始数据分析与导入、元数据编写、数据文档整理）进行详细阐述。

5.1 数据表设计

根据以上科学标本资源共性描述规范，利用Excel和Access软件，建立离线数据录入模版，定义Excel表格的数据格式和内嵌数据选项，并按照表1列出的八大类资源的数据格式及数据选项，尽量规范中国古人类遗址数据库提供方录入数据时的内容和格式。数据模板定义了需要填写的数据项字段，对于有数据限制的字符型字段，运用下拉式菜单，直接从可选数据项中选择。如“共享方式”，下拉式菜单分别为完全开放共享、协议共享、暂不共享。

5.2 原始数据分析与导入

中国古人类遗址数据主要包括古人类化石数据、古猿化石数据和文化遗物数据共3类标本数据。古人类化石数据表共有40个数据项，古猿化石数据表共有34个数据项，文化遗物数据表共有53个数据项，大部分数据项能与共性描述规范匹配且可直接取用；缺乏国家、省、库存位置号、保存资源类型、图像信息，需补充采集；

部分字段名称及计量单位表述不同，如分类、经纬度，需按照规范要求进行转换。按照规范进行补充完善后，导入到共性描述规范表。其他与共性描述规范不匹配的数据项不作处理，如古人类化石数据表中的遗址名称、保存部位、文献资料、颅长、颅宽、颅高、面宽、上面高、眶宽、眶高、鼻宽、鼻高、颅周长、颅正中矢状弧、颅横弧等。

5.3 元数据编写

按照科技基础性资料元数据标准，对中国旧石器时代遗址标本数据库元数据进行描述，如表2所示。

5.4 数据文档整理

为了使数据文档易核对、易管理、易使用，需要对数据文档进行规范化整理。将所有数据文件整理到一个文件夹下，文件夹以项目编号2007FY110200命名，此文件夹为一级目录。在这个文件夹中，要建立以元数据标识命名的子文件夹，每条元数据一个文件夹，为二级目录。在以元数据标识命名的子文件夹中，应有每一条元数据所对应的标本数据文件夹，为三级目录。标本数据文件夹中应包含共性描述规范表、原始数据表及标本图片文件夹，为四级目录，如古人类化石数据资源描述规范表、古人类化石数据及古人类化石图像，需对每一条元数据的共性描述规范表、原始数据及缩略图进行整理（图2）。

6 结论与讨论

随着我国对科技创新活动投入大量科研经费，科学标本资源及数据的产出和积累效果明显，然而，资源的利用效率还有待提高。建立科学标本资源共性描述规范是实现跨领域跨项目标本数据整合及资源高效利用的必要条件。本文研究了科学标本资源汇交要素、基础性资料元数据标准、共性描述规范的体系，详细阐述了植物种质资源、动物种质资源、微生物菌种资源、人类遗传资源、生物标本资源、岩矿化石标本资源、实验材料资源及标准物质八大类科学标本资源共性描述信息的内容，以及描述方法、计量单位

表 2 元数据示例

资源标识	2007FY110200-05-2014101105		资源学科分类	生物学
中文名称	中国旧石器时代遗址标本数据库		英文名称	
资源描述摘要	以科学规范为原则，在项目执行过程中将发掘与调查过程中所得的遗址标本进行测量与数据录入，并建立数据库，其内容主要包括古人类化石、古猿化石与文化遗物等测量数据，目前已经录入古人类化石数据185例，其中完整头骨24例、头骨碎片68例、下颌骨25例、牙齿45例、前肢骨6例、后肢股17例；古猿化石数据牙齿化石8例；文化遗物数据共16693条，其中石核1173条，石片10640条，石器827条，残片212条，断块3051条，其他石制品790条。涉及的遗址有水洞沟遗址、大窑遗址二道沟地点、大窑遗址四道沟地点、金斯太遗址等。			
关键词	中国，旧石器时代，遗址，古人类化石，古猿化石，文化遗物			
资源类型	标本资源		资源格式	Excel
资源时间	2009—2012年		资源地点	中国
最新修订时间	2012-12-20 0:00:00		共享方式	完全开放共享
资源质量描述	在人骨化石与古猿化石栏，根据人类化石标本的描述和测量等基本信息，制订了人类标本描述基本信息15条，具体如下：资源名称、资源外文名称、产地、省、资源形成的时代、地质产状或层位、形态特征描述：化石标本的主要鉴定特征、测量、经度、纬度、图像、保存单位、标本编号、资源状态、文献资料。在文化遗物栏，根据遗物的类型，分为石核、石片、石器、微屑/碎屑、断块、骨角木遗物、其他石制品等，测量内容包括编号、类型、原型、岩性、预制台面与否、剥片技术、文化时代、出土时间、台面数量、台面关系、剥片面数量、剥片面关系、遗址名称、遗址所在省、遗址所在县、标本来源、标本入库标号、遗址地质时代、遗址考古时代、地层堆积、自然地层、文化层、水平层、探方号、标本保存单位等。数据均来源于仪器测量，主要为游标卡尺与高精度计量称，精度为0.01厘米，测量人员对每一个数据均测量3次，仔细认真校对，取其平均值，然后逐一录入。			
在线链接地址	http://124.16.245.193/Main/Login.aspx			
缩略图	 头骨后肢骨石器			
来源项目	项目编号	2007FY110200	项目名称	东北森林植物种质资源专项调查
	项目负责人	刘*	第一承担单位	中国科学院古脊椎动物与古人类研究所
资源负责方	负责人姓名	高*	电子邮箱	g*****@ivpp.ac.cn
	联系电话	010-883*****	传真	010-683*****
	所在单位名称	中国科学院古脊椎动物与古人类研究所		
	地址	北京市西外大街142号	邮编	100044
资源管理方	单位名称	中国科学院古脊椎动物与古人类研究所		
	联系人姓名	刘*	电子邮件	l*****@ivpp.ac.cn
	联系电话	010-883*****	传真	010-683*****
	地址	北京市西外大街142号	邮编	100044
元数据管理信息	编写人姓名	高*	联系电话	010-883*****
	电子邮箱	g*****@ivpp.ac.cn	元数据最新更新时间	2015-2-2 15:14:48

等,结合“中国古人类遗址”典型案例介绍了科学标本数据整编的方法,初步探讨了挖掘科学标本资源数据的价值及提高数据可用性的手段。

本文的研究成果不仅适用于科技基础性工作项目标本数据汇交,而且对科学标本资源整合起

到了指导和示范作用。基于元数据、共性描述规范和原始数据文档规范化整编的标本数据库,在标本数据资源价值的挖掘和相关性分析方面具有广阔的应用前景。

(1) 数据的关联性。关联数据是提升数据

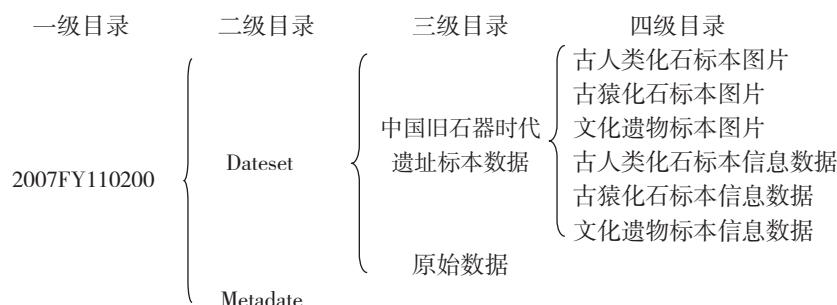


图2 科学标本数据文档整理示意图

可用性的重要前提，同时避免了数据资源的碎片化。平台将数据与相关的标本、文献、专利、科研成果等资料相关联，使用户在利用数据库系统进行查询时，可获取与之相关的科学资料。例如：在检索生物资源时，根据关键词，与生物多样性、生物志书、各种生物量数据以及生物资源种类与分布等数据联系起来，考虑用户检索阅读习惯，首先推荐与其相关联的信息。

（2）数据的挖掘。在标本大数据库的基础上进行数据挖掘，需要从海量的数据中，将分散的数据集合起来，并且在数据集中找到内在规律和联系。例如：利用生物标本数据结合地理信息数据、环境数据等其他生物多样性数据，建立数学模型，分析已知物种的地理分布、时空动态、种群数量、物种进化史、物种与环境之间及物种与物种之间的相关作用^[10]，将成为大数据时代生物多样性研究的重要手段。

（3）数据的可视化。数据可视化在数据挖掘中发挥着关键的作用，即用图形或报表的方式将数据直观简洁地显示出来，使数据中的复杂信息一目了然，有助于用户寻找数据分布规律^[11-12]。例如：在数据中叠加数字化的地质图、地形图、地理图等使数据利用的意义更加显著，可了解各资源的地理分布状况，包括经纬度及海拔高度分布和标准物质研制单位分布，分析某个生物类群在某个地层中的分布特点，预测岩矿化石可能出露的位置以及地质产状或地层的空间分布等。

参考文献

- [1] 许哲平, 覃海宁, 马克平, 等. 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, 2012, 44(1): 27-33. DOI: 10.3772/j.issn.1674-1544.2012.01.005.
- [2] 黄国勤, 黄雪梅, 宋高堂, 等. 江西省自然资源的现状、问题及战略研究[J]. 2005.
- [3] 曹一化, 刘旭, 第三作者??, 等. 自然资源资源共性描述规范[M]. 北京: 中国科学技术出版社, 2006.
- [4] 周彦. 矿物标本资源描述标准研究制定[D]. 北京: 中国地质大学(北京), 2005.
- [5] 胡序怀. 我国人类资源信息描述规范的研究制定与初步应用[D]. 北京: 中国协和医科大学, 2006.
- [6] LY/T 2192—2013, 林业种质资源共性描述规范[S]. 2013.
- [7] 石蕾, 袁伟. 建立科技计划资源汇交长效机制的思考[J]. 中国科技资源导刊, 2012, 44(4): 2-5. DOI: 10.3772/j.issn.1674-1544.2012.04.002.
- [8] 黄如花, 邱春艳. 国内外科学数据元数据研究进展[J]. 图书与情报, 2014(6): 102-108.
- [9] 何明跃, 吴淦国. 矿物、岩石、矿石标本资源及矿床描述标准[M]. 北京: 地质出版社, 2013.
- [10] 张健. 大数据时代的生物多样性科学与宏生态学[J]. 生物多样性, 2017, 25(4): 355-363. DOI: 10.17520/biods.2017037.
- [11] 樊隽轩, 张华, 侯旭东, 等. 古生物学和地层学研究的定量趋势: GBDB数字化科研平台的建设及其意义[J]. 古生物学报, 2011, 50(2): 141-153.
- [12] 樊隽轩, 侯旭东, 陈中阳, 等. 基于GBDB数据库的地层学研究与应用[J]. 地层学杂志, 2013, 37(4): 400-409.