

CSST 科学数据处理系统月报

第 50 期

主办方：中国空间站巡天望远镜科学工作联合中心

时间：2025 年 5 月 15 日



数据处理软件集成模块研制工作进展

近期，数据处理软件集成模块完成了数据查询、任务管理、DAG 管理三个关键前端界面的设计，详细介绍如下：

数据查询界面用于管理编排数据与流水线数据，界面左侧展示数据目录，右侧展示选中目录下的数据列表，列表中展示数据的关键字段信息，支持多条件的数据筛选。针对流水线数据，支持查看数据详情，并提供 fits 图像的在线预览功能。

任务管理界面用于发起重处理任务与管理任务记录，用户在界面点击“新建处理任务”，填写 DAG 组、dataset、obs_group、obs_type、batch_id 等参数信息，随后点击“提交”，发起的重处理任务将展示在任务广场列表，列表中实时展示已提交任务的运行状态，支持通过任务 ID、batch_id、DAG 组、dataset 字段对列表中的记录进行筛选。

DAG 管理界面用于查看各后端模块下各具体 DAG 的详情与运行日志信息。界面左侧展示所有

DAG 列表，右侧界面展示选中 DAG 的详情，包含基础信息、执行信息、执行记录、DAG 节点执行记录。执行记录列表展示了 DAG 每次执行的实例信息，选中某条实例后，下方 DAG 节点执行记录列表展示该实例下各节点的执行记录，点击列表右侧的“查看日志”，可查看该节点的执行日志详情。

数据查询、任务管理、DAG 管理这三个前端界面的设计完成，是数据处理软件前端的重要功能，将为运行及 QC 人员提供更直观、高效的操作平台，还将优化整个数据处理过程中的任务调度、数据管理和全流程监控，大幅提升工作效率。

数据系统研制进展

| 序号 | 模块名称 | 月度进展 |
|----|---------|---|
| 1 | 多色成像 | 使用宽场数据完成 Mosaic 测试；基本完成深场流水线功能模块；星系模型测光取得进展；完成对 25 平方度处理数据的宇宙线探测结果评估工作。 |
| 2 | 无缝光谱 | 针对新版透过率曲线导致零级像强度变化情况，系统分析了以零级像作为波长定标基准所引入的红移测量误差，以及以无缝光谱图像 WCS 定标为基准所引入的红移测量误差；在此基础上进一步优化抽谱和光谱红移测量算法；针对硬件团队提供的 GI 滤光片透过率曲线向蓝端延伸的调整方案，开展其对 0 级像强度增强的效果以及可能引入的污染的测试评估工作；配合开展流水线集成和配置项测试等相关工作。 |
| 3 | 数据流管理 | 根据 0 级数据定义的更新修改 DFS 服务和接口；开展 0 级数据导入和编排数据关联工作。 |
| 4 | 流水线运行管理 | 完成在公有云部署 scalebox 平台；根据 DAG 的调整修改流水线运行相关代码；开展流水线集成界面的讨论，根据需求完成 gRPC 接口，并提供接口说明文档。 |

| | | |
|----|------------|--|
| 5 | 多通道成像仪 | 完成一级流水线固化；初步完成配置项测试大纲，开展新版数据仿真，为测试做准备；MCI 模块仿真根据新版 0 级数据定义更新并接入编排；在长光进行了三次地面测试，在现场进行初步分析，目前在上海台开展集同分析。 |
| 6 | 积分视场光谱仪 | IFS 模块仿真根据 0 级数据定义更新并接入新的编排；一级流水线算法非线性效应根据新的测试结果更新；初步完成数据处理软件配置项测试计划大纲的编写工作。 |
| 7 | 星冕仪 | 完成 0 级数据修订和观测仿真程序的更新，完成两篇观测仿真论文编写；完成测试项的设计和测试数据的准备工作；参加星冕仪鉴定件工程指标和科学定标测试并负责数据处理任务。 |
| 8 | 太赫兹 | 完成太赫兹模块数据处理软件配置项测试方案的进一步细化并持续推进测试相关工作；完成太赫兹模块 0 级数据文件接口以及 0 级数据 toml 文件更新；完成太赫兹编排数据文件定义；开展 1 级数据文件接口以及 1 级数据 toml 文件更新工作；开展新版太赫兹 0 级仿真数据和编排数据文件生成工作；完成在轨定标处理方面指向定标算法研制以及算法仿真验证，持续推进太赫兹模块在轨定标数据处理流水线工作。 |
| 9 | 天体测量 | 完成不同的定心算法的精度分析；开展数据处理软件配置项测试图像仿真工作。 |
| 10 | 观测数据仿真软件 | 持续优化仿真软件，更新仿真数据的头文件模块；完成主巡天仿真技术文章，其中 3 篇已经投稿在审，2 篇完成初稿内审。 |
| 11 | 数据可用性标记 | 完成 1 级数据定义文档改版工作并执行会签工作。 |
| 12 | 在轨定标 | 完成在轨定标方案修改完善工作；开展定标参考数据定义更新工作；完成定标星场输入星表更新并重跑仿真。 |
| 13 | 观测需求编排 | 根据最新讨论的编排与 DFS 接口调整数据库结构，修改编排软件中相关的数据处理及输出部分；同步梳理各后端编排程序，形成独立软件包，使得各后端可更专注于各自的编排逻辑，也更便于接入到编排系统中；一般观测申请更新前端框架，从 jinja2 升级为 react，进一步拆分观测提案申请的提交、评审、反馈、管理、外部工具共计五个核心业务；机遇观测工具进行了 Kafka 消息系统优化，测试多 group 及多 topic 下的响应能力。 |
| 14 | 数据处理软件集成 | 完成 DFS 与编排软件、0 级数据定义的对齐；完成数据处理任务发起的逻辑设计与部分代码实现；完成数据查询、任务管理、DAG 管理三个关键前端界面的设计。 |
| 15 | 存储与计算 | codespace 支持使用隔离的 podman 容器；完成 z100 集群网络硬件升级。 |
| 16 | 天文算子 | CSST 基础模型 2.0 使视觉模型支持实例学习的同时能够进行多任务学习，本月首先在最基础也是最难解决的分割任务上进行实验，尝试在自然图像数据集 VOC 上复现一系列自然图像的实例学习方案，并进一步改进算法，目前正在复现精度过程中。 |
| 17 | CSST 计算工作站 | 完成在线协作开发功能整体开发，目前正在联调中并预计 5 月上旬发布上线；完成科研工作台对接国台 git 仓库；开展 1 级流水线日志读取和 scalebox 对接的功能开发工作。 |
| 18 | 数据融合 | 基于星表数据融合框架完成多波段图像融合分析服务架构设计，针对无缝光谱分块重组中的长图像片段拼接问题，设计张量维度重组替代传统图像拼接操作，实现零拷贝图像拼接；完成多波段星表数据融合服务中基础云服务 ODPS 的迁移与调测。 |



近期节点和计划

| 时间节点 | 计划任务安排 | 时间节点 | 计划任务安排 |
|----------|--------------------------|-----------|----------------------|
| 4 月 30 日 | 完成 1 级流水线的全部开发和版本固化 | 6 月 30 日 | 完成 2 级流水线的全部开发和版本固化 |
| 4 月 30 日 | 完成定标产品生成流水线和定标产品的生成工具包固化 | 8 月 30 日 | 完成数据处理软件 1 级流水线配置项测试 |
| 4 月 30 日 | 完成 50 平方度深场仿真 | 9 月 30 日 | 完成数据处理软件 2 级流水线配置项测试 |
| 4 月 30 日 | 基本完成 1000 平方度仿真 | 8 月 30 日 | 完成数据处理软件新功能开发 |
| 4 月 30 日 | 在轨定标方案评审 | 12 月 30 日 | 完成编排软件配置项测试 |
| 5 月 30 日 | 完成《软件设计说明》评审 | 26 年 1 月 | 实现共有云环境下的部署，开展系统测试 |

编辑：王海

审核：罗常青 刘江

签发：刘建海