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).

My OBELICS Goals

- WP 3.3 D-INT (Data integration) - Development of **STOA** to prototype workflow system for future projects
- WP 3.4 D-ANA (Data analysis) – Next generation source finding and characterisation for radio astronomy - **Basc**

Background - STOA

- Find additional value in already processed data – initially the ALMA archive
- Do surveys across multiple archived projects
- Compare with observations at other wavelengths not considered in initial project

ALMACAL I: FIRST DUAL-BAND NUMBER COUNTS FROM A DEEP AND WIDE ALMA SUBMM SURVEY, FREE FROM COSMIC VARIANCE

I. Oteo[1,2], M. A. Zwaan[2], R. J. Ivison[2,1], I. Smail[3], and A. D. Biggs[2]
[1]Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ UK
[2]European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany and
[3]Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE UK
Draft version August 24, 2015

ABSTRACT

We have exploited ALMA calibration observations to carry out a novel, wide and deep submm survey, ALMACAL. These calibration data comprise a large number of observations of calibrator fields in a variety of frequency bands and array configurations. Gathering together data acquired during multiple visits to many ALMA calibrators, it is possible to reach noise levels which allow the detection of faint dusty, star-forming galaxies (DSFGs) over a significant area. In this paper we outline our survey strategy and report the first results. We have analysed data for 69 calibrators, reaching depths of $\sim 25\,\mu$Jy beam^{-1} at sub-arcsec resolution. Adopting a conservative approach based on $\geq 5\sigma$ detections, we have found eight and 11 DSFGs in ALMA bands 6 and 7, respectively, with flux densities $S_{1.2\mathrm{mm}} \geq 0.2$ mJy. The faintest galaxies would have been missed by even the deepest *Herschel* surveys. Our cumulative number counts have been determined independently at 870 μm and 1.2 mm, from a sparse sampling of the astronomical sky, and are thus relatively free of cosmic variance. The counts are lower than reported previously by a factor of at least 2×. Future analyses will yield large, secure samples of DSFGs, with redshifts determined via detection of submm spectral lines. Uniquely, our strategy then allows morphological studies of very faint DSFGs – representative of more normal star-forming galaxies than conventional submm galaxies (SMGs) – in fields where self-calibration is feasible, yielding milliarcsecond spatial resolution.

Subject headings: galaxy evolution; submm galaxies; dust emission; number counts

Background - STOA

- Important metadata about observations is not carried forward into ALMA image products
- In order to do batch work across heterogenous archive, need a uniform metadata format
- Products are always contained in a single folder. Use XML file to direct later stages of the pipeline
- Groups FITS files into ones that pertain to a single image and includes identifying information

```
<project>
<group key="Products">
  <group key="AGN39_CS_AGN51">
    <file FILETYPE="correctedImage"  NAXIS1="300" NAXIS2="300" NAXIS3="1"
    BMAJ="0.00039497964912" BMIN="0.000282475451628" BPA="-81.8802185059"
    BTYPE="Intensity" OBJECT="CS_AGN51" CTYPE1="RA---SIN"
    CRVAL1="149.955829167" CDELT1="-5.55555555556e-05" CRPIX1="151.0"
    CUNIT1="deg" CTYPE2="DEC--SIN" CRVAL2="2.02806388889"
    CDELT2="5.55555555556e-05" CRPIX2="151.0" CUNIT2="deg" CTYPE3="FREQ"
    CRVAL3="3.43494693736e+11" CDELT3="16015783495.4" CRPIX3="1.0" CUNIT3="Hz"
    DATE-OBS="2014-12-31T09:42:21.840000">AGN39_CS_AGN51.image.pbcor.fits</file>
    <file FILETYPE="primaryBeam"  NAXIS1="300" NAXIS2="300" NAXIS3="1" BMAJ=""
    BMIN="" BPA="" BTYPE="Intensity" OBJECT="CS_AGN51" CTYPE1="RA---SIN"
    CRVAL1="149.955829167" CDELT1="-5.55555555556e-05" CRPIX1="151.0"
    CUNIT1="deg" CTYPE2="DEC--SIN" CRVAL2="2.02806388889"
    CDELT2="5.55555555556e-05" CRPIX2="151.0" CUNIT2="deg" CTYPE3="FREQ"
    CRVAL3="3.43494693736e+11" CDELT3="16015783495.4" CRPIX3="1.0" CUNIT3="Hz"
    DATE-OBS="2014-12-31T09:42:21.840000">AGN39_CS_AGN51.flux.fits</file>
</group>
```

This observation consists of a primary beam corrected image and the flux profile used to correct it. Some FITS header information is also reproduced

Background - Basc

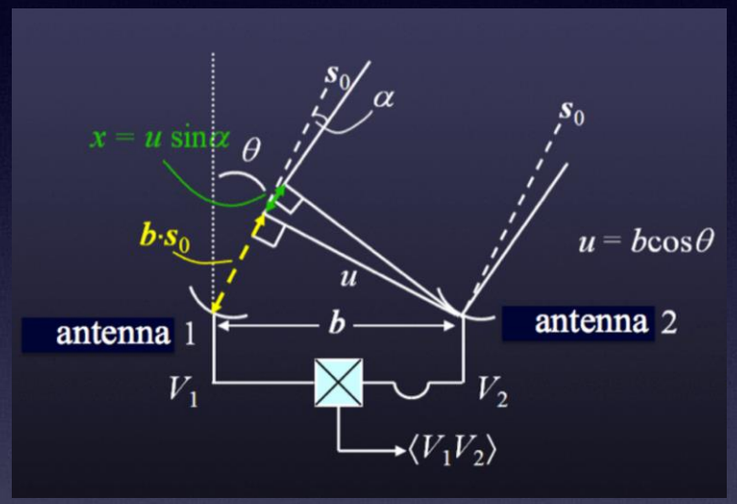
From the SExtractor manual:

DETECT_TYPE	CCD	*keyword*	Type of device that produced the image:
		CCD	– linear detector like CCDs or NICMOS,
		PHOTO	– photographic scan.

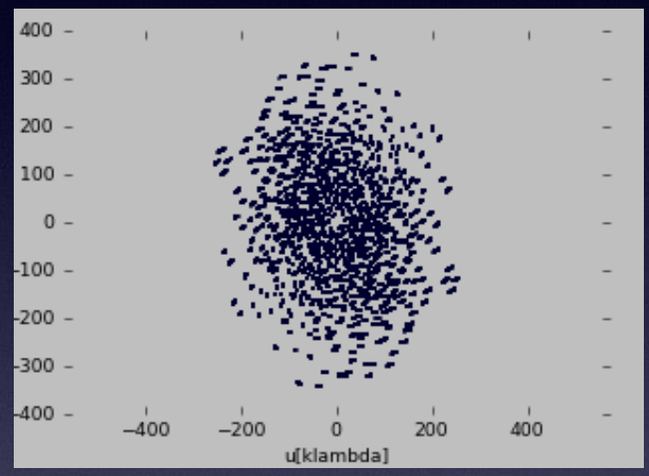
https://www.astromatic.net/pubsvn/software/sextractor/trunk/doc/sextractor.pdf

Background - Basc

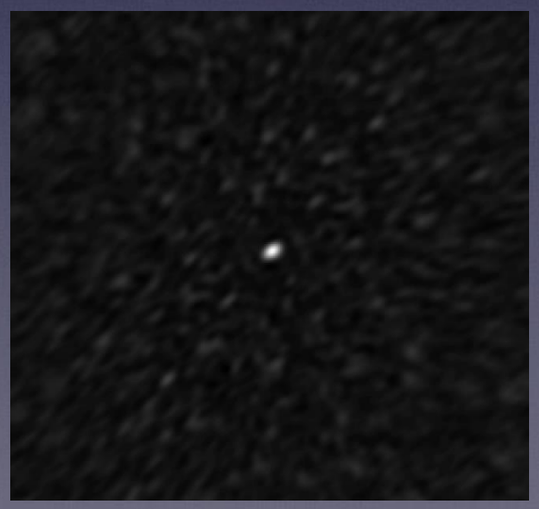
Each pair of antennas gives a component of the Fourier transform of the sky brightness



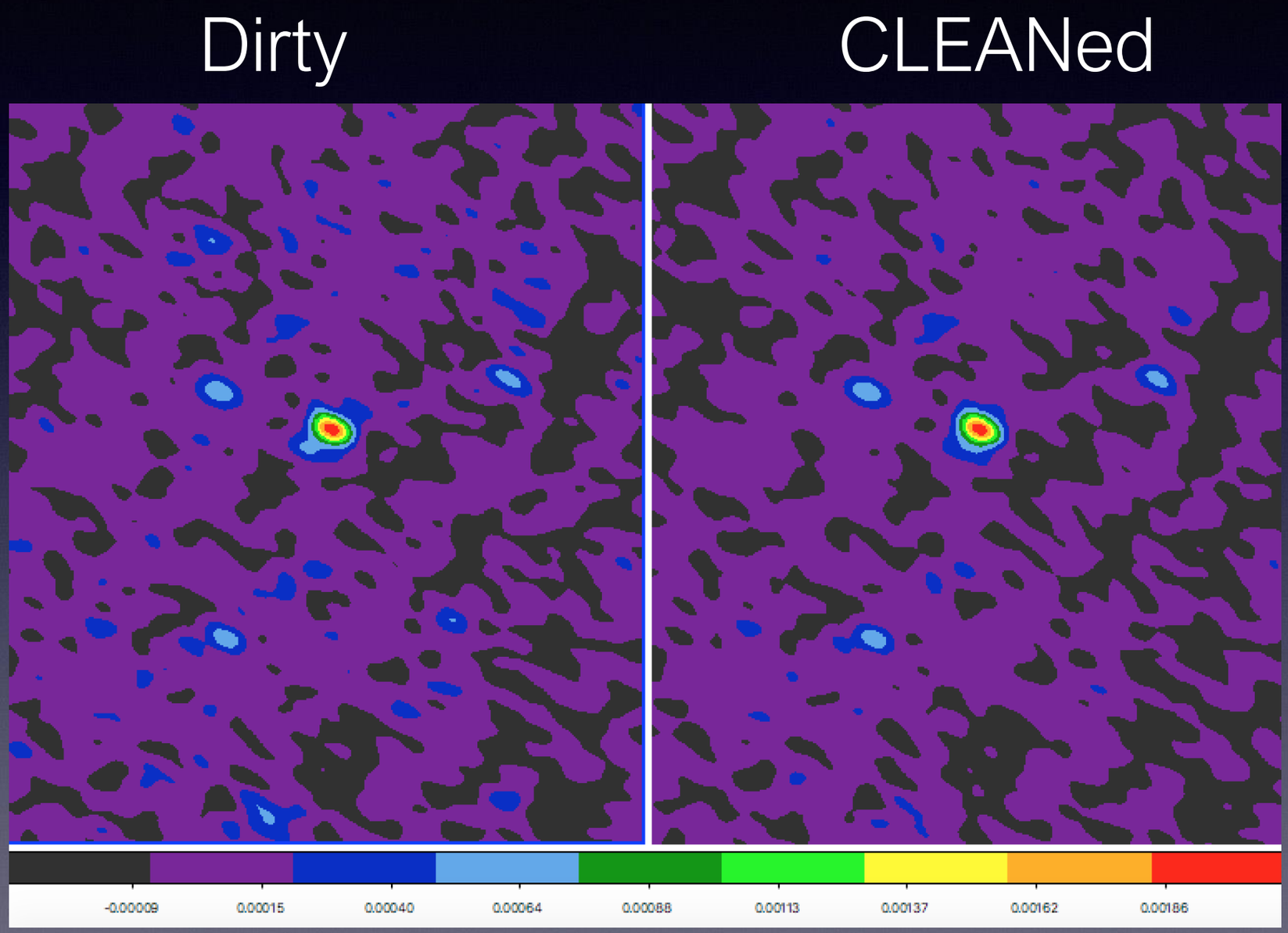
Use many antennas to get good coverage…



And invert it to get a map of the sky



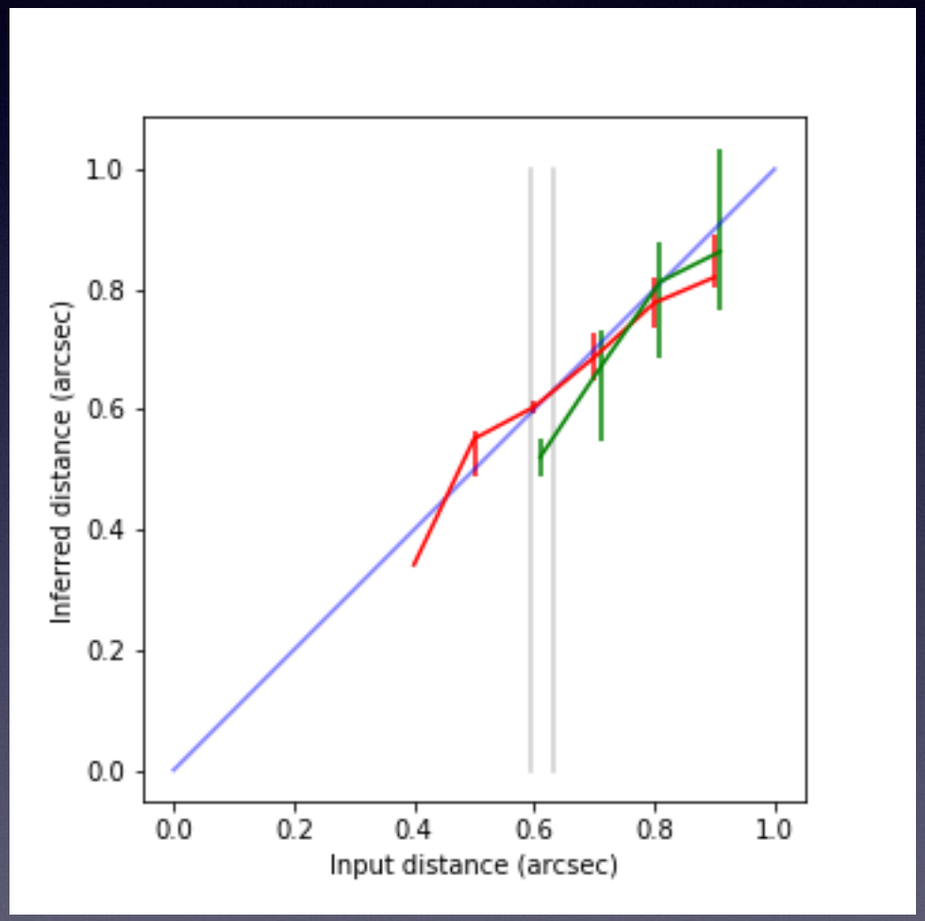
Background - Basc



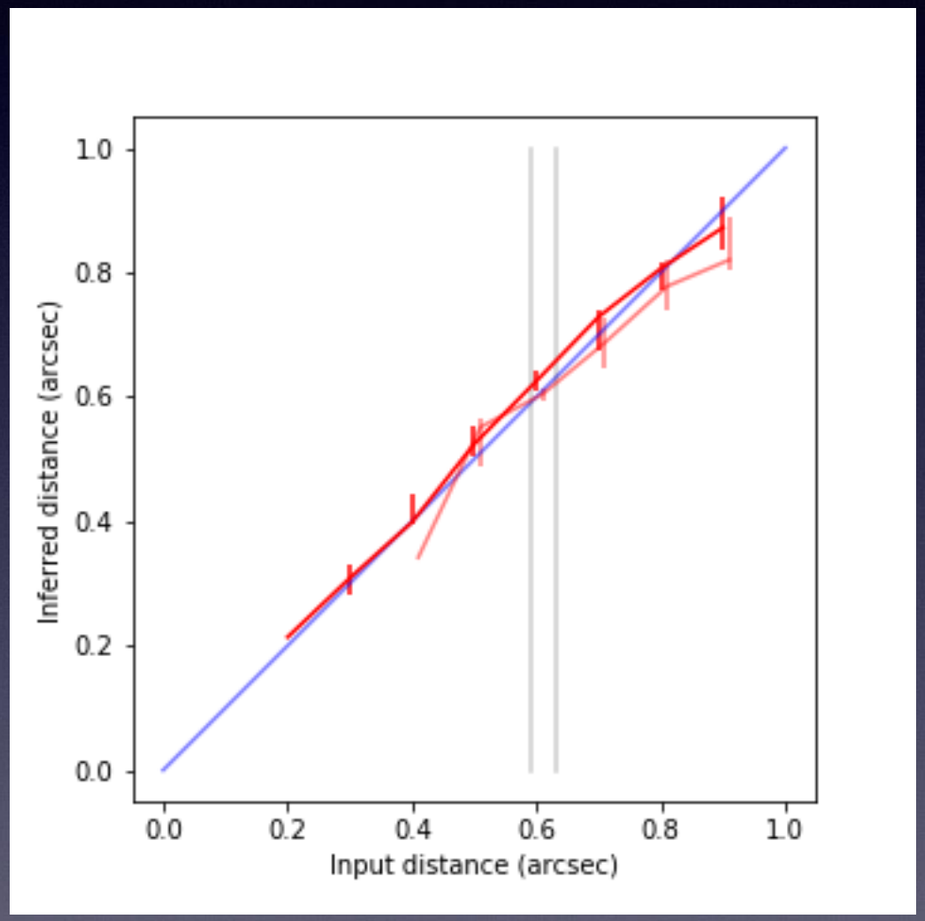
Data from project 2013.1.00911.S, PI Dragan Salak

Background - Basc

Point source discrimination



vs. SExtractor

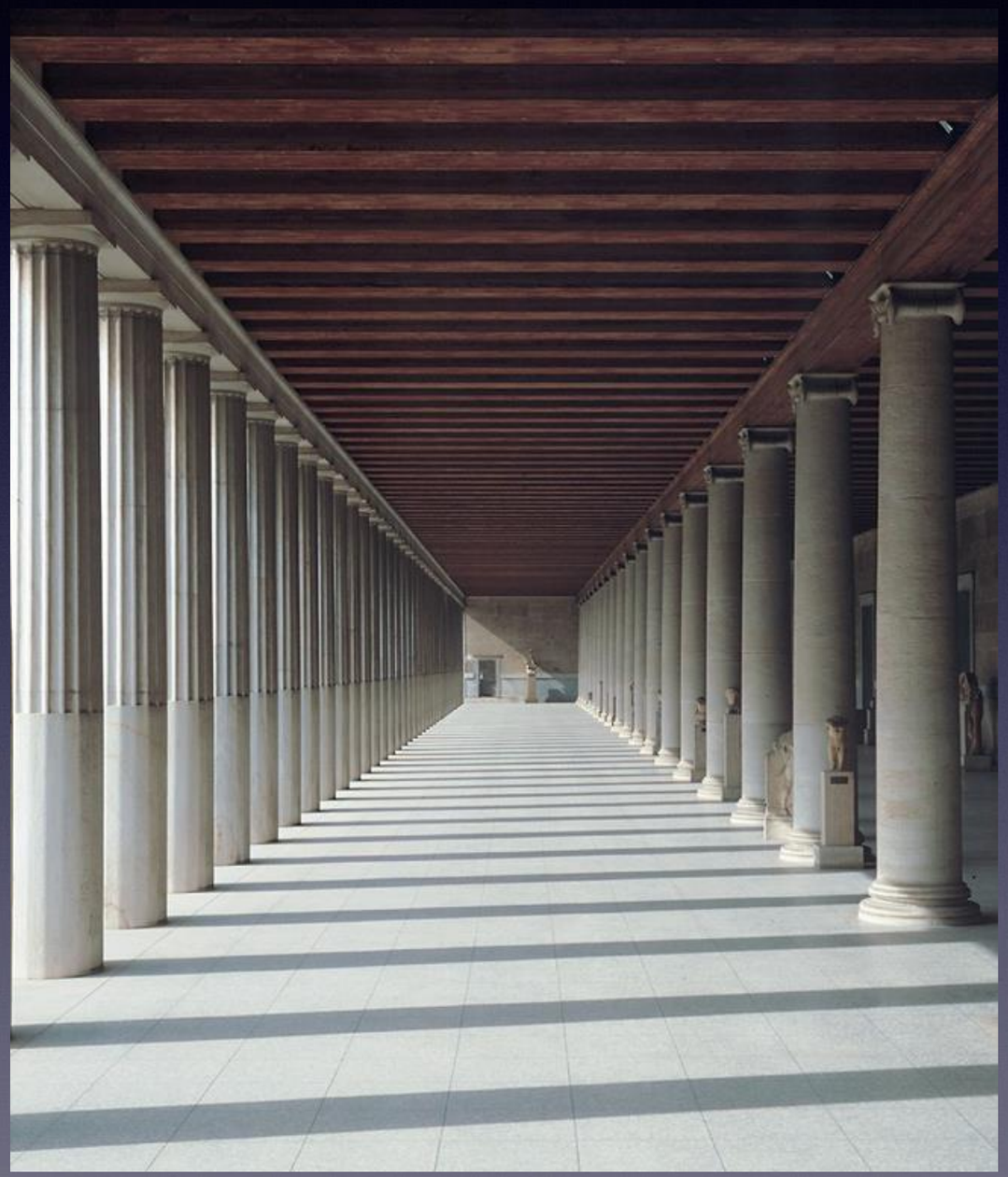


40x difference in flux

Hague et al in preparation

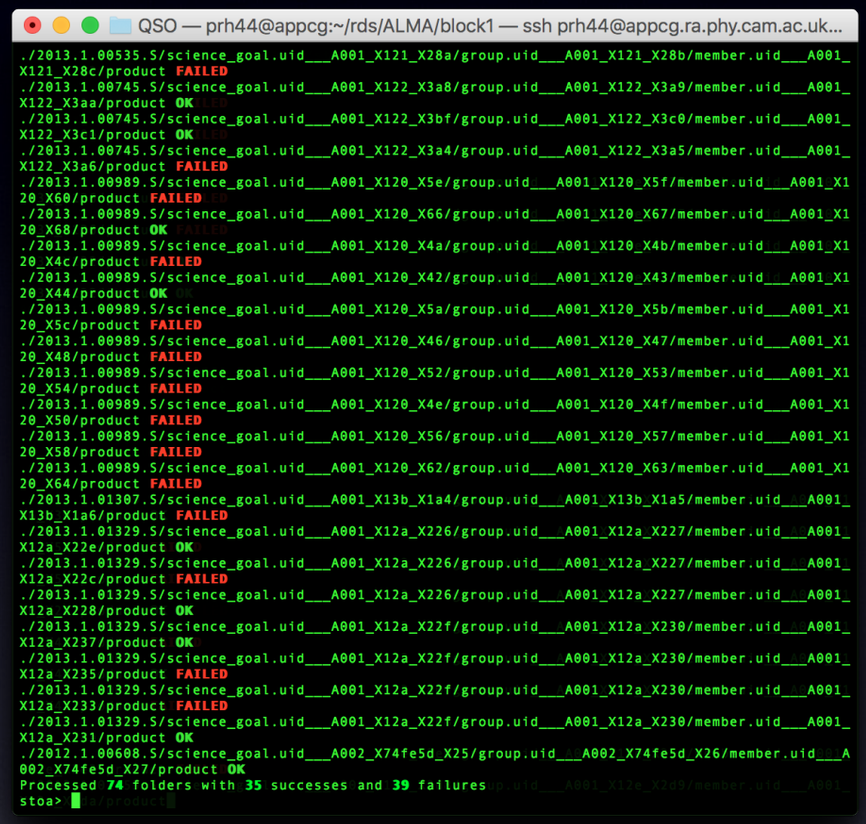
STOA - Script Tracking for Observational Astronomy

- Process management system
- Runs scripts on multiple sets of data, each time with different parameters and a different environment
- Collaboration features - can flag and annotate data and products
- Interfaces with existing astronomy software (e.g. TOPCAT)

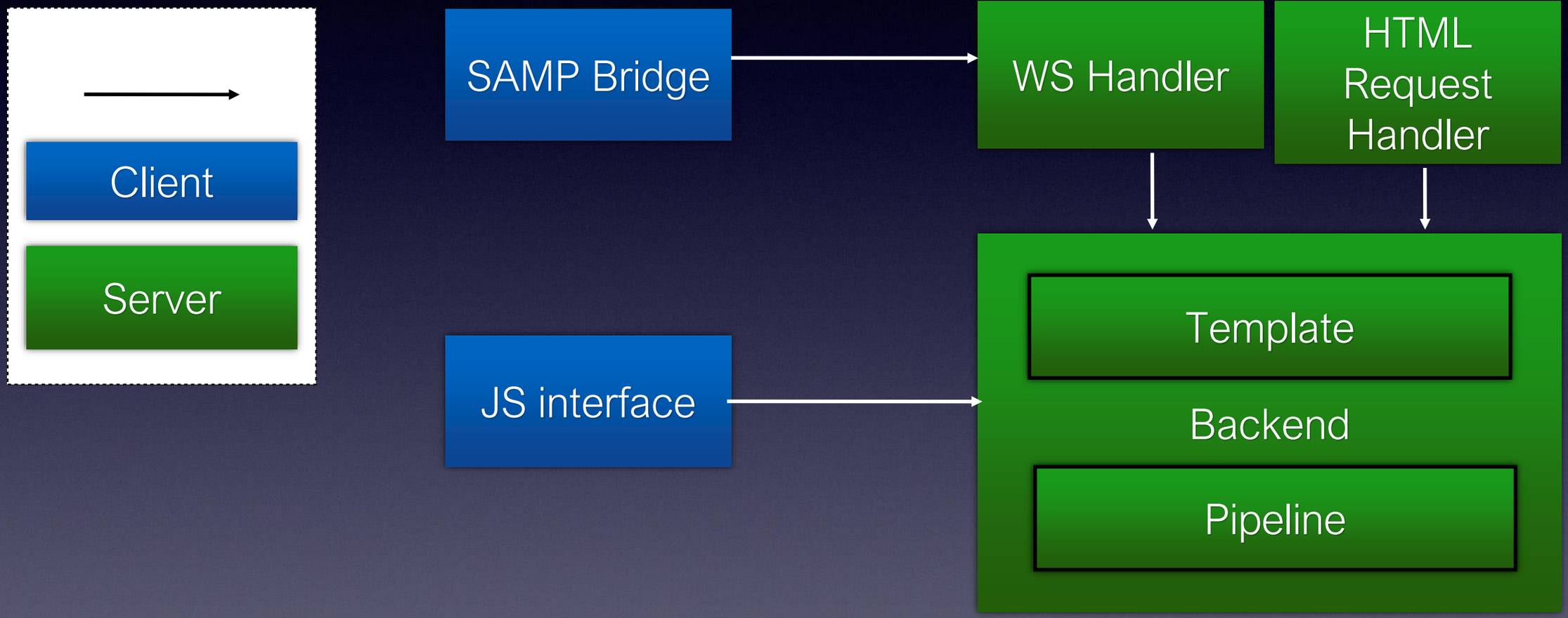


Design

- Command line written in Python, with web interface using Tornado
- Run a process on multiple targets, review targets (e.g. products folder) that have failed and flag targets after manual inspection of data
- Modify process and/or parameter. Rerun process on failed and flagged targets

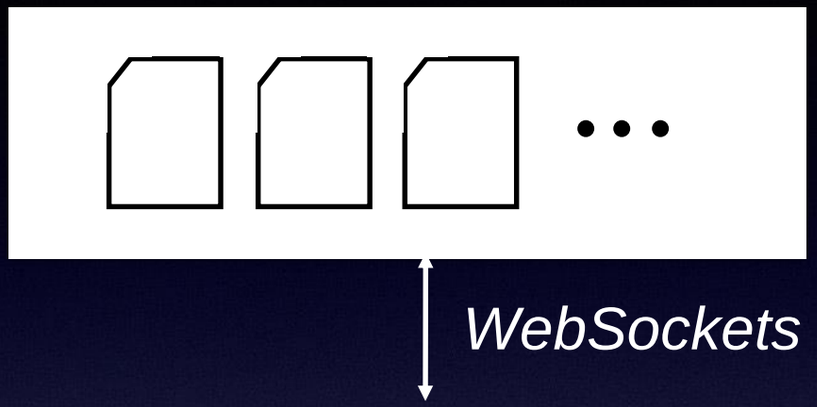


Design


Client
Server
SAMP Bridge
JS interface
WS Handler
HTML Request Handler
Template
Backend
Pipeline

Design

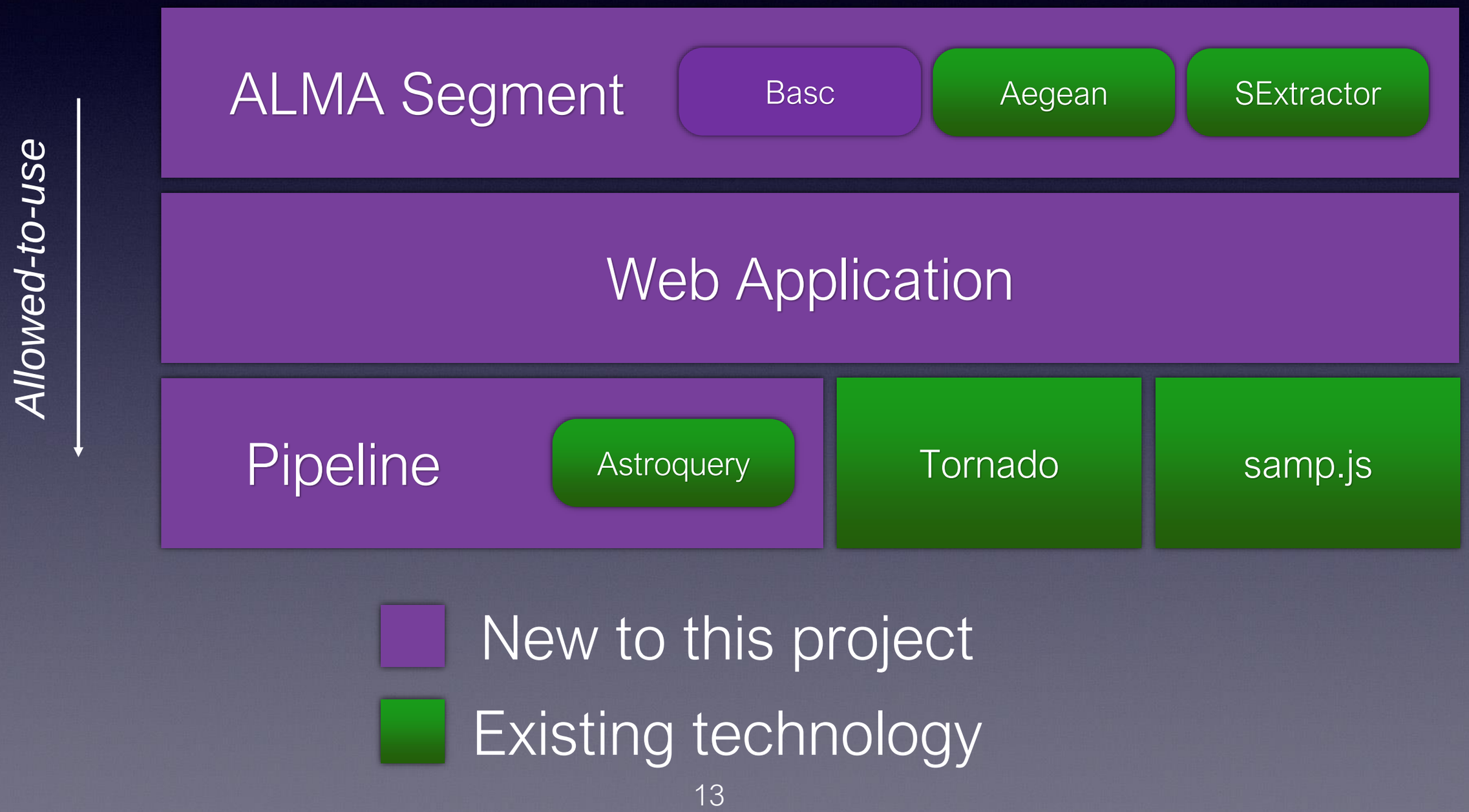
- Can organise the results by observation, by matches with existing catalogue.
- Multiple users can work on the same data set and can flag data for others inspection, and add comments
- Minimum re-computation – don't rerun the entire batch, or even the entire observation, to update the output.
- Allows sharing of data tables with local apps via SAMP bridge



WebSockets

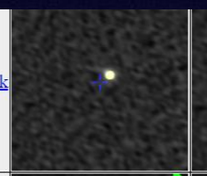
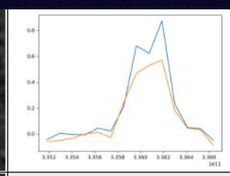
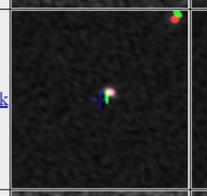
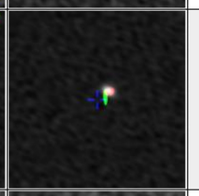
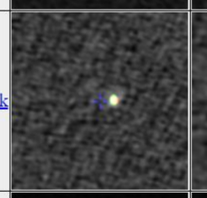
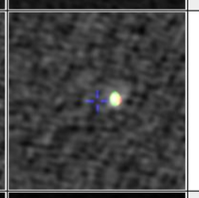
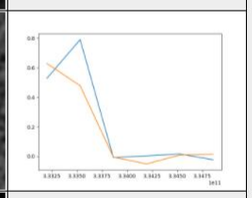
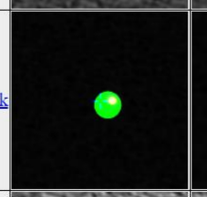
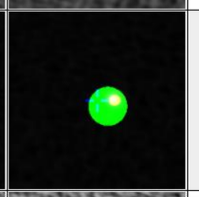
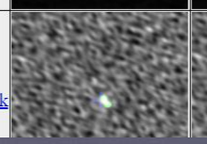
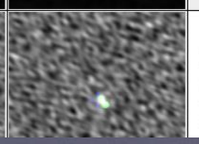
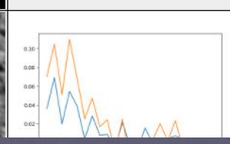
SAMP + WS

Design


ALMA Segment
Basc
Aegean
SExtractor
Web Application
Pipeline
Astroquery
Tornado
samp.js
Allowed-to-use
New to this project
Existing technology

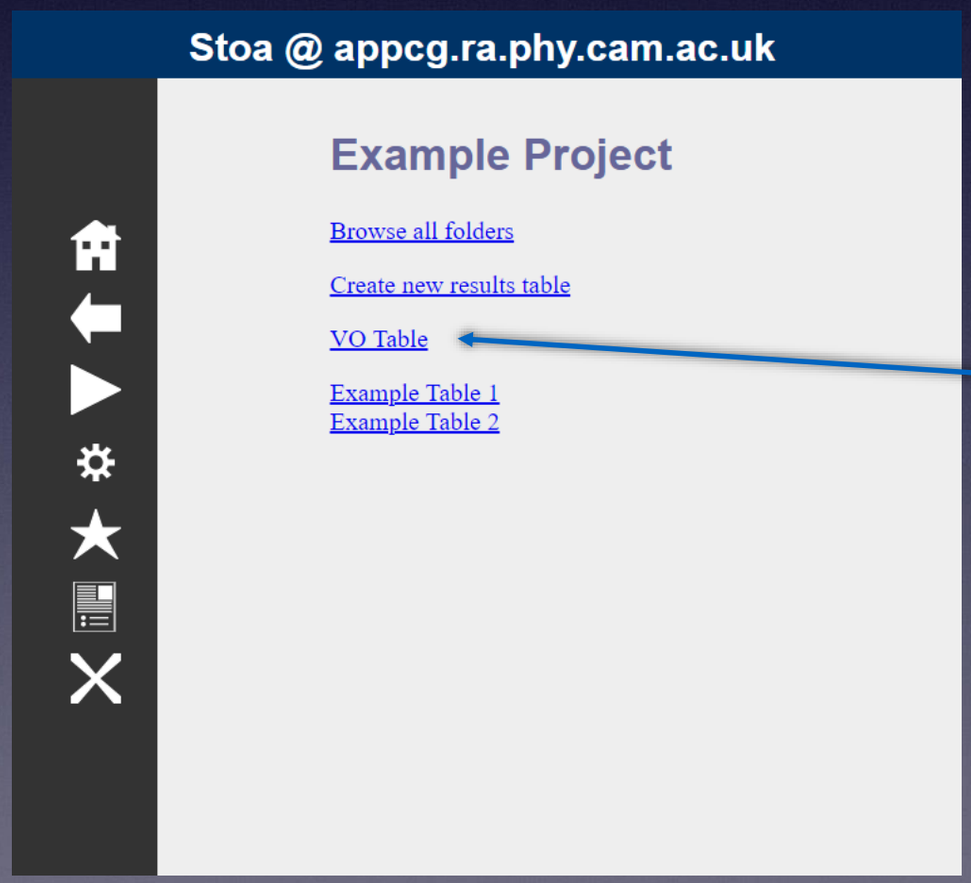
Tables

- Tables both describe the output and also control workflows
- Each row, either explicitly in a cell or implicitly, should describe how the workflow behind it operates
- Still under development…

SDSS J09230+0247	140.764583333333332	2.7941666666666665	Image	NED Link			
			Image	NED Link			
SDSS J13415+0141	205.392499999999998	1.6991666666666665	Image	NED Link			
			Image	NED Link			
SDSS J09351+0801	143.78541666666663	8.020555555555555	Image	NED Link			

Generic Fork

Available from
https://github.com/petehague/stoa



Dockerfile for quick demo

Example VO interaction

Will be porting my own projects ASAP

Current Decisions

- What is the best way to make this widely available?
- How should it connect to other services?
- How can I avoid generating superfluous standards
- What is the best data/workflow representation?