WEAVE data products and overall dataflow

David Murphy
CASU, Institute of Astronomy
Cambridge University

WEAVE Science Verification Community Workshop
Tenerife 14th-15th November 2019

This session: *aim* and overview

- *Build on the broad outline of WEAVE dataflow given by Shoko*

- The main components within the WEAVE data model
- How they interact
- WEAVE data products: Input and Output
- How you get them

This session: *aim* and overview

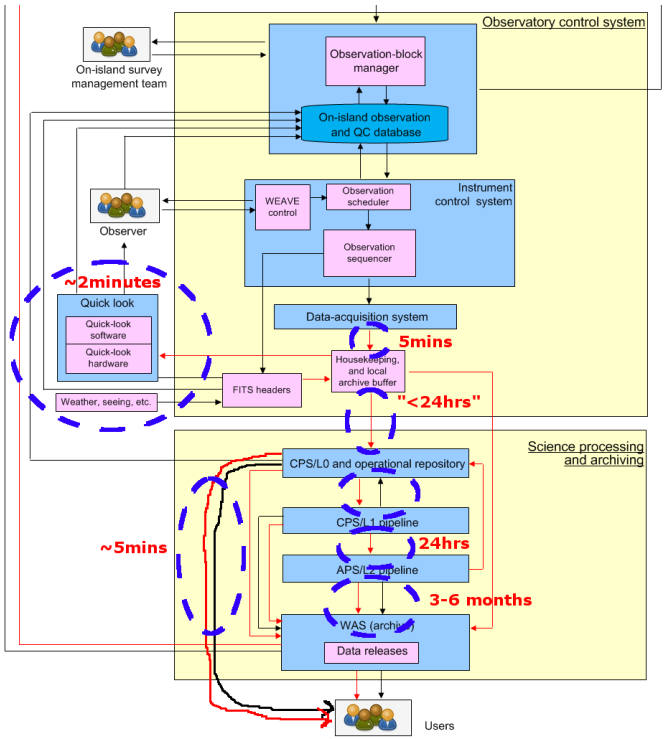
- *Build on the broad outline of WEAVE dataflow given by Shoko*

- The main components within the WEAVE data model
- How they interact
- WEAVE data products: Input and Output
- How you get them

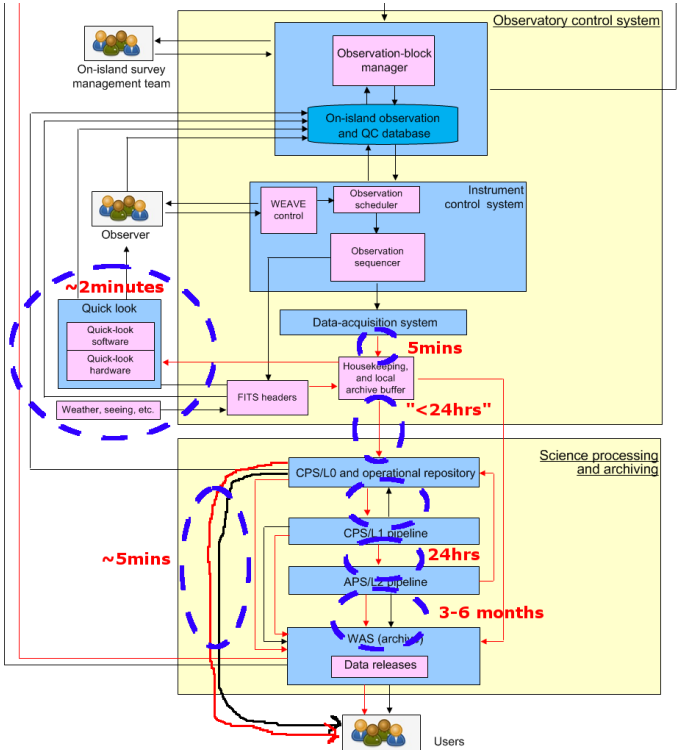
- This isn't a lecture. Ask questions, ask for clarification etc etc

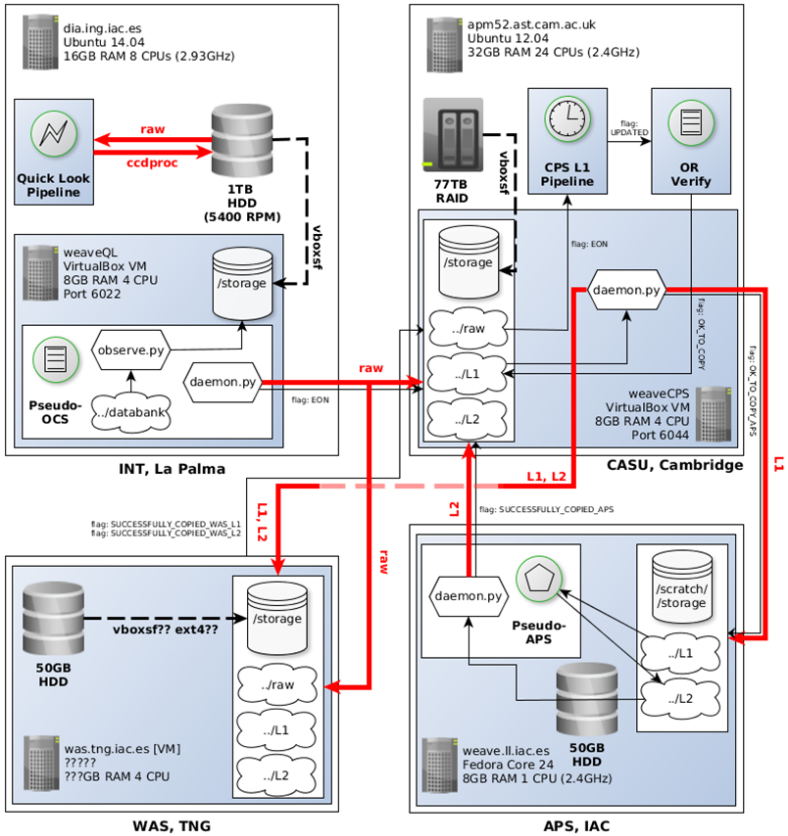
- The WEAVE dataflow model is complicated

- The WEAVE dataflow model is complicated

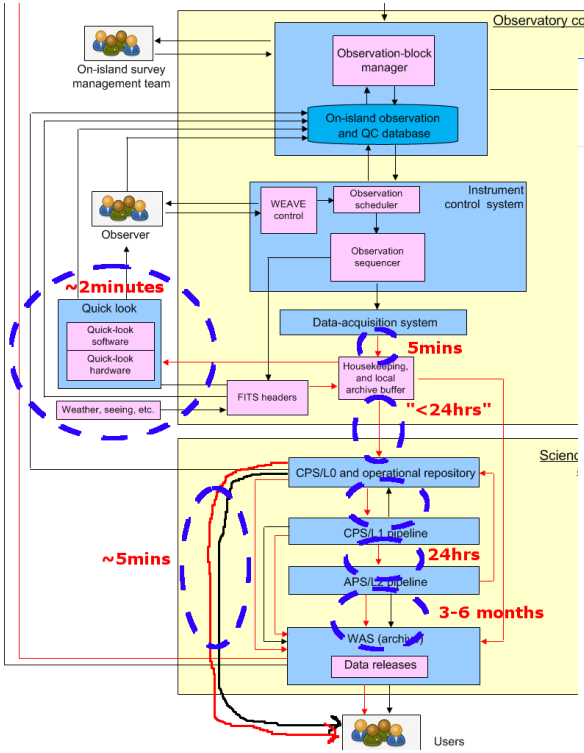


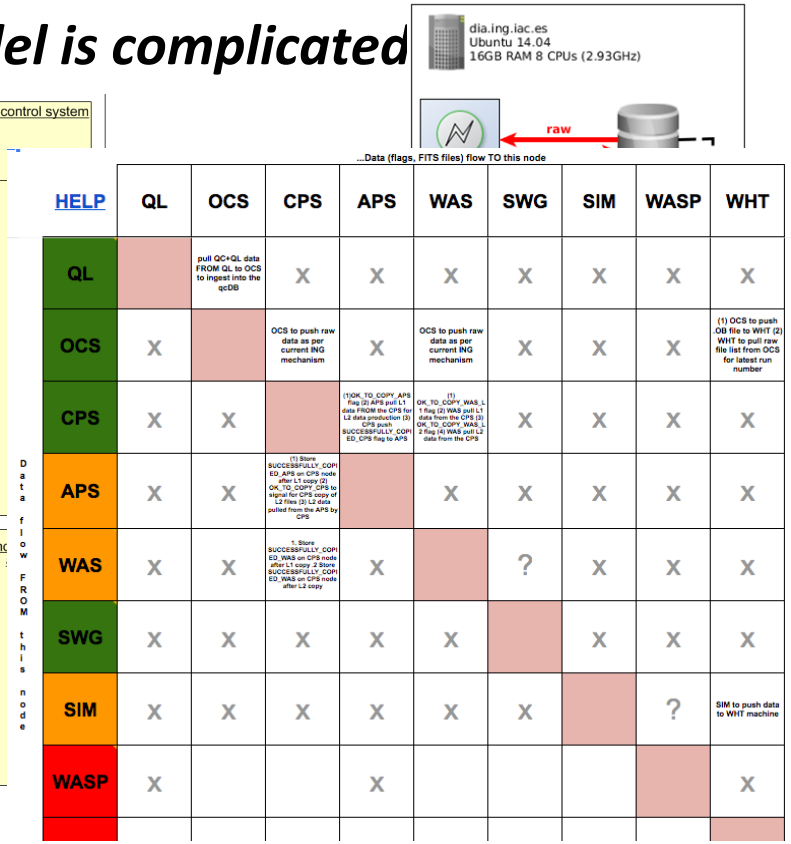
- *The WEAVE dataflow model is complicated*

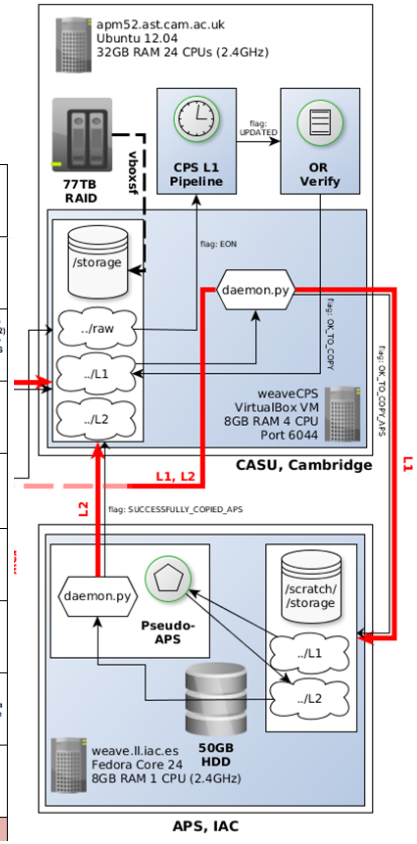




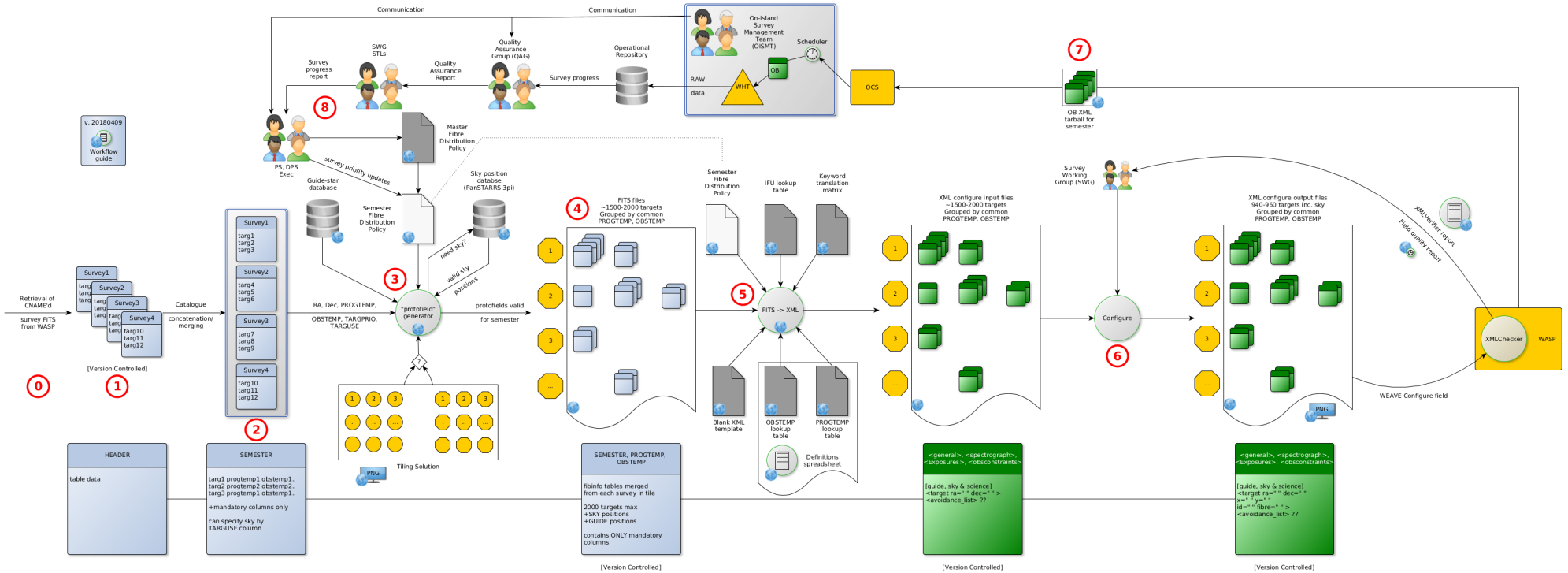
- ***The WEAVE dataflow model is complicated***



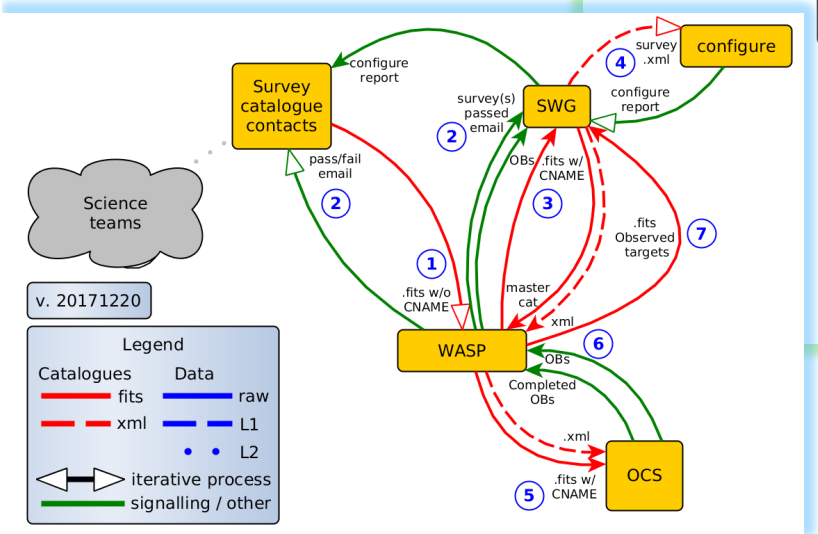


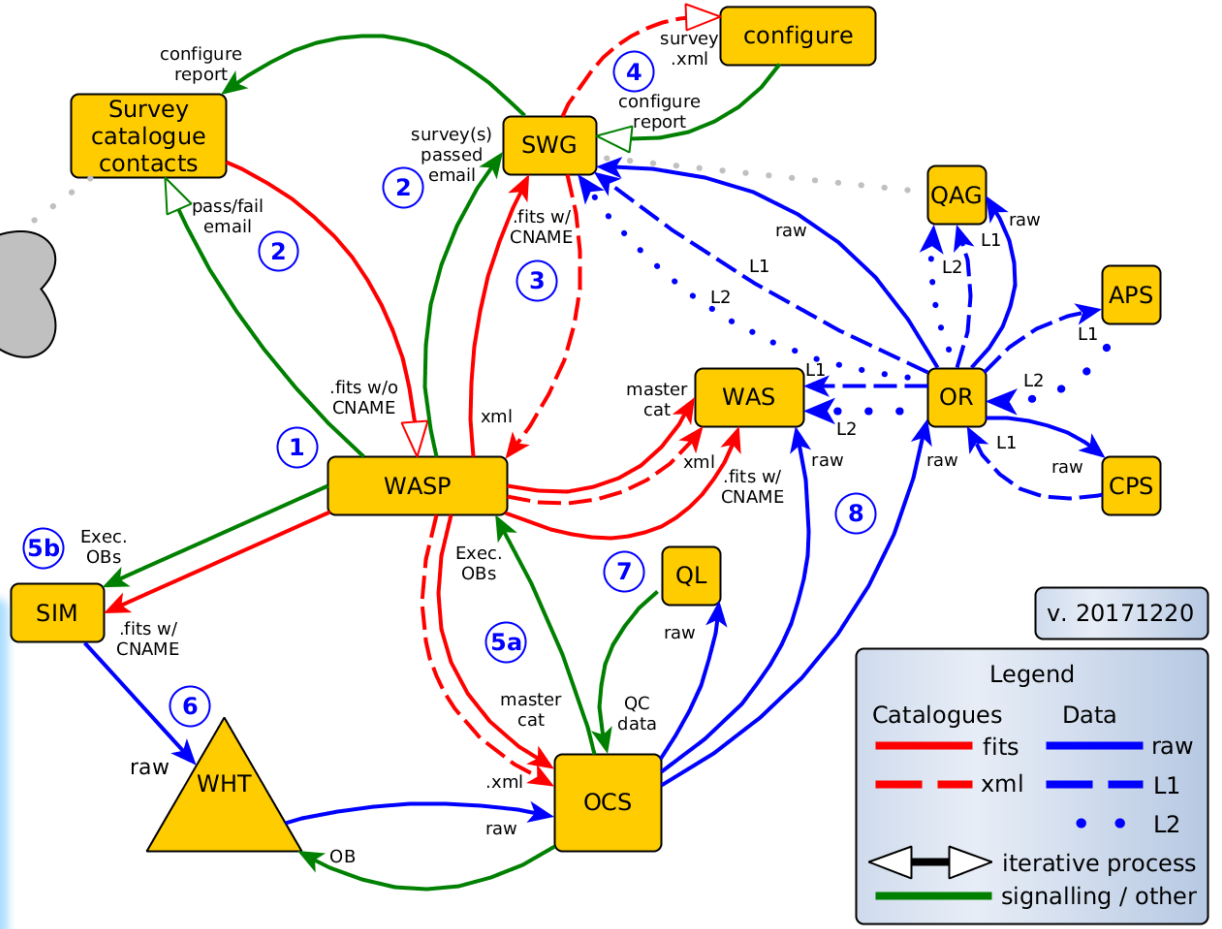


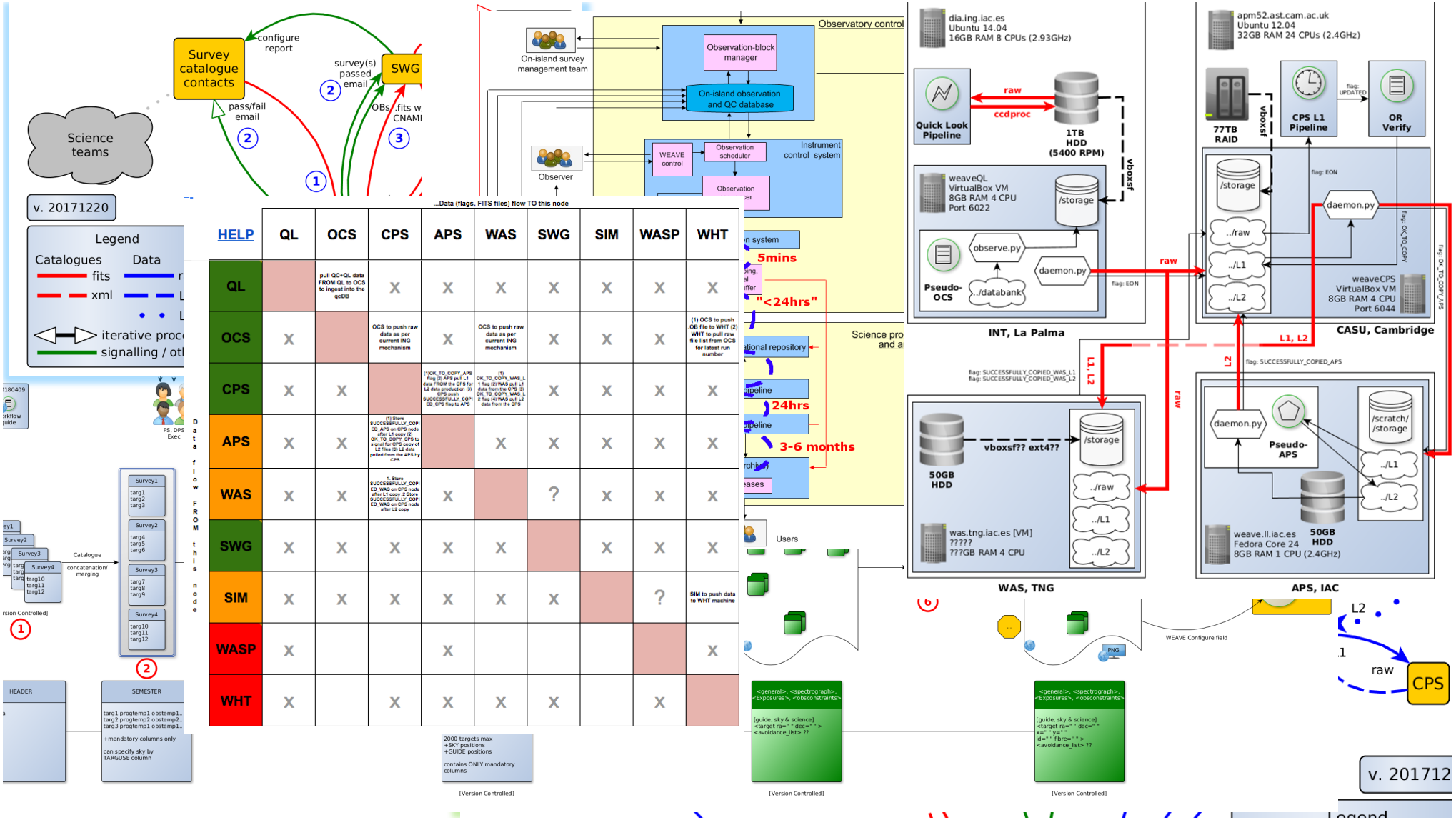
- ***The WEAVE dataflow model is complicated***



- ***THE WEAVE DA***

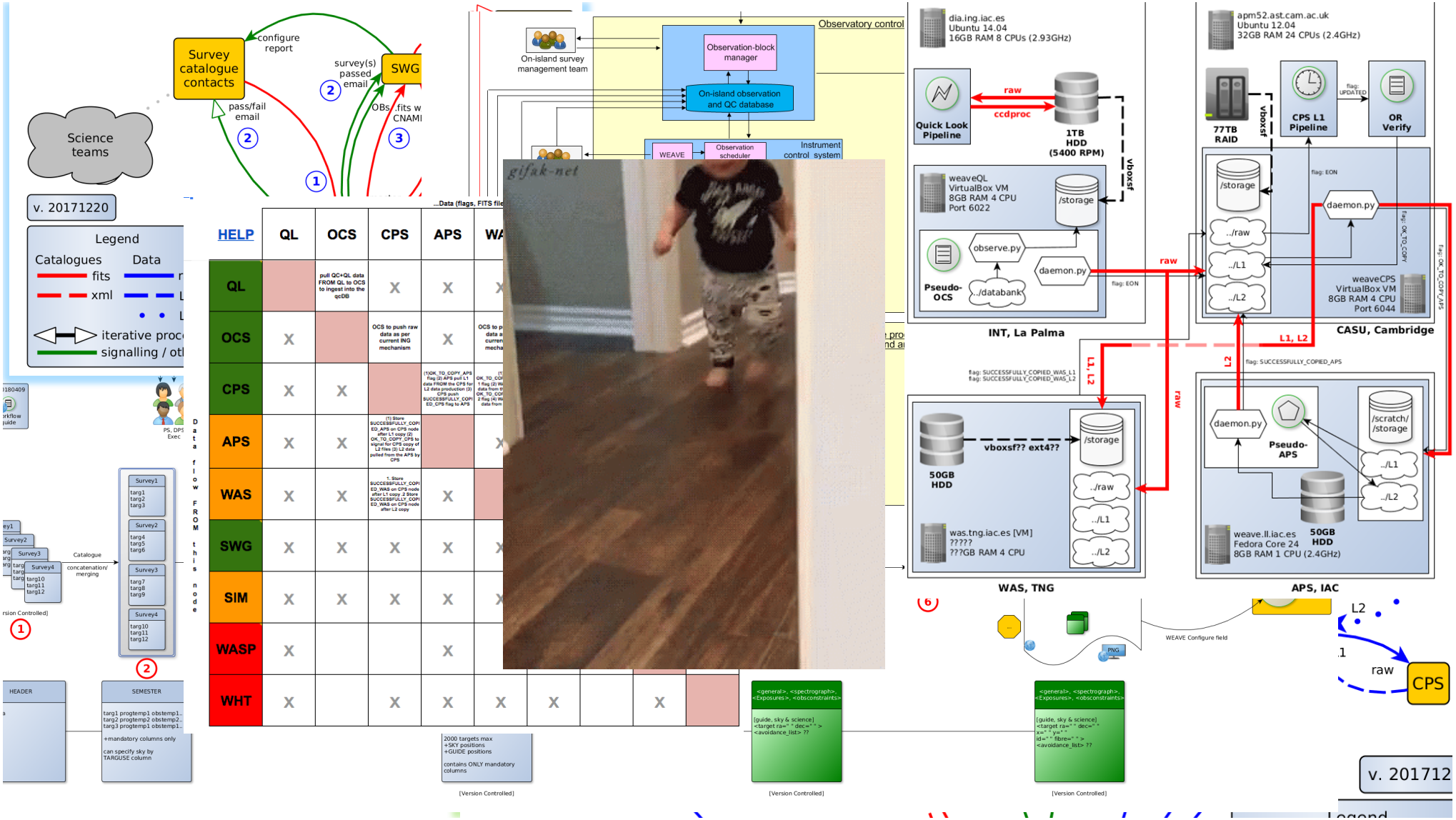

Science teams
Survey catalogue contacts
configure
configure report
survey .xml
SWG
survey(s) passed email
pass/fail email
.fits w/o CNAME
.fits w/ CNAME
WASP
OBs
master cat
xml
.fits Observed targets
Completed OBs
OCS
v. 20171220
Legend
Catalogues
Data
fits
raw
xml
L1
L2
iterative process
signalling / other


Science teams
Survey catalogue contacts
configure
configure report
survey .xml
SWG
survey(s) passed email
pass/fail email
.fits w/ CNAME
.fits w/o CNAME
WASP
xml
master cat
WAS
OR
QAG
APS
CPS
raw
L1
L2
SIM
Exec. OBs
.fits w/ CNAME
QL
QC data
WHT
OB
OCS
.xml
v. 20171220
Legend
Catalogues
Data
fits
raw
xml
L1
L2
iterative process
signalling / other



...Data (flags, FITS files) flow TO this node

HELP	QL	OCS	CPS	APS	WAS	SWG	SIM	WASP	WHT
QL		pull QC+QL data FROM QL to OCS to ingest into the qcDB	X	X	X	X	X	X	X
OCS	X		OCS to push raw data as per current ING mechanism	X	OCS to push raw data as per current ING mechanism	X	X	X	(1) OCS to push .OB file to WHT (2) WHT to pull raw file list from OCS for latest run number
CPS	X	X		(1)OK_TO_COPY_APS flag (2) APS pull L1 data FROM the CPS for L2 data production (3) CPS push SUCCESSFULLY_COPIED_CPS flag to APS	(1) OK_TO_COPY_WAS_L1 flag (2) WAS pull L1 data from the CPS (3) OK_TO_COPY_WAS_L2 flag (4) WAS pull L2 data from the CPS	X	X	X	X
APS	X	X	(1) Store SUCCESSFULLY_COPIED_APS on CPS node after L1 copy (2) OK_TO_COPY_CPS to signal for CPS copy of L2 files (3) L2 data pulled from the APS by CPS		X	X	X	X	X
WAS	X	X	1- Store SUCCESSFULLY_COPIED_WAS on CPS node after L1 copy .2 Store SUCCESSFULLY_COPIED_WAS on CPS node after L2 copy	X		?	X	X	X
SWG	X	X	X	X	X		X	X	X
SIM	X	X	X	X	X	X		?	SIM to push data to WHT machine
WASP	X			X					X
WHT	X		X	X	X	X		X	


Science teams
Survey catalogue contacts
configure report
survey(s) passed email
pass/fail email
SWG
v. 20171220
Legend
Catalogues
Data
fits
xml
iterative proc
signalling / ot
HELP
QL
OCS
CPS
APS
WAS
SWG
SIM
WASP
WHT
pull QC+QL data FROM QL to OCS to ingest into the qcDB
OCS to push raw data as per current ING mechanism
Observatory control
On-island survey management team
Observation-block manager
On-island observation and QC database
WEAVE
Observation scheduler
Instrument control system
dia.ing.iac.es
Ubuntu 14.04
16GB RAM 8 CPUs (2.93GHz)
Quick Look Pipeline
raw
ccdproc
1TB HDD (5400 RPM)
vboxsf
weaveQL VirtualBox VM
8GB RAM 4 CPU
Port 6022
/storage
observe.py
daemon.py
Pseudo-OCS
../databank
INT, La Palma
apm52.ast.cam.ac.uk
Ubuntu 12.04
32GB RAM 24 CPUs (2.4GHz)
77TB RAID
CPS L1 Pipeline
OR Verify
flag: UPDATED
flag: EON
../raw
../L1
../L2
weaveCPS VirtualBox VM
8GB RAM 4 CPU
Port 6044
CASU, Cambridge
L1, L2
L2
flag: SUCCESSFULLY_COPIED_APS
flag: SUCCESSFULLY_COPIED_WAS_L1
flag: SUCCESSFULLY_COPIED_WAS_L2
vboxsf?? ext4??
50GB HDD
was.tng.iac.es [VM]
????
???GB RAM 4 CPU
WAS, TNG
Pseudo-APS
/scratch/ /storage
weave.ll.iac.es
Fedora Core 24
8GB RAM 1 CPU (2.4GHz)
50GB HDD
APS, IAC
WEAVE Configure field
PNG
CPS
Catalogue concatenation/ merging
Survey1
Survey2
Survey3
Survey4
HEADER
SEMESTER
+mandatory columns only
can specify sky by TARGUSE column
2000 targets max
+SKY positions
+GUIDE positions
contains ONLY mandatory columns
[Version Controlled]
[guide, sky & science]
[Exposures], [obsconstraints]
[Version Controlled]
[guide, sky & science]
[Version Controlled]
v. 201712
Legend

Our aims for this workshop

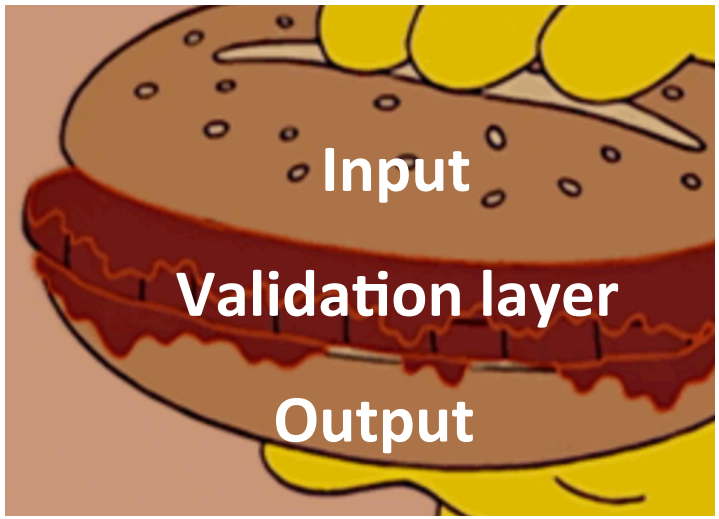
- The WEAVE dataflow model is complicated
- Our mission should be to make using WEAVE as easy to use as possible
- However… WEAVE **is** a complicated instrument – multiple modes, spectrograph arms, resolutions, varied 1D, 2D, 3D data products, complex interplays between different systems
- The overall dataflow system must incorporate all of these, but under a "survey-mode" philosophy that demands science can be done even in adverse conditions, with fast data turnaround and often conflicting survey requirements.

WEAVE dataflow context

- The WEAVE dataflow model is complicated
- We have already implemented the dataflow design, and simulated realistic WEAVE operations under this model
- Implementation encompasses required inputs and the outputs generated by pipelines and analysis packages
- In devising the overall system, necessary balance between flexibility/ features for would-be users vs. the need for streamlined “survey-mode observing” and data treatment
- It’s always the small niggles that catch you (us)…. No questions are stupid, so please ask: you might poke a hole in our model (good!)

WEAVE I/O – an overview

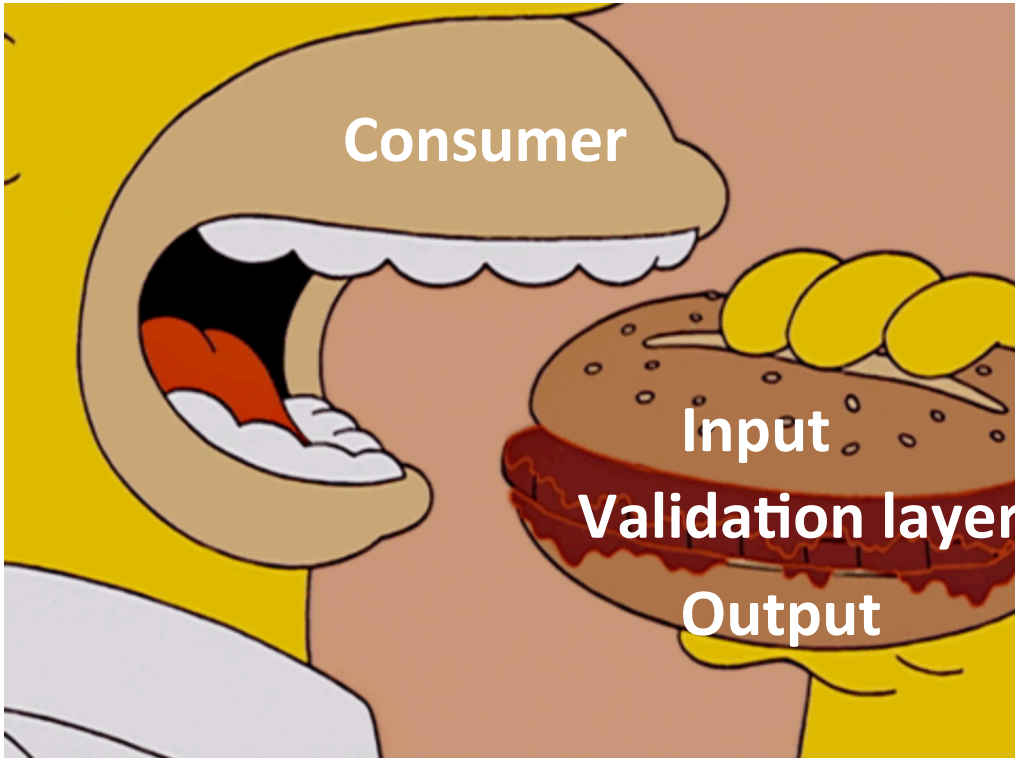
- We will discuss key parts of the dataflow today; in-depth look at SV-relevant components tomorrow.
- The WEAVE data sandwich:



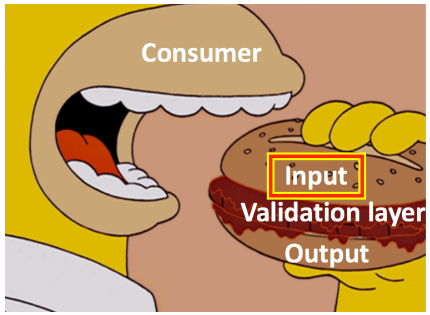
- Basic concept is that user-generated input data are subject to validation against the WEAVE data model
- Data passed to WHT, observations made, output data are also validated (just in case) against the model…

WEAVE I/O – an overview

- … and then made available for consumption.
- There is more than one way to eat the WEAVE data sandwich – we will cover these today, and highlight the differences.
- Let's now look at these 4 components

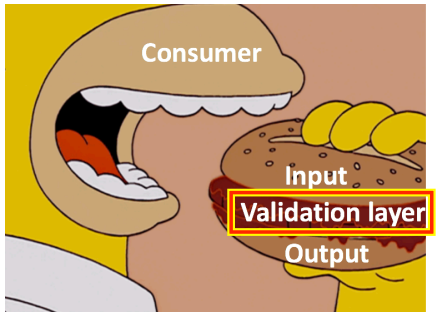


WEAVE I/O – an overview



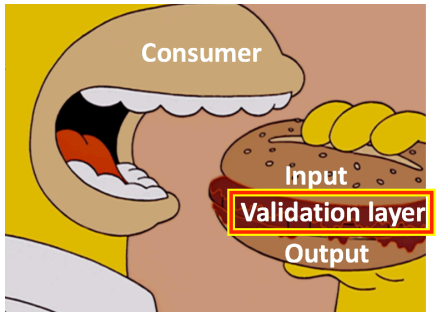
- Input (what users must provide)
  - Input FITS-format catalogues describing the targets you wish to observe
  - XML-format Observing Blocks (OBs) describing the fiber positions, instrument configuration, minimum observing conditions, etc
- For Science Verification:
  - PIs will generate their FITS catalogues
  - WEAVE members will help generate OBs on your behalf
  - PIs will review generated OBs to agree they meet aims of the proposal
- Under normal operations
  - PIs are responsible for FITS and OB submissions, provided with tools to help

WEAVE I/O – an overview



- Validation (how we make sure what's been provided makes sense)
  - Upload facility to a validation platform (think ESO p2pp)
  - Pass / Fail with a generated report: both FITS and XML
  - Tells you the products are legal, but can't check for "common sense"
    - "I only want to allocate 5 of these 940 fibers to my targets" (you might want to!!)
    - "I meant to use NGC in my target names, but used NCG by mistake"
    - "I stuck that mIFU bundle on the wrong galaxy"
    - "This 20th mag galaxy can totally be observed with a sky brightness of 17.5 mags"
  - Warnings *are* issued in some cases, but no substitute for common sense!
  - This facility acts as a gatekeeper between users and the instrument

WEAVE I/O – an overview

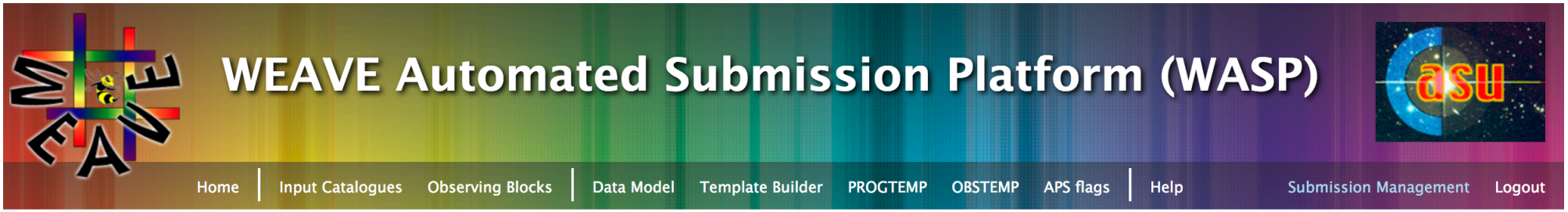


- Validation (how we make sure what's been provided makes sense)

Users → Catalogues → FITS:XML → OBs → WHT → Science-ready data

The WEAVE Automated Submission Platform (**WASP**)

Teaser slides for tomorrow



JobID 1573684536

#	ID	File	User	Field	RA	Dec	PROGTEMP	OBSTEMP	Plate	Sci fibres %	Process State	Result	Report	Download
1	2	3365.xml	weave	PS1_3557	27.0658	27.9602	41331	XECEC	LIFU	<LIFU>	COMPLETED	FAILED	x	[XML]
2	3	3367.xml	weave	PS1_3557	27.0658	27.9602	41331	XECEC	LIFU	<LIFU>	PROCESSING			
3	4	3366.xml	weave	PS1_3557	27.0658	27.9602	41331	XECEC	LIFU	<LIFU>	PENDING			
4	5	3368.xml	weave	PS1_3557	27.0658	27.9602	41331	XECEC	LIFU	<LIFU>	PENDING			

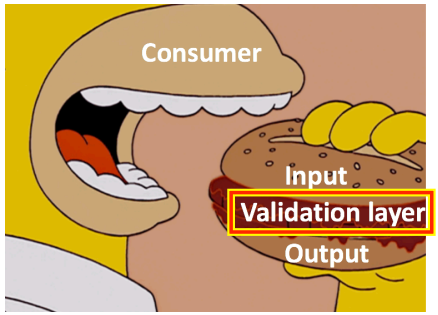
Cambridge Astronomy Survey Unit – Institute of Astronomy
The WEAVE Consortium

casuhelp@ast.cam.ac.uk

XML report for 3367.xml (jobID 1573684536, uid 3)

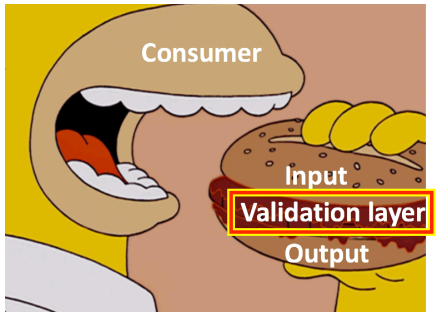
```
================================================================================
Result for file 3367.xml: Failed
# of invalid entries: 2013
# of warnings: 1651
================================================================================
XML File Verification for WASP Compliance
Current Version = OpR3d
Current trimester = 2019A2
Started at: Wed Nov 13 22:37:01 2019
0/ FileIO and XML parsing
================================================================================
VALID: File passed preliminary IO and parsing checks
1/ XML validation
================================================================================
INPUTFILE: 3367.xml
TEMPLATEFILE: 'BlankXMLTemplate.xml' (version 1.13)
WARNING: Version of submitted XML (1.10) does not match version of current template (1.13)
2/ Baseline structure check against template XML
================================================================================
VALID: Submitted XML matches structure and content of template XML
3/ Attribute check against template XML (based on LIFU observations)
================================================================================
INVALID: For IFU submissions, XML should contain XMLIFU code version in a root:comment attribute. This attribute was not found
INVALID: Attribute 'casuid' is missing from element 'observation' in submitted XML file
INVALID: Attribute 'linkedgroup' is missing from element 'observation' in submitted XML file
INVALID: Attribute 'obsgroup_validity' is missing from element 'observation' in submitted XML file
INVALID: Attribute 'trimester' is missing from element 'observation' in submitted XML file
INVALID: Illegal attribute 'semester' exists in element 'observation' in submitted XML file
INVALID: Immutable attribute 'max_sky' (value:'100') in element 'observation:configure' does not match template XML value ('603')
4/ PROGTEMP verification: 41231
```

WEAVE I/O – an overview



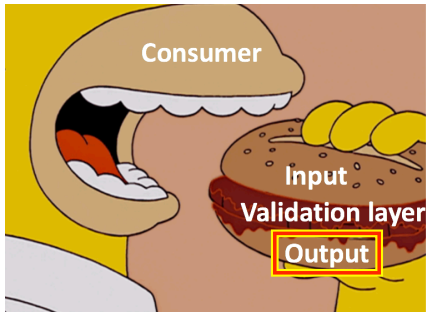
- Validation (how we make sure what's been provided makes sense)
  - Upload facility to a validation platform (think ESO p2pp)
  - Pass / Fail with a generated report: both FITS and XML
  - Tells you the products are legal, but can't check for "common sense"
    - "I only want to allocate 5 of these 940 fibers to my targets" (you might want to!!)
    - "I meant to use NGC in my target names, but used NCG by mistake"
    - "I stuck that mIFU bundle on the wrong galaxy"
    - "This 20th mag galaxy can totally be observed with a sky brightness of 17.5 mags"
  - Warnings *are* issued in some cases, but no substitute for common sense!
  - This facility acts as a gatekeeper between users and the instrument

WEAVE I/O – an overview



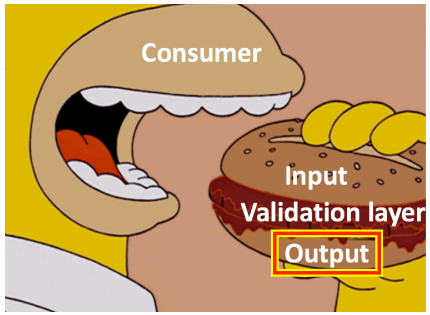
- Validation (how we make sure what's been provided makes sense)
  - Upload facility to a validation platform (think ESO p2pp)
  - Pass / Fail with a generated report: both FITS and XML
  - Tells you the products are legal, but can't check for "common sense"
  - Warnings *are* issued in some cases, but no substitute for common sense!
  - This facility acts as a gatekeeper between users and the instrument
  - Validation platform is called WASP – developed and hosted by CASU
  - Comes with a variety of useful tools for preparation and management of observations
  - Test channel is available to check submissions well in advance of deadlines

WEAVE I/O – an overview



- Output (what data you will have access to)
  - Data from the instrument, including calibration data (arcs, flats)
  - Science frames processed with the WEAVE reduction pipeline:
    - Bias subtracted, flatfielded (“2D image processing”)
    - Spectral extraction, wavelength resampling, flux calibration, sky subtraction
    - Creation of deep stacks from exposures within the 1hr OB and even-deeper stacks from observations over multiple OBs
    - Generation of data cubes for IFU observations (both mini-IFU and Large IFU)
  - Data products derived from analysis of these reduced products:
    - Target classification
    - Model template fitting (stellar and galaxy)
    - Redshift estimation, stellar parameter estimates (Teff, log g), abundances, indices
    - User selection of modules that allow analysis through additional software packages

WEAVE I/O – an overview



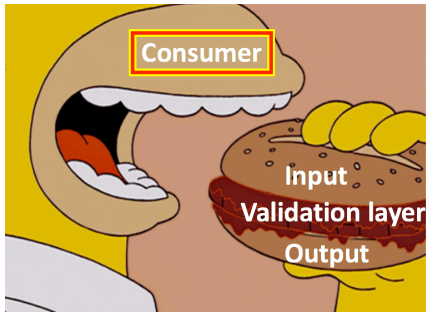
- Output (what data you will have access to)

"RAW"
- Data from the instrument, including calibration data (arcs, flats)
- Science frames processed with the WEAVE reduction pipeline:
  - Bias subtracted, flatfielded ("2D image processing")
  - Spectral extraction, wavelength resampling, flux calibration, sky subtraction
  - Creation of deep stacks from exposures within the 1hr OB and even-deeper stacks from observations over multiple OBs
  - Generation of data cubes for IFU observations (both mini-IFU and Large IFU)

CPS "L1"
- Data products derived from analysis of these reduced products:
  - Target classification
  - Model template fitting (stellar and galaxy)
  - Redshift estimation, stellar parameter estimates (Teff, log g), abundances, indices
  - User selection of modules that allow analysis through additional software packages

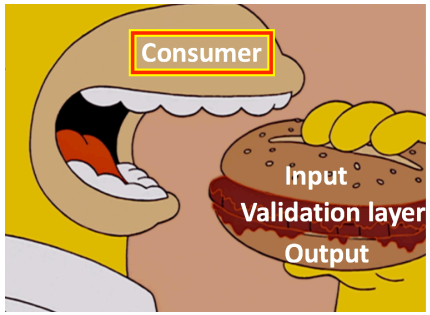
APS "L2"

WEAVE I/O – an overview



- Consumption / access (how you get the data)
  - There are two access points for WEAVE data, but they are different in design and purpose. SV users will have access to both.
  - Operational Repository is hosted at CASU
    - Tracking survey progress, OB status, data processing status
    - Access to most recent data available (RAW, L1, L2), driven by files, not targets
    - **NOT** part of the data release cycle: data is subject to change!
  - WEAVE Archive System hosted at Telescopio Nazionale Galileo (TNG)
    - Provides target-driven searches of **stable** data releases
    - Customisable and flexible searches, user interface NOT designed by David Murphy
    - Data visualisation, bulk downloading
  - SV papers **must** use data from the WAS, not the OR

WEAVE I/O – an overview



- Consumption / access (how you get the data)
  - There are two access points for WEAVE data, but they are different in design and purpose. SV users will have access to both.
  - Operational Repository is hosted at CASU
    - Tracking survey progress, OB status, data processing status
    - Access to most recent data available (RAW, L1, L2), driven by files, not targets
    - **NOT** part of the data release cycle: data is subject to change!

"OR"

  - WEAVE Archive System hosted at Telescopio Nazionale Galileo (TNG)
    - Provides target-driven searches of **stable** data releases
    - Customisable and flexible searches, user interface NOT designed by David Murphy
    - Data visualisation, bulk downloading

"WAS"

  - SV papers **must** use data from the WAS, not the OR

WEAVE I/O – what do the products look like?

FITS Extensions	RAW (one file / arm)	L1 (CPS; 1 file / arm)	L2 (APS; 1 file / exposure)*	
	FITS file, image + tables	FITS file, image + tables	FITS file, tables	
	PHU	PHU (inherited)	PHU (inherited)	
	Image data Det1	Sky-subtracted 1D spectra	Fibre information table	
	Image data Det2	inv. variance	Classification spectral fit	Classification extracted parameters
		No sky-subtracted 1D spectra	FERRE stellar spectral fit	FERRE extracted parameters
		No sky-subtracted inv. Variance	RVSPECFIT stellar spectral fit	RVSPECFIT extracted parameters
	Fibre information table	Fibre information table (extras added)	Galaxy model spectral fit	Galaxy extracted parameters
		IFU: white light image for cube	* In some instances, per-arm too	

WEAVE I/O – what do the products look like?

FITS Extensions

RAW (one file / arm)	L1 (CPS; 1 file / arm)	L2 (APS; 1 file / exposure)*	
FITS file, image + tables	FITS file, image + tables	FITS file, tables	
PHU	PHU (inherited)	PHU (inherited)	
Image data Det1	Sky-subtracted 1D spectra	Fibre information table	
Image data Det2	inv. variance	Classification spectral fit	Classification extracted parameters
	No sky-subtracted 1D spectra	FERRE stellar spectral fit	FERRE extracted parameters
	No sky-subtracted inv. Variance	RVSPECFIT stellar spectral fit	RVSPECFIT extracted parameters
Fibre information table	Fibre information table (extras added)	Galaxy model spectral fit	Galaxy extracted parameters
	IFU: white light image for cube	* In some instances, per-arm too	

WEAVE I/O – the fibre information table

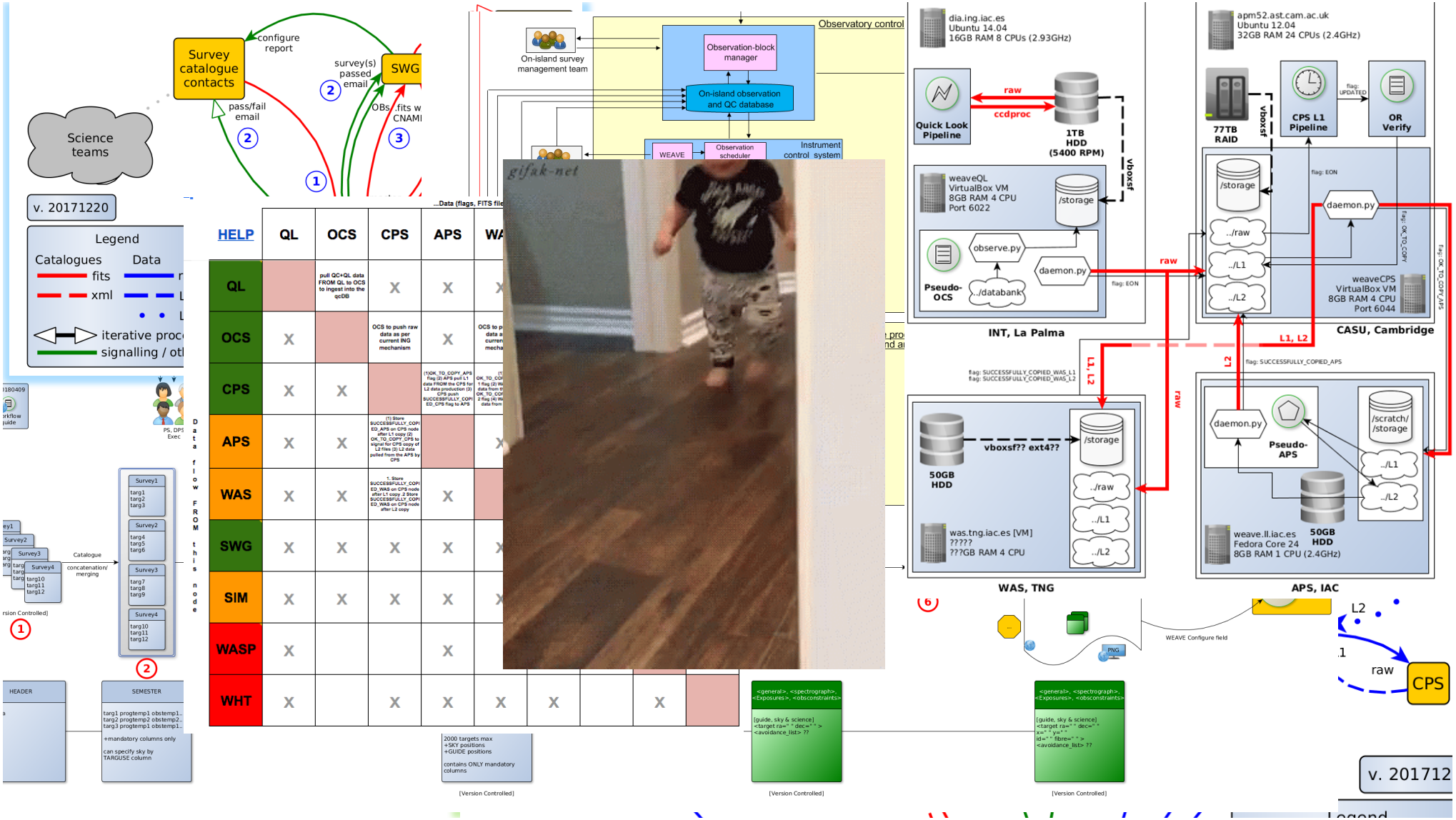
- Describes what the fibers are pointing at in each data product
- Inherits important data from the FITS target catalogue the user supplies
- This information will be ingested into the (target driven) WEAVE Archive System, and to a lesser extent the Operational Repository
- The **fib**re **info**rmation (fibinfo) table must include important columns that allow us to identify:
  - What target?
  - How was it observed? What observing conditions were requested?
  - How was it analysed?

WEAVE Targets: useful things to care about

- CNAME – maps position on sky → ID
  - All science targets, calibration targets and sky fibers have a CNAME
  - 01:33:50.9 +30:39:35.8 → WVE_01335090+3039358
- Instrument configuration: “PROGTEMP” how was this target observed?
- Observing constraints: “OBSTEMP” what were the minimum conditions requested?
- Analysis flags: “APS_FLAG” are extra L2 analysis packages required?
- Applies to exposures, OB-stacks, “super-stacks”*
  - *when a target is observed over multiple Obs, this is a stack of all observations
- A search in the WEAVE Archive for a target can yield >1 result. Why?
  - The same target could be observed in MOS in low-res AND high-res
  - Target “uniqueness” is generally:
    - CNAME + configuration (+survey in some cases)

Let’s tie this all together….

- …with some simple plots

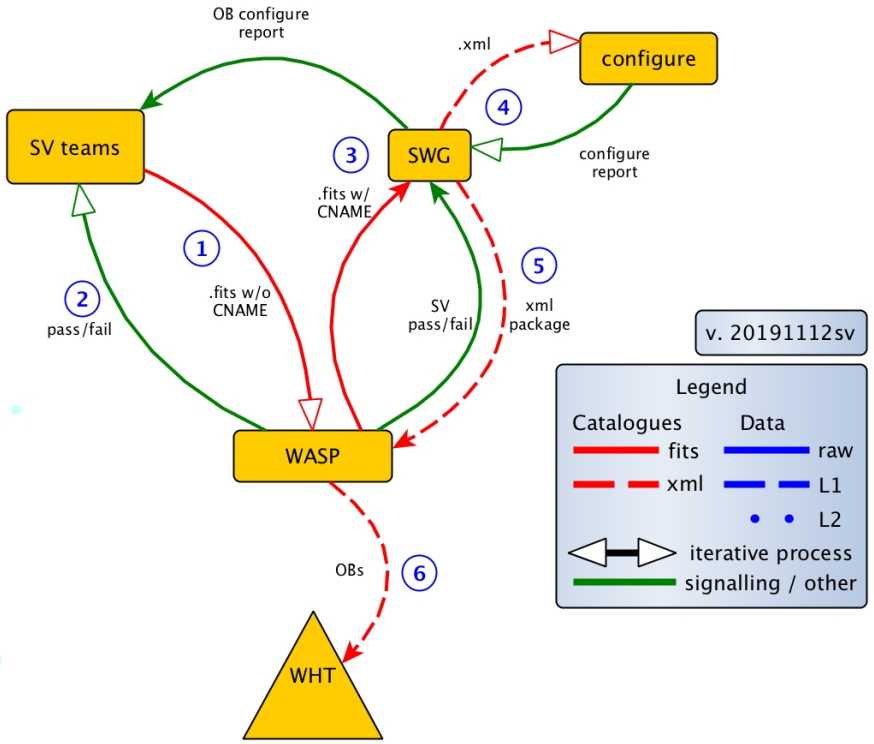

Science teams
Survey catalogue contacts
configure report
survey(s) passed email
pass/fail email
SWG
v. 20171220
Legend
Catalogues
Data
fits
xml
iterative proc
signalling / ot
HELP
QL
OCS
CPS
APS
WAS
SWG
SIM
WASP
WHT
pull QC+QL data FROM QL to OCS to ingest into the qcDB
OCS to push raw data as per current ING mechanism
Observatory control
On-island survey management team
Observation-block manager
On-island observation and QC database
WEAVE
Observation scheduler
Instrument control system
dia.ing.iac.es
Ubuntu 14.04
16GB RAM 8 CPUs (2.93GHz)
Quick Look Pipeline
raw
ccdproc
1TB HDD (5400 RPM)
vboxsf
weaveQL VirtualBox VM
8GB RAM 4 CPU
Port 6022
/storage
observe.py
daemon.py
Pseudo-OCS
../databank
INT, La Palma
apm52.ast.cam.ac.uk
Ubuntu 12.04
32GB RAM 24 CPUs (2.4GHz)
77TB RAID
CPS L1 Pipeline
OR Verify
flag: UPDATED
flag: EON
../raw
../L1
../L2
weaveCPS VirtualBox VM
8GB RAM 4 CPU
Port 6044
CASU, Cambridge
L1, L2
L2
flag: SUCCESSFULLY_COPIED_APS
flag: SUCCESSFULLY_COPIED_WAS_L1
flag: SUCCESSFULLY_COPIED_WAS_L2
vboxsf?? ext4??
50GB HDD
was.tng.iac.es [VM]
????
???GB RAM 4 CPU
WAS, TNG
Pseudo-APS
/scratch/ /storage
weave.ll.iac.es
Fedora Core 24
8GB RAM 1 CPU (2.4GHz)
50GB HDD
APS, IAC
WEAVE Configure field
PNG
CPS
Catalogue concatenation/ merging
Survey1
Survey2
Survey3
Survey4
HEADER
SEMESTER
+mandatory columns only
can specify sky by TARGUSE column
2000 targets max
+SKY positions
+GUIDE positions
contains ONLY mandatory columns
[Version Controlled]
[guide, sky & science]
[Exposures], [obsconstraints]
[Version Controlled]
[guide, sky & science]
[Version Controlled]
v. 201712
Legend

WEAVE I/O – tying this all together

- Targets → FITS → WASP → FITS+ → XML → WASP → XML(OB) → WHT
- RAW → L1 → L2

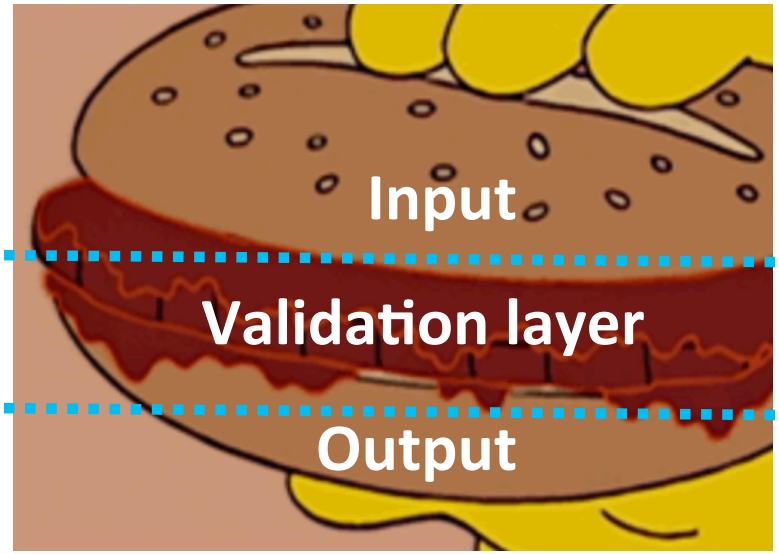
WEAVE I/O – tying this all together

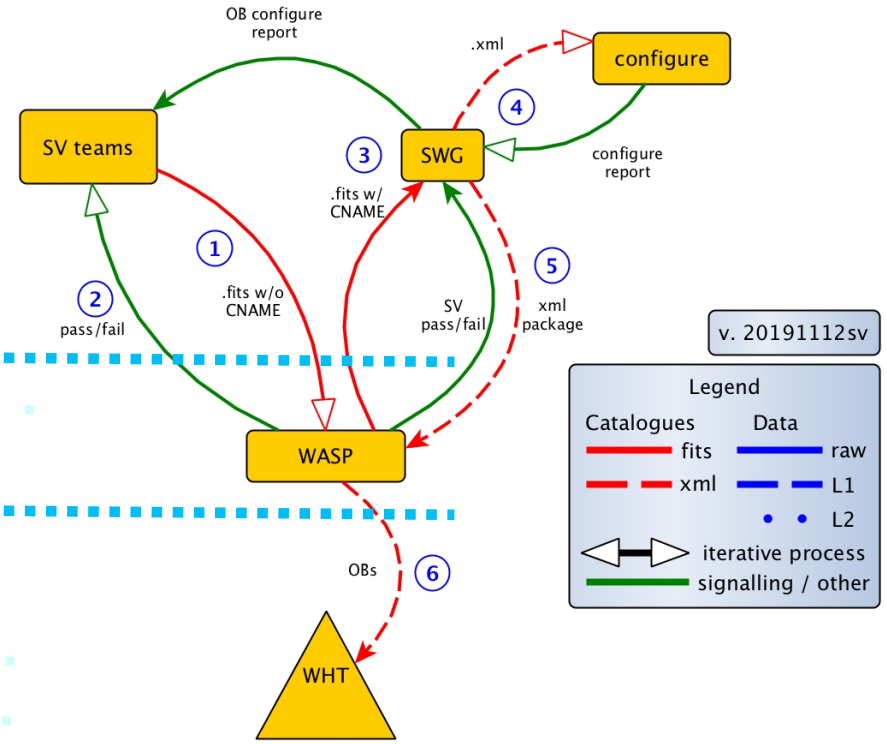
- Targets → FITS → WASP → FITS+ → XML → WASP → XML(OB) → WHT
- RAW → L1 → L2



WEAVE I/O – tying this all together

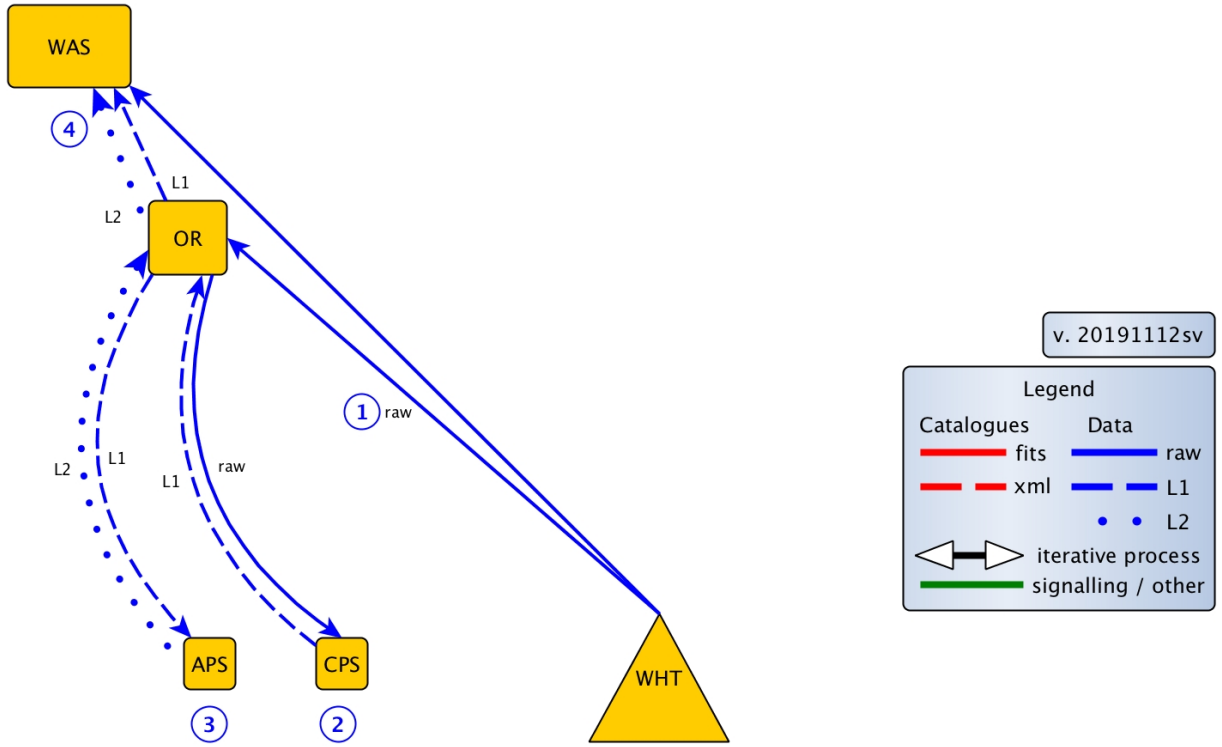
- Targets → FITS → WASP → FITS+ → XML → WASP → XML(OB) → WHT
- RAW → L1 → L2





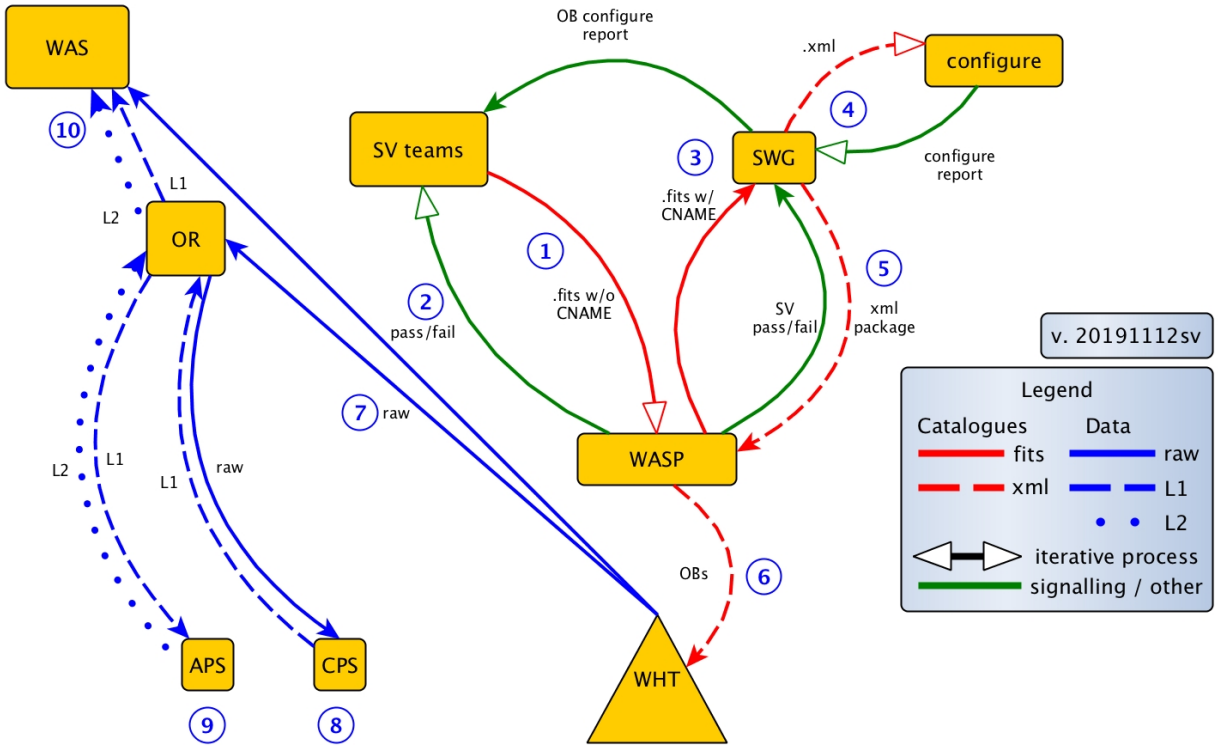
WEAVE I/O – tying this all together

- Targets → FITS → WASP → FITS+ → XML → WASP → XML(OB) → WHT
- RAW → L1 → L2

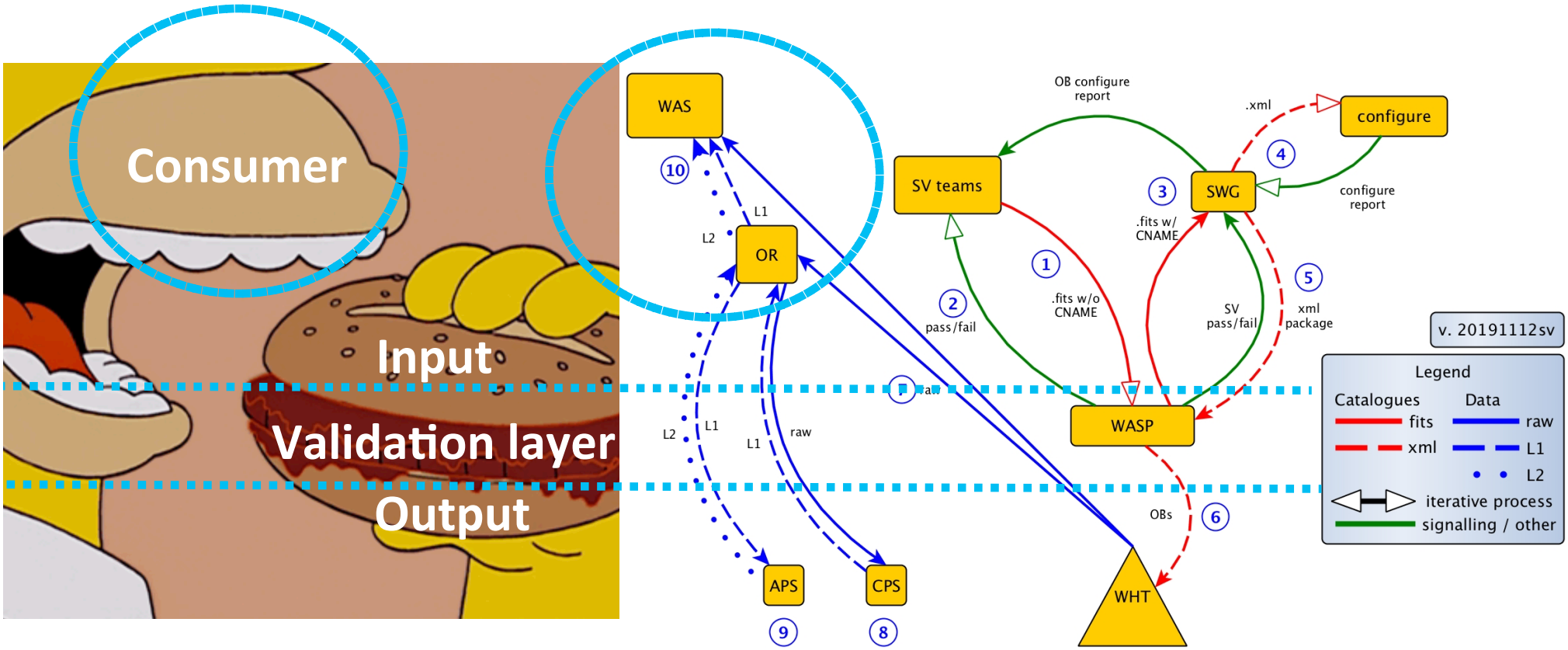


WEAVE I/O – tying this all together

- Targets → FITS → WASP → FITS+ → XML → WASP → XML(OB) → WHT
- RAW → L1 → L2



WEAVE I/O – tying this all together



Quick recap

- Don’t worry, stay hydrated!
- WEAVE(!) been doing this for quite a while, and it’s still challenging to us

Inputs	Validation	Outputs	Access
FITS catalogues (targets)	WASP platform	RAW data	Operational Repository (survey/processing progress, bleeding-edge data)
XML observing blocks	WASP platform	L1 processed data (“spectra”)	WEAVE Archive System (proper data releases, designed for science end-users)
		L2 analysis (“spectral analysis, redshifts, abundances)	

Quick recap

- Don’t worry, stay hydrated!
- WEAVE(!) been doing this for quite a while, and it’s still challenging to us

Inputs	Validation	Outputs	Access
FITS catalogues (targets)	WASP platform	RAW data	Operational Repository (survey/processing progress, bleeding-edge data)
XML observing blocks	WASP platform	L1 processed data (“spectra”)	WEAVE Archive System (proper data releases, designed for science end-users)
		L2 analysis (“spectral analysis, redshifts, abundances)	

Next session

Quick recap

- Don’t worry, stay hydrated!
- WEAVE(!) been doing this for quite a while, and it’s still challenging to us

Inputs	Validation	Outputs	Access
FITS catalogues (targets)	WASP platform	RAW data	Operational Repository (survey/processing progress, bleeding-edge data)
XML observing blocks	WASP platform	L1 processed data (“spectra”)	WEAVE Archive System (proper data releases, designed for science end-users)
		L2 analysis (“spectral analysis, redshifts, abundances)	

Next session

This afternoon

Quick recap

- Don’t worry, stay hydrated!
- WEAVE(!) been doing this for quite a while, and it’s still challenging to us

Inputs	Validation	Outputs	Access
FITS catalogues (targets)	WASP platform	RAW data	Operational Repository (survey/processing progress, bleeding-edge data)
XML observing blocks	WASP platform	L1 processed data (“spectra”)	WEAVE Archive System (proper data releases, designed for science end-users)
		L2 analysis (“spectral analysis, redshifts, abundances)	

Tomorrow

Next session

This afternoon

Quick recap

- Don’t worry, stay hydrated!
- WEAVE(!) been doing this for quite a while, and it’s still challenging to us

Inputs	Validation	Outputs	Access
FITS catalogues (targets)	WASP platform	RAW data	Operational Repository (survey/processing progress, bleeding-edge data)
XML observing blocks	WASP platform	L1 processed data (“spectra”)	WEAVE Archive System (proper data releases, designed for science end-users)
		L2 analysis (“spectral analysis, redshifts, abundances)	

Tomorrow (w/ WASP demos)

Next session

This afternoon

This session: *aim* and overview

- *Build on the broad outline of WEAVE dataflow given by Shoko*

- The main components within the WEAVE data model
- How they interact
- WEAVE data products: Input and Output
- How you get them