

# STOA – Script Tracking for Observational Astronomy

**Peter Hague - University of Cambridge**

H2020-Astronomy ESFRI and Research  
Infrastructure Cluster (Grant  
Agreement number: 653477).



H2020-Astronomy ESFRI and Research Infrastructure Cluster  
(Grant Agreement Number: 653477).

# Background

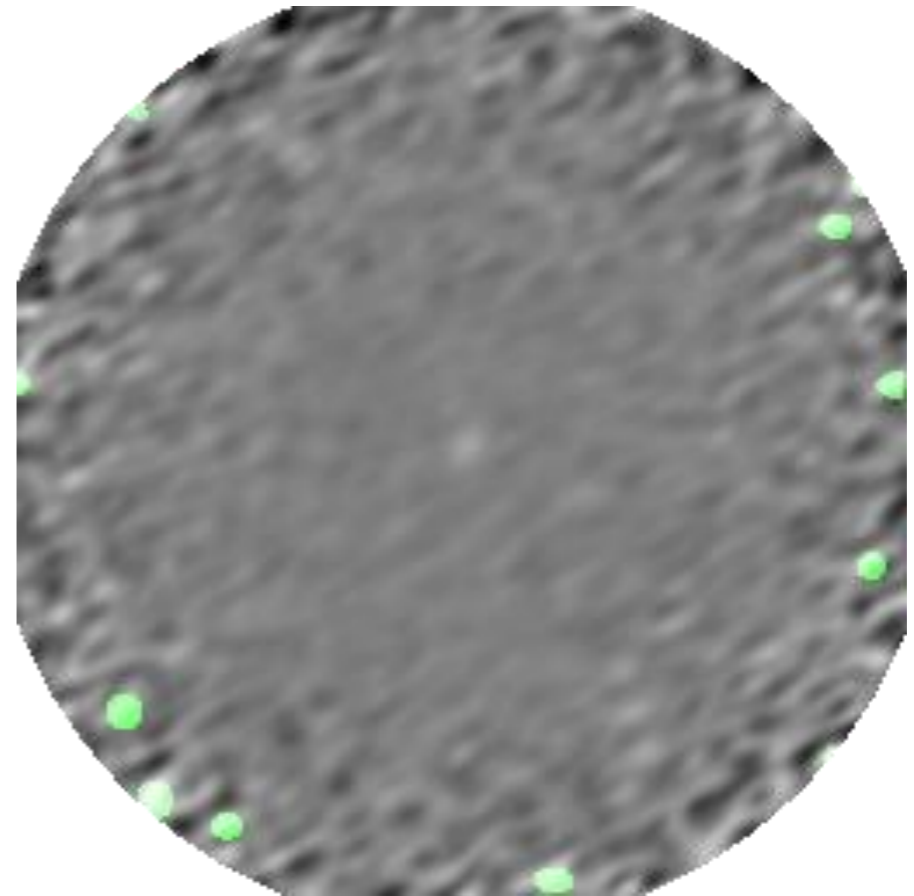
- **BaSC** and other projects required batch operations on ALMA archive
- **STOA** emerged from my efforts to streamline and automate these operations
- Now a standalone web application:  
<https://github.com/petehague/stoa>

# Objectives

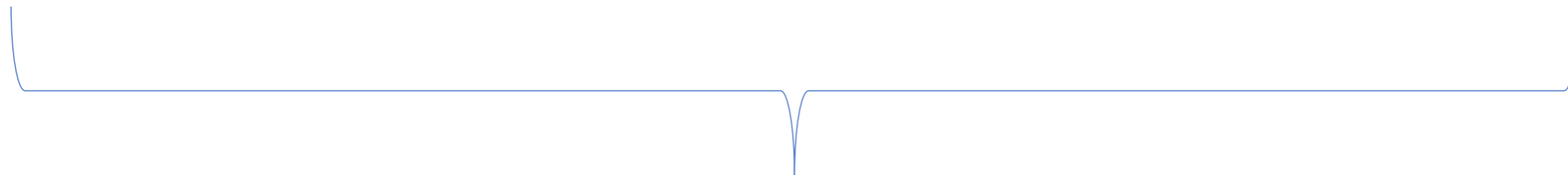
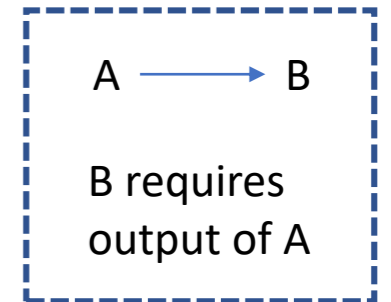
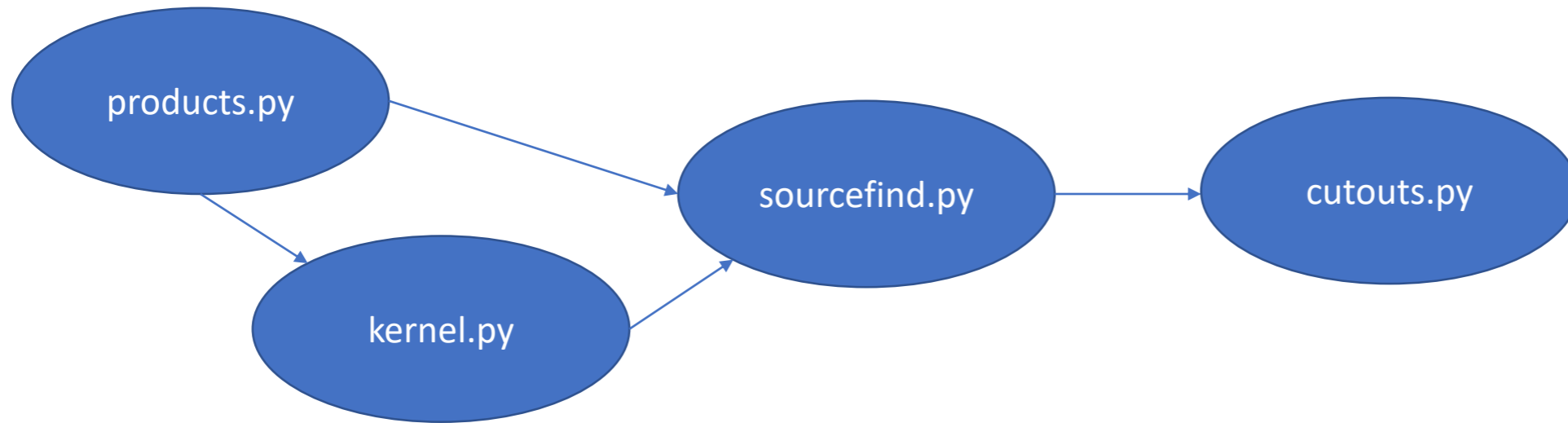
- Want to perform actions in batches
- Each instance of an action may need to run differently
- Must track provenance of result – ensure we know precisely which version of each script led to a specific output
- Must be able to share data, processes and outputs
- Final results should be easily accessible to others

# Example task

- Want to run SExtractor on every observation that matches a certain criteria
- Ideally, archive should contain clean image, primary beam correction, and the corrected image
- One or more may not be present. May also not be immediately clear from metadata which is which, so algorithm has to guess
- Sometimes guesses wrong...



# Workflows




COMMON  
WORKFLOW  
LANGUAGE

# Worktables

- Each row is an execution of the workflow
- Writable inputs, read-only outputs
- Software tracks status (e.g. if currently shown outputs correspond to inputs or if new outputs are pending)
- Worktables connected in relational style; automatically triggering recomputation

*	Inputs	Outputs



\* Hidden fields – tracking processing status etc.

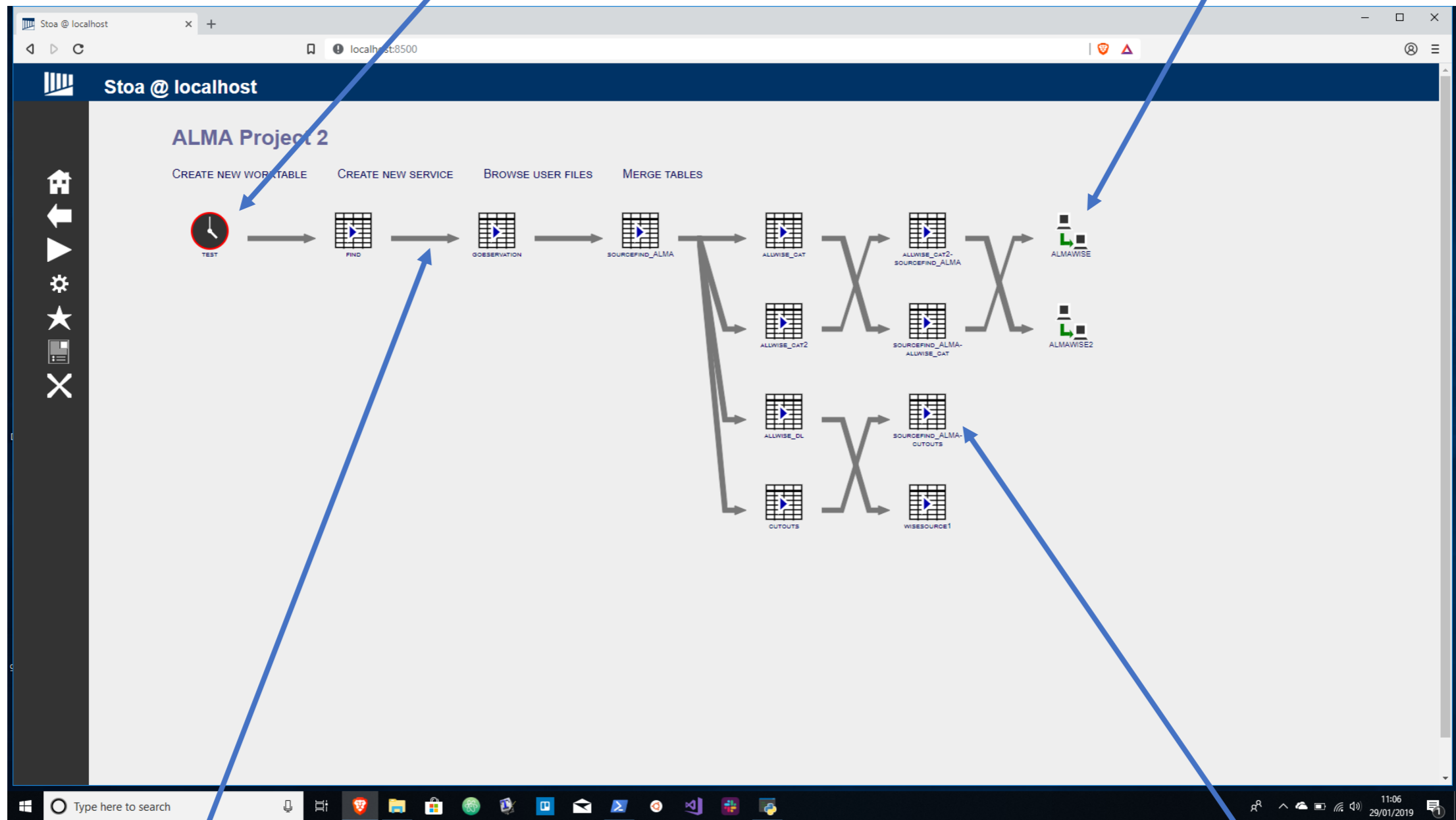
# Worktables

- **STOA** allows access to a worktable as Owner, Collaborator, and Reader.
- Owner can initiate computation, collaborator can flag rows and add comments, and reader sees worktable as a simple read only table
- Worktables can be served up as *.fits* downloads or through VO cone search protocol.

# Example

Time trigger

Download/VO Service

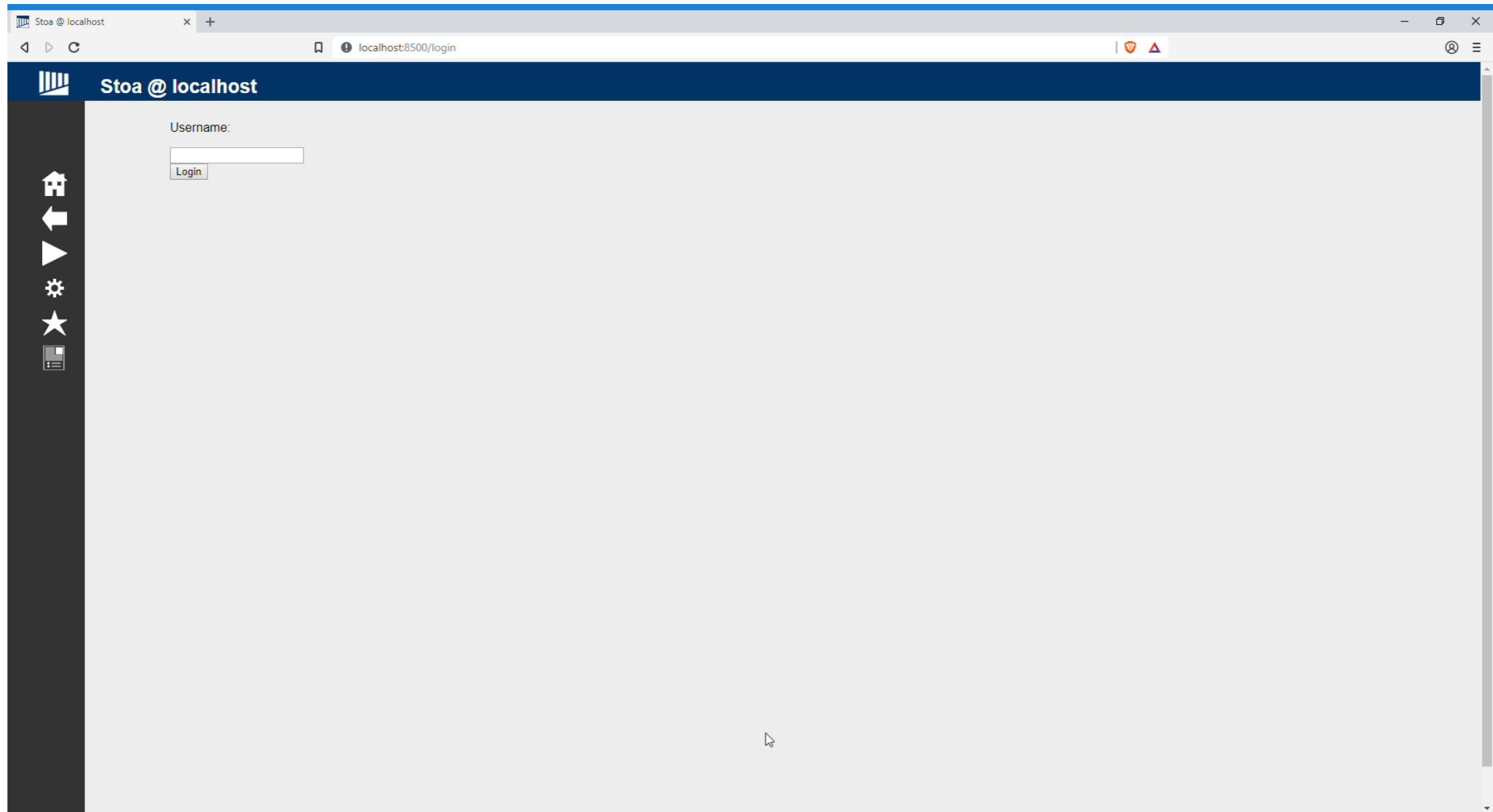


Using output of a table as key for another

Join Operation



# Example



# Example

```
1 #!/usr/bin/env cwl-runner
2
3 cwlVersion: v1.0
4 class: CommandLineTool
5 baseCommand: ["python", "grabcoords.py"]
6
7
8
9 inputs:
10   pathname:
11     type: string
12     inputBinding:
13       position: 1
14
15 outputs:
16   ra:
17     type: array
18     items: double
19   dec:
20     type: array
21     items: double
```



Stoa @ astro-vm1.vss.cloud.private.cam.ac.uk

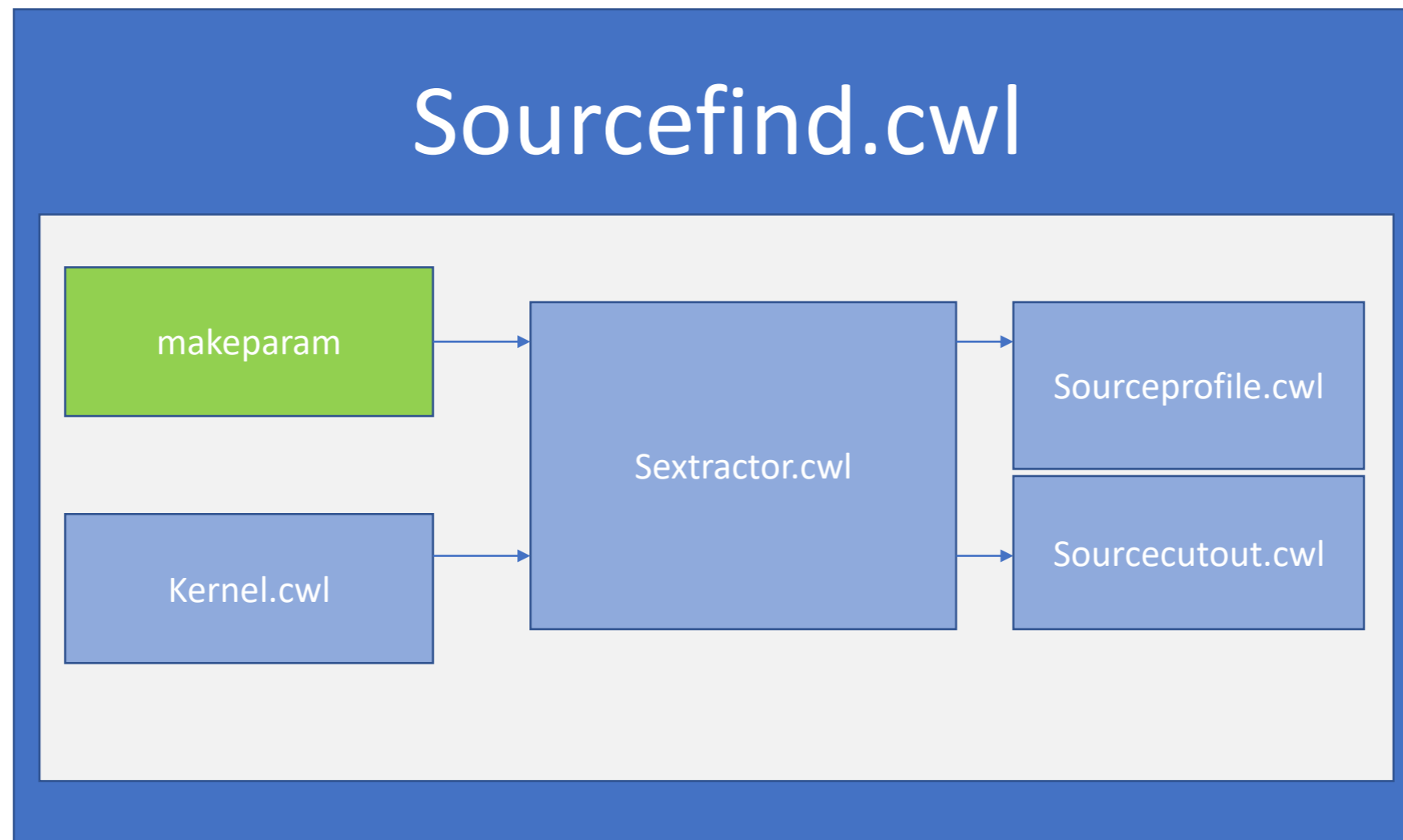
Worktable: grabcoords.wtx

RUN ENTIRE TABLE CLEAR OUTPUT DELETE TABLE RENAME TABLE  
EDIT TABLE

	pathname	ra	dec
	I_str	O_float	O_float
UCD			
RUN	/home/prh44/r...mber.uid_A001_X2fb_X3e4/product	65.4085416667	-26.9543333333
RUN	/home/prh44/r...mber.uid_A001_X2fb_X3bc/product	9.73549999996	-10.4316388889
RUN	/home/prh44/r...mber.uid_A001_X2fb_X3f8/product	167.641583333	-13.496
RUN	/home/prh44/r...mber.uid_A001_X2fb_X424/product	340.204166667	-18.6621666667
RUN	/home/prh44/r...mber.uid_A001_X2fb_X43c/product	159.225791667	-2.54386111111
RUN	/home/prh44/r...mber.uid_A001_X2fb_X438/product	65.504125	-19.4579722222
RUN	/home/prh44/r...mber.uid_A001_X2fb_X430/product	359.135208333	-6.3831388889
RUN	/home/prh44/r...mber.uid_A001_X2fb_X42c/product	349.637916667	-30.4926111111
RUN	/home/prh44/r...mber.uid_A001_X2fb_X3f4/product	162.079541667	-1.16119444444
RUN	/home/prh44/r...mber.uid_A001_X2fb_X420/product	332.801625	-32.1035833334
RUN	/home/prh44/r...mber.uid_A001_X2fb_X408/product	183.112416667	5.09263888889
RUN	/home/prh44/r...up_ouss_id/member_ouss_id/product	178.088625	0.926861111111
RUN	/home/prh44/r...mber.uid_A001_X2fb_X3b8/product	7.02733333328	4.95713888889
RUN	/home/prh44/r...mber.uid_A001_X2fb_X404/product	181.906	6.50288888889
RUN	/home/prh44/r...mber.uid_A001_X2fb_X3e8/product	73.5074583333	-44.8086388889
RUN	/home/prh44/r...mber.uid_A001_X2fb_X414/product	227.424083333	-17.8241111111
RUN	/home/prh44/r...mber.uid_A001_X2fb_X418/product	308.041666667	-21.2339722222
RUN	/home/prh44/r...mber.uid_A001_X2fb_X3ec/product	130.621791667	12.3133888889
RUN	/home/prh44/r...mber.uid_A001_X2fb_X3c0/product	11.5985416667	-28.6298055556
RUN	/home/prh44/r...mber.uid_A001_X2fb_X3f0/product	157.612916667	5.41527777778
RUN	/home/prh44/r...mber.uid_A001_X2fb_X3c4/product	25.6822083333	-33.4626388889
RUN	/home/prh44/r...mber.uid_A001_X2fb_X440/product	231.657666667	-20.8335277778
RUN	/home/prh44/r...mber.uid_A001_X2fb_X40c/product	196.534416667	3.94063888889
RUN	/home/prh44/r...mber.uid_A001_X2fb_X410/product	217.089125	-16.0453611111
RUN	/home/prh44/r...mber.uid_A001_X2fb_X3fc/product	177.013708333	7.03563888889
RUN	/home/prh44/r...mber.uid_A001_X2fb_X41c/product	315.227583333	-17.25625
RUN	/home/prh44/r...mber.uid_A001_X2fb_X428/product	349.576458333	-31.2295555556
+			

# Example

CWL allows much more complex workflows, invoking multiple steps



# Templates/Built in actions

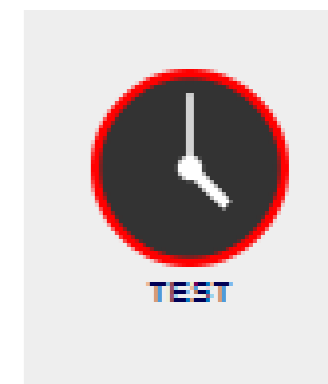
Basic operations (e.g. bash 'find')



Access to services



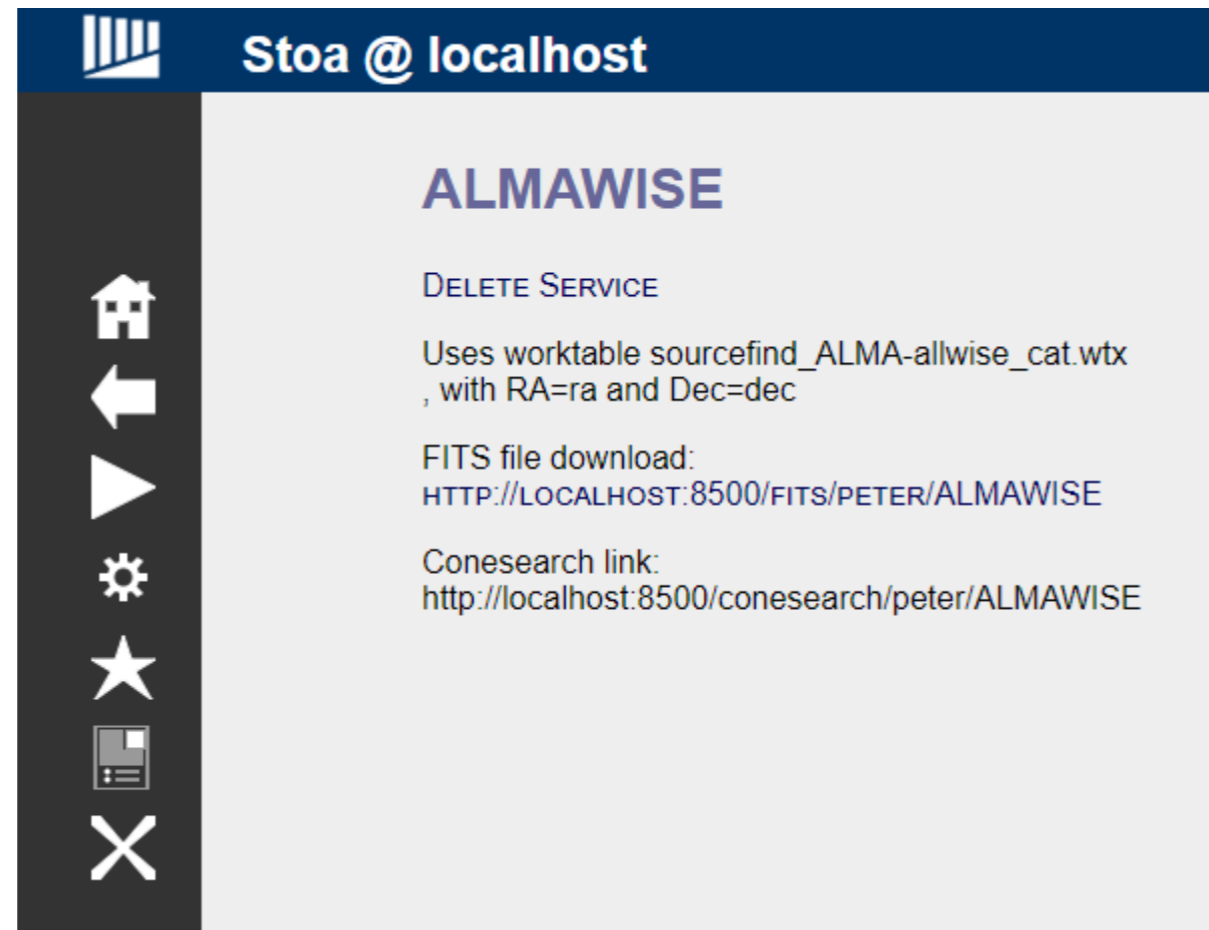
Time trigger – initiate computation at regular intervals



Public table server



# Example



Current public table interface

# The Future

- Authentication – can't have public STOA services without it!
- Figure out how STOA fits in with current services/registries
- Expand the action library
- Help users to build worktables more easily
- Look at more integration with other software – specifically for provenance and minimal recomputation

# Summary

<https://www.github.com/petehague/STOA>

Available under APACHE license

Contact: [prh44@cam.ac.uk](mailto:prh44@cam.ac.uk)