



文献 CSTR:

32001.14.11-6035.csd.2024.0184.zh



文献 DOI:

10.11922/11-6035.csd.2024.0184.zh

文献分类: 生物学

收稿日期: 2024-11-07

录用日期: 2025-09-08

发表日期: 2025-12-19

植物多样性科学数据 APIs 的建设

邱金水^{1,2,3*}, 张建文¹, 金涛¹, 王朋¹, 杜宁¹, 黄蓉¹, 庄会富^{1,2,3*}

1. 中国科学院昆明植物研究所科学数据中心, 昆明 650201

2. 植物化学与天然药物全国重点实验室, 昆明 650201

3. 国家重要野生植物种质资源库, 昆明 650201

摘要: 随着大数据技术和人工智能技术的高速发展, 科学数据已成为推动科研范式变革的核心科技资源。生物多样性科学数据与生物资源本身一样, 已成为国家的重要战略资源, 是国际科技与产业的竞争热点和战略制高点。针对当前国内植物多样性科学数据在开放共享方面与国际主流数据库之间存在差距的问题, 本研究提出并实施建设植物多样性科学数据应用程序编程接口 (Application Programming Interfaces, APIs)。基于中国科学院昆明植物研究所科学数据中心已收集并整理的植物名称、物种信息、图片元数据、种质元数据、DNA 条形码、标本元数据、图书和文献元数据等植物科学数据, 通过对各类植物科学数据进行编目, 设计了 APIs 的业务流程和标准规范, 采用了模块化的开发方法研发建成了植物多样性科学数据 APIs, 实现了对各类植物科学数据的任意属性字段组合查询、数据高效检索、数据安全以及接口权限控制等功能, 从而助力推进国内植物多样性科学数据接口标准的建立与完善, 提升我国植物多样性科学数据的共享水平, 促进植物多样性的研究, 为生物多样性保护提供数据和技术支撑, 同时也可以为其他学科领域的科学数据 APIs 的建设提供参考。

关键词: 植物多样性; 科学数据; 数据接口; 数据开放共享; 数据中心建设; 数据要素

引言

数据是新的生产要素, 是基础性资源和战略性资源, 也是重要生产力, 在确保数据安全的基础上让数据“供得出、流得动、用得好”才能更好地发挥数据作为生产要素的作用。科学数据是科研工作者在从事科技活动或研究过程中产生的一系列数据, 包括通过野外调查、野外试验、野外观测、实验室实验和检测检验等方式取得的用于科学研究的原始数据, 以及通过加工、集成、分析、计算和模拟仿真等方式产生的衍生数据^[1-2]。随着大数据时代的到来, 尤其是生成式人工智能技术的高速发展, 科技创新越来越依赖对海量数据的集成、分析和挖掘利用, 科学数据已成为国家科技创新和经济社会发展的基础性战略资源^[3]。2019 年, 党的十九届四中全会首次将数据与土地、劳动力、资本、技术并列作为重要的生产要素^[4-5]。在生物多样性领域, 生物多样性科学数据与生物资源本身一样, 已成为国家的重要战略资源, 是国际科技与产业的竞争热点和战略制高点^[6]。基于生物

* 论文通信作者

邱金水: qiujinshui@mail.kib.ac.cn

庄会富: zhuanghuifu@mail.kib.ac.cn

多样性数据能为科学研究和产业发展提供数据和技术支撑,如利用物种图像数据能助力新物种的发表^[7],利用标本数据可以分析标本采集现状、预测采集趋势以及推断人类活动影响^[8-9],利用 DNA 条形码数据可以实现物种的鉴定以及协助海关检疫,对多模态、多源异构和开放共享的物种信息数据进行集成整合和挖掘分析,能够解决更为复杂的生物领域知识库建设^[10]、物种名录编目^[11]以及物种的起源、演化与适应机制^[12]等科学问题。

为了加强对科学数据的管理,规范科学数据的汇交,保障科学数据安全和提高科学数据的开放共享水平,2018 年国务院印发《科学数据管理办法》,提出科学数据分级管理、安全可控、充分利用的原则和“开放为常态、不开放为例外”的共享理念,标志着我国科学数据管理和共享进入新的历史阶段^[13]。2021 年,国家标准化管理委员会正式发布《科技计划项目形成的科学数据汇交通用代码集》《科技计划项目形成的科学数据汇交通用数据元》《科技计划项目形成的科学数据汇交技术与管理规范》3 项国家标准,这些标准规范了科学数据的汇交管理,推动了科学数据的开放共享^[14]。2023 年,国家数据局正式揭牌成立,主要负责协调推进数据基础制度建设,统筹数据资源整合共享和开发利用等。在国家科学数据管理办法、标准的发布以及相关管理部门的成立等推动下,近几年,国内的科学数据领域取得了长足的发展,一批国家级、省部级、学科领域以及机构数据中心陆续建立,覆盖植物科学的数据汇交、数据共享、数据处理和挖掘利用等各个环节的科学数据服务平台陆续上线发布^[15-18],各个数据中心的数据存储容量和开放共享水平迅速得到提升。

但由于国内科学数据的发展和数据中心的建设起步较晚,在科学数据管理和开放共享等方面仍然与发达国家存在差距和不足^[19]。在植物多样性科学数据领域,国际主流的科学数据共享服务平台不仅提供科学数据的在线检索服务,同时还提供科学数据的应用程序编程接口(Application Programming Interfaces, APIs)服务^[20-22],但国内的植物多样性科学数据共享服务平台主要提供科学数据的在线检索服务,除少数提供 APIs 服务外^[23],大多不提供 APIs 服务^[24-26],这在一定程度上限制了国内科学数据的开放共享,也阻碍了国内科学数据共享服务平台跨出国门走向世界的进程。因此,本研究基于中国科学院昆明植物研究所科学数据中心的实际情况,探索构建植物多样性科学数据 APIs,从而助力推进国内植物多样性科学数据接口标准的建立与完善,提升我国植物多样性科学数据的共享水平,促进植物多样性的研究,为生物多样性保护提供数据和技术支撑,同时也可以为其他学科领域的科学数据 APIs 的建设提供参考。

1 材料与方法

1.1 科学数据

1.1.1 数据资源

中国科学院昆明植物研究所科学数据中心是中国科学院首批成立的 13 个所级科学数据中心之一,主要负责承担国家重大科技项目的科学数据汇交和植物多样性科学数据的收集、整合、管理和共享等任务。目前昆明植物研究所科学数据中心已收集并整理包括植物的物种名称、物种信息、图片元数据、种质元数据、DNA 条形码、标本元数据、图书和文献元数据等各类植物科学数据,基于已整理的各类植物科学数据,数据中心已建成并发布了国家重要野生植物种质资源库共享服务平台、植物名称服务平台、植物物种分布预测服务平台、植物图像库、植物物种信息数据库和植物 DNA 条形码数据库等一批植物多样性科学数据共享服务平台。但目前这些科学数据共享服务平台主要以网

页的方式为用户提供在线检索和浏览服务，几乎不提供应用程序编程接口（APIs）服务，因此无法满足用户对植物多样性科学数据的共享需求。为进一步丰富昆明植物研究所科学数据中心的科学数据共享服务模式，本研究基于科学数据中心已整理好的部分核心植物多样性科学数据资源，如植物名称、物种信息、图片元数据、种质元数据、文献元数据、标本元数据、物种分布预测信息和 DNA 条形码等，探索构建植物多样性科学数据 APIs。

1.1.2 数据编目

植物多样性科学数据按不同类别分成植物的名称、物种信息、图片元数据和种质元数据等，同类科学数据也存在多源异构的情况，因此需要为各类科学数据制订各自统一的数据编目，这不仅为了实现对多源异构的数据进行集成整合，同时也是为制订各类数据接口的标准规范提供参考。数据接口的主要功能是实现不同的系统之间进行数据交互，从而实现数据共享、交换、迁移和同步等功能，因此为了确保数据接口的正常、稳定和可靠，数据接口的标准规范应该是通用和稳定的，即数据编目也应该尽量是通用和稳定的，为此，我们参考了 Darwin Core 制定了各类植物科学数据的数据编目。Darwin Core 由生物多样性信息标准化组织(TDWG)发布和维护，是一个被国际生物多样性研究领域广泛接受和使用的生物多样性信息学数据标准，它定义了一系列可能被称为属性、元素、字段、列、属性或概念的术语，用于描述生物的观察记录、标本、样品以及一些其他相关信息^[27]。

数据编目的过程同时也是理清科学数据中心的数据家底的过程，昆明植物研究所科学数据中心多年来致力于植物多样性领域的科学数据的收集、整理和开放共享，因此本文科学数据的一级编目确定为植物多样性科学数据；科学数据的二级编目根据科学数据中心目前已经收集和整理的植物多样性科学数据进行分类确定，目前二级编目分为植物名称、物种信息、图片元数据、种质元数据、文献元数据、标本元数据、物种分布预测和 DNA 条形码共 8 类；科学数据的三级编目可以理解为数据的属性或字段，即用于描述某个对象的属性集合，在数据库设计过程中则体现为数据库表的字段，三级编目的制定规则主要包括三个：一是目前科学数据中心在某个类别的科学数据收集，已经收集了哪些属性；二是还有哪些未被收集的属性是应该被收集或重点关注的；三是以上属性的描述是否符合 Darwin Core 的规则。基于以上过程和规则完成了植物多样性科学数据的编目，植物名称科学数据的编目如表 1 所示，其他植物多样性科学数据的编目见附录 1（详见 <https://doi.org/10.57760/sciencedb.33786>）。

表 1 植物名称科学数据的数据编目

Table 1 Cataloguing of scientific data on plant names

一级编目 Primary cataloging	二级编目 Secondary cataloging	三级编目 Tertiary cataloging
Scientific data for plant diversity (植 物多样性科 学数据)	Taxon(植物 名称)	id(ID)、datasetID(数据集 ID)、taxonID(分类 ID)、scientificName(学名)、scientificNameAuthorship(作者信息)、taxonomicStatus(分类状态)、taxonRank(分类等级)、higherClassification(分类列表)、parentNameUsage(父级名称)、parentNameUsageID(父级名称 ID)、acceptedNameUsage(接受名)、acceptedNameUsageID(接受名 ID)、originalNameUsage(原始名称)、originalNameUsageID(原始名称 ID)、family(科)、genus(属)、specificEpithet(种加词)、infraspecificEpithet(种下加词)、taxonRemarks(分类注释)、nameAccordingTo(来源参考)、namePublishedIn(发表地方)、namePublishedInYear(发表年份)、references(引用信息)、

一级编目 Primary cataloging	二级编目 Secondary cataloging	三级编目 Tertiary cataloging
		fullscientificName(拉丁全名)、canonicalName(拉丁简名)、vernacularName(俗名)

1.2 技术方法

1.2.1 业务流程设计

中国科学院昆明植物研究所科学数据中心目前已建成了一批植物多样性科学数据库，但这些数据库是分散独立的，因此建设统一的植物多样性科学数据 APIs 需要将存储在各个数据库的科学数据汇聚到同一个数据中台数据库，数据中台数据库通过定期从各个分散的科学数据库抽取科学数据的方式，实现科学数据的同步和更新。数据中台数据库是植物多样性科学数据的资源池，用于存储各类植物科学数据，在 APIs 业务处理流程中处于核心的位置，它既要能够实现从不同数据库中快速汇聚科学数据，也要提供高效的检索功能。由于科学数据中心前期建设的科学数据库使用了多种不同的数据库管理系统，包括 SQL Server 和 MySQL 等，因此数据中台数据库需要同时兼容不同的数据库管理系统且确保更高的检索效率，本文使用了 Apache Solr 搜索引擎作为数据中台数据库，它能够通过配置不同的数据库连接方式实现从不同的数据库管理系统中快速拉取数据，并且提供了全量更新和增量更新的功能，从而实现快速的科学数据汇聚、同步和更新。用户通过调用科学数据 APIs 请求应用服务器，应用服务器根据 APIs 调用参数进行权限控制和业务处理，并查询数据中台数据库，应用服务器从数据中台数据库获取到结果后对数据进行加工处理，并将结果以 Json 格式返回给用户，从而完成 APIs 调用与处理流程，业务流程设计如图 1 所示。

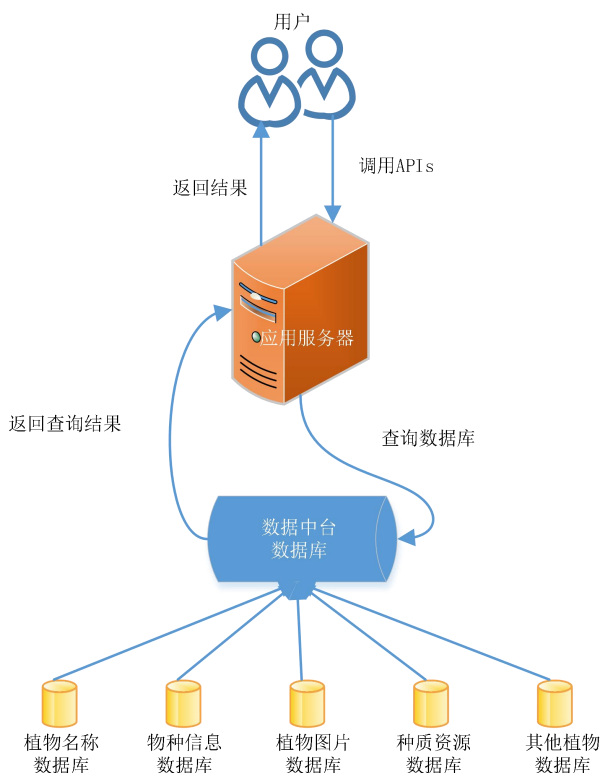


图 1 业务流程设计

Figure 1 Business process design

1.2.2 模块化开发

植物多样性科学数据的类别(二级编目)数量众多, 各类所包含的属性字段(三级编目)则更多, 如果每个类别的每个属性字段都单独进行处理, APIs 的研发效率将会非常低下。因此, 我们采用了模块化的开发方法, 通过提高模块的内聚性和降低模块之间的耦合性, 从而提高 APIs 的研发效率。具体而言, 根据数据编目的结果, 将每类数据接口确定为一个模块, 各个模块即数据接口之间相互独立互不干扰, 各个接口可以根据各自个性化需求进行相关的业务处理, 但各个模块内部均调用了统一的权限控制函数和数据查询函数, 统一的权限控制函数能根据权限参数对各个接口的访问权限进行统一控制, 统一数据查询函数能根据各类数据的属性字段和查询参数相结合的方式统一生成查询语句进行查询并返回查询结果, 而无需单独处理各个属性字段。

模块化开发不仅能提高 APIs 的研发效率, 同时由于各个 API 相对独立因此对 APIs 进行更新维护的效率会更高, 例如新增 API 的简要流程和方法: (1) 对科学数据进行编目, 并根据编目结果在系统中建立类(对象)和属性(字段)的模型; (2) 在 Apache Solr 搜索引擎中配置模型属性与数据库字段的映射关系、Solr 索引字段以及数据库连接方式等信息, 从数据库中拉取数据到 Solr 中存储; (3) 在 APIs 访问控制列表中配置访问新 API 的权限控制参数; (4) 在系统中新增 API 请求处理函数(接口地址), 通过调用统一的权限控制函数和数据查询函数等模块处理用户的请求并返回查询结果。

1.2.3 字段组合查询

为了尽可能地满足用户调用 APIs 的个性化需求, 植物多样性科学数据 APIs 实现了各个属性字段的任意组合查询, 用户可以根据自己的需求选择相应的字段进行组合作为查询过滤条件, 而且各个字段均支持精确查询和模糊查询, 模糊查询主要通过使用*字符模糊匹配任意字符, 而对于某些属性字段的内容主要是长文的, APIs 还做了特殊处理从而支持用户进行分词查询, 即 APIs 具备从长文本中检索同时满足给定多个关键词的能力, 因此植物多样性科学数据 APIs 能为用户提供相对灵活和自由的接口调用服务。

字段组合查询能方便用户通过多个条件组合进行数据查询和获取查询结果, 用户可以通过访问中国科学院昆明植物研究所科学数据中心-植物多样性科学数据接口服务平台(<https://api.iflora.cn>)发布的 API 文档获取相应的接口地址和请求/返回参数, 请求/返回参数中给出了字段名称、中文备注、字段类型、模糊检索通配符以及可选过滤参数等信息, 用户可以根据以上信息以字段组合查询的方式调用 API 接口, 如检索《中国生物物种名录》中槭属下的已接受种有哪些可通过调用植物名称 API 接口“<https://api.iflora.cn/API/Taxon?datasetID=COLChina&genus=Acer&taxonRank=Species&taxonomicStatus=accepted>”来获取相应的查询结果。

1.2.4 数据高效检索

本研究研发的植物多样性科学数据 APIs 类型丰富、数据资源庞大, 某些类别的数据资源记录数量已达千万级别, 而且为了满足任意属性字段的组合查询, 甚至是模糊查询, 因此数据检索的效率要求较高。面向海量数据的多条件组合查询和模糊查询, 我们采用了基于 Apache Solr 高性能搜索引擎构建的数据中台数据库, 并通过将各个类别的所有属性字段都建立索引从而实现植物多样性科学数据的高效检索。Apache Solr 是一个基于 Lucene 全文检索技术和采用 Java 程序设计语言开发的高性能开源搜索引擎, 它通过 `ata-config.xml` 文件来配置数据交互的方式和字段对应关系的映射, 通过

managed-schema 文件来配置字段的类型、存储、索引和排序等功能。APIs 服务平台通过将用户的查询请求处理成 Lucene 查询语句,然后将查询语句输入 SolrNet,通过 SolrNet 实现 .Net 平台与 Apache Solr 的交互,最终实现基于 Apache Solr 的高效数据检索。

1.2.5 接口安全控制

科学数据的应用开放应该同时兼顾安全保护和开放共享的平衡,需要确保其可能带来的安全风险不被用来危害我国经济社会的发展^[28]。因此,科学数据的开放共享既要满足《科学数据管理办法》提出的“开放为常态、不开放为例外”的共享理念,又要做到科学数据的安全可控,前提是做好科学数据的分级分类管理。本研究涉及的植物多样性科学数据为一般科学数据,但也存在一些较为敏感的数据内容,为此,我们在数据汇聚阶段,即从各个数据库抽取科学数据到数据中台数据库时,已对数据进行脱敏处理,因此 APIs 返回的结果数据均为可开放共享的数据。此外,为了确保 APIs 的可用性,我们为各个 API 设定了每天可开放调用的默认次数,当然如果用户有特殊需求,也可以与我们联系申请提高调用次数。具体而言,我们通过使用 APIs 访问控制列表对用户申请备案的 IP 地址设置单独的调用次数限制,如表 2 所示,从而实现对科学数据和接口权限的安全管理。

表 2 APIs 访问控制列表

Table 2 Access control list of APIs

IP 地址 IP address	植物名称接口 Taxon API	物种信息接口 Flora API	图片元数据接口 Photo API	种质元数据接口 Germplasm API	其他接口 Other API
210.72.89.4	1000	1000	1000	500	...
210.72.89.35	2000	2000	2000	500	...
210.72.89.57	3000	1000	1000	500	...
210.72.88.224	4000	1000	1000	500	...
210.72.88.235	5000	5000	5000	500	...
...	...	...	...	...	...

2 结果

基于中国科学院昆明植物研究所科学数据中心收集和整理的植物多样性科学数据,通过对各类植物科学数据进行分类编目,采用基于模块化的开发方法,我们研发了覆盖植物的名称、物种信息、图片元数据、种质元数据等植物多样性科学数据 APIs,各类植物科学数据接口地址如表 3 所示。

表 3 植物多样性科学数据 APIs

Table 3 Scientific data APIs for plant diversity

序号 No	接口名称 API name	接口地址 API address
1	植物名称接口	https://api.iflora.cn/API/Taxon
2	物种信息接口	https://api.iflora.cn/API/Flora
3	图片元数据接口	https://api.iflora.cn/API/Photo
4	种质元数据接口	https://api.iflora.cn/API/Germplasm
5	文献元数据接口	https://api.iflora.cn/API/Literature

序号 No	接口名称 API name	接口地址 API address
6	标本元数据接口	https://api.iflora.cn/API/Specimen
7	物种分布预测接口	https://api.iflora.cn/API/Prediction
8	DNA 条形码接口	https://api.iflora.cn/API/DNABarcode

各个 API 均实现了任意属性字段的组合查询、数据高效检索、数据安全以及接口权限控制等功能，植物多样性科学数据 APIs 已在中国科学院昆明植物研究所科学数据中心植物多样性科学数据接口服务平台(<https://api.iflora.cn/>)上线发布，如图 2 所示。



图 2 植物多样性科学数据接口服务平台

Figure 2 Plant diversity scientific data interface service platform

用户通过该平台可以详细了解各个接口的标准规范，调用相关接口并快速获取植物多样性科学数据，如用户调用植物名称接口 <https://api.iflora.cn/API/Taxon> 后，系统以 Json 格式返回植物名称数据，如图 3 所示。

```
{
  "Message": "Success",
  "totalCount": "8315217",
  "pageIndex": "1",
  "pageSize": "10",
  "Data": [
    {
      "id": "57AD89F6-8DE9-4541-BEC5-8EFD99BC929B",
      "datasetID": "GBIF",
      "taxonID": "11286856",
      "scientificName": "Neuroscapha guillemineana Tul.",
      "scientificNameAuthorship": "Tul.",
      "taxonomicStatus": "synonym",
      "taxonRank": "species",
      "higherClassification": "Plantae > Tracheophyta > Magnoliopsida > Fabales > Fabaceae > Lonchocarpus Kunth",
      "parentNameUsage": "Lonchocarpus Kunth",
      "parentNameUsageID": "2968975",
      "acceptedNameUsage": "Lonchocarpus cultratus (Vell.) A.M.G.Azevedo & H.C.Lima",
      "acceptedNameUsageID": "3908683",
      "originalNameUsage": "",
      "originalNameUsageID": null,
      "family": "Fabaceae",
      "genus": "Lonchocarpus",
      "specificEpithet": "guillemineana"
    }
  ]
}
```

图 3 调用植物名称接口以 Json 格式返回数据

Figure 3 Calling the plant name interface to return data in Json format

同时，基于各个植物多样性科学数据 APIs，以植物拉丁简名为检索关键词(如图 2 右上角检索框)，

我们开发了一个检索植物多样性科学数据的应用示范页面，见附录 2（（详见 <https://doi.org/10.57760/sciencedb.33786>））所示。

3 讨论

当前，开放科学在全球范围内已达成广泛共识，科学数据既是科技创新活动的重要产出，也是新一轮科技创新的重要基础，因此具有巨大的开发和应用价值，尤其在当前人工智能推动科研范式变革以及国内各行各业进行的“数据要素 X”行动中，科学数据已成为最为核心的科技资源，各级机构开始重视科学数据的汇交、管理和共享等工作^[29-30]。2018 年国务院发布《科学数据管理办法》后，先后有 17 个省市发布了相应的地方性的科学数据管理实施细则(试行)，2019 年科技部、财政部优化调整了原有的国家科技资源共享服务平台，批准设立了 20 个国家科学数据中心和 30 个国家生物种质与实验材料资源库；同年，中国科学院启动了科学数据中心体系建设工作，2021 年科学数据中心体系建设已初步完成，首批成立了 1 个总中心、18 个学科中心和 13 个所级中心。中国科学院昆明植物研究所是首批成立所级中心的单位，也是中国科学院植物科学数据中心（学科中心）的参建单位。

科学数据的价值在于开发和应用，如何盘活沉睡的科学数据使其开放共享流动起来，从而充分地对其进行挖掘利用是一个关键而艰巨的课题。国内科学数据中心的建设尚处于起步探索阶段，在管理机制、技术研发以及数据生态等方面都还比较薄弱，尤其在科学数据的开放共享方面与国际主流数据库还存在巨大的差距，因此国内在科学数据共享服务平台的建设方面仍然需要向国际主流数据库学习和借鉴良好的管理经验和先进的技术方法。本研究从植物多样性科学数据 APIs 建设的维度探讨国内科学数据的开放共享问题，可以为生物多样性科学数据的开放共享提供相关参考，以期进一步推动国内科学数据的开放共享进程，从而让科学数据更好地服务于我国的科技创新和科技进步。

致 谢

本研究受云南省技术创新人才（202405AD350053）、中国科学院技术支撑人才（KIB202303）、中国科学院网络安全和信息化专项（CAS-WX2022SDC-SJ01）、云南省科技人才与平台计划（202305AH340005）、云南省重大科技专项（202402AE090039）、中国科学院青年创新促进会会员支持项目（2022397）资助。

作者分工职责

邱金水（1987—），男，广东省韶关市人，硕士学历，高级工程师，研究方向为生物多样性信息学。主要承担数据收集、数据编目、需求分析、系统设计、系统实现和论文撰写等工作。

张建文（1980—），男，湖南省临湘市人，博士学历，高级工程师，研究方向为植物多样性。主要承担数据整理、数据编目和论文撰写等工作。

金涛（1987—），男，云南省昆明市人，硕士学历，高级工程师，研究方向为网络安全与信息化。主要承担系统测试、系统运维和论文撰写等工作。

王朋（1987—），男，山东省菏泽市人，硕士学历，工程师，研究方向为信息技术。主要承担数据整理、数据编目和论文撰写等工作。

杜宁（1980—），女，山东省聊城市人，硕士学历，高级工程师，研究方向为文献情报学。主要承担

数据整理、数据编目和论文撰写等工作。

黄蓉（1989—），女，云南省昆明市人，硕士学历，工程师，研究方向为文献情报学。主要承担数据整理、数据编目和论文撰写等工作。

庄会富（1985—），男，山东省日照市人，博士学历，高级工程师，研究方向为生物多样性信息学。主要承担数据收集、数据编目、需求分析、系统测试和论文撰写等工作。

参考文献

- [1] 马荣华, 许金朵, 侯渲, 等. 中国湖泊科学数据共享的路径和方法[J]. 中国科学数据, 2023, 8(4): 210 – 223. [MA R H, XU J D, HOU X, et al. Strategies and methodologies for facilitating lake data sharing in China[J]. China Scientific Data, 2023, 8(4): 210 – 223.]
- [2] 宗倩倩. 科学数据共享利用的知识产权保护机制研究[J]. 图书馆杂志, 2024, 43(5): 42 – 51. DOI: 10.13663/j.cnki.lj.2024.05.005. [ZONG Q Q. Research on intellectual property protection mechanism of scientific data sharing and utilization[J]. Library Journal, 2024, 43(5): 42 – 51. DOI: 10.13663/j.cnki.lj.2024.05.005.]
- [3] 冷疏影, 苏天赞, 张亮. 海洋科学数据开放共享实践与思考[J]. 科学通报, 2023, 68(17): 2239 – 2249. [LENG S Y, SU A, ZHANG L. Practice and reflections on open sharing of marine scientific data[J]. Chinese Science Bulletin, 2023, 68(17): 2239 – 2249.]
- [4] 汪洋, 郑晓欢, 班艳, 等. 深化落实数据要素政策, 建立健全科学数据生态圈[J]. 中国科学数据, 2023, 8(1): 127 – 138. DOI: 10.11922/11-6035.csd.2023.0049.zh. [WANG Y, ZHENG X H, BAN Y, et al. Establishing a sound scientific data ecosystem through the implementation of a data element policy[J]. China Scientific Data, 2023, 8(1): 127 – 138. DOI: 10.11922/11-6035.csd.2023.0049.zh.]
- [5] 王呈珊, 王瑞丹. 基于政策导向的科学数据权属问题分析[J]. 中国科技资源导刊, 2024, 56(3): 26 – 33. WANG C S, WANG R D. Analysis of the ownership of scientific data under policy guidance[J]. China Science & Technology Resources Review, 2024, 56(3): 26 – 33.]
- [6] 马克平, 朱敏, 纪力强, 等. 中国生物多样性大数据平台建设[J]. 中国科学院院刊, 2018, 33(8): 838 – 845. DOI: 10.16418/j.issn.1000-3045.2018.08.011. [MA K P, ZHU M, JI L Q, et al. Establishing China infrastructure for big biodiversity data[J]. Bulletin of Chinese Academy of Sciences, 2018, 33(8): 838 – 845. DOI: 10.16418/j.issn.1000-3045.2018.08.011.]
- [7] YE J F, CHEN L, SHAN Z J, et al. *Pternopetalum paucifoliolatum* (Apiaceae), a new Critically Endangered species from Sichuan, China[J]. Phytokeys, 2020, 166: 97-103.
- [8] 姜承勇, 余卫星, 杨婷, 等. 基于中国馆藏标本数据分析全国植物标本采集现状及采集趋势预测[J]. 科研信息化技术与应用, 2018, 9(5): 94 – 101. DOI: 10.11871/j.issn.1674-9480.2018.05.011. [JIANG C Y, YU W X, YANG T, et al. Geographical sampling: current status analysis and prediction in China based on Chinese herbaria specimen records[J]. e-Science technology & application, 2018(5): 94 – 101. DOI: 10.11871/j.issn.1674-9480.2018.05.011.]
- [9] XU W B, SVENNING J C, CHEN G K, et al. Human activities have opposing effects on distributions of narrow-ranged and widespread plant species in China[J]. Proceedings of the National Academy of Sciences of the United States of America, 2019, 116(52): 26674 – 26681. DOI: 10.1073/pnas.1911851116.

- [10] 朱艳华, 胡良霖. 基于领域知识关联的集成服务系统研究: 以植物学领域为例[J]. 中国科技资源导刊, 2013, 45(1): 51 – 56. DOI: 10.3772/j.issn.1674-1544.2013.01.010. [ZHU Y H, HU L L. Research of integrated service system based on domain knowledge connection by botany field[J]. China Science & Technology Resources Review, 2013, 45(1): 51 – 56. DOI: 10.3772/j.issn.1674-1544.2013.01.010.]
- [11] 钱宏, 张健, 赵静超. 世界上已知维管植物有多少种? 基于多个全球植物数据库的整合[J]. 生物多样性, 2022, 30(7): 33 – 37. DOI: 10.17520/biods.2022254. [QIAN H, ZHANG J, ZHAO J C. How many known vascular plant species are there in the world? An integration of multiple global plant databases[J]. Biodiversity Science, 2022, 30(7): 33 – 37. DOI: 10.17520/biods.2022254.]
- [12] LU L M, MAO L F, YANG T, et al. Evolutionary history of the angiosperm flora of China[J]. Nature, 2018, 554(7691): 234 – 238. DOI: 10.1038/nature25485.
- [13] 张洋, 肖燕珠. 生命周期视角下《科学数据管理办法》解读及其启示[J]. 图书馆学研究, 2019(15): 37 – 43, 13. DOI: 10.15941/j.cnki.issn1001-0424.2019.15.006. [ZHANG Y, XIAO Y Z. Interpretation and enlightenment on the rules of scientific data management from the perspective of life cycle[J]. Research on Library Science, 2019(15): 37 – 43, 13. DOI: 10.15941/j.cnki.issn1001-0424.2019.15.006.]
- [14] 温珂, 宋大成, 游玗怡, 等. 政策计量视角下我国科学数据开放共享政策演进与体系构建[J]. 科学通报, 2024, 69(9): 1175 – 1182. [WEN K, SONG D C, YOU D Y, et al. Features and current situation of the evolution of China's scientific data open sharing policy from the perspective of policy bibliometric[J]. Chinese Science Bulletin, 2024, 69(9): 1175 – 1182.]
- [15] 邱金水, 庄会富, 王朋. 科学实验数据共享平台的设计与实现[J]. 实验室研究与探索, 2022, 41(11): 277 – 281. DOI: 10.19927/j.cnki.syyt.2022.11.056. [QIU J S, ZHUANG H F, WANG P. Design and implementation of scientific experiment data sharing platform[J]. Research and Exploration in Laboratory, 2022, 41(11): 277 – 281. DOI: 10.19927/j.cnki.syyt.2022.11.056.]
- [16] 邱金水, 王亚楠, 庄会富. 中国生物多样性在线数据处理平台的构建[J]. 生物多样性, 2022, 30(11): 218 – 228. DOI: 10.17520/biods.2022356. [QIU J S, WANG Y N, ZHUANG H F. Construction of the Chinese biodiversity online data processing platform[J]. Biodiversity Science, 2022, 30(11): 218 – 228. DOI: 10.17520/biods.2022356.]
- [17] QIU J S, ZHANG J W, JIN T, et al. PNSS: An online plant name service system[J]. Biodiversity Data Journal 13: e142973. <https://doi.org/10.3897/BDJ.13.e142973>.
- [18] QIU J S, ZHANG J W, WANG Y N, et al. PPDC: an online platform for the prediction of plant distributions in China[J]. Journal of Plant Ecology, 2024: rtac094. DOI: 10.1093/jpe/rtac094.
- [19] 张耀南, 吴亚敏, 张彩荷, 等. 欧盟科学数据发展政策与规划[J]. 中国科学数据, 2024, 9(1): 1 – 15. DOI: 10.11922/11-6035.csd.2023.0140.zh. [ZHANG Y N, WU Y M, ZHANG C H, et al. Policies and planning for advancing scientific data development in the EU[J]. China Scientific Data, 2024, 9(1): 1 – 15. DOI: 10.11922/11-6035.csd.2023.0140.zh.]
- [20] GBIF. Global Biodiversity Information Facility[EB/OL]. [2024-07-18]. <https://www.gbif.org>.
- [21] Catalogue of Life[EB/OL]. [2024-07-18]. <https://www.catalogueoflife.org>.
- [22] BHL. Biodiversity Heritage Library[EB/OL]. [2024-07-18]. <https://www.biodiversitylibrary.org>.
- [23] Catalogue of Life China[EB/OL]. [2024-07-18]. <http://www.sp2000.org.cn>.
- [24] 邱金水, 蔡杰, 杨娅娟, 等. 国家重要野生植物种质资源库共享服务平台设计与实现[J]. 植物遗

- 传资源学报, 2022, 23(5): 1536 – 1545. DOI: 10.13430/j.cnki.jpgr.20220316001. [QIU J S, CAI J, YANG Y J, et al. Design and implementation of sharing service platform for national wild plant germplasm resource center[J]. Journal of Plant Genetic Resources, 2022, 23(5): 1536 – 1545. DOI: 10.13430/j.cnki.jpgr.20220316001.]
- [25] National Specimen Information Infrastructure[EB/OL]. [2024-07-18]. <http://www.nsii.org.cn>.
- [26] Plant Photo Bank of China[EB/OL]. [2024-07-18]. <http://ppbc.iplant.cn>.
- [27] WIECZOREK J, BLOOM D, GURALNICK R, et al. Darwin Core: an evolving community-developed biodiversity data standard[J]. PLoS One, 2012, 7(1): e29715. DOI: 10.1371/journal.pone.0029715.
- [28] 廖方宇, 胡良霖, 王健, 等. 科学数据安全标准研究与工作建议[J]. 科学通报, 2024, 69(9): 1142 – 1148. DOI: 10.1360/TB-2023-0228. [LIAO F Y, HU L L, WANG J, et al. Research and suggestions on scientific data security standards[J]. Chinese Science Bulletin, 2024, 69(9): 1142 – 1148. DOI: 10.1360/TB-2023-0228.]
- [29] 郭华东, 陈和生, 闫冬梅, 等. 加强开放数据基础设施建设, 推动开放科学发展[J]. 中国科学院院刊, 2023, 38(6): 806 – 817. DOI: 10.16418/j.issn.1000-3045.20230208001. [GUO H D, CHEN H S, YAN D M, et al. Strengthening open data infrastructure and promoting open science[J]. Bulletin of Chinese Academy of Sciences, 2023, 38(6): 806 – 817. DOI: 10.16418/j.issn.1000-3045.20230208001.]
- [30] 廖方宇, 李婧, 龙春, 等. 开放科学背景下科学数据开放共享安全挑战及我国对策思考[J]. 农业大数据学报, 2024, 6(2): 146 – 155. DOI: 10.19788/j.issn.2096-6369.000027. [LIAO F Y, LI J, LONG C, et al. Security challenges and countermeasures on open sharing of scientific data in the context of open science[J]. Journal of Agricultural Big Data, 2024, 6(2): 146 – 155. DOI: 10.19788/j.issn.2096-6369.000027.]

论文引用格式

邱金水, 张建文, 金涛, 等. 植物多样性科学数据 APIs 的建设[J/OL]. 中国科学数据, 2025, 10(4). (2025-12-19). DOI: 10.11922/11-6035.csd.2024.0184.zh.

Construction of scientific data APIs for plant diversity

QIU Jinshui^{1,2,3*}, ZHANG Jianwen¹, JIN Tao¹, WANG Peng¹,
DU Ning¹, HUANG Rong¹, ZHUANG Huifu^{1,2,3*}

1. Scientific Data Center, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, P.R. China

2. State Key Laboratory of Phytochemistry and Natural Medicines, Kunming 650201, P.R. China

3. National Wild Plant Germplasm Resource Center, Kunming 650201, P.R. China

*Email: qiu jinshui@mail.kib.ac.cn (QIU Jinshui); zhuanghuifu@mail.kib.ac.cn (ZHUANG Huifu)

ABSTRACT: With the rapid development of big data technology and artificial intelligence technology, scientific data have become one of the core scientific and technological resources driving the transformation of scientific research paradigm. Biodiversity scientific data, like biological resources

themselves, have become an important national strategic and a focal point of international competition in science, technology, and industry. To address the gap between domestic plant diversity scientific data and international mainstream databases in terms of openness and sharing, Application Programming Interfaces (APIs) for plant diversity scientific data are proposed and implemented in this study. Based on plant scientific data, including plant name, species information, picture metadata, germplasm metadata, DNA barcode, specimen metadata, as well as book and literature metadata collected and organized by the Scientific Data Center, Kunming Institute of Botany, Chinese Academy of Sciences, the APIs business process and standard specifications were designed through cataloging of diverse plant scientific data. Using a modular development method, the APIs for plant diversity scientific data were developed and built, enabling functions such as flexible combination queries of arbitrary attribute fields, efficient data retrieval, data security assurance, and interface access control. These APIs contribute to the establishment and improvement of domestic standards for plant diversity scientific data interfaces, enhance the level of data sharing level of plant diversity scientific data in China, promote research on plant diversity, and provide data and technical support for biodiversity conservation. In addition, the proposed framework can serve as a reference for the construction of scientific data APIs in other disciplinary domains.

Key words: plant diversity; scientific data; APIs; openness and sharing of data; data center construction; data element