

# 基于生命周期理论的农业科学 数据中心化管理模式

高飞<sup>1,2</sup>, 周国民<sup>2,3</sup>, 满芮<sup>2,4</sup>

1. 中国农业科学院作物科学研究所, 北京 100081;
2. 国家农业科学数据中心, 北京 100081;
3. 中国农业科学院农田灌溉研究所, 河南 新乡 453002;
4. 中国农业科学院农业信息研究所, 北京 100081

## 摘要

农业科学数据资源的建设、管理与共享工作得到了世界各国政府、科研机构和科学家的高度重视, 相关国际组织和农业科学数据平台格外活跃, 为农业科技创新提供了有力支撑。从科学数据资源建设的现状来看, 科学数据管理中心化是未来的发展趋势。对农业科学数据中心化管理模式进行探索, 构建了农业科学数据生命周期管理模型, 并从科学数据管理标准、数据汇交与加工、数据长期保存和数据共享服务4个方面对模型进行解释, 力图为科学数据管理、共享与服务提供参考。

## 关键词

农业科学数据; 生命周期; 管理模式; 数据共享

中图分类号: G203

文献标志码: A

doi: 10.11959j.issn.2096-0271.2022003

## *Centralization of agricultural scientific data management model based on life cycle theory*

GAO Fei<sup>1,2</sup>, ZHOU Guomin<sup>2,3</sup>, MAN Rui<sup>2,4</sup>

1. Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China
2. National Agriculture Science Data Center, Beijing 100081, China
3. Farmland Irrigation Research Institute of Chinese Academy of Agricultural Sciences, Xinxiang 453002, China
4. Agricultural Information Institute of Chinese Academy of Agricultural Sciences, Beijing 100081, China

## *Abstract*

Governments, research institutions and scientists all over the world have attached great importance to the construction, management and sharing of agricultural science data resources. Relevant international organizations and agricultural science data platforms are particularly active, providing strong support for agricultural science and technological innovation. From the current situation of science data resource construction, science data management centralization is the future development trend. The agriculture science data management centralization mode was explored, the agricultural science data life cycle management model was constructed, and the model was explained from four aspects: science data

management standard, data collection and processing, long-term data preservation and data sharing service. This study tries to provide reference for science data management, sharing and service.

### Key words

agricultural science data, life cycle, management model, data sharing

## 0 引言

近年来, 农业科学数据资源的建设、管理与共享工作得到了世界各国政府、科研机构 and 科学家的高度重视, 相关国际组织和农业科学数据平台格外活跃, 农业科学数据管理工作持续推进, 为农业科技创新提供了有力支撑。联合国粮食及农业组织 (Food and Agriculture Organization, FAO) 发布了农业环境指标、农业科技指标、土地利用、渔业资源等10多个数据库, 积极促进农业科技创新研究。国际橡胶研究组织 (International Rubber Study Group, IRSG) 搭建了世界天然橡胶产业数据库, 数据覆盖世界主要天然橡胶的生产面积、产量、库存量、贸易量、市场价格以及主要天然橡胶消费国的消费量、进出口量、进出口价格等。国际椰子共同体 (International Coconut Community, ICC)、国际可可组织 (International Cocoa Organization, ICO)、国际胡椒组织 (International Pepper Community, IPC) 等分别搭建了对应的热带作物产业数据库, 为热带作物经济研究提供了数据支撑。我国科学数据平台建设始于21世纪初<sup>[1]</sup>, 2014年建成地球系统科学、人口与健康、农业等8个领域的国家科技资源共享平台, 2019年科学技术部、财政部对原有国家平台进行优化调整, 形成了20个国家科学数据中心, 推进相关领域科技资源向国家平台汇聚与整

合。从科学数据资源建设现状来看, 科学数据中心化管理是未来发展趋势, 但数据中心如何管理科学数据、进行数据共享还需探索。

## 1 模型与理论基础

### 1.1 科学数据生命周期

科学数据产生于科学实验、调查、观测等科学研究活动, 服务于科研工作者, 其生命周期与科研工作流程紧密相关。从现有文献来看, 根据研究内容的不同, 科学数据生命周期各阶段划分略有不同, 大致可分为数据计划、数据获取 (生产)、数据处理、数据存储、数据共享 (重用) 5个阶段。杨传汶等人<sup>[2]</sup>在此基础上增加了数据更新阶段, 并提出了基于科研动态的数据服务, 如协助制订数据计划、设计元数据、提供保存工具、提供领域专家信息、提供数据检索服务、数据评价交流和协助数据更新完善等。储节旺等人<sup>[3]</sup>将数据共享细分为共享、分析、再利用3个环节, 构建了科学数据管理体系, 包括科学数据管理制度、管理风险防控体系、数据质量评估和基于科学数据生命周期的数据资源配置、技术支持、人才队伍建设和信息素养培育。夏义堃等人<sup>[4]</sup>从学科特性和学术伦理角度出发, 认为应从基础层 (政策标准、基础设施、数据能力、资金保证)、流程控制层 (数据管理计划、采集、组织、保存、

共享利用)和主体层(资助机构、研究机构、出版商、数据平台)3个层面对生命科学数据的质量进行控制。陈欣等人<sup>[5]</sup>进行了社会科学数据特征研究,将数据生命周期简化为创建、分析、公开3个阶段。姚占雷等人<sup>[6]</sup>基于人文社科数据生命周期各阶段特点,构建了数据管理平台,满足人文社科研究学者的科研需求。生命周期是一种有效的分析工具,可以清晰地反映数据创建后的各阶段特点,本文采用此方法进行农业科学数据中心化管理模式研究。

## 1.2 农业科学数据生命周期管理模型

本文基于国家农业科学数据中心(以下简称数据中心)的数据管理实例,将农业科学数据管理活动分为数据资源建设、汇交、加工、长期保存、共享5个阶段,如图1所示。本文的农业科学数据资源主要包括科技项目科学数据、长期观测科学数据和学科领域自建科学数据。数据汇交从广义上来说是数据拥有者将科学数据提交

到科学数据管理机构的过程,除了科技项目科学数据、长期观测科学数据,学科领域自建科学数据也可被提交至科学数据管理机构。例如,在发表论文时,出版社往往要求论文使用的数据同时发表,但并不是所有出版社都能提供数据存储服务。这时,作者可以将数据提交至科学数据管理机构进行公开发表。数据加工主要针对元数据和数据实体,通过添加科学数据资源标识、智能分类、融汇等控制数据质量,进行规范化管理。数据长期保存要对数据进行分级分类管理,同时保证数据存储环境安全,保证数据可以重复使用,实现对科学研究过程的追溯。数据管理的最终目的是实现数据共享,数据中心应将元数据在网络上进行共享,并提供数据检索服务。

数据管理标准是贯穿整个数据生命周期的,包括数据采集、汇交、格式交换、质量控制、元数据等标准规范。同时,由于数据以几何倍数增长,科学数据管理离不开基础设施(分析工具、存储系统、管理工具)的辅助。

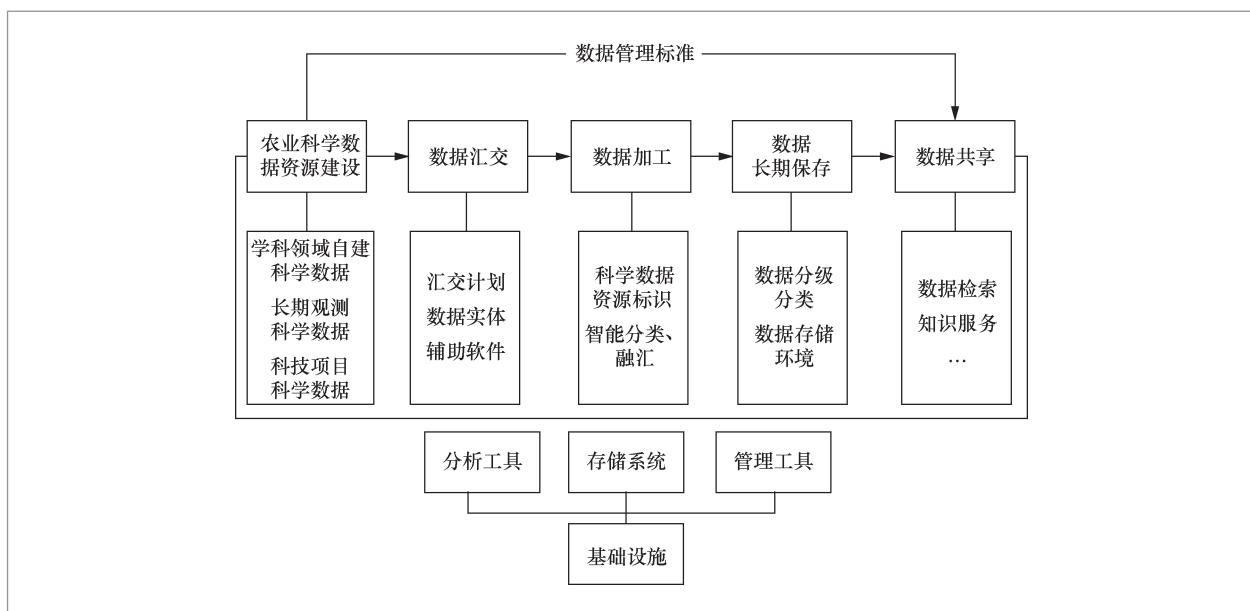


图1 农业科学数据生命周期管理模型

## 2 农业科学数据管理的主要问题

### 2.1 农业科学数据资源缺乏长期管护

数字时代的到来使数据存储更加便捷,但可靠性却大大降低。存储设备过时、读取设备淘汰都可能导致数据无法使用。而与传统的文献信息相比,由数字技术支持的数字信息在存储、传输和持久保存方面存在一系列与生俱来的问题。数字信息的存活和使用必须要得到特别的维护和管理,以确保数字信息长期存活和数字信息真实可信,能够被未来的使用者理解和应用。但农业科学数据,特别是原始数据,大量分散在相关行业的政府部门、研究机构、农业企业等,无法进行统一管理,缺乏有效的数据管护,不利于科研人员使用,难以发挥农业科学数据在研究中的作用。

### 2.2 农业科学数据质量参差不齐

农业数字资源规模庞大、增长快速,但质量参差不齐,主要表现为数据重复保存、数据丢失、分类不合理、缺少相关质量说明文档等。原始数据由项目团队保存,而项目团队往往缺乏管理意识,造成数据质量降低。例如,研究人员根据自身经验创建元数据,会出现元数据冗余、丢失、编码错误、前后不一致、版本混乱等问题。

### 2.3 农业科学数据共享率低

农业数字资源数据结构复杂、多变,内容组分权属边界模糊,获得与使用条件烦琐复杂,导致我国科学数据开放共享

程度不够理想,与社会大众的需求差距较大。有学者调查研究发现,科学数据的权益不清是我国科学数据开放共享程度不够理想的重要原因之一,科学数据持有者担心开放共享数据后带来权益纠纷等不良影响。此外,我国科学数据流失严重,国际上很多有影响力的杂志要求论文发表前必须提交支撑论文的基础科学数据,学者为了发表论文常常在没有将数据汇交到国内管理机构的情况下向国外提供数据。

## 3 农业科学数据中心化管理模式

### 3.1 数据管理标准

数据中心需要制定相关的管理标准以便计算机识别和处理数据,使数据更容易被发现和重用。

#### (1) 农业科学数据采集标准

农业科学数据采集标准定义了农业数据采集各阶段需要遵循的基本规则,对采集对象、采集方法、原始数据的获得与记录等过程进行了定义和结构化,使数据采集无歧义,便于计算机理解与记录。数据采集对象应明确其位置属性、时间属性和基本特点。采集指标需经过数据化标准专家、领域专家的标准化处理,形成可直接组成独立于语法数据交换格式的、可重用的数据采集指标。采集的原始数据不允许修改,以确保数据记录的连续性和完整性。

#### (2) 农业科学数据质量控制

数据的质量是影响科学数据重用的关键性因素之一,而农业科学数据由于其内容的广阔性、结构的复杂性,数据质量控制尤为重要。针对农业科学数据的特点,数据中心制定了农业科学数据质量检查与

控制规范。农业科学数据质量应从定量与非定量标准两方面进行控制。数据质量定量评估标准主要包括：①完整性，数据集是否存在冗余数据或缺少数据；②逻辑一致性，即数据概念是否符合概念模式规则、值是否在值域范围内、数据存储与数据集物理结构是否一致、数据集拓扑关系是否一致；③位置精度，包括绝对精度、相对精度和栅格数据位置精度；④时间精度，包括时间测量精度、时间一致性、时间正确性；⑤专题精度，即数据分类是否正确、非定量属性（如数据集标题、关键字、数据版本等）描述是否正确、数值属性精度是否准确。数据质量非定量评估标准主要包括：数据集创建目的是否说明、数据用途是否填写，以及数据志（数据从收集、获取、汇编到现状的整个生命周期）是否记录清晰。在数据控制管理过程中，这两种类型的数据质量评估结果都应当被提供，每个数据质量结果都有一个数值类型，且这个结果可以被计算机识别。

### (3) 农业科学数据元数据标准

农业科学数据元数据标准对完整描述数据对象的数据项集合、著录规则进行

了定义，适用于资料共享、数据发布、数据集编目、数据交换和网络查询服务等。农业科学数据元数据应提供标识、内容、分发、质量、表现、参照、图示表达、扩展、限制和维护等信息。元数据标准体系分为标准和引用两部分。标准部分包括标识信息、内容信息、分发信息、数据质量信息、数据表现信息、参照系信息、图示表达目录信息、元数据扩展信息、应用模式信息、限制信息和维护信息；引用部分包括覆盖范围信息以及引用和责任方信息。元数据内容框架如图2所示，每一部分信息都用统一建模语言（unified modeling language, UML）包表示。元数据实体可按需要聚集或重复，以满足标准规定的必选要求和领域的其他要求。对于公共元数据来说，标识信息为必选项，其他信息为可选项。

针对农业科学数据生命周期管理，数据中心还制定了数据汇交管理办法、农业科学数据服务规范、数据发布管理规则、农业科学数据信息安全管理规范、农业科学数据中心用户管理规范、农业科学数据集集成和访问规范等标准规范。

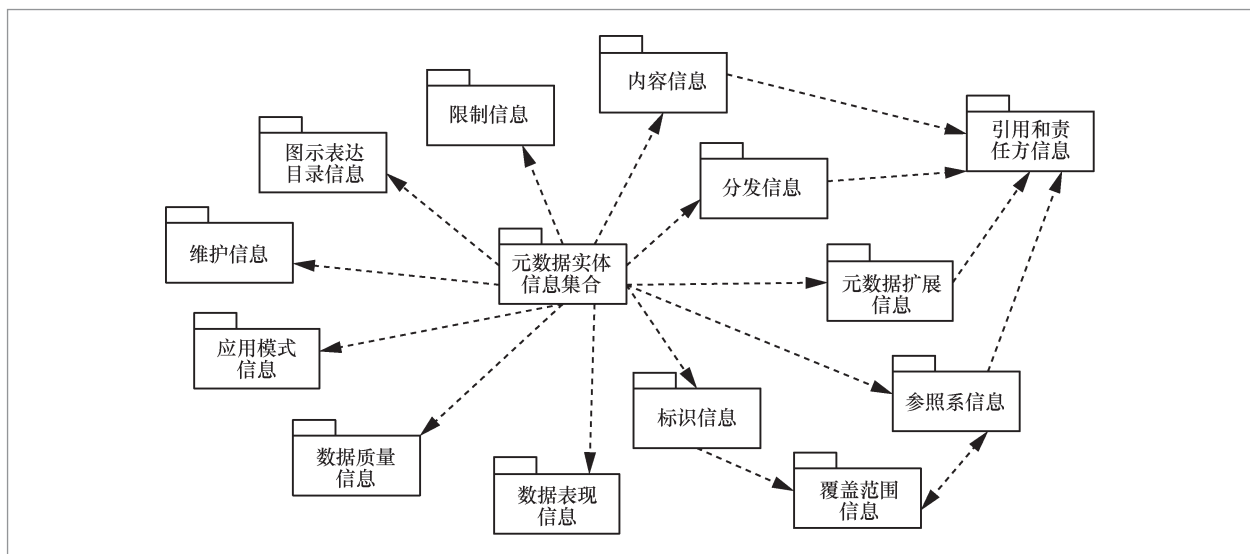


图2 元数据内容框架

### 3.2 数据汇交与加工

#### (1) 数据汇交管理

数据中心作为数据管理机构对汇交数据进行管理与加工维护。农业科学数据汇交内容包括汇交方案、质量自查报告和科学数据。汇交方案应明确以下内容：数据汇交义务人、数据的种类和范围、数据产生方式、数据格式、数据管理机构、数据质量说明、汇交形式和进度、数据的科学价值和使用领域、数据保护期限和其他说明事项。项目承担单位对数据的真实性、完整性、一致性进行自查后提交质量自查报告。科学数据是项目数据汇交的核心，应包含实体数据、数据描述信息和辅助工具软件。国家农业科学数据中心配备专门的数据保护和管理（保管）人员，采取现代化的手段保存数据，保证汇交数据的安全；同时积极创造条件，保证农业科学数据的合理利用，推动数据共享。数据中心对汇交的科学数据进行分类、分级存储和管理，确保数据的物理安全。数据中心在数据验收后及时公布项目汇交科学数据元数据，在保护项目承担单位合法权益的基础上，做好数据共享和服务工作。

#### (2) 数据加工

数据中心保存的原始数据不能直接在网络上共享，还需要经过一定的加工处理。在元数据层面，对照科学数据元数据标准，补全元数据的必选项，对于可选项，则根据学科领域要求进行适当的扩展或删除。例如：①元数据不全问题，常见于缺少描述信息、地址信息、邮编地址信息等字段，需要进行补全；②实体数据格式问题，常见于格式错误、格式可读性差（以PDF报告提供数据、以图片形式提供表格数据）等，需要进行修正、识别和提取；③实体数据字段问题，常见于字段定义不

规范、字母字段缺少含义解释等，需要进行修改和标注；④数据可用性不足，常见于提交的论文、报告、证书等不可用，需要进行修改。

数据中心还需为科学数据添加数据身份标识。科技资源标识是科技资源实体唯一的身份编码，例如：海南岛热带作物种质资源考察库的科技资源标识为CSTR:17058.11.E0015.20210616.00.ds.0385。其中，CSTR为中国科技资源代号，17058为国家农业科学数据中心代码，11表示该资源类型为科学数据，E0015表示数据生产者所在单位为中国热带农业科学院科技信息研究所，20210616表示数据创建日期，00两个数字分别表示数据来源为调查、数据类型为数值型，ds表示该资源为数据集合，最后4位为数据流水编码，其结构如图3所示。

对于数据实体，需进行一致性检测（生成MD5码）和智能分类、融汇。例如：①基于同一种作物，对不同项目产生的汇交数据中涉及本作物的内容进行提取、标注和统一量度，并进行跨数据集融合，以构建针对本作物的数据专题；②基于某地理区域进行坐标计算后，针对多个数据集，对涉及本地地理区域范围的数据进行提取，

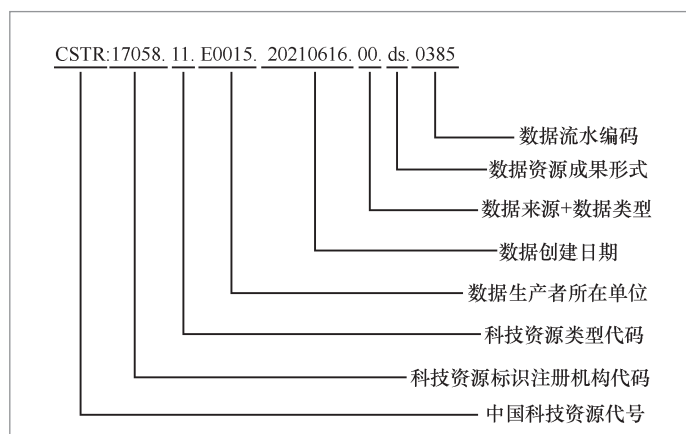


图3 农业科技资源标识符结构示意图

并进行跨数据集融合,以构建针对本区域的数据专题。

### 3.3 数据长期保存

#### (1) 数据分类

数据资源长期保存是各国数据平台高度关注的战略问题,数据分类是长期保存的重要环节。数据中心对现有农业科学数据资源进行调研,然后结合科学数据用户需求,对农业领域的数据资源目录信息进行重新编制,归并总结出14大类58小类的全新资源分类体系(见表1)。新的资源分类体系不仅使科学数据资源分类更加明晰,也方便用户进行索引和检索。由于农业科研和生产高速发展,数据资源分类体系需要随时进行微调以适应新的需求。

#### (2) 数据分级

由于农业科学数据包含的学科领域众多,涉及粮食安全、种子安全、生物安全、

经济安全等方面,必须根据数据的重要程度和涉及的安全问题进行分级管理。数据安全分级指按照数据遭受破坏后造成的影响进行安全等级划分,以达到对不同安全等级的数据实施不同安全防护的目的。农业科学数据分级根据数据遭受破坏后造成的影响等从高到低分为5级、4级、3级、2级、1级等指导性的分级初始值,各级判断准则如下。

- 5级数据判断准则:遭受破坏后,对国家安全产生较大影响的农业数据,通常包括地形地貌、遥感影像、气候资源等;数据安全性遭到破坏后,对公众权益或农业企业利益造成严重影响的数据,如科技成果、转基因库等。

- 4级数据判断准则:数据遭到破坏后,对公众权益造成一般影响,或对个人隐私或农业企业的合法权益造成严重影响,但不影响国家安全,如农业科研项目投资、农业金融与投资等。

表1 农业科学数据资源体系

一级分类	二级分类
作物科学	作物种质资源、作物遗传育种、作物耕作与栽培、作物分子生物学
动物科学与动物医学	动物种质资源、动物营养与饲养、动物医学、动物分子生物学
热带作物科学	热带作物种质资源、热带作物遗传育种、热带作物耕作与栽培、热带作物分子生物学、渔业分子生物学
渔业科学	渔业种质资源、渔业遗传育种、渔业水域资源环境、水产养殖、海洋与渔业捕捞
草地与草业科学	品种资源与遗传育种、草地资源、草业生产、草地监测
农业资源与环境科学	农业气象学、农业土壤学、农业土地利用、植物营养与肥料学、农业水资源学、农业生态与环境学、农业遥感、农业规划与生产布局
农业微生物科学	微生物资源、饲用微生物、加工微生物、厌氧微生物、食用微生物
食品营养与加工科学	食品加工、营养品质、标准检测、贮藏保鲜
植物保护科学	植物病虫害、生物防治、杂草鼠害
农业工程	农业机械、灌溉工程
农业农村经济科学	生产经济、贸易价格、市场经济
农业科技基础	科技成果、科技机构、科技人才、科技项目
果树科学	园艺科学、果树病害、果树虫害、果品加工
生物安全	基因安全、物种安全、动植物疫病

● 3级数据判断准则：数据用于部分场景，一般针对特定人员公开，且仅被必须知悉的对象访问或使用，如产品追溯、产地追溯等；数据遭到破坏或数据安全性遭到破坏后，对公众权益造成轻微影响，或对个人隐私或农业企业的合法权益造成一般影响，但不影响国家安全，如种质资源等。

● 2级数据判断准则：只对部分受限用户公开，通常指内部管理且不宜广泛公开的数据，如农业区划等；数据的安全性遭到破坏后，对个人隐私或农业企业的合法权益造成轻微影响，但对国家安全、公众权益的影响极小，如农产品质量追溯等。

● 1级数据判断准则：数据一般可被公开或可被公众获知、使用，如组织机构等；农业组织或农业科学数据管理者主动公开的信息，如生产许可等。数据遭到破坏或数据安全性遭到破坏后，可能不会对个人隐私或农业企业的合法权益造成影响，或仅造成微弱影响，但不影响国家安全、公众权益，如商品信息等。

数据长期保存需要使用质量较好的存

储介质，以延长数据保存时间。不管存储在什么介质里，都需要对数据进行定期检查、修复，实现定期转存。备份方案根据数据安全级别进行区分，重要数据至少保存3份。数据保存使用标准的、可互相兼容的或开放的、无损的数据格式。如文本文件应选择ODF格式，表格文件应选择ASCII格式，视频文件应选择MPEG-4格式，图片文件应选择TIFF或JPEG2000格式，网页应选择XML或PDF格式。

### 3.4 数据共享服务

科学数据共享指科学数据不受其拥有单位的限制，可以在更大范围内被利用的一种业务合作与共享方式。数据中心通过建设科学数据共享平台，实现数据的长期保存，帮助科技工作者有效地管理数据、统一数据的引用标识符、提高数据的可发现性。国家农业科学数据中心平台架构如图4所示。

国家农业科学数据中心门户网站是农

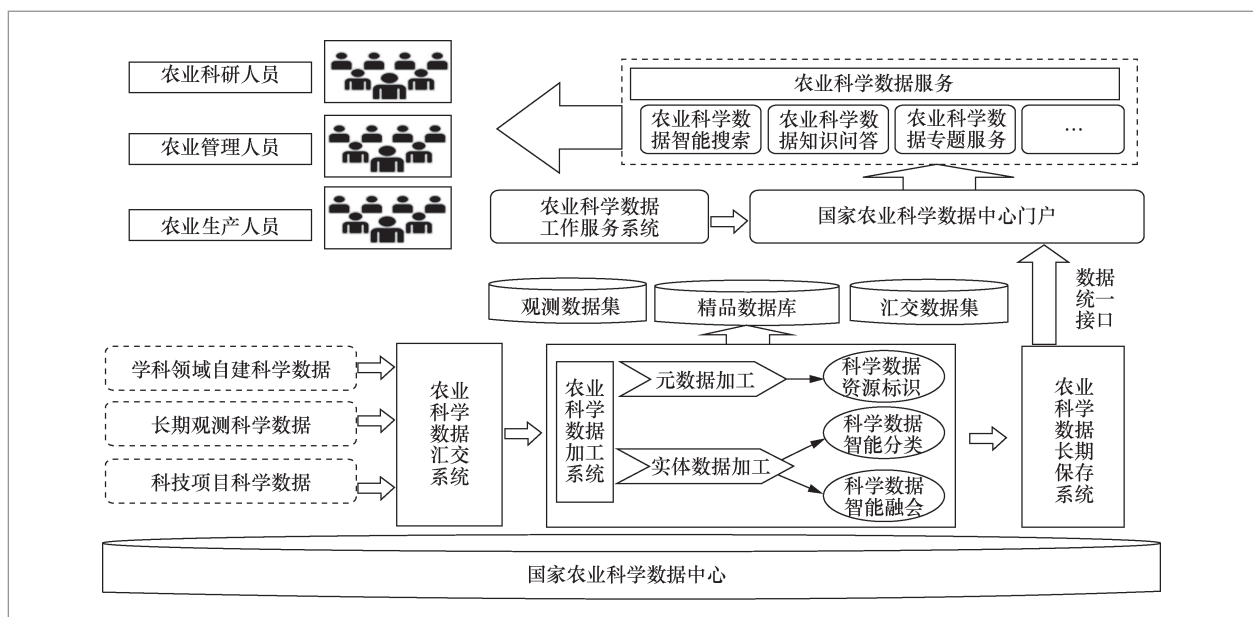


图4 国家农业科学数据中心平台架构

业科学数据展示、查询、宣传的总平台，用户交互的总门户，具有数据汇交、资源浏览、智能搜索、知识问答、专题服务等多项功能。其可向全社会广大用户提供高效、便捷的农业科学数据资源目录和数据实体服务。

数据平台建有4个应用系统：农业科学数据汇交系统、农业科学数据加工系统、农业科学数据长期保存系统和农业科学数据工作服务系统。农业科学数据汇交系统主要对科技项目科学数据、长期观测科学数据和学科领域自建科学数据三大类数据进行收集，实现科学数据汇交计划和汇交内容的提交、审核、跟进、反馈和审批。农业科学数据加工系统对原始数据的元数据和实体数据进行加工，添加科学数据资源标识，进行科学数据智能分类、融汇。加工后的数据被保存在农业科学数据长期保存系统，由数据中心进行管理维护。农业科学数据工作服务系统汇聚了各学科领域数据中心信息，可为中心工作提供便捷的统计、监测和管理的窗口。

农业科学专题数据库由加工系统处理后形成的观测数据集、精品数据集、汇交数据集组成，是中心为用户提供的主要数据产品。用户可以通过平台门户获取需要的专题科研数据。

## 4 开发应用

### 4.1 汇交系统

国家农业科学数据中心开发了农业科学数据汇交系统，用于农业科学数据汇交计划和数据实体的提交、审核、跟进、反馈。本系统有五大功能模块：系统首页、科学数据汇交计划、自查质量信息报告、科学数据汇交内容、基本信息管理，如图5所示。

- 系统首页：提醒用户需要办理的业务，并及时了解已办业务的状态。
- 科学数据汇交计划：包含更新任务书，在线填写和导入科学数据汇交计划。
- 自查质量信息报告：用于对科学数据汇交内容进行自查。
- 科学数据汇交内容：用于填写科学数据描述信息，上传汇总数据集、使用软件。
- 基本信息管理：可以修改密码、单位名称、联系人、联系电话、电子邮箱等基本信息。

### 4.2 加工系统

国家农业科学数据中心开发了农业科



图5 农业科学数据汇交系统

学数据加工系统(如图6所示),根据统一的格式,对科技计划项目科学数据汇交审核系统、长期性数据汇交系统、总中心门户、分中心门户、实验站门户等其他系统收集的数据资源进行加工处理,以满足数据共享的规范及要求。系统功能主要包含:原始数据查看、元数据加工、数据审核、数据分布、加工任务分配、用户管理、个人信息管理等。

### 4.3 长期保存系统

农业科学数据长期保存系统实现了PB级农业科学数据对象的长期安全存储,由国家科技计划项目形成的科学数据支撑数据的长期保存提供一流的仓储。农业科学数据长期保存系统整体分为前后端两部分:后端保存系统、前端展示系统。后端保存系统主要用于与其他业务系统进行后台数据同步,不直接对外提供服务;前端展示系统主要用于对外展示保存系统内的元数据信息,提供简单的检索和查看服务,也提供可供其他系统调用的数据接口(如图7所示)。系统主界面展示了4项统计信息和几条最近更新的资源。点击资源名称可跳转到对应的资源详情页查看资源元数据信息。



图7 农业科学数据长期保存系统

### 4.4 数据共享服务

国家农业科学数据中心门户网站提供丰富的数据服务,除了基础的数据检索服务,还包括参考咨询服务、数据挖掘分析、数据配套工具导航、数据库(集)收录认证与查询、用户卡、用户反馈等,如图8所示。

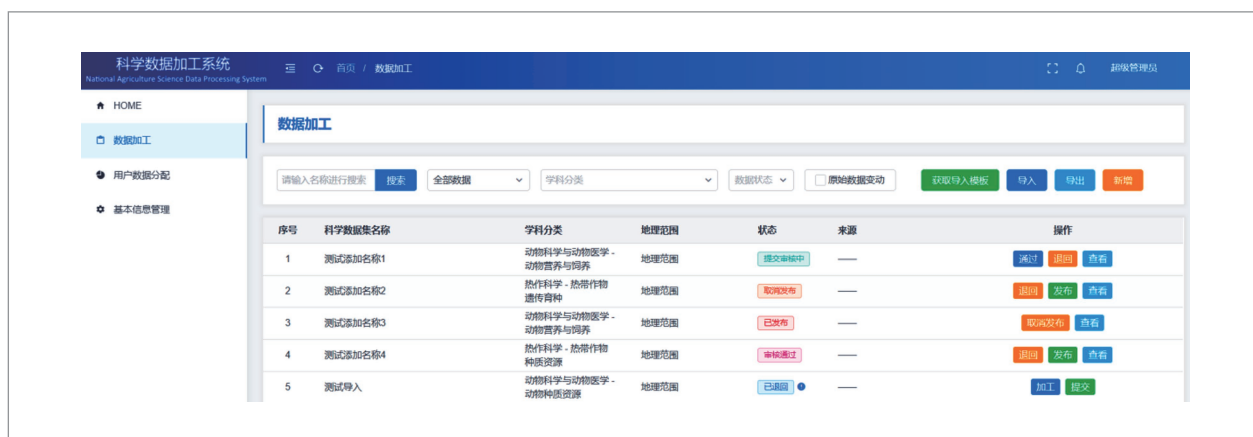


图6 数据加工示例

### 参考咨询服务

**选择分类**

- 全部分类
- 作物科学
- 动物科学与动物医学
- 植物科学
- 渔业科学
- 草地与畜牧业科学
- 农业资源与环境科学
- 植物保护科学
- 农业微生物科学
- 食品营养与加工科学
- 农业工程
- 农业科技科学
- 农业科技基础
- 农林科学
- 生物安全

**当前选择分类:**

请输入需求关键字, 多个关键字间请以逗号隔开

数据需求详细说明

电子邮箱:  联系电话:  手机/微信:

是否愿意公开:  是  否

### 数据挖掘分析

**服务流程及说明**

1. 用户可以根据自己实际需要, 发起数据挖掘分析服务请求, 需按要求填写服务单信息。
2. 国家农业科学数据中心收到服务请求后, 服务人员根据服务请求的先后顺序来联系用户进行沟通。
3. 服务人员沟通了解具体的需求后, 会安排相应的数据工程师进行数据分析服务。
4. 数据工程师会在约定的时间进行数据分析, 并把结果发送用户进行验收。
5. 用户对解决的问题进行验收。
6. 服务结束后请您对服务进行评价, 以便我们进行持续的完善与提升。

### 数据配套工具导航

● 本站提供了数据配套工具主页导航, 请用户遵照相关许可协议使用。

数据库工具

MySQL数据库

数据库工具

MongoDB

数据库工具

Kettle

### 用户卡

1. 用户卡是什么?  
用户卡是注册用户申请国家农业科学数据中心提供数据服务的电子凭证。
2. 如何申请用户卡?  
申请用户卡的过程非常简单, 包括如下3个步骤:  
(1) 注册成为国家农业科学数据中心的用户;  
(2) 登录到个人中心, 完善自己的姓名、单位、专业等认证信息, 并提交用户卡申请;  
(3) 国家农业科学数据中心人工审核用户申请材料, 确认申请个人信息无误后, 发放用户卡。
3. 用户卡有什么用?  
国家农业科学数据中心将为用户卡用户提供更加精准的信息:  
(1) 申请下载国家农业科学数据中心的实体数据;  
(2) 提供更加符合自己的专业背景的检索结果;  
(3) 订阅信息用户会不定期收到相关数据集的推荐信息;  
(4) 优先受邀参加国家农业科学数据中心举办的线下活动。

### 数据库(集)收录认证

申请类型:  个人  单位

申请人:

联系电话:

电子邮箱:

数据库(集):

认证凭证:  请上传身份证照片; 支持文件格式: 图片; 大小限制: 5M

申请说明:

### 数据库(集)收录查询

请输入数据库(集)名称关键字  请输入资源标识符

已有98个数据库(集)完成收录认证

### 用户反馈

您可在评论区对国家农业科学数据中心建设的发展和改进, 数据管理, 产品和数据工具的优化等。

标题:

反馈内容:

邮箱:

图8 数据共享服务

## 5 结束语

数据管理的目的是整合数据, 促进数据重用, 引导知识发现和创新<sup>[7]</sup>。良好的数据管理可以提高数字出版物的质量, 简化数据发现、评估、重用的过程。英国数据档案(UK data archive, UKDA)作为数据整合与重用研究实践的先驱, 针对数据收集、数据清理、数据录入、数据保存、数据访问建立了一系列管理标准。澳大利亚统

计局(Australian Bureau of Statistics, ABS)与政府、研究机构和企业合作, 整合社会、经济和环境数据集, 并构建了5个安全框架进行数据管理。科学数据管理是研究热点, 我国学者在借鉴国外先进经验的基础上, 结合学科领域数据共享需求, 提出了科学数据管理政策建议<sup>[8-14]</sup>。农业科学数据与农业科技活动紧密相关, 数据采集、保存、利用都需具备一定的专业素养, 因此需要数据管理机构(科学数据中心)进行管理。本文从数据管理主体的角度对农业科学数据管理模式进行研究, 构建了农

业科学数据生命周期管理模型,对各数据阶段的任务、机制进行解析,并根据理论研究进行了初步应用开发。但通过计算机发现、访问、集成和分析与任务相关的科学数据仍是科学数据管理的一大挑战,本文对数据智能管理算法与实现方面的探索较少,还需进一步研究。

## 参考文献:

- [1] 赵瑞雪, 赵华, 朱亮. 国内外农业科学大数据建设与共享进展[J]. 农业大数据学报, 2019, 1(1): 24-37.  
ZHAO R X, ZHAO H, ZHU L. Progress in the development and sharing of big data in agricultural science between China and foreign countries[J]. Journal of Agricultural Big Data, 2019, 1(1): 24-37.
- [2] 杨传汶, 徐坤. 基于生命周期的动态科学数据服务模式研究[J]. 图书馆论坛, 2015, 35(10): 82-87.  
YANG C W, XU K. Study of dynamic scientific data service model based on life cycle[J]. Library Tribune, 2015, 35(10): 82-87.
- [3] 储节旺, 夏莉. 嵌入生命周期理论的科学数据管理体系构建研究——牛津大学为例[J]. 现代情报, 2020, 40(10): 34-42.  
CHU J W, XIA L. Research on scientific Data management construction based on life cycle theory[J]. Journal of Modern Information, 2020, 40(10): 34-42.
- [4] 夏义堃, 管茜. 基于生命周期的生命科学数据质量控制体系研究[J]. 图书与情报, 2021(3): 23-34.  
XIA Y K, GUAN Q. Biological science data quality control system based on life cycle[J]. Library & Information, 2021(3): 23-34.
- [5] 陈欣, 詹建军, 叶春森, 等. 基于高校科学数据生命周期的社会科学数据特征研究[J]. 情报科学, 2021, 39(2): 86-95.  
CHEN X, ZHAN J J, YE C S, et al. Research on features of social scientific data based on life cycle of university scientific data[J]. Information Science, 2021, 39(2): 86-95.
- [6] 姚占雷, 谷俊, 许鑫. 全生命周期视域下人文社科研究数据管理平台的设计与实现[J]. 图书情报工作, 2021, 65(7): 25-37.  
YAO Z L, GU J, XU X. Design and implementation of management platform for humanities and social sciences research data in the perspective of full life cycle[J]. Library and Information Service, 2021, 65(7): 25-37.
- [7] WILKINSON M D, DUMONTIER M, AALBERSBERG I J, et al. The FAIR guiding principles for scientific data management and stewardship[J]. Scientific Data, 2016, 3: 160018.
- [8] 邢文明. 我国科研数据管理与共享政策保障研究[D]. 武汉: 武汉大学, 2014.  
XING W M. Research on policy guarantee of scientific data management and sharing in China[D]. Wuhan: Wuhan University, 2014.
- [9] 陈丽君. 基于生命周期模型的科学数据服务研究[J]. 图书馆研究与工作, 2016(3): 16-19.  
CHEN L J. Study of scientific data service based on life cycle model[J]. Library Science Research & Work, 2016(3): 16-19.
- [10] 魏悦, 刘桂锋. 基于数据生命周期的国外高校科学数据管理与共享政策分析[J]. 情报杂志, 2017, 36(5): 153-158.  
WEI Y, LIU G F. Analysis of research data management and sharing policy in foreign universities based on data lifecycle[J]. Journal of Intelligence, 2017, 36(5): 153-158.
- [11] 黄铭瑞, 李国庆, 李静, 等. 国家科学数据中心管理模式的国际对比研究[J]. 农业大数据学报, 2019, 1(4): 14-29.  
HUANG M R, LI G Q, LI J, et al. International comparative study on management mode of national science data center[J]. Journal of Agricultural Big Data, 2019, 1(4): 14-29.
- [12] 武彤. 基于数据生命周期的美国研究图书馆科学数据开放共享服务研究[J]. 图书与情报,

2019(1): 135-144.

WU T. Study of American academic library's open sharing service based on data life cycle[J]. Library & Information, 2019(1): 135-144.

- [13] 杨乐, 颜石磊, 李洪波. 科研数据生命周期研究和数据知识库理论架构[J]. 图书情报工作, 2019, 63(1): 91-97.

YANG L, YAN S L, LI H B. Theoretical study of research data lifecycle and data

repository[J]. Library and Information Service, 2019, 63(1): 91-97.

- [14] 崔磊. 农业大数据建设的需求、模式与单品种全产业链路推进路径[J]. 大数据, 2019, 5(5): 100-108.

CUI L. Demand and model of agricultural big data construction in China and the way to promote the whole industry chain[J]. Big Data Research, 2019, 5(5): 100-108.

#### 作者简介



高飞(1988-), 女, 博士, 中国农业科学院作物科学研究所助理研究员, 主要研究方向为科学数据管理。



周国民(1969-), 男, 博士, 中国农业科学院农田灌溉研究所所长、研究员, 国家农业科学数据中心主任, 主要研究方向为农业信息化、农业科学数据管理。



满芮(1985-), 女, 博士, 中国农业科学院农业信息研究所助理研究员, 主要研究方向科学数据管理。

收稿日期: 2021-10-29

通信作者: 高飞, gaofei20120214@163.com

基金项目: 农业农村部行业标准制定项目(No.NYB-21071)

**Foundation Item:** Industry Standard Formulation Project of Ministry of Agriculture and Rural Affairs (No.NYB-21071)