# 国家农业科学数据共享中心资源建设探析*

朱亮，孟宪学，赵瑞雪，赵华
（中国农业科学院农业信息研究所，北京 100081）

摘要：国家农业科学数据共享中心是科技部首批认定的23家国家科技基础条件平台之一，其目标是有效盘活、挖掘、抢救和保存农业科学数据资源，实现农业科学数据的共享与集成应用，为农业科技创新与发展提供基础支撑与保障。本文重点对国家农业科学数据共享中心的农业科学数据资源建设情况进行阐述，主要包括农业科学数据资源整合体系与策略、标准规范、整合内容、精品数据集、对策及建议等。

关键词：国家农业科学数据共享中心；农业科学数据；信息资源建设
中图分类号：G250
DOI：10.3772/j.issn.1673-2286.2017.11.003

## 1 引言

随着以数据密集型科学发现为主要科学研究特征的科学研究第四范式的兴起与发展，以及世界数据中心、国际科学数据委员会的成立，催生了科学数据这一新兴学科领域[1]，以科学数据获取、共享、利用等为主要内容的科学数据组织管理研究和实践得到广泛开展。而作为基础性工作的科学数据长期保存也引起各类型科研相关机构的高度关注[2]，如欧盟资助的“欧洲科学数据长期保存计划”项目旨在探索如何保障科学研究过程中产生的原始数据及最终成果等科学数据资源的长期保存问题，以发掘科学数据资源长期可存取、可利用、可理解的有效方法和途径[3]。科学数据是人类社会科技活动长期积累的或通过其他方式获取的反映客观世界的本质、特征、变化规律等原始性、基础性数据，以及根据不同科技活动需要进行系统加工整理的各类数据集合[4-5]。农业科学数据是各类农业科技活动的基本产出之一，是农业科学研究得以持续发展和科学技术不断创新的资源宝库，它不仅直接反映一个国家的整体农业科技基础水平，还影响农业科技实力能否持续、稳定、长久地发展和提高[6]，具有重要的保存和开发利用价值。

在农业科学数据建设方面，国际组织、欧美发达国家较早开展相关实践，已建成一批重要的农业科学数据库。如联合国粮农组织是国际上收集、整理和发布农业数据的重要机构，其主导建设了农业统计数据库、水生物种引进数据库，以及性别和土地权利数据库、TERRASTAT数据库、水资源及农业信息系统、农业市场信息系统、家畜饲料资源信息系统、水产科学及渔业信息系统等；知名的基因银行数据库最初由美国能源所负责建设，是世界最大的公共生物信息库，收集来自不同物种的DNA序列；联合国环境计划署支持建设了全球资源信息数据库；澳大利亚、英国、美国分别建设了世界牧草属数据库、世界草业数据库和牧草资源数据库；美国植物保护中心和国际作物保护协会创建病虫害综合治理资源数据库；加拿大、美国、英国、法国、澳大利亚、日本等国家基于地理信息系统技术先后建设本国的土壤数据库。在我国，自20世纪80年代，随着“国家科技攻关计划”“国家高技术研究发展计划”等科技计划的实施，国家投入并积累了一大批包括科学数据在内的具有持续利用价值的各类科技资源，为我国科技进步奠定物质技术基础[7]。然而，在科技资源整体规模不断扩大的同时，其暴露的条块分割、部门封闭、单位所有、利用率低、共享机制缺乏等问题日益突

* 本研究得到国家科技基础条件平台专项“农业科学数据共享中心”（编号：2005DKA31800）和“中国农业科学院科技创新工程”项目（编号：CAAS-ASTIP-2017-AII）资助。

显。就农业科学数据而言，早期虽有部分农业科学数据资源以专题数据库的形式得到加工和存储，但其加工标准不统一，存储载体和存储技术系统各异，无法实现真正意义上的资源整合与共享利用。为改变此现状，以2005年科技部和财政部正式启动的“国家科技基础条件平台建设”专项为契机[8]，由中国农业科学院农业信息研究所牵头，联合中国水产科学研究院、中国热带农业科学院、中国农业科学院专业研究所等单位共同参与实施了国家农业科学数据共享中心建设，其目标是有效盘活、挖掘、抢救和保存我国农业科学数据资源，实现农业科学数据的共享与集成应用，为农业科技创新与发展提供基础支撑与保障。本文将重点对国家农业科学数据共享中心的资源整合体系与策略、标准规范、整合内容、精品数据集、对策及建议等进行阐述。

## 2 国家农业科学数据共享中心资源建设

### 2.1 农业科学数据资源整合体系与策略

面对农业科学数据类型多样、专业众多且跨度大、存储分散的现状，国家农业科学数据共享中心采用“以学科为龙头”的资源整合策略，建立包括作物科学、动物科学与动物医学、农业区划科学、草地与草业科学、渔业与水产科学、热带作物科学、农业科技基础、农业生物技术与生物安全、农业微生物科学、农业资源与环境科学、食品工程与农业质量标准、农业信息与科技发展的12大类核心学科资源整合框架，形成由数据主中心、数据分中心、数据节点三个层次组成的资源整合体系。在实际资源整合过程中，针对每类学科资源的专业领域特点，国家农业科学数据共享中心选择拥有资源最多、技术力量最强的单位作为资源整合的依托单位，具体承担该类学科资源整合的组织和实施，学科资源整合完毕后先汇交到数据分中心，再由数据分中心提交到数据主中心。

为保障农业科学数据资源整合的质量，国家农业科学数据共享中心建立了一套行之有效的资源整合策略。首先，制定农业科学数据学科分类框架以及数据制作标准规范；其次，组织数据拥有者（数据节点）按照统一的标准规范进行专业数据的规范化加工、整理，建立本地农业科学数据集以及相应文档（数据源、元数据、数据字典、数据使用说明等），并通过统一的在线系统完成数据及文档汇交；最后，由数据主中心负责对汇交数据进行审核验收、二次标引、网络化组织、管理与发布。

目前，国家农业科学数据共享中心按照“学科-主体数据库-数据集”三级模式整合了62个农业核心主体数据库，700个数据集，数据总量超400TB，已成为我国农业科学数据最大的“蓄水池”和“集散地”。已整合资源如表1所示。

### 2.2 农业科学数据资源整合标准规范

资源建设，标准先行。标准规范是实现农业科学数据资源高效整合与共享的基础和保障。根据农业科学数据资源整合的实际需要，国家农业科学数据共享中心通过参考借鉴，形成了包括《农业科学数据汇交管理办法》《农业科学数据检查与质量控制管理办法》等管理规范，《农业科学数据元数据标准》《农业科学数据标引规范》《农业科学数据范畴分类编码标准》《农业科学数据集分类规范》《农业科学数据采集规范》《农业科学数据著录规范》《农业科学数据加工流程规范》等数据制作、数据组织和管理方面的通用标准，以及多项农业专业领域标准的农业科学数据整合规范体系。这些标准规范的制定和实施确保国家农业科学数据共享中心资源建设工作的科学、规范开展，如在资源建设过程中，项目参建单位可依据相关标准高效、顺利地完成农业科学数据资源的采集、整理、分类、加工、著录、标引、元数据制作、汇交等工作，保证不同学科、不同类型数据集建设的规范性和流程的统一性，为下一步多学科资源的集成与应用奠定基础。

### 2.3 农业科学数据资源整合内容

国家农业科学数据共享中心资源整合主要包含两部分内容，即农业科学数据集元数据整合和农业科学数据集实体数据整合。农业科学数据集实体数据整合指完成农业科学数据资源的采集、规范化加工、质检和入库。农业科学数据集元数据主要对科学数据外部形式和内部特征进行详细描述，其主要目标是提供科学数据资源的全面指南，以便用户对数据资源进行准确、高效、充分地开发与利用[9]。农业科学数据集元数据整合的依据是国家农业科学数据共享中心制定的《农业科学数据元数据标准》，包括核心元数据和农业领域扩展元数据。农业科学数据核心元数据指唯一标识一个数据集所需的最少元数据内容，其为用户提供数据的

表 1 农业科学数据资源整合情况

| 学科分类 | 主体库名称 | 学科分类 | 主体库名称 |
|---|---|---|---|
| 作物科学 | 作物遗传资源数据库 | 动物科学与动物医学 | 动物资源与遗传育种数据库 |
| | 作物育种数据库 | | 中国饲料养分数据库 |
| | 作物栽培数据库 | | 国际饲料养分数据库 |
| | 作物分子生物学数据库 | | 动物营养需要数据库 |
| | 作物生产数据库 | | 动物医学数据库 |
| | 作物生理生化数据库 | | 动物科学基础数据库 |
| 农业区划科学 | 农业区划数据库 | 草地与草业科学 | 草地数据库 |
| | 农业资源调查与评价数据库 | | 牧草数据库 |
| | 农业土地利用数据库 | | 草业生产与经济数据库 |
| | 农业区域规划与生产布局数据库 | | 草原区生态背景数据库 |
| | 农业遥感监测数据库 | | 草业动态监测管理信息库 |
| 渔业与水产科学 | 渔业水域资源与生态特征数据库 | 热带作物科学 | 热带作物遗传资源数据库 |
| | 渔业物种资源与生物基础特征数据库 | | 热带作物育种数据库 |
| | 渔业生物资源野外观测调查数据库 | | 热带作物栽培数据库 |
| | 渔业生态环境野外调查数据库 | | 热带作物生物学数据库 |
| | 渔业生产与经济管理数据库 | | 热带作物基础数据库 |
| 农业科技基础 | 农业科技统计数据库 | 农业生物技术与生物安全 | 植物基因组数据库 |
| | 农业科技管理数据库 | | 微生物基因组数据库 |
| | 农业科技动态与发展数据库 | | 农作物转基因数据库 |
| | 农业科技专题数据库 | | 植物生物反应器数据库 |
| | 农业科技信息资源导航库 | | 生物安全数据库 |
| 农业微生物科学 | 农作物病原真菌数据库 | 农业资源与环境科学 | 全国灌溉试验数据库 |
| | 农作物病原细菌数据库 | | 全国数字土壤数据库 |
| | 农作物病毒数据库 | | 全国土壤肥料数据库 |
| | 植物检疫性微生物数据库 | | 全国农田生态环境数据库 |
| | 生物防治微生物数据库 | | 全国农业昆虫数据库 |
| 食品工程与农业质量标准 | 农产品质量标准数据库 | 农业信息与科技发展 | 全国粮食生产数据库 |
| | 农产品质量检测数据库 | | 全国畜牧业生产数据库 |
| | 农作物加工品质数据库 | | 全国农业信息基础设施数据库 |
| | 农产品加工工艺与设备数据库 | | 全国农业经济统计数据库 |
| | 农产品加工质量安全控制数据库 | | 国外农业生产统计数据库 |

最基本信息（如数据内容、数据分类、数据存储与访问信息、数据提供单位信息以及数据更新信息等），便于用户查询检索。国家农业科学数据共享中心要求整合的所有数据集必须完成核心元数据的建设。农业科学数据核心元数据如表2所示。

## 2.4 精品数据集

农业科学数据资源建设，资源质量是核心。高质量精品资源常具有更高的保存和使用价值，也更能体现一个国家科学数据资源整合和共享的整体水平。一直以来，国家农业科学数据共享中心在全面稳步提升数据资源质量的同时，还根据自身专业领域特点重点打造了以下具有广泛影响的精品数据库（集）。

### 2.4.1 作物遗传资源特性评价鉴定数据库

该数据库收集了我国约35万份农作物遗传资源的特性评价鉴定数据，包括资源的护照信息、农艺性状、品质性状、抗病、抗虫等信息，主要用于作物遗传资源

表 2 农业科学数据核心元数据

| 名 称 | 英文名称 | 标 识 | 定 义 |
|---|---|---|---|
| 元数据实体信息 | Metadata | Metadata | 定义有关数据资源元数据的根实体 |
| 元数据标识符 | Name | FormatName | 元数据的唯一标识 |
| 元数据语种 | Version | FormatVer | 元数据使用的语言 |
| 元数据创建日期 | MetadataCreateDate | MetaCreDate | 创建元数据的日期 |
| 元数据字符集 | MetadataCharaterSet | MetaChrSet | 元数据使用的字符编码标准 |
| 数据集标识信息 | DataIdentification | Ident | 唯一标识数据集的基本信息 |
| 数据集名称 | Title | Title | 数据集名称 |
| 数据集英文名称 | EnglishTitle | EngTitle | 数据集的英文名称 |
| 数据集语种 | DatasetLanguage | DataLang | 数据资源采用的语言 |
| 数据集字符集 | CharacterSet | DataChar | 数据集使用的字符编码标准 |
| 数据集完成日期 | DatasetDate | resrefdate | 数据集制作完成的日期 |
| 数据集分类 | TopicCategory | TpCat | 用于农业科学数据整合与共享工程中的数据集分类 |
| 关键词 | Keyword | Keywords | 列出描述数据的可搜索的关键词 |
| 数据集摘要 | Abstract | Abstract | 数据集的简要说明 |
| 数据来源 | DataSource | DataSource | 数据来源 |
| 数据集进展状况 | DatasetStatus | DatasetStatus | 数据集生产与完成情况 |
| 在线链接地址 | Linkage | Linkage | 使用URL地址或类似地址模式进行在线访问的地址 |
| 离线存储介质 | MediumName | MediumName | 数据存储所采用的介质 |
| 维护和更新频率 | UpdateRate | UpdateRate | 数据集维护和更新的频率 |
| 数据集使用局限性 | ApplicationLimitation | AppLimit | 影响数据集适用性的限制，如“不适于某个地区”等 |
| 数据集安全限制分级 | Classification | Class | 对数据处理时安全限制分级的名称 |
| 数据集使用限制 | UseLimitation | UseLimimit | 影响数据集使用的一般限制 |
| 数据集访问限制 | AccessConstraints | AccessConst | 用于确保隐私权或保护知识产权的访问限制，和获取数据时的任何特殊的约束或限制 |
| 负责单位 | OrganisationName | OrgName | 数据资源负责单位名称 |
| 负责人 | IndividualName | IndName | 数据资源负责人姓名 |
| 职责 | Role | Role | 负责单位的职责 |
| 国家 | Country | Country | 负责单位或负责人所在国家 |
| 城市 | City | City | 负责单位或负责人所在的城市 |
| 地址 | DeliveryPoint | DelPoint | 负责单位或负责人的详细地址 |
| 邮政编码 | PostalCode | Postcode | 负责单位或负责人的邮政编码 |
| 电子邮件地址 | ElectronicMailAddress | EmailAdd | 负责单位或负责人的电子邮件地址 |
| 电话 | Voicephone | VoiceNum | 负责单位或负责人的联系电话 |
| 传真 | Facsimile | FaxNum | 负责单位或负责人的联系传真 |

管理、研究，以及作物新品种选育和种植等领域。

### 2.4.2 中国饲料原料数据库

该数据库收集整理了自1981年以来我国饲料原料营养价值的评定结果，收集了家畜、家禽及其他家养动物采食的常规饲料及非常规饲料原料，入库的养分项目200余项，是设计各种畜禽饲料配方的最主要参考数据源。以此为基础，每年发布新版《中国饲料成分及营养价值表》。

### 2.4.3 海洋生物资源动态监测鱼类数据库

该数据库收集了黄渤海区、东海区、南海区重要渔

业水域的现场监测数据及多个相关项目的资源监测数据，包括调查船名、航次、海区、采样水深、渔区、鱼类中文名、鱼种尾数等28项参数信息。

#### 2.4.4 农业区划数据库

该数据库是基于我国1949年以后开展的三次大规模农业资源与区划调查成果形成的，包括大量综合农业区划报告、自然区划报告、专业区划报告，及农业统计资料、农业生产和布局图片等。

#### 2.4.5 草地植被观测数据库

该库收集了呼伦贝尔站、锡林浩特站、甘德站、玛曲站等13个草地野外台站1980年以来的土壤观测数据，包括采样地点、经度、纬度、生殖苗平均高度、叶层平均高度、绿色鲜重、根系深度等多个指标信息。

#### 2.4.6 热带作物病虫害数据库

该数据库系统收集整理了热带地区常见的病虫害信息，包括病虫害中英文名称、病原信息、寄主信息、侵染部位、症状、发病规律、防治方法等。数据库配以大量的彩色图片，直观地介绍病虫害的发生症状及其病原等信息。

## 3 问题分析与对策

农业科学数据资源建设是一项基础性、长期性任务，为实现我国农业科学数据战略资源的长期保存和共享利用，国家农业科学数据共享中心在今后资源建设方面需着重做好以下方面工作。

### 3.1 扩大整合资源学科覆盖面

多年来，国家农业科学数据共享中心在现有资源建设框架上重点对作物科学、动物科学与动物医学等学科资源进行整合，虽然在农业领域其他学科资源建设方面也有实践，但多是以单个数据集的形式进行，未在学科层面进行总体设计与系统实施。未来，国家农业科学数据共享中心需做好补缺工作，以响应国家农村农业发展方向（如农业供给侧改革、美丽乡村建设等）和社会发展热点（如食物安全、转基因等）为重点，以建设专题数据集为手段，不断拓展整合资源的学科覆盖面，尽早实现农业科学数据资源的全学科领域布局。

### 3.2 创新农业科学数据资源整合模式

国家农业科学数据共享中心以科技部财政立项支持的方式开始建设，项目牵头单位通过任务书的形式将各参建单位组织起来开展农业科学数据资源整合，这种模式给数据资源整合的范围、深度和规模都带来一定局限。为此，国家农业科学数据共享中心可采取资源交换（如与科学数据出版商建立资源互换渠道）、用户自主资源注册（如利用积分权益对等的方式鼓励用户提交其农业科学数据资源）等措施创新和完善现有资源整合模式，不断增加农业科学数据资源整合的参与主体。

### 3.3 加强资源的深度整合与集成

目前，国家农业科学数据共享中心建设的大多数资源均是相互独立的数据集，未能实现不同数据集间的关联，这是国家农业科学数据共享中心在下一步资源建设工作中需要重点解决的问题，对此可充分利用关联数据、知识组织体系、数据融合等技术手段对不同农业科学数据集间的关联关系进行挖掘、揭示和再组织，实现农业科学数据资源的深度整合和集成。

### 3.4 继续坚持数据精品化策略

今后，在提升资源质量、打造精品数据集方面，国家农业科学数据共享中心需重点做好两方面工作：一是通过多渠道经费支持、协同众包等方式重点建设基础性、长序列、领域特色鲜明的农业科学数据集；二是完善农业科学数据资源质量评审机制，并结合市场化运营收费等方式切实吸引更多优质农业科学数据资源。

## 参考文献

[1] 李慧佳,马建玲,王楠,等.国内外科学数据的组织与管理研究进展[J].图书情报工作,2013,57(23):130-136.

[2] 庄晓喆.国外高校科学数据保存政策调查与思考[J].图书馆学研究,2015(16):68-72,76.

[3] 欧盟科学数据长期保存计划:PARSE.Insight[EB/OL].[2017-10-15]. http://www.nlc.cn/newtsgj/gtqk/tyck/2008nzml/102/.

[4] 中华人民共和国国家质量监督检验检疫总局中国国家标准化管理委员会.科技平台 通用术语:第2部分:术语和定义GB/T 31075—2014[S].北京:中国标准出版社,2014.

[5] 司莉,邢文明.国外科学数据管理与共享政策调查及对我国的启示[J].情报资料工作,2013(1):61-66.

[6] 赵瑞雪.国家农业科学数据共享中心建设实践与展望[J].农业网络信息,2009(6):4-5,12.

[7] 上海市科技基础条件资源调查[EB/OL].[2017-09-25].http://www.docin.com/p-118151460.html.

[8] 科技部发展计划司等.整合共享创新——国家科技基础条件平台建设回顾与展望[M].北京:中国科学技术出版社,2009.

[9] 赵华,王健.科学数据元数据功能与内容分析[J].科技管理研究,2015,35(17):232-235.

## 作者简介

朱亮，男，1981年生，博士，副研究馆员，研究方向：文献计量、情报分析、科学数据建设与管理研究，E-mail：zhuliang@caas.net.cn。
孟宪学，男，1955年生，博士，研究员，研究方向：农业信息管理、情报学，E-mail：mengxianxue@caas.net.cn。
赵瑞雪，女，1968年生，博士，研究员，博士生导师，研究方向：信息管理与信息系统、信息资源管理、知识组织与数字图书馆，E-mail：zhaoruixue@caas.cn。
赵华，女，1980年生，博士研究生，助理研究员，研究方向：科学数据管理，E-mail：zhaohua02@caas.cn。

Research on Resource Construction for National Agricultural Science Data Sharing Center

ZHU Liang, MENG XianXue, ZHAO RuiXue, ZHAO Hua
(Agricultural Information Institute of CAAS, Beijing 100081, China)

Abstract: National Agricultural Science Data Sharing Center is one of the 23 science and technology infrastructures of China, its goal is to effectively revitalize, excavate, rescue and preserve agricultural science data resources, promotes the sharing and integration of agricultural science data. This paper focuses on theresource construction of National Agricultural Science Data Sharing Center, which mainly includes the integration system and strategy, integration standard specification, integration content, quality data set, countermeasure and suggestion.

Keywords: National Agricultural Science Data Sharing Center; Agricultural Science Data; Information Resource Construction

（收稿日期：2017-10-31）

书 讯

# 《汉语主题词表》（工程技术卷）

《汉语主题词表》自1980年问世以后，经1991年进行自然科学版修订，在我国图书情报界发挥了应有的作用，曾经获得了国家科学技术进步二等奖。为了适应网络环境下知识组织与数据处理的需要，2009年由中国科学技术信息研究所主持，并联合全国图书情报界相关机构，完成《汉语主题词表（工程技术卷）》的重新编制工作。

全书共收录优选词19.6万条，非优选词16.4万条，等同率0.84。在体系结构、词汇术语、词间关系等方面进行改进创新。为了方便工程技术领域不同专业用户使用，《汉语主题词表》（工程技术卷）按专业分13个分册出版，同时建立《汉语主题词表》服务系统，提供在线概念检索和辅助标引服务，通过可视化技术展示各类概念关系，是图书馆、档案馆、出版社、期刊杂志社、文献信息中心等专业工作者及科研、教育及工程技术领域人员必备的参考书。

《汉语主题词表（工程技术卷）》已于2014年由科学技术文献出版社出版，全书2 300余万字，总定价3 880元，可分册购买。