

《农业科学数据在线分析挖掘技术规范》 编制说明

一、目的和意义

农业科学数据是支撑国家农业科技创新和农业农村经济发展的重要基础性、战略性资源。随着人工智能、大数据为代表的数字时代的来临，数据已作为重要生产要素之一，被正式纳入国家所定义的要素市场化配置之中，蕴含重大意义。同时，数据资源已成为与粮食、能源同等重要的战略资源，数据驱动的研究模式给传统科学研究领域带来了新的研究方法 with 科学研究范式，推动以农业科学数据为重要内核的农业农村大数据科技创新，也为农业农村现代化发展带来了新的机遇。

随着数据驱动科学研究范式的发展，农业科学数据在科技创新中的作用越来越突出，随之而来的是，农业科学数据分析挖掘和应用的方法与技术研究，以及相关工具软件的研发，受到了科技界和产业界的高度重视。国外欧美发达国家在农业领域科学数据挖掘技术和软件研究与应用上积累雄厚，总体处于领先地位，国内在农业领域科学数据挖掘技术和软件研究与应用上进步迅速，但总体处于追赶的状态。

场景驱动的农业科学数据挖掘方法、技术、系统的研究与应用是当前的热点，并具有良好的发展前景。一是，从单一尺度到大数据尺度的跨尺度挖掘。随着技术发展，尤其是随着基因组、表型组、遥感技术的发展，不同尺度的数据不断积累，针对不同场景的不同尺度数据，诞生了不同挖掘工具和挖掘方法，形成了跨尺度挖掘的发展趋势。二是，从

单模型到多种模型组合的跨模型分析。农业科学数据是多维复杂数据，不同方向具有不同学科背景的模式，单一模型已经不足以满足数据分析和挖掘的需求。如作物表型数据挖掘中，科学家大多开始使用多种模型相互组合，首先使用图像检测模型寻找定位作物表型器官，然后对相应的器官进行图像识别和计数，通过图像分割、精准测量、智能评估等多种模型进行处理，这些模型的组合使用，表明农业科学数据的挖掘已经迈入跨模型分析阶段。

随着数据驱动科学研究范式的发展，农业科学数据在科技创新中的作用越来越突出，随之而来的是农业科学数据分析挖掘和应用方法与技术研究也不断发展。围绕农业科学数据分析挖掘还存在数据语义孤岛严重，以及数据挖掘工具不全、不配套与场景适应性差等突出问题。

本技术标准的作用体现在农业科学数据在线分析挖掘技术领域的规范化、标准化，突破“数据资源-分析工具-应用场景”衔接不畅的问题，形成集数据资源、分析模型、组件工具、场景分析和标准流程于一体的在线分析挖掘应用环境，支撑从“数据聚合--挖掘分析链--在线分析--场景应用”的农业科学数据在线分析挖掘全过程，实现不同场景分析应用的在线计算分析。

1、有利于农业科学数据标准化和一致性：技术规范定义了数据的格式、编码和标准，确保不同来源的数据可以无缝集成和比较。这有助于消除数据孤岛，提高数据的互操作性和一致性。

2、有利于农业科学数据共享和集成：技术规范定义了数据共享的协议和标准，促进不同机构和用户之间的数据交换和共享。这有助于打

破数据壁垒，实现数据的广泛利用。

3、规范农业科学数据在线分析挖掘的算法和模型：规范了一系列适用于农业科学数据的分析和挖掘算法，如场景管理、在线分析、自动化挖掘等，帮助研究人员和用户选择合适的工具和技术。

4、推动农业科学进步：通过制定技术标准，促进了对农业科学数据在线分析挖掘的研究和了解。这有助于推动农业科学的进步，为未来的农业科学数据共享和研究提供更多知识和方法。

二、依据与任务来源

农业科学数据是支撑国家农业科技创新和农业农村经济发展的重要基础性、战略性资源。农业科学数据分析挖掘是推动现代农业发展的重要手段，我国高度重视这一领域的发展，并出台了一系列政策措施来支持和促进农业科学数据的采集、分析和应用。

在国家战略规划方面。国务院办公厅关于印发《科学数据管理办法的通知》国办发〔2018〕17号提出进一步加强和规范科学数据管理，加强科学数据的分级分类、加工整理和分析挖掘，更好支撑国家科技创新、经济社会发展和国家安全。《国民经济和社会发展第十四个五年规划和2035年远景目标纲要》明确提出要加快数字化发展，建设数字中国，其中包括推进农业数字化转型，加强农业科学数据的采集和分析。《数字乡村发展战略纲要》：强调利用大数据、物联网、人工智能等技术，推动农业现代化，提升农村信息化水平。

在农业农村部专项政策文件方面。《农业农村部关于推进农业农村大数据发展的实施意见》提出到2025年，实现农业产业链、价值链、

供应链的联通，大幅提升农业生产智能化、经营网络化、管理高效化、服务便捷化的能力和水平。《国家农业科学观测工作管理办法（试行）》：加强和规范农业科学观测工作，为推动农业科技创新提供基础支撑，为农业农村绿色发展和管理决策提供科学依据。

在农业数据共享与开放方面。《国家农业科学数据总中心管理办法》提出推动农业科学数据的共享和开放，建立统一的数据汇交、存储和管理平台，促进数据的互联互通。《农业科学数据汇交管理办法》规范农业科学数据的汇交工作，确保数据的完整性和一致性，实现数据共享和利用。

三、起草过程

1、收集国内外相关标准、资料

目前有 30 多项国内与农业大数据在线挖掘相关的国家标准、行业标准以及国外标准主要包括：（1）YD/T 3762-2020 《大数据 数据挖掘平台技术要求与测试方法》提供了数据挖掘平台产品的设计、研发、测试、评估和验收、科学大数据系统的测试。（2）YDT 3760-2020 《大数据 数据管理平台技术要求与测试方法》本标准提供了适用于大数据数据管理平台产品的研发、测试、评估和验收等方面的建议和指导。（3）NY/T 4061-2021 《农业大数据核心元数据》本标准提供了适用于农业大数据采集、存储、处理以及信息共享和发布活动中的数据管理。（4）ISO/IEC TR 20547-1 《Big Data Reference Architecture》为大数据的参考架构提供一个通用的框架和应用指南，帮助组织理解和实施大数据解决方案。

2、标准编写情况

编写组在标准编制前，按照标准编制规范，制定了标准制定工作计划，明确了各阶段的任务与目标，根据查阅资料，以及调研走访相关部门以及相关农业大数据企业及科研院所，编写组共同商议征求意见稿内容。

3、征求意见及审查会

2024年 月，全面征集各方意见，汇总整理，工作组全方位征求相关企业、单位、行业主管部门、专家以及企业的意见并进行修改。对标准草案征求意见进行集中处理，对内容进行了讨论和修改，明确标准制定中农业科学数据在线分析挖掘技术的数据、算子组件、分析挖掘等规范以及数据、算子组件、核心能力、分析挖掘能力、性能等主要板块标准。

四、编制原则和主要技术内容确定

1、编制原则

总体把握实用、普适、兼容的原则，同时遵循保住底线并适当超前总体要求，对经济、环保、安全等重要指标项进行量化。编制本标准遵循如下原则：

(1) 本标准的编制遵循 GB/T 1.1-2020《标准化工作导则 第一部分：标准化文件的结构和起草规则》；

(2) 在收集国内农业大数据相关标准基础上，广泛征求企业、管理部门的意见，结合农业大数据实际情况，确定标准内容；

(3) 通过调研与验证，使制定的标准做到系统性、科学性、实用

性、协调性和可操作性。

2、主要技术内容确定

本文件规定了农业科学数据在线分析挖掘技术的数据、算子组件、分析挖掘等规范以及数据、算子组件、核心能力、分析挖掘能力、性能等测试标准的要求。

本文件适用于农业科学数据在线分析挖掘平台的建设。

(1) 术语和定义

本标准明确规定了农业科学数据在线分析挖掘的数据管理、算子管理、模型分析、场景分析的术语和定义。

(2) 数据管理

本标准明确规定了农业科学数据在线分析挖掘技术应具备数据源管理的功能，包括：数据领域、数据格式、数据大小、数据描述、数据上传、数据下载、数据修改、数据删除。

(3) 算子组件管理

本标准明确规定了农业科学数据在线分析挖掘技术应具备算子组件管理的功能，包括：算子组件类别、算子组件格式、算子组件大小、算子组件描述、算子组件封装、算子组件上传、算子组件下载、算子组件修改、算子组件删除。

(4) 分析挖掘

本标准明确规定了农业科学数据在线分析挖掘技术应具备分析挖掘管理的功能，包括：数据调用、算子组件调用、在线分析引擎调用、算子分析流程、场景分析流程。

(5) 农业科学数据在线分析挖掘技术数据及算子组件管理能力测试标准

本标准明确规定了农业科学数据在线分析挖掘平台各类数据管理的测试方法，包括数据上传、数据下载、数据修改、数据删除以及算子组件的上传、下载、修改和删除管理。

(6) 农业科学数据在线分析挖掘技术核心能力测试标准

本标准明确规定了农业科学数据在线分析挖掘平台各类算子组件管理的测试方法，包括统计分析类算子组件、机器学习类算子组件、深度学习类算子组件和专业类算子组件。

(7) 农业科学数据在线分析挖掘技术平台分析挖掘能力测试标准

本标准明确规定了农业科学数据在线分析挖掘平台分析挖掘能力的测试方法，包括数据调用、算子组件调用和在线分析引擎调用。

《农业科学数据在线分析挖掘技术规范》团体标准起草工作组

2024. 12. 22