JSTUC Data Analysis Advisory Group
Interim Report

June 8, 2020

George Becker (chair) (UC Riverside)
Jennifer Andrews (U of Arizona)
Zach Berta-Thompson (U Colorado Boulder)
Marusa Bradac (UC Davis)
Stephane Charlot (IAP/CNRS-Sorbonne U)
Casey Papovich (Texas A&M University)
Johan Richard (CRAL)
Mike Wong (UC Berkeley)

Summary

STScI is making good progress towards developing a suite of DA tools that will enable JWST science. The scope and structure of the effort are well thought out, and early progress is in step with the group's internal timeline.

STScI should enhance its efforts by working more closely with community groups (ERS and GTO teams) to address high-priority needs for early JWST science. Areas to focus on should include (i) simulated data and the connection between DRP products and DA tools, (ii) coordination and prioritization of DA tool development between groups, and (iii) functionality of existing DA tools.

Charge

From the JSTUC:

The JSTUC Data Analysis Advisory Group will work with STScI to ensure that data processing tools most needed and desired by the community will be in place in time to enable maximum scientific return of JWST. The DAAG will <u>identify the most high-priority needs of the community</u> (e.g. image creation, multi-object spectroscopy, photometry, spectral line detection, extraction, etc.) and <u>monitor tool creation</u> to ensure that there are no critical gaps. A priority is to <u>support the provision of simulated data</u> to members of ERS teams and the community. We envision the role to consist of high level guidance to STScI on priorities, reporting back to the JSTUC with a progress assessment.

Guiding Questions

- What is the status of DA tool development at STScI? Is there a well-considered plan to produce tools that will support early JWST science? Is the development proceeding at a reasonable pace?
- How is DA tool development at STScI connected with development efforts by community groups, particularly the ERS and GTO teams?
- How is the DA tool development connected to the data reduction pipeline (DRP) development? Is DA tool development adequately supported by access to realistic simulated data?

Initial Tasks

1. Understand the DA Tool development effort currently underway at STScI
   - Organization, goals, progress to date

2. Assess the connection between data efforts by STScI and those of other JWST stakeholders
   - Communicate with ERS, GTO teams
   - Identify the most urgent needs

DAAG Activities Jan–June 2020

- Received overview of STScI DAWG efforts (January 2020)
- Reviewed DA tools development roadmap
- Attended quarterly STScI JWST Data Analysis Demos
- Received STScI JWST Data Analysis Build Plan briefing (April 2020)
- Communicated with ERS teams
- Held monthly telecons (with some interruption due to COVID-19)

DA Tool Development @ STScI

- Generally very positive
- Significant expansion of previous efforts (e.g., Qt tools) enabled partly by the launch delay
- Structure and scope well thought out
  - Overall approach (e.g., focus on Python routines organized into Jupyter notebooks) largely guided by data collected in a user survey
  - Intended to be forward looking in terms of adoption and support
  - Tools intended for broad access (e.g., Jupyter notebook functionality to be incorporated into web-based and stand-alone apps)
- Centers on science cases derived from planned GTO and ERS observations
- Significant contribution to astropy core libraries
- Timeline: 22 Jupyter notebooks to be delivered on a quarterly schedule. Keeping pace as of latest quarterly DA tool demo.

DA Tool Development @ STScI (cont'd)

- Opportunity for input: Prioritization of science use cases
  - DAAG has not yet considered these in detail, though we urge communication with community groups (see later slides)
- Use of simulated data
  - Many STScI DA tools start with simulated data that is not highly specific to DRP-processed data, although some include pipeline-processed data or include DRP steps
- Limited inclusion of existing tools, particularly those not written in Python
  - "Off-ramps" considered

ERS Team Perspective

- DAAG communicated with 9/13 ERS teams
- Goal was to get a broad sense of how the ERS contributions relate to STScI efforts, assess their DA priorities, and identify common concerns
- Programs generally involve multiple instruments and modes, and therefore presenting a wide range of DA challenges
- Common feedback:
    - Concern over access to and/or reliability of simulated data.
    - Need for DA tools that address JWST-specific challenges (e.g., MSA, IFU analysis; IFU visualization (cubeviz), working with intermediate DRP products)
    - Expectation that existing tools would play an important role in early science (i.e., limited support for prioritizing replacements of legacy tools)
    - Progress often substantially aided by direct connection to STScI staff or GTO team, particularly in technically challenging areas (e.g., slitless spectroscopy)

Simulated Data

- Critical for developing DA tool compatible with and specific to JWST data
- ERS teams noted:
  - Difficulty in obtaining simulated data that had been processed through the DRP
  - Concerns that DRP products may not be stable in terms of format, header keywords, etc.
- Simulated data from Mirage (NIRCam and NIRISS) and MIRIsim being used successfully by some ERS teams
- NIRSpec is a particular concern.
  - Simulations are done by instrument development team (IDT) / GTO team
  - IDT/GTO team willing to provide simulated data and has trained two ERS teams on simulations
  - Some teams may not be aware that IDT/GTO help is available
  - IDT/GTO team has had difficulty running simulated count maps though STScI DRP

Initial Recommendations

1. Increase communication between STScI, ERS, and GTO teams
    - Expertise relevant to JWST DA is highly distributed across teams and institutions
    - High potential for cross-fertilization, particularly in multi-group forums (telecons, workshops)
    - Topics to consider:
        - Status of/access to simulated data
        - Coordination of DA tool development
        - Prioritization of STScI DA science cases
2. Address deficits in simulated data
    - Coordinate with ERS teams to determine what is needed in terms of DRP-processed data
    - Work towards a tighter integration of DRP and DA efforts
    - Consider a repository of simulated data
        - Serve as a common reference point for DA tools development
        - Enable tracking of changes to DRP products (quarterly?)

Initial Recommendations (cont'd)

3. Prioritize DA tools that address JWST-specific needs
    - e.g., MSA, IFU analysis and visualization tools
    - Seek further input from ERS, GTO teams (others?)
4. Facilitate existing DA tools that are either widely used or difficult to replace
    - Work with ERS & GTO teams to identify and, if possible, resolve data formatting issues for high priority tools, including compatibility with ASDF format
    - May encourage further scrutiny of simulated data in some cases
    - Note:
        - A light-touch approach may be sufficient, with STScI mainly responding to queries from community groups and incorporating any feedback into its DRP and DA tools
        - STScI should not be expected to <u>support</u> legacy tools outside of its development ecosystem.

Summary

STScI is making good progress towards developing a suite of DA tools that will enable JWST science. The scope and structure of the effort are well thought out, and early progress is in step with the group's established timeline.

STScI should enhance its efforts by working more closely with community groups (ERS and GTO teams) to address high-priority needs for early JWST science. Areas to focus on should include (i) simulated data and the connection between DRP products and DA tools, (ii) coordination and prioritization of DA tool development between groups, and (iii) functionality of existing DA tools.

Upcoming DAAG Tasks (partial list)

- Better understand the DRP status and the interface between DRP and DA development.
- Examine computing requirements for producing simulated data and running the DRP
- Look at plans for DA tool organization, distribution, and training