

第三批“数据要素×”典型案例 ——科技创新领域

科技创新领域案例之一

青藏高原科学大数据 助力地球系统研究与高原绿色发展

青藏高原被誉为“世界屋脊”和“亚洲水塔”，是全球气候变化的敏感区和生态安全屏障。然而，青藏高原科学数据治理方面仍面临多重挑战，数据稀疏异构与碎片化严重、观测网络薄弱以及数据流通不顺畅制约了科学研究的深度。中国科学院青藏高原研究所创新性地构建了“数据驱动、技术赋能、机制创新”三位一体的解决方案，探索大数据与人工智能在地球科学中的应用，助力地球系统科学研究。

一是构建空-天-地一体化观测体系，实现极端环境数据的全面采集与汇聚

构建卫星、无人机、野外站等多尺度观测系统，研发仿生鹰眼设备，实现雅江流域、拉鲁湿地等区域亿级像素超高清影像的采集与监测；搭建高寒区物联网监测信息系统，以满足极端环境下的观测需求，保障数据可靠传输，实现多源数据的有效汇聚。截至 2025 年 3 月，已成功整合青藏高原及周边地区的 7200 多个科学数据集，总数据量超过 500TB。

二是发展数据智能制造能力，研发高质量数据产品，驱动地球系统研究新范式

将人工智能方法引入青藏高原跨圈层数据整合，显著提升了高原气候、冻土环境、冰缘地貌等关键要素数据的空间连续性和

时间分辨率，实现从传统点状、断面观测到广域、长时序跨越，形成自动化、高精度、可扩展的数据智能制造能力，研发青藏高原高质量数据产品超过 100 个，驱动地球系统建模与精细模拟，支持数据-模型-决策全链条应用。

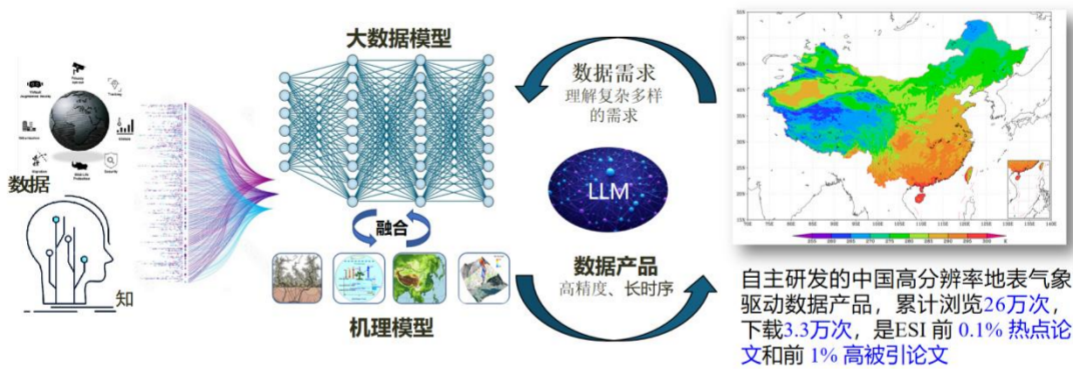


图 数据智造能力

三是成为国际科学数据仓储，推动开放共享，服务国家重大工程与高原绿色发展

秉持 FAIR 原则，成为 Springer Nature、AGU、ESSD 等认可的国际科学数据仓储中心，实现科研数据成果的广泛共享，保障数据价值最大化；优质数据产品广泛应用于重大工程和区域管理，包括第二次青藏科考、青藏高原生态保护等国家重大工程，多项政策建议被国家和地方采纳并获领导批示。目前，本项目已服务全球 63 个国家和地区数十万用户，年数据下载量达 1.8PB，与 WMO（世界气象组织）、ICIMOD（国际山地综合发展中心）等合作，推动全球科研协作。

数据赋能林学和森林生态学科研范式变革

森林是巨大的碳汇，每年抵消人类排放 CO₂ 的 28%，是实现碳中和的关键。传统林学和森林生态学研究面临数据采集管理方式落后、野外调查效率低、关键模型缺失等难题，限制了森林结构监测、碳汇计量和过程解析，成为该领域知识创新的瓶颈。中国科学院沈阳应用生态研究所联合中国科学院计算机网络信息中心，依托院重大科研装置——科尔塔群，融合物联网、近地面遥感、人工智能和大数据等信息技术，打造数据采集、自动传输、智能分析和决策全链条平台，建立森林结构、碳汇多要素数据库，应用于林学/森林生态学研究，赋能科研范式转型。

一是建立“天-空-塔-地”一体化观测体系，改变传统研究数据的获取方式，实现信息化采集、传输与多源异构数据库构建

以激光雷达为核心、以近地面遥感（无人机+塔基+地基）为主要平台，创立“天-空-塔-地”一体化观测研究体系与数据融合方法。通过建设 4G/5G 基站，铺设万兆光纤，把科尔塔群、水文站和科研样地内的仪器设备采集到的数据汇聚到数据中心。建立森林结构、碳汇多源异构数据库，涵盖样地调查、碳通量数据、激光雷达点云、无人机遥感等多种森林碳汇数据，夯实林学和森林生态学研究的数据基础。

二是利用“数字模拟+人工智能”技术，变革复杂地形下森林碳汇计量模型，提高森林碳汇计量精度

在数据赋能的支持下，模拟森林边界层气象过程，采用智能数据质量控制方法，优化编程算法，自动、智能、准确、集成地完成通量数据质量控制，实时获取高质量的碳通量数据。提出新地形复杂度因子，建立基于“多塔协同-人工智能”的“森林碳汇计量模型簇”，变革复杂地形下森林碳汇计量模型，开发人工智能算法的计算程序，形成了可验证、可核算、可共享的森林碳汇数据产品，搭建森林碳-水通量智能监测子平台，服务森林碳汇核算。

三是运用“大数据+人工智能”手段，促进林学和森林生态学的知识创新，服务森林增汇管理

开发“森林碳汇功能三维结构信息化监测与展示系统”，实现森林单木-林分-流域-区域多维度结构的自动分析与管理，单次作业人工耗时从传统方法的2—3年缩短至约2周，并在东北森林应用。运用大数据与人工智能方法，促进了树种叶片功能性状、耐荫性规律、森林演替和反照率等方面的科学发现；提出诱导人工纯林成为混交林，增加人工林降温-固碳协同效应等管理建议，形成了数据赋能推动的林业新质生产力，促进林学和森林生态学的知识创新。

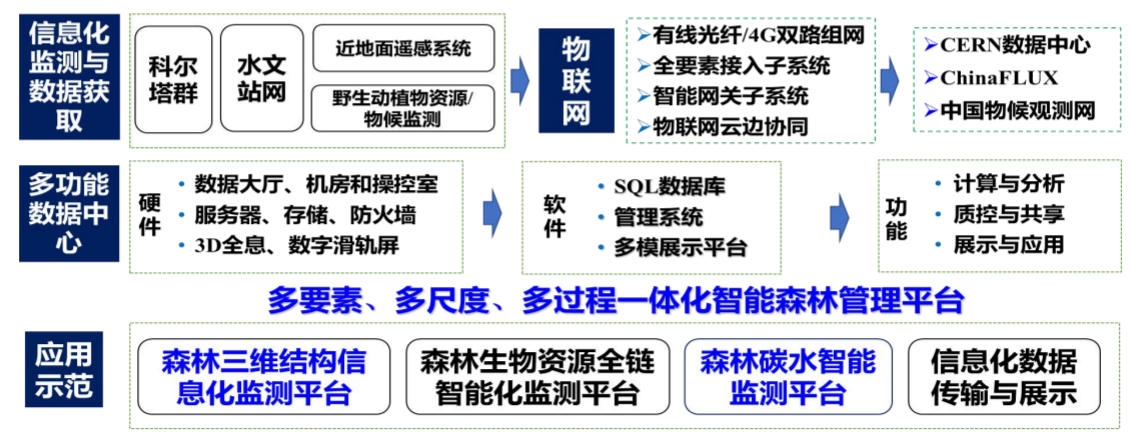


图 一体化森林管理平台

生物制造数据融合赋能产业创新发展

我国生物资源丰富，企业需求旺盛，但生物制造相关数据资源分散、标准不统一、共享机制不畅，制约产业进一步发展。中国科学院微生物研究所联合中国科学院天津工业生物技术研究所以共同搭建生物制造科学数据共享服务平台，通过整合利用微生物菌株数据、组学数据、基础分子（元件、酶等）数据、反应途径数据、文献专利数据、仪器产出数据等资源，突破产业发展中工业菌株培植、关键技术、催化效率、成本等方面的瓶颈。

一是构建数据库，强化数据资源整合

通过国家科技计划项目数据汇交等途径获取，经统一质量控制整合，构建菌种资源及组学数据库、元件数据库、工业菌种数据库、精密发酵库和产品库等多个专业数据库，构建起覆盖生物制造全生命周期的数据体系。

二是促进数据安全共享

开发面向终端用户的门户系统，支持用户浏览、下载并使用平台公开共享的数据，同时允许在线使用相关工具进行微生物基因组分析与研究。系统采用隐私计算、区块链等技术，确保数据在跨主体流通中的安全性；通过数据访问权限管理，保障数据的安全共享；建立多级备份机制、防御网络攻击和实施灾难恢复策略，全面加强数据安全保障。

三是赋能生物制造产业

基于平台制备了高产 L-高丝氨酸的重组大肠杆菌，该菌株成为我国自主知识产权高产菌株；通过系统代谢工程改造和发酵工艺改进，高丝氨酸的实验室发酵水平已突破 100g/L，转化率达到 50%。

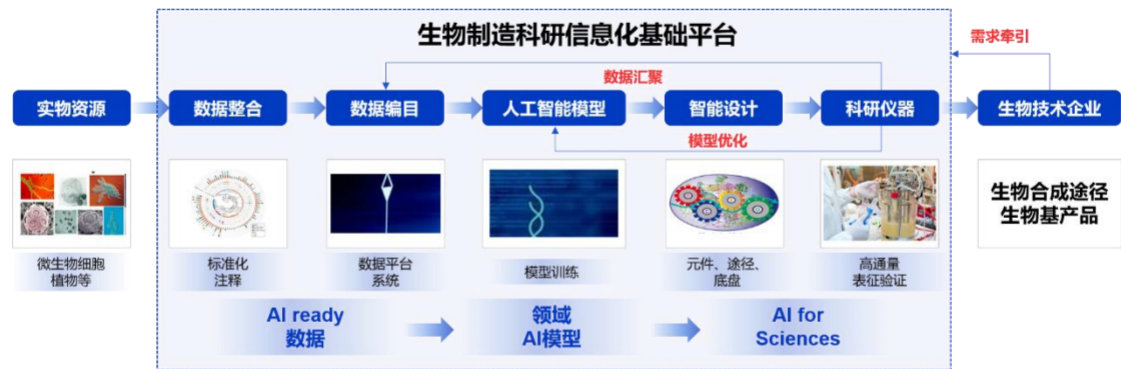


图 生物制造科研基础平台

气象数据融合应用 推动多领域场景创新

气象数据是地球系统科学的基础资源，对防灾减灾和经济社会发展具有重要意义。当前气象数据存在“数据孤岛”、新兴技术应用不充分、跨领域融合不足、安全监管能力薄弱等问题。为提升数据要素配置效率和服务能力，青海省气象信息中心基于青海气象大数据云平台，推动气象数据汇聚融合、复用增效和高效流通。

一是强化数据基础，汇聚多源数据

通过内部“应收尽收”及外部共建共享、数据交换、共享服务等多元形式，推动多主体海量数据资源汇聚整合和融合应用。目前基于青海气象大数据云平台已汇聚 12 大类 972 种数据，在线数据量达 2.3PB，构建起具有青海气象特色的数据资源体系。

二是应用算法工具，推动降本增效

提供基础设施资源、智能算法工具及应用组件，支持快速调用数据，通过容器云一键部署、流水线算法调度优化流程，结合综合业务监控和智能化运维，降低资金、开发、时间及人力成本，支撑业务向“轻量级、智能化、定制化、开放化”转型。

三是深化数据融合应用，服务产业发展

基于“云+端”模式，推动数据、算法、工具等在多场景的复用共享与融合创新。推进气象数据与水利、生态、林草、测绘等跨部门跨领域数据融合应用，服务防灾减灾、智慧农业、公路交

通、清洁能源、金融服务及绿色生态等场景。

四是构建安全体系，促进合规流通

从制度体系、软硬设备、技术防护、运行机制等多维度加强安全防护，建立高价值数据产品业务准入、数据产品和服务溯源登记等机制，构建全生命周期数据安全管理体系，促进气象数据合规高效流通。

加强整合利用 推动时空勘测数据价值释放

时空勘测数据包含全域覆盖、高精度、实时更新的地理空间信息，蕴藏着巨大的应用价值。但当前这类数据存在多源异构、标准不统一、更新滞后、共享渠道不畅等问题，导致数据资源的潜在价值难以充分发挥。为有效解决这些问题，新疆兵团勘测设计院集团股份有限公司打造了时空勘测大数据处理及服务平台，以此实现勘测信息资源的高效整合与利用。

一是构建统一标准体系，夯实数据要素应用基础

针对勘测数据多源异构、标准不统一等问题，制定了涵盖数据汇交、质量管控、数据服务的标准规范体系。通过建立统一的数据资源目录，实施标准化处理流程，有效保障数据的准确性和可用性。

二是研发关键技术，支撑时空勘探数据整合与共享

研发了基于“版本-增量”的时空数据动态更新技术，实现了时空勘探数据的高效管理和实时更新；研发了多源异构数据跨数据库融合处理技术，打通了时空勘探数据整合与共享的壁垒；创新推出了组件化地图服务自动发布技术，大幅提升了时空勘探数据服务的响应效率。

三是构建一体化服务平台，释放时空勘探数据的乘数效应

通过打造集数据管理与服务管理于一体的综合服务平台，实现时空勘探数据资源的全生命周期精细化管理。平台深度整合数

据汇聚、管理、服务及应用等全链条能力，支持勘测数据实现“存量整合、增量更新、一图尽览、按需调用”。目前，该平台已沉淀了 30 余类勘测数据，提供 10 多种数据处理模型，服务单位超过 50 家，有效推动了时空勘探数据在水利、农业、应急等多个领域的创新应用。

在现代水利工程中，将勘测数据应用于水利设施的监测和维护，为精准灌溉、洪水预警和水资源管理等提供了支持。在应急管理领域，借助勘测数据的空间信息，助力快速应对各种突发事件和自然灾害。目前，相关应用已覆盖兵团 14 个师市、100 多个团场的项目，直接节约数据生产费用约 3000 万元，提升数据生产效率约 70%，节约人力资源成本约 500 万元。

生态系统长期观测数据赋能生态系统研究领域创新

高质量的生态系统长期观测数据是掌握生态系统变化机制、评估系统功能动态及服务碳中和目标与生态治理的重要基础。但我国生态系统地面观测数据分散于多部门，存在观测技术规范不一致、数据质量参差不齐、数据产品体系不健全等问题，制约了生态学过程机理的跨站点分析和生态系统变化的精准评估。国家生态科学数据中心依托中国科学院地理科学与资源研究所标准化规范不同台站生态系统长期观测数据，构建覆盖国土空间、生态系统类型齐全的定位观测数据集，支撑陆地碳氮水循环研究成果产出

一是建立标准规范体系

构建生态系统领域长期观测数据采集、质量控制、数据汇交、资源描述、数据服务等标准体系，制定统一观测规范、数据产品加工技术规程。从源头上保证生态系统和生物多样性领域国家野外站观测数据采集的规范化和标准化，保障数据的可比性、连续性，推动基于观测的联网研究。

二是加强长期定位观测数据资源供给

建立生态网络云（EcoCloud）平台，提供数据采集、传输、存储、质控、开发、共享及应用一体化服务，提升野外台站长期观测数据的采集管理分析和水平。构建多级数据产品体系，制定数据产品加工技术规程，指导野外站生态系统观测数据开发

利用。依托“台站-专业分中心-数据中心”三级质控体系保障数据加工质量。目前已建立覆盖 53 个生态站、2000 万条数据的生态系统定位观测数据集，为生态系统研究、资源环境的保护利用与治理以及农、林、牧、渔业相关生产活动提供重要的数据支撑。

三是推动特色数据资源建设

围绕陆地生态系统碳循环和全球变化领域等热点问题，获取农田、草地、湿地、荒漠、森林、喀斯特和养殖塘等不同下垫面的碳氮水通量观测数据，通过数据加工处理和质量控制，构建中国典型生态系统碳-氮-水通量数据集，数据量累计 14TB，覆盖 60 多个通量观测塔，时间跨度为 2003—2020 年，为揭示生态系统过程研究和模型验证以及遥感产品地面真实性检验等提供支撑。

四是构建生态领域数据分析方法体系

融合深度学习、生态模型与数据同化技术，构建生态数据分析方法体系；研发数据质量控制、长时序数据重构、碳循环实时同化预测、生态系统评估等专业化软件工具，支持生态数据深度分析与挖掘；研制生态系统评估模型 CEVSA-ES，实现固碳功能与其他重要生态系统服务功能的一体化估算，为定期评估国家生态系统重要服务功能提供支撑。

连续 20 年的长期观测数据有力支撑了陆地生态系统碳氮水耦合循环研究，累计为 400 多个国家重大项目、1700 余家机构提供数据和知识服务，助力碳氮水循环机制理论创新，相关成果为国家碳管理决策提供理论支持，也为我国生态碳汇功能评估等工作提供数据与方法论支撑。

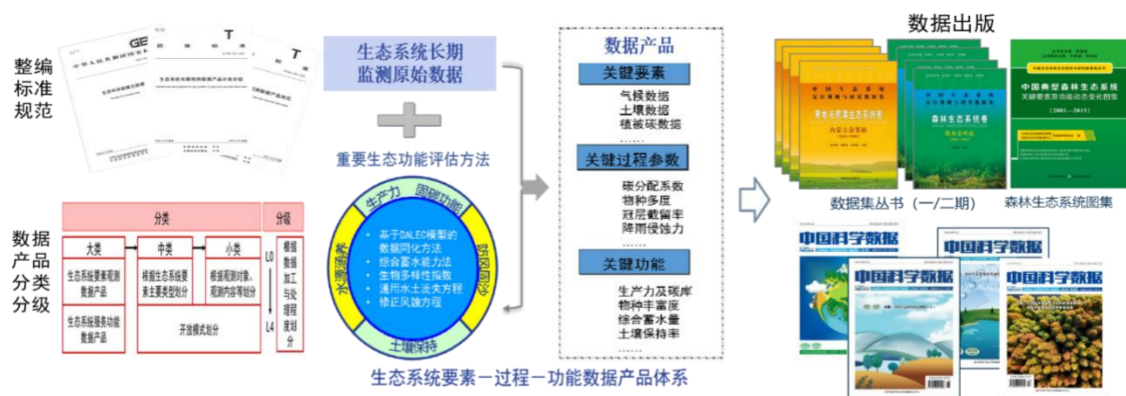


图 1 生态系统长期定位观测数据产品生产和出版

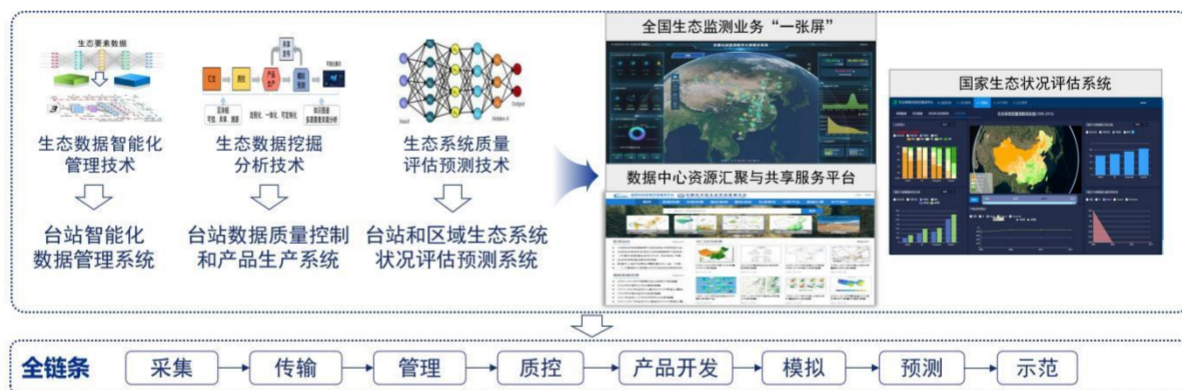


图 2 全链条“生态网络云（EcoCloud）”平台

高质量遥感数据赋能可持续发展

针对科学数据的数据密集、计算密集等挑战，鹏城国家实验室在国产算力上打造自主可控的科学数据计算与服务一体化平台，针对遥感观测数据碎片化、离散化等导致数据质量参差不齐、数据共享难等问题，构建高质量多源陆表信息数据集，研发了在线计算与共享服务系统，推动遥感数据高效流通与利用。

一是破解观测数据时空应用难题

针对全球变化、可持续发展的科学应用需求，研究高效的时空一致性校正方法，结合多源数据，构建全球近 40 年陆表密集时空无缝数据立方体；支持定制生产全球地表覆盖、居民地扩张、水面动态、湿地动态等信息数据产品，支持高质量数据集的开放共享。

二是构建定制计算数据共享服务系统

为满足用户对无缝数据立方体定制化开发需求，研制计算数据共享服务系统，提供自动化工具集和定制化计算方案，支持用户根据具体需求灵活选择配置参数，帮助用户快速实现个性化数据处理任务，有效提升数据利用效率。

三是研发基于委托计算的科学数据分析体系结构

研究基于大模型的自动化数据探索程序生成，基于数据解释器的科学数据分析管道构建、基于自反馈的数据分析程序修正等技术生成委托程序，构建基于委托计算的科学数据分析体系结构，

实现对仿真数据分析和计算；基于隐私计算等技术，保障数据不可见，实现高效、快速地完成数据开发。

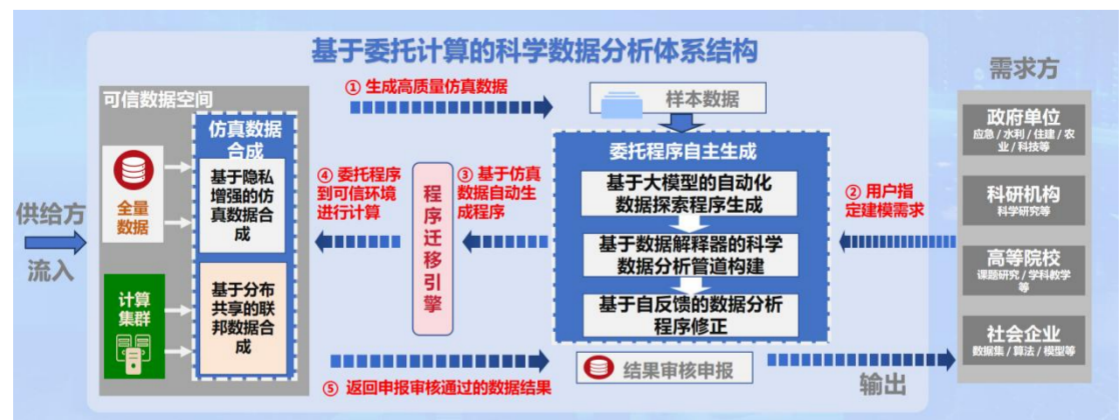


图 基于委托计算的科学数据分析体系结构

AI 赋能天文“探索”与“教育”融合

作为探索宇宙奥秘的前沿学科，天文学拥有丰富且独特的数据资源，具有极高的科研价值与探索意义。然而，海量数据中高效识别并发现具有重要科学价值的稀有天体目标日趋困难。为此，太原理工大学与中国科学院国家天文台携手合作共同开发了数据要素利用平台——“星系马戏团”(GalaxyCircus)，有效满足了专业科研的数据分析需求与公众参与科学探索的热情，提升了天文研究的效率与社会参与度。

一是构建一体化数据平台，汇聚天文数据

通过整合国内外知名巡天项目的天文目标图像资源，构建涵盖 750 余万个天文目标的多源天文图像数据集。基于智能预处理模块和自动化数据工作流，实现了数据清洗、融合与标准化的全流程自动化。与传统方法相比，显著提升了数据处理效率，为后续人工智能模型的开发和平台应用奠定基础。

二是建立天文图像大模型与智能分析系统

基于天文图像视觉大模型及智能分析系统，支撑星系形态自动分类、行星状星云搜寻、强引力透镜系统发现等多项天文学前沿研究。面向科研用户开放模型的应用程序接口，提供图像增强、目标检测、特征提取等图像处理工具链，推进了天文学研究的智能化水平。

三是多层次协同，完善天文知识体系

结合人工智能，构建由公众用户、科研人员及领域专家组成的多层次协作体系，该体系以“AI 初筛-公众复核-专家确认”的闭环机制为核心，将前沿科研与公众科普进行深度融合。设计了类似连连看的游戏模式，让用户轻松参与到天体的标注与发现任务中。集成了天文大语言模型，用户可以进行天体检索，获得可视化的结果溯源和相关天文知识讲解。通过“一人一星”认领计划、发现者命名认证以及荣誉激励制度等激励机制，激发公众的参与热情。

目前已实现对天体候选体的自动提纯与优选，识别效率显著提升；基于平台模型与协作体系，成功发现 25 个小尺寸行星状星云、24 个强引力透镜候选体，并在 Euclid 数据中发现 25 个全新引力透镜新候选体，支撑了我国在前沿天文学领域的科技研究。



图 总体流程图