JWST Data Analysis Training

From now until the first year of science operations

Klaus Pontoppidan

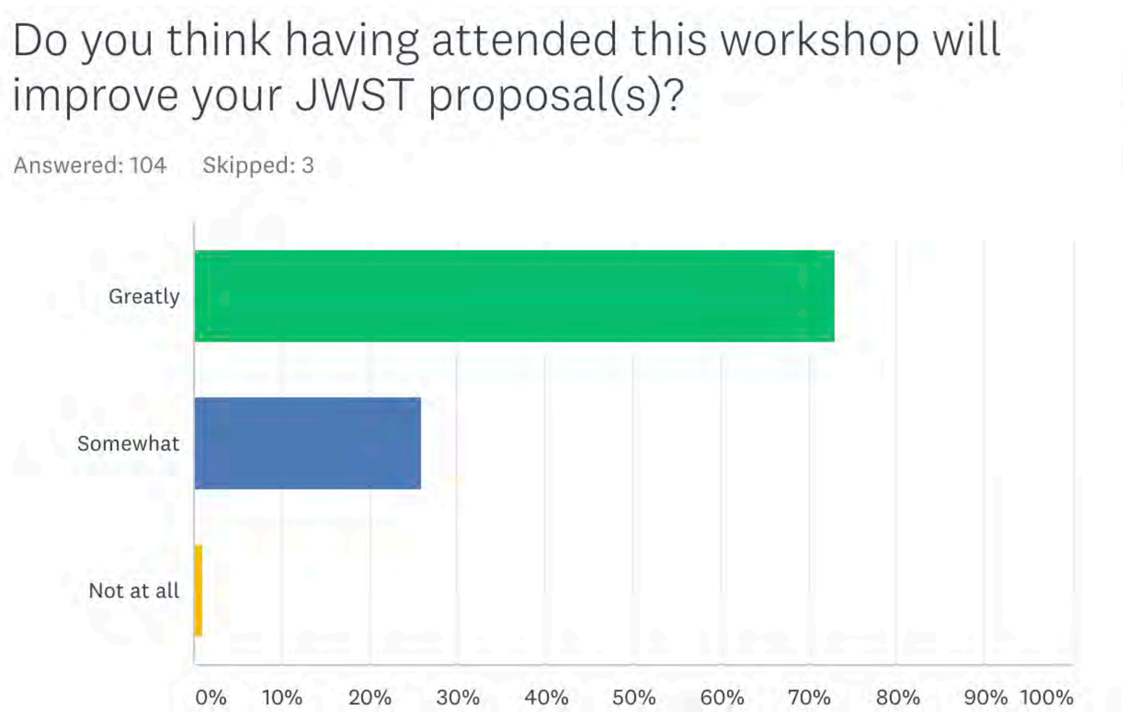
JSTUC, June 8, 2020

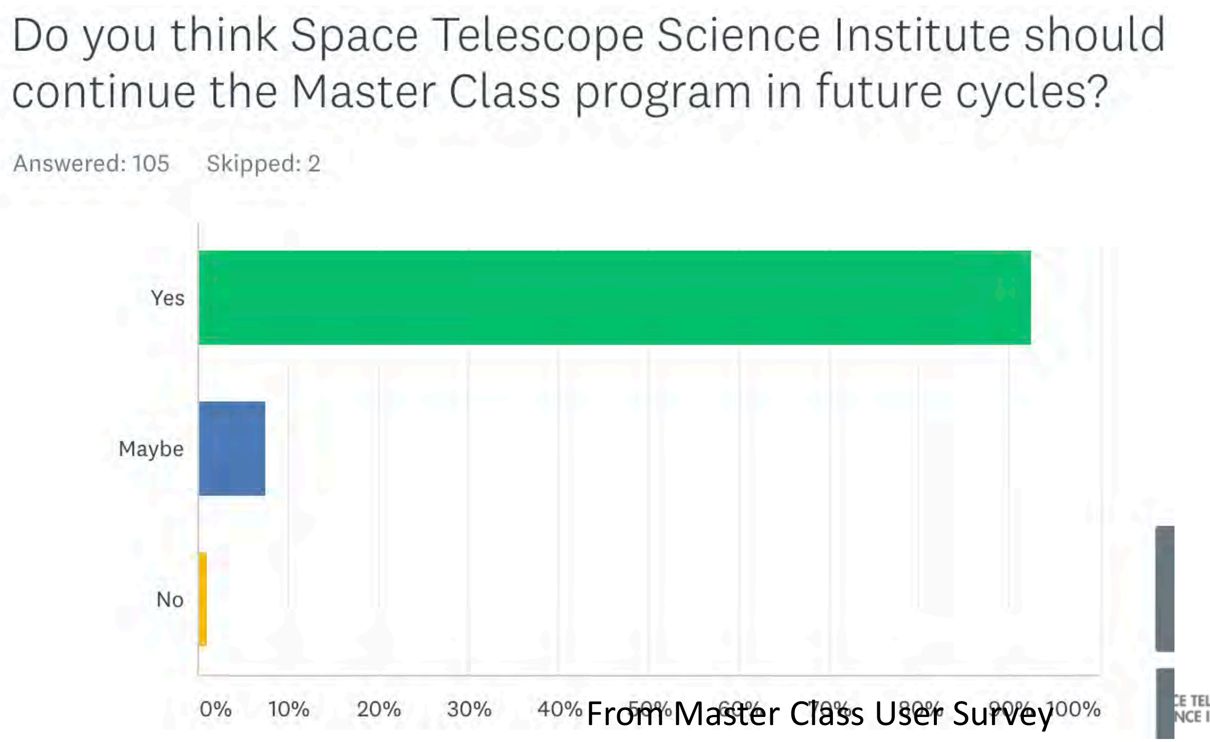
Background

- A data analysis user survey was conducted from June 7-August 2, 2019
- Main findings:
    1. Python is the most common language, but more than 50% of respondents also use other major platforms
    2. Almost 70% of users anticipate starting data analysis efforts from individual, calibrated exposures.
    3. Almost 80% of users expect to download their JWST data from the archive and analyze it on a local machine.
    4. Respondents prioritize training specific to JWST, as opposed to general Python training and non-specific tools.
    5. Respondents identified a critical need for training materials that are available when needed (written documentation, help desk, colleagues, video tutorials, etc.).
- Concepts for data analysis engagement developed in collaboration with JWST User Support, the Data Science Mission Office, and the project at GSFC.

Data Analysis Training concepts

- Hold a Master Class for data analysis?
  - The proposal planning Master Class was largely successful
  - Reached a significant fraction of the community at low cost
  - Shipped 1600 Pocket Guides
  - Received very positive feedback
  - Last part of the program was affected by the covid 19 pandemic





From Master Class User Survey

Arguments against a Data Analysis Master Class

- Held discussions with internal and external stakeholders on the possibility of a Data Analysis Master Class:
  - Real life science use cases are very specialized
    - Master class participants will not be comfortable teaching use cases outside of their area of expertise.
    - A large number of sessions would be required
  - All relevant data analysis tools should be complete before holding a Master Class
    - Not consistent with the development schedule.
  - Two distinct epochs for training are needed; pre-launch and science operations
    - Not consistent with a single pre-launch Master Class.
- A different training concept, optimized for data analysis, is needed.

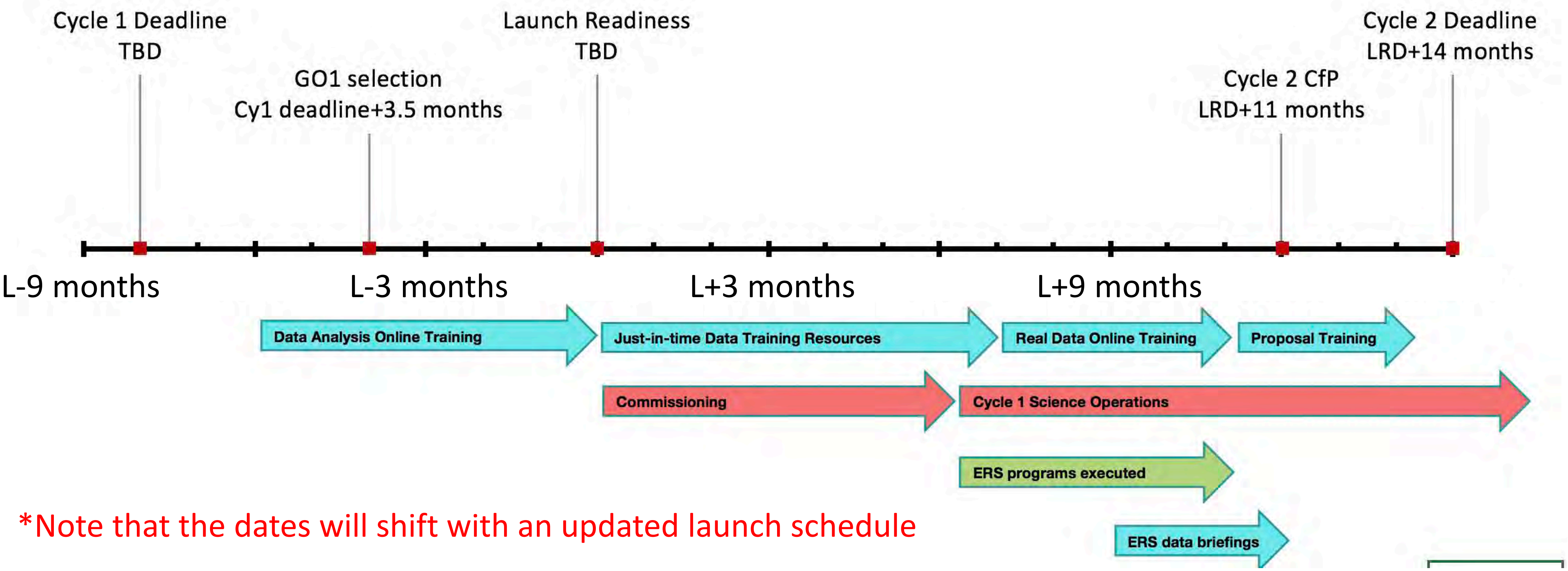
Objectives for a Data Analysis Training Concept

1. **Reaches much of the community**
   - Number of participants and science area
2. **Has a flexible schedule**
   - So it can be coordinates with active data analysis tools development and mission status
3. **Is fully online**
   - So it is resistant to impacts from covid-19, also low carbon.
4. **Is accessible to partners**
   - International partners can access events and/or use materials
5. **Offers asynchronous material**
   - Should lead to production of training materials that can be used for self-teaching and reference.
6. **Supports diversity and inclusion**
   - By lowering barriers to participation

The *JWebinar* series

- The *JWebinars* will be a series of online training sessions.
  - **Short** (2-4 hours) to make it easy to find time to attend
  - **Topical** (typical focus on one science use case) to make use of a well-defined subset of the data analysis tools
  - **Interactive** to keep the audience engaged
  - **Regular** (up to 20) to maximize opportunity and grow audience
  - **Linked** to offline resources to provide access to all
- The *JWebinars* will be organized by a local committee
  - Representation from STScI and possibly ERS teams
  - Two distinct *campaigns:*
  - Pre-launch, using current notebooks and simulated data
  - Science Operations, using real data, in particular ERS data sets.

Training timeline



*Note that the dates will shift with an updated launch schedule

Likely prerequisites for the JWebinars

- **Shared, online data analysis environment**
  - Jupyterhub is a likely platform
  - Use experience from previous workshops (e.g., TESS)
  - Avoid users having to install software on their own machine in order to attend a JWebinar.
- **Simulated data**
  - During the pre-launch campaign, many JWebinars will likely use a simulated data set
  - Existing development of Data Analysis Jupyter Notebooks include such data sets
  - JWebinars will increase discoverability
- **Resources for developing and teaching individual *JWebinars***
  - Staff is needed to write and teach individual *JWebinars*, but each event likely only requires 2-3 teachers, depending on subject.
  - Depending on interest, *JWebinars* may be repeated.

Notional examples of JWebinars

ID	Topic	Time frame
1	JWST Pipeline installation and first steps	Pre-launch
2	Multi-band imaging and photometry	Pre-launch
3	MOS Spectroscopy with NIRSpec	Pre-launch
4	IFUs and spectral cube modeling	Pre-launch
5	Galaxy spectroscopy	Pre-launch
6	MIRI MRS Spectroscopy of point sources	Pre-launch
7	NIRISS time-series observations	Pre-launch
8	First results from ERS programs and lessons learned	Science Ops
9	JWST Spectro-photometric calibration	Science Ops
10	JWST Astrometric calibration	Science Ops
	Etc… (Up to 20 JWebinars to be prioritized)	

Conferences and workshops

- The *JWebinars* may not be for everyone, so in the interest of a balanced portfolio, we will still support major professional events

Event	Venue	Dates	Activities
DPS 2020	Spokane, WA	10/23/20 - 10/30/20	• STScI booth
AAS 237	Phoenix, AZ	1/3/21 - 1/7/21	• JWST Town Hall • STScI booth • Workshop • 2-3 Splinter Meetings
Cycle 1 Deadline	N/A	No earlier than 10/15/20	• Milestone
Pre-launch JWebinars	Virtual	9/1/20 - Launch	• Series of webinars
Launch Readiness Date	N/A	TBD	• Milestone
AAS 238	Anchorage, AK	6/6/21 - 6/10/21	• **Data Analysis Workshop** • Information Session/Town Hall? • Exhibit
First Science Observations	N/A	L+6 months	• Milestone
DPS 2021	Providence, RI	10/3/21 - 10/8/21	• STScI booth
Science Operations JWebinars	Virtual	L+8 months	• Webinar
Cycle 2 CfP	N/A	L+8 months	• Milestone
AAS 239	Salt Lake City, UT	1/9/22 - 1/13/22	• JWST Town Hall • First JWST science or ERO session? • **Data Analysis Workshop**

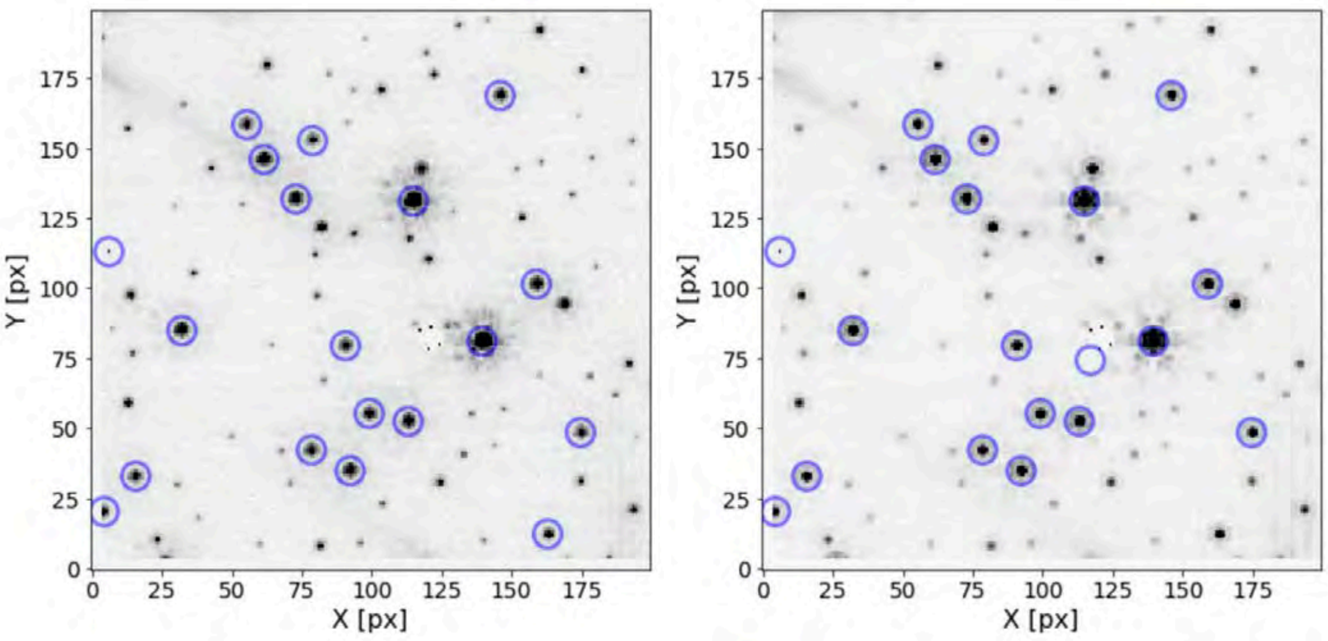
DD-ERS engagement

- ERS teams are committed to supporting the community in the analysis of the ERS data sets.
- This is accomplished by briefings, the three first have already been held

	Event	Venue	Date	Resource/recording
✓	Announcement session	231st Winter AAS	January 2018	youtube/JWSTObserver
✓	Progress briefing	232nd Summer AAS	June 2018	N/A
✓	Webinar series	Online	March 2020	youtube/JWSTObserver
	Deliverable update	Online	Fall 2020	jwst.stsci.edu
	Results and lessons briefing workshop	Online / AAS?	Cycle 2 CfP, 2022	Documentation and youtube/JWSTObserver

Just-in-time asynchronous resources

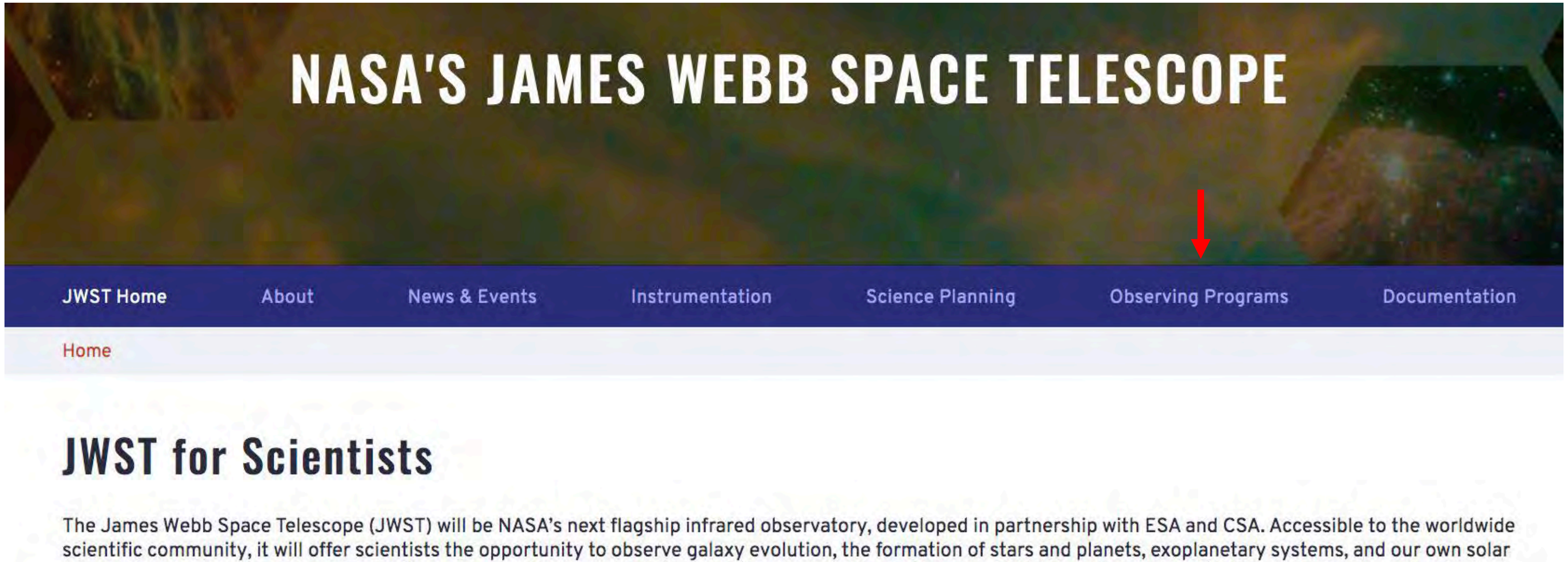
- Science use case Jupyter Notebooks
  - Already under development, with 7 produced so far
- Data analysis training videos
  - Scope of these to be determined
  - Examples may include visualization tools tutorials, data retrieval from MAST, use of selected Jupyter notebooks.
- Simulated data
  - As used by Jupyter Notebooks, but also available separately
- JDox data analysis documentation
  - Data formats, instrument and mode special considerations, ecosystem overview, etc.



Source detection in simulated NIRCam field

Data Analysis web presence and social media

- jwst.stsci.edu will have a central web area for data analysis to collect links to all appropriate resources
  - Notebooks, tools download and installation, JWebinars and other training events, documentation on JDox and readthedocs, etc.
  - Will also contain a curated summary list of simulated data sets.
  - News about data analysis tools will be disseminated on JWSTObserver
  - The @JWSTObserver social media accounts will highlight news items and events.



Questions to the JSTUC

1. Will the training timeline address the most important needs of the community?

2. Will the format / types of planned training reasonably satisfy the broad community?