



# 4MOST – 4m Multi-Object Spectroscopic Telescope

## Pipelines and Data Products

Nicholas Walton (IoA, University of Cambridge)

Mike Irwin, David Murphy, Luis Peralta, Clare Worley

8 May 2019

[www.4MOST.eu](http://www.4MOST.eu)



We remember our friend and colleague

Jim Lewis (1959 – 2019)

# Ten 4MOST Consortium Surveys ...



No	Survey Name	Survey (Co-)PI
S1	Milky Way Halo LR Survey	Irwin (IoA) , Helmi (RuG)
S2	Milky Way Halo HR Survey	Christlieb (ZAH)
S3	Milky Way Disk and Bulge LR Survey	Chiappini, Minchev, Starkenburg (AIP)
S4	Milky Way Disk and Bulge HR Survey	Bensby (LU), Bergemann (MPIA)
S5	Galaxy Clusters Survey	Finoguenov (MPE)
S6	AGN Survey	Merloni (MPE)
S7	Galaxy Evolution Survey (WAVES)	Driver (USW), Liske (HHU)
S8	Cosmology Redshift Survey	Richard (CRAL), Kneib (EPFL)
S9	Magellanic Clouds Survey	Cioni (AIP)
S10	Time-Domain Extragalactic Survey (TiDES)	Sullivan (Southampton)

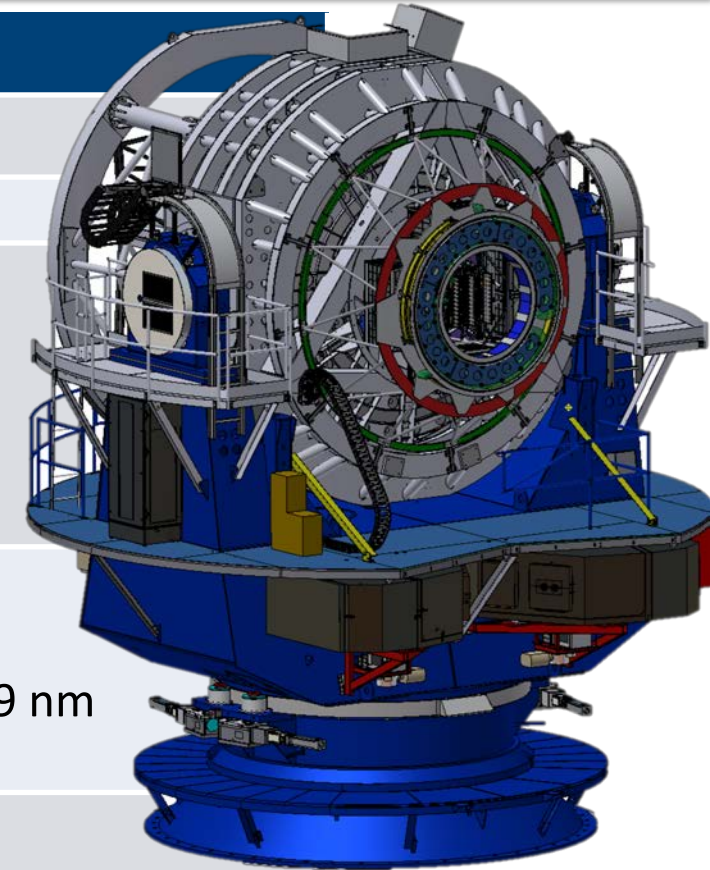
# ...inform the Science Requirements ... leads to ...



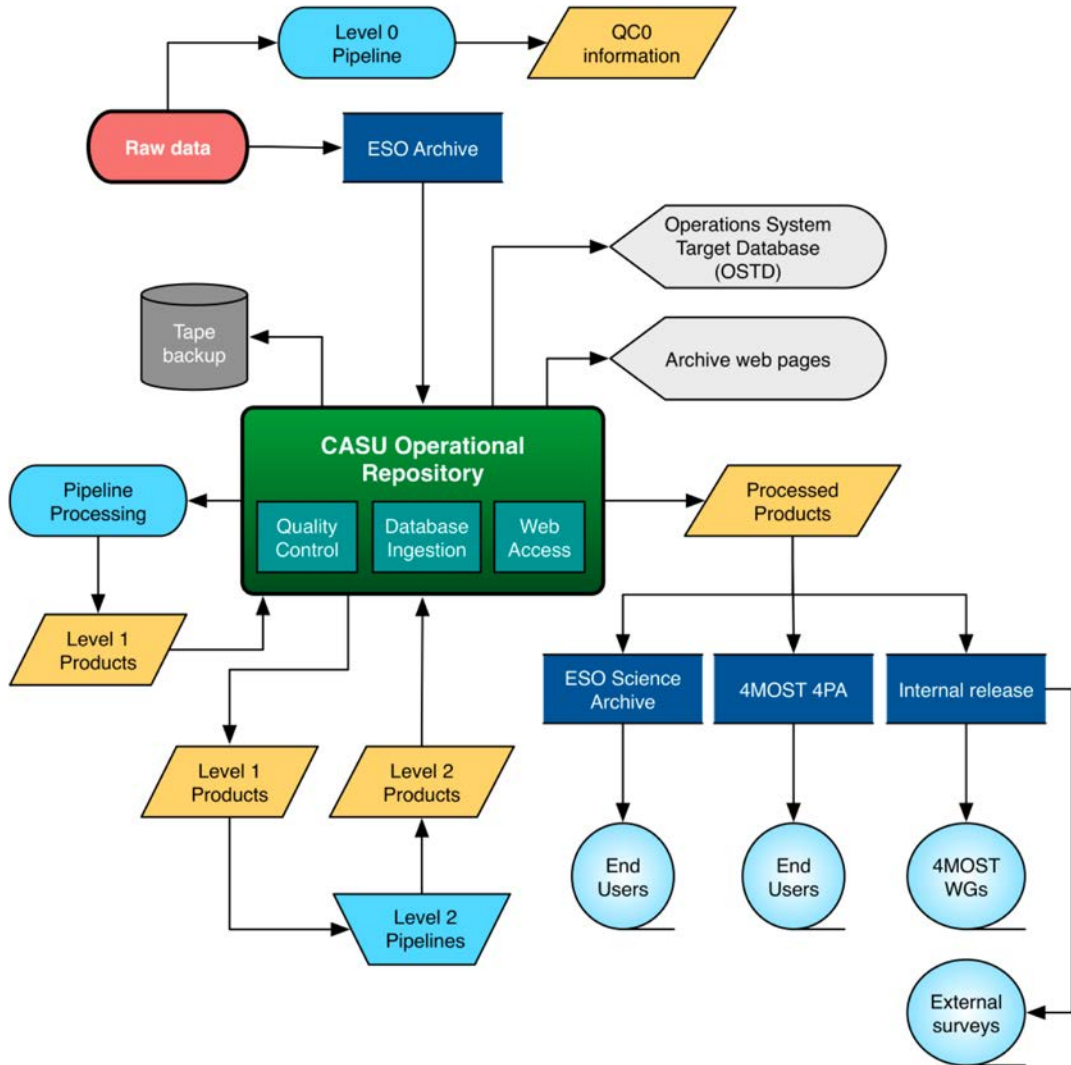
- 4MOST shall be able to obtain:
  - Redshifts of AGN and galaxies (also in clusters)
    - R~5000 spectra of 22 r-mag targets with S/N=5/Å with >3 targets in  $\phi=2'$
  - Radial velocities of  $\leq 1$  km/s accuracy and Stellar parameters of  $< 0.15$  dex accuracy of any Gaia star
    - R~5000 spectra of 20 r-mag stars with S/N=10 per Ångström
  - Abundances of up to 15 chemical elements
    - R~20000 spectra of 16 V-mag stars with S/N=140 per Ångström
- In a 5 year survey 4MOST shall obtain:
  - 15 (goal 30) million targets at R~5000
  - 1.0 (goal 3.0) million targets at R~20,000
  - 16,000 (goal 23,000) degree<sup>2</sup> area on the sky at least two times

# ... 4MOST Instrument Specification

Specification	Design value
Field-of-View (hexagon)	$\sim 4.1 \text{ degree}^2 (\varnothing > 2.6^\circ)$
Multiplex fiber positioner	2436
Medium Resolution Spectrographs (2x)	$R \sim 4000\text{--}7500$
# Fibres	812 fibres (2x)
Passband	370-950 nm
Velocity accuracy	$< 1 \text{ km/s}$
High Resolution Spectrograph (1x)	$R \sim 20,000$
# Fibres	812 fibres
Passband	392.6–435.5 nm, 516–573 nm, 610–679 nm
Velocity accuracy	$< 1 \text{ km/s}$
# of fibers in $\varnothing = 2'$ circle	$> 3$
Fibre diameter	$\varnothing = 1.42 \text{ arcsec}$
Area (first 5 year survey)	$> 2\text{h} \times 18,000 \text{ deg}^2$
Number of science spectra (5 year)	$\sim 75 \text{ million of } 20 \text{ min}$



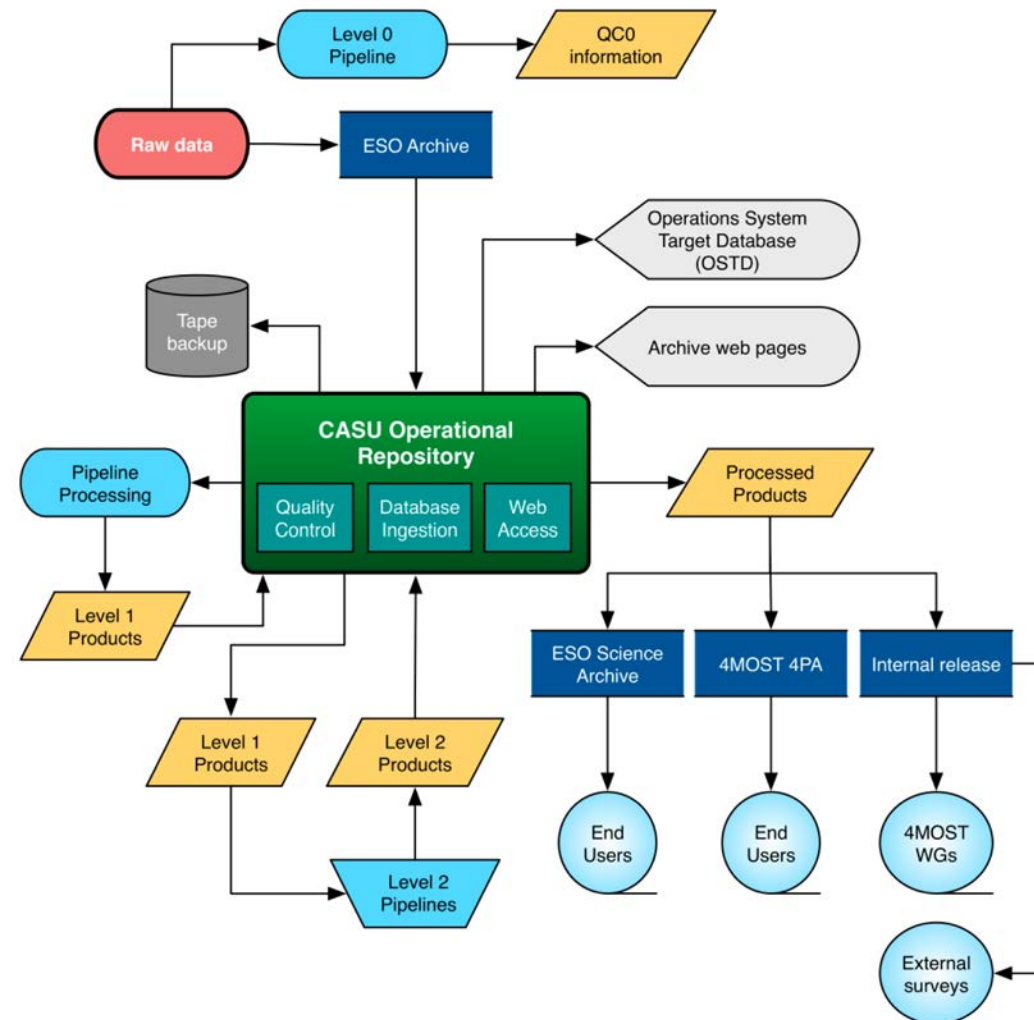
# Data Management System: Meeting the 4MOST Data Challenge



- Mass scale spectral processing
- Many interfaces
- Development complexity
  - Technical
  - Organisational
- Design for operation effectiveness
- Limited resources
  - But expectations high



# 4MOST Data System: The Pipeline Acronyms



- QC0: QC0 Pipeline and Database: Paranal
- QC1: QC1 Pipeline and Database: Garching
- 4L1: Level 1 pipeline: CASU
- 4GP: Galactic Pipeline: Lund
- 4XP: ExtraGalactic Pipeline: UWA
- 4CP: Classification Pipeline:
- 4SP: Selection function Pipeline:
- 4OR: Operational Repository: CASU
- 4PA: 4MOST Public Archive: AIP
- SAF: Science Archive Facility: Garching

# DMS DFDR (2017) and FDR (2018)

major design reviews passed



Full set of DMS design documents

- Data Management System Development Plan
- Data Management System Requirements Specification
- Data Management System Archive Design Report
- Data Management System Design Report
- QC + Pipeline Data Reduction Description
- Data Management System-L2 Pipelines ICD
- 4MOST-ESO Back-end ICD



## Design and Analysis Report

### Data Management System

Document No.: VIS-DER-4MOST-47116-1400-0001

Issue No.: 2.0

DocuShare URL: <https://ds-web.aip.de/docushare/dsweb/View/Collection-1423>

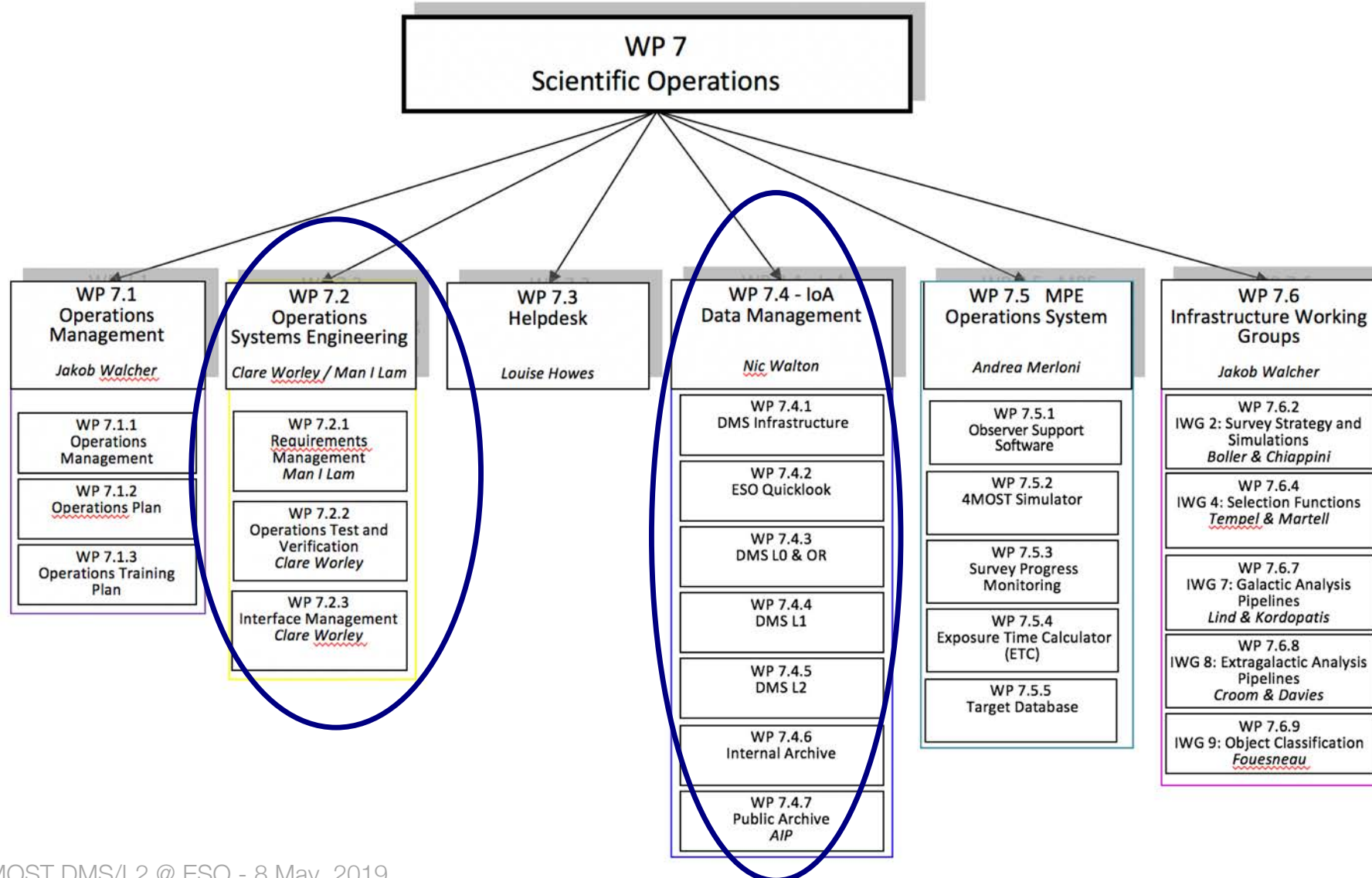
Prepared by \_\_\_\_\_ 2017-02-27  
*Nicholas Walton*

Approved by \_\_\_\_\_ 2017-02-27  
*James R Lewis*

Released by \_\_\_\_\_ 2017-02-27  
*Joar Brynnel*



# Science Operations Organisation: DMS WP7.4



# WEAVE as a 4MOST development pilot

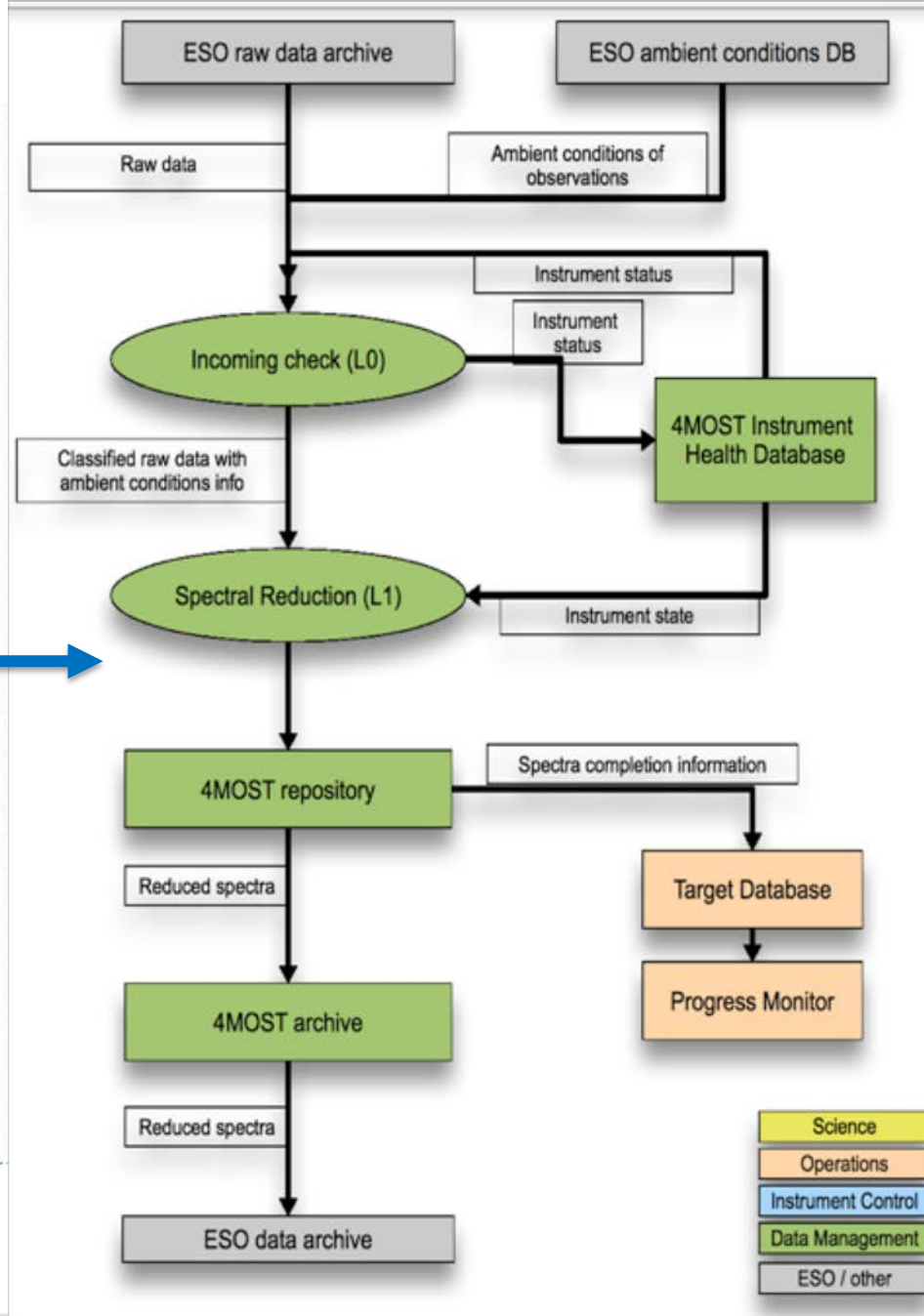
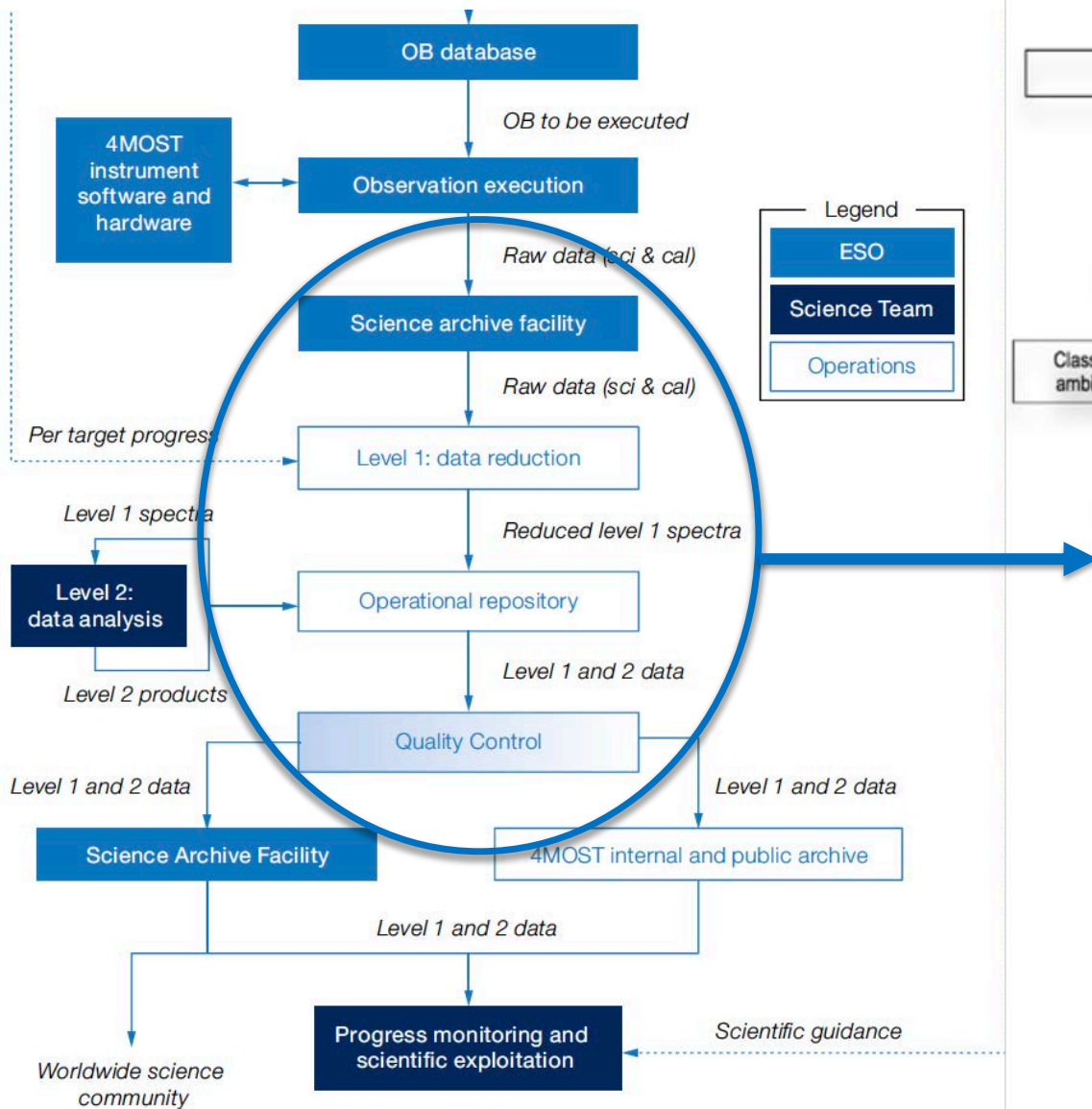


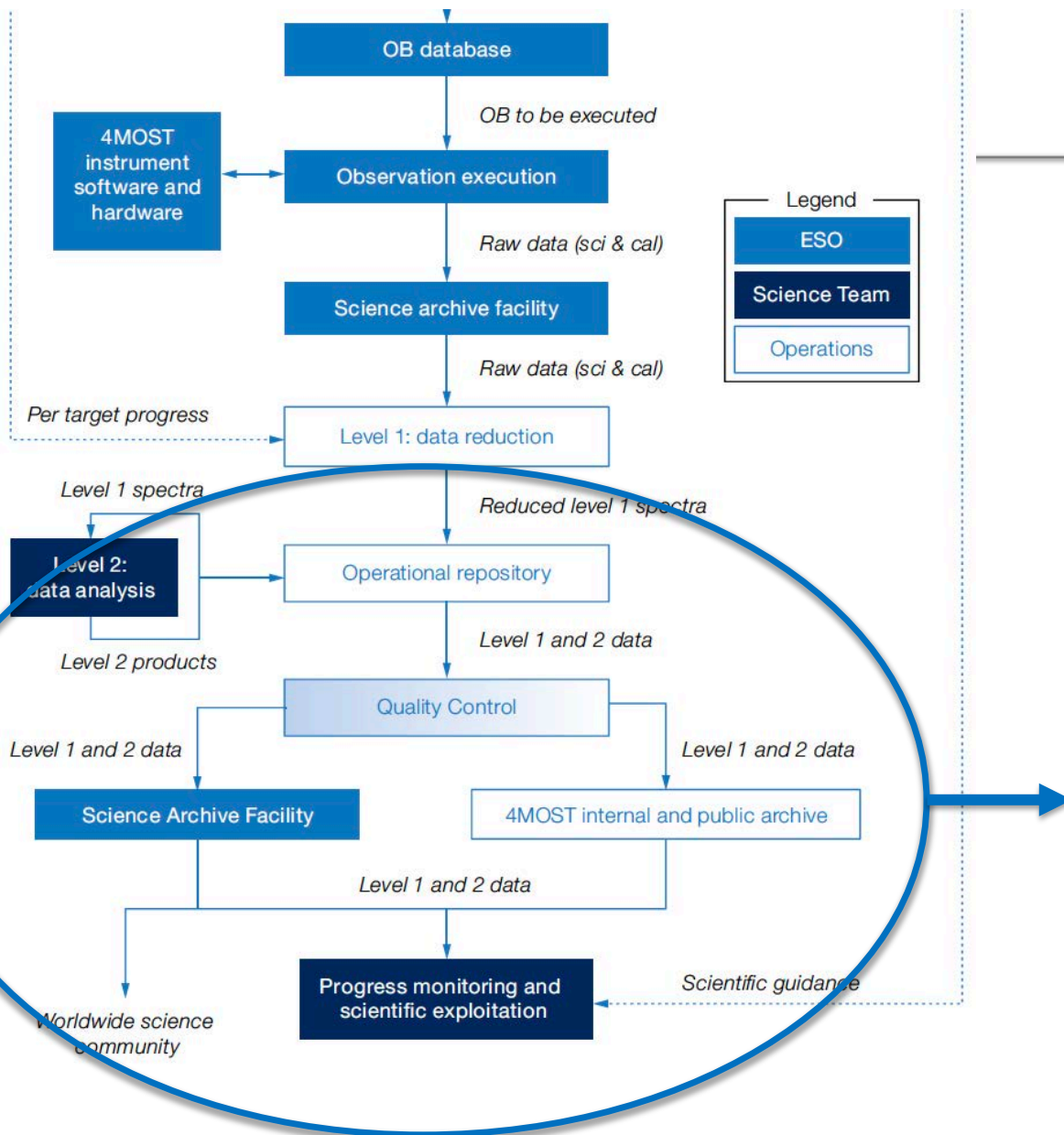
CASU development of the WEAVE pipeline provides operational assessment of flow system – acts as a 4MOST pilot. Recent relevant work includes:

- definition of output data formats and data model
- Improvements to baseline spectral extraction method (tested on GES + WYFFOS data)
- Improvements to scattered light correction
  - Integrated spectra extraction/ cross-talk removal as single operation is giving improved error estimates
  - Investigating additional PCA analysis of residual skyline errors

## Data Management L0/L1 Data Flow

Operations and data flow becomes more complex with iterative loops and health monitoring





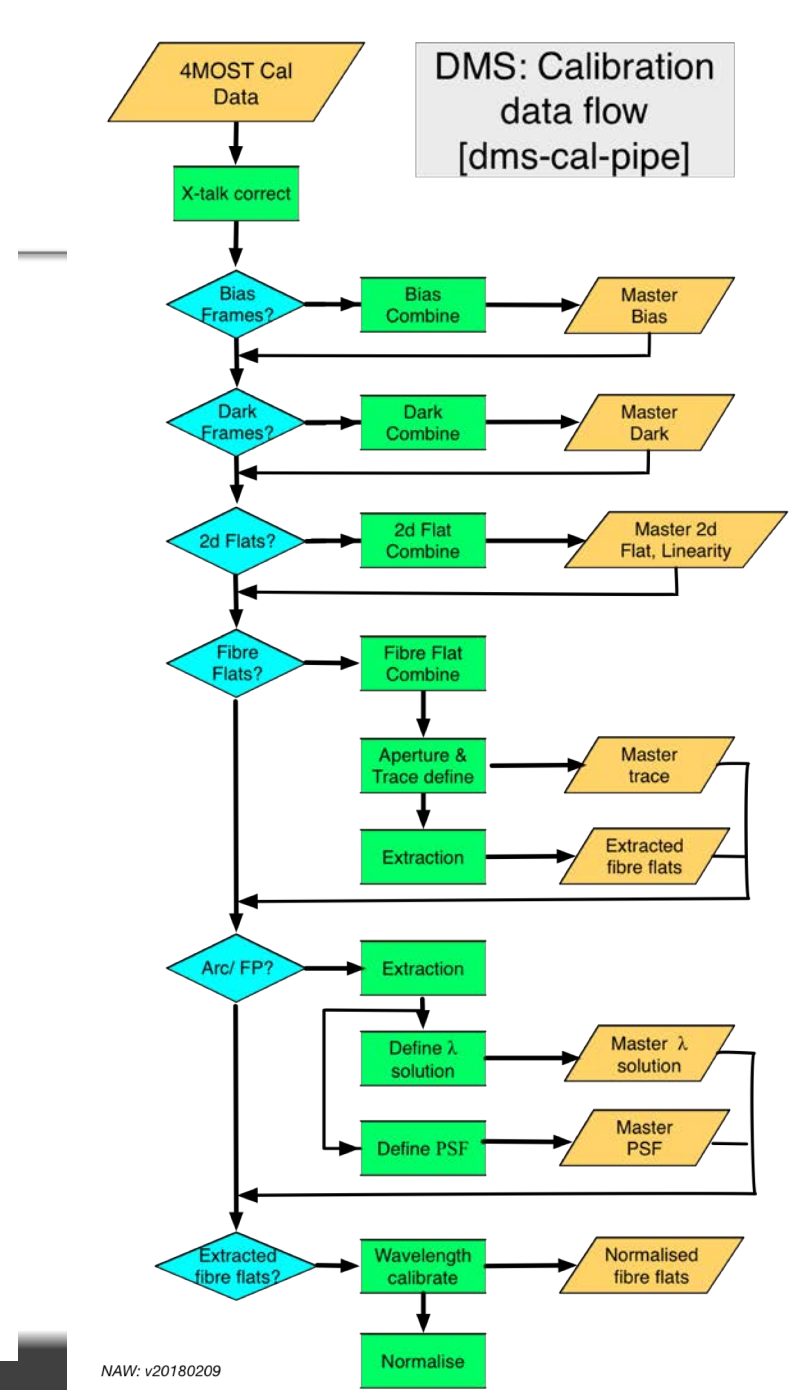
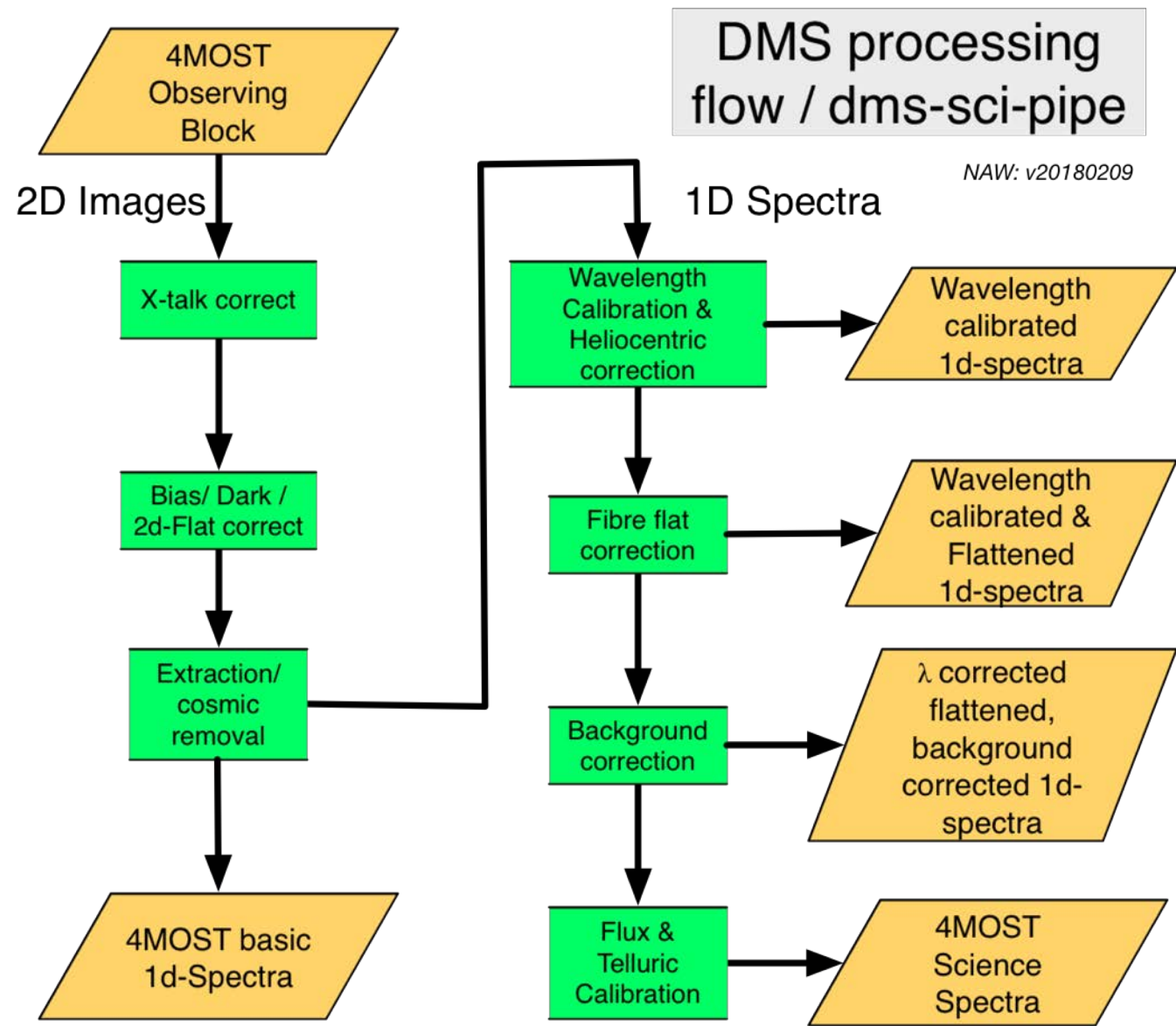
# Data Management L2/Release Data Flow

Role of Science WGs & Infrastructure WGs defining requirements for data release

Outputs from the IWG L2 pipelines subject to quality control prior to release to ESO – involves 4MOST Science Team



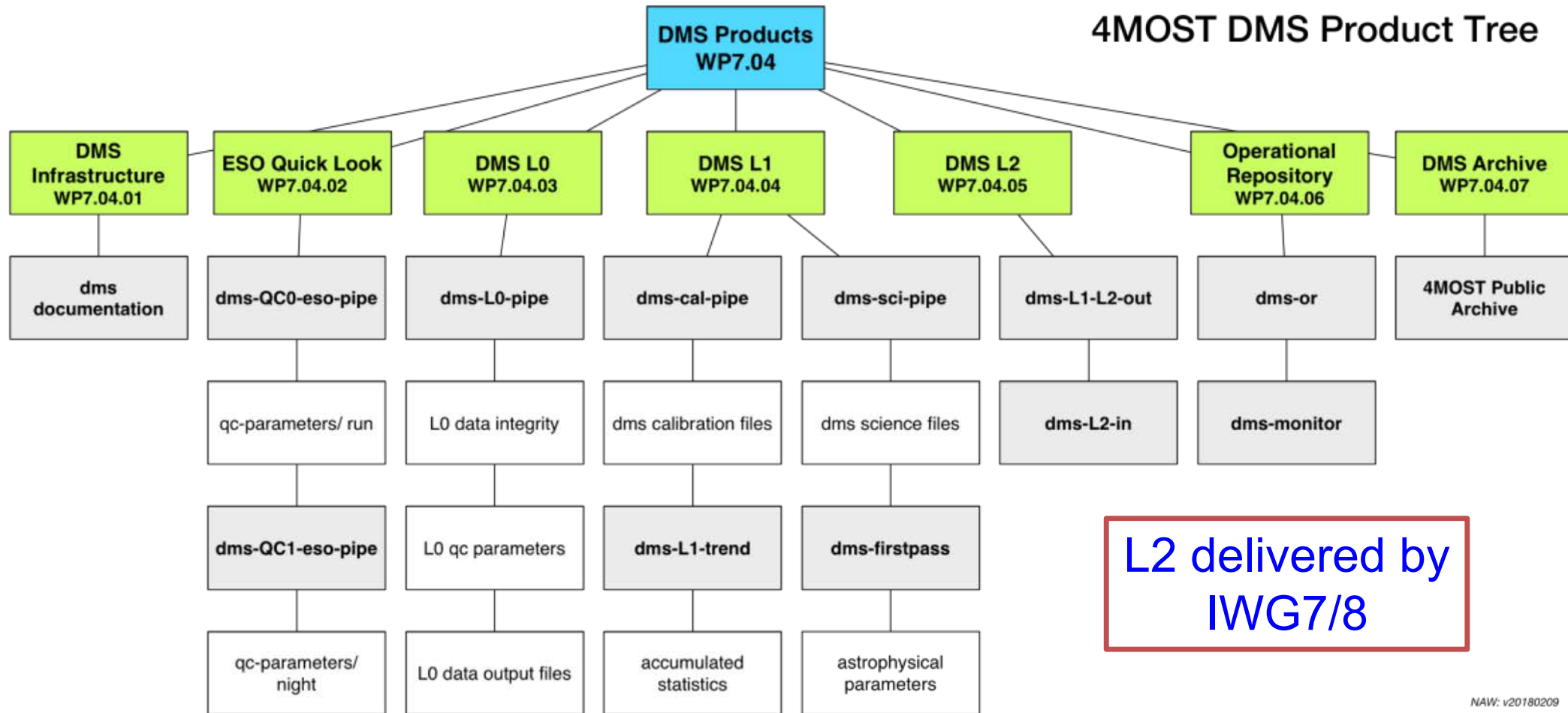
# L1 Science Data Flow



# DMS Product Tree



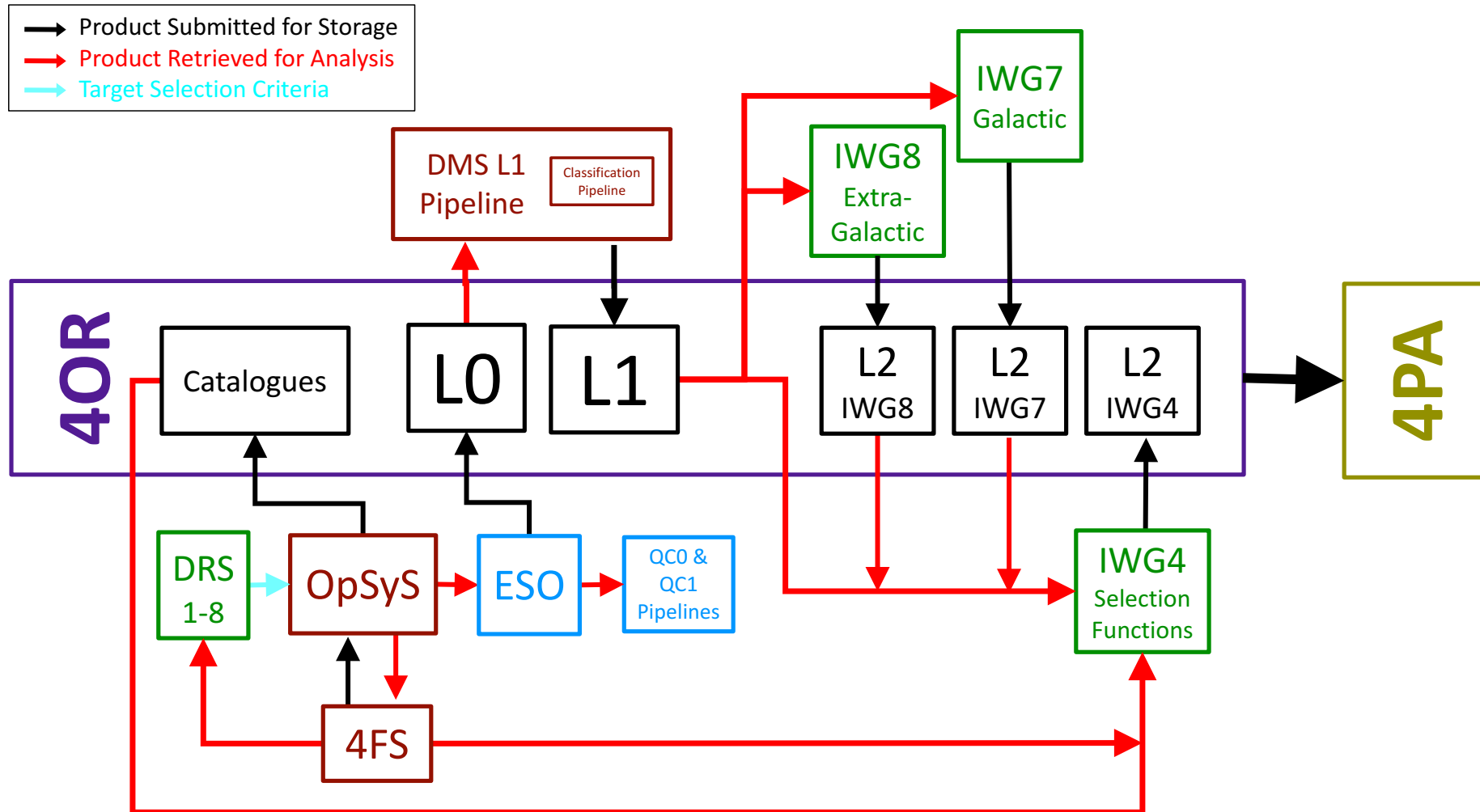
4MOST DMS Product Tree



NAW: v20180209



# DMS Interfaces and Data Flow: Upstream and Downstream



# L1 Data Products:

## 4MOST L1 output and input to the L2 pipelines

L1 data products  
shown here are  
current baseline



- Resampled, flux-calibrated spectra
- Non-stacked spectra from every exposure
- Stacked spectra per OB
- Metadata describing the used stacking and weighting files
- Inverse variance arrays for all of above
- Flux units and scaling
- The throughput curve that was used to correct the shape of the spectrum.
- Wavelength solutions for all of the above.
- Dead pixel and cosmic ray mask encoded in the inverse variance arrays.

- Heliocentric correction used for each spectrum
- History of the reduction process
- All target catalogue information of the target that was contained in the input OB.
- Version of the L1 pipeline used.
- Observatory parameters
- Position(s) in the originating L0 2d data file(s)
- References to (IDs of) the first and second neighbouring spectra
- Telluric correction spectra

# TiDES-Live Transients: Rapid Data Access



- 4MOST survey 10: Time Domain ExtraGal
  - Three components: SN Cosmology, AGN Reverberation and Live Transients
- TiDES Live Transients (LT) requires ‘rapid’ classification of transient candidates
  - Up to 5% of LR fibres allocated
  - Classification timescale: “night+24 hrs”
- TiDES-LT Spectra extraction from ‘standard’ QC0 pipeline
  - Wavelength cal to  $2\text{\AA}$ , relative flux cal to 20%

current baseline  
planning

# 4GP: The L2 Galactic Pipeline



- Low resolution: 12M targets @  $V < 20$
- High resolution: 4M targets @  $G < 17$

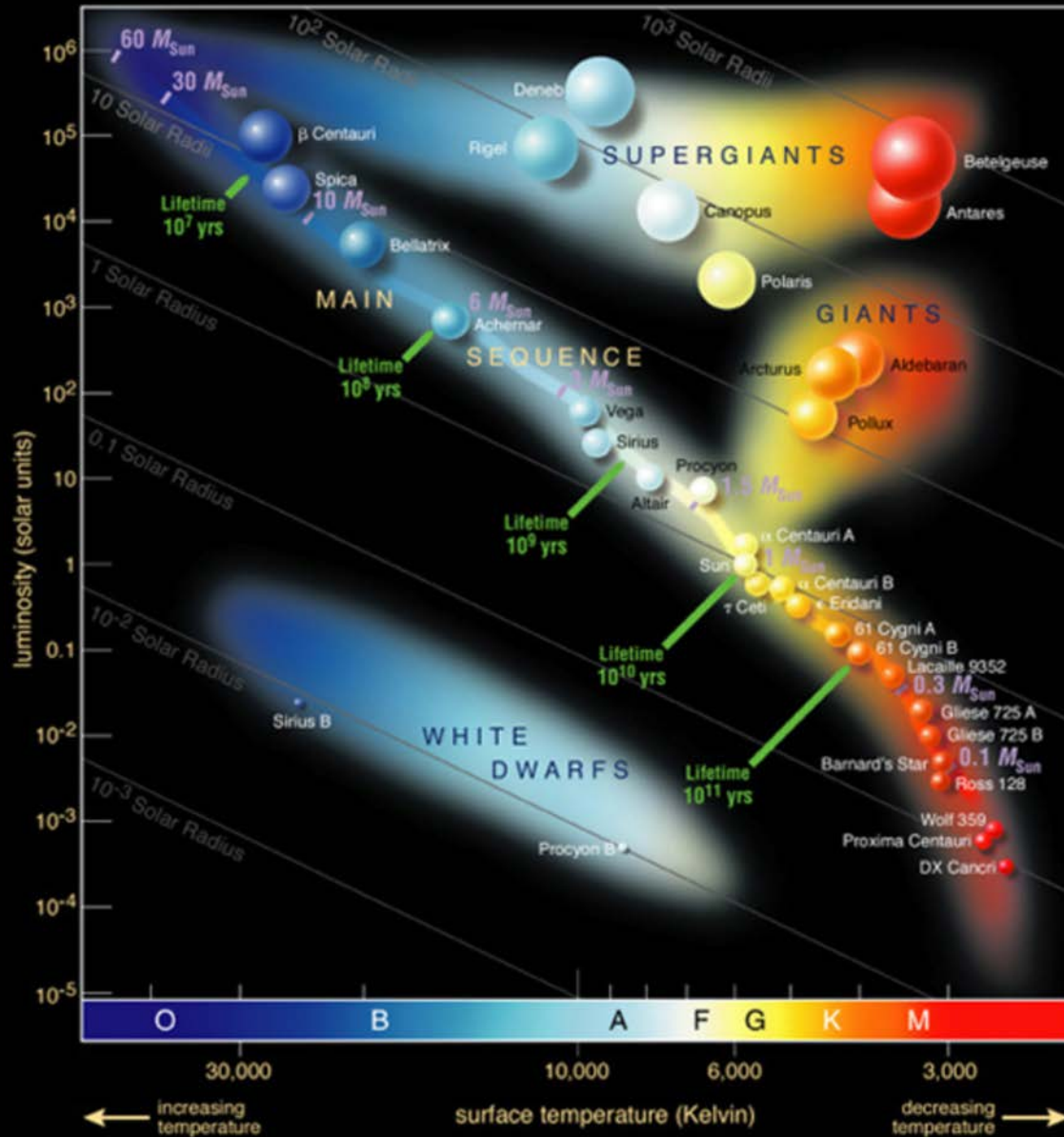
Consortium Survey	Brightness range (magnitudes)	Targets (millions)
S1 Milky Way Halo LR	$15.0 \leq G \leq 20.0$	1.5
S2 Milky Way Halo HR	$12.0 \leq G \leq 17.0$	1.5
S3 Milky Way Disc and Bulge LR (4MIDABLE-LR)	$4.0 \leq G \leq 19.0$	10.0
S4 Milky Way Disc and Bulge HR (4MIDABLE-HR)	$10.0 \leq G \leq 15.5$	2.5
S5 Galaxy Clusters	$18.0 \leq r \leq 22.0$	1.7
S6 AGN	$18.0 \leq r \leq 22.8$	1.0
S7 Galaxy Evolution (WAVES)	$18.0 \leq r \leq 22.5$	1.6
S8 Cosmology Redshift Survey	$20.0 \leq r \leq 23.9$	8.0
S9 Magellanic Clouds (1001MC)	$10.5 \leq G \leq 19.5$	0.5
S10 Transients (TIDES)	$18.0 \leq r \leq 22.5$	0.3
Total		> 28

Table 2. The minimal number and typical magnitude range of targets that each Consortium Survey expects to observe in the first five-year survey of 4MOST.

## 4GP products:

- Radial Velocities
- Stellar parameters
- Abundances of selected elements together with their precisions and accuracies

Slide from:  
Lind & Kordopatis



# Galactic and SMC/LMC Survey Targets

Slide from:  
Lind & Kordopatis

# Galactic Survey Requirements

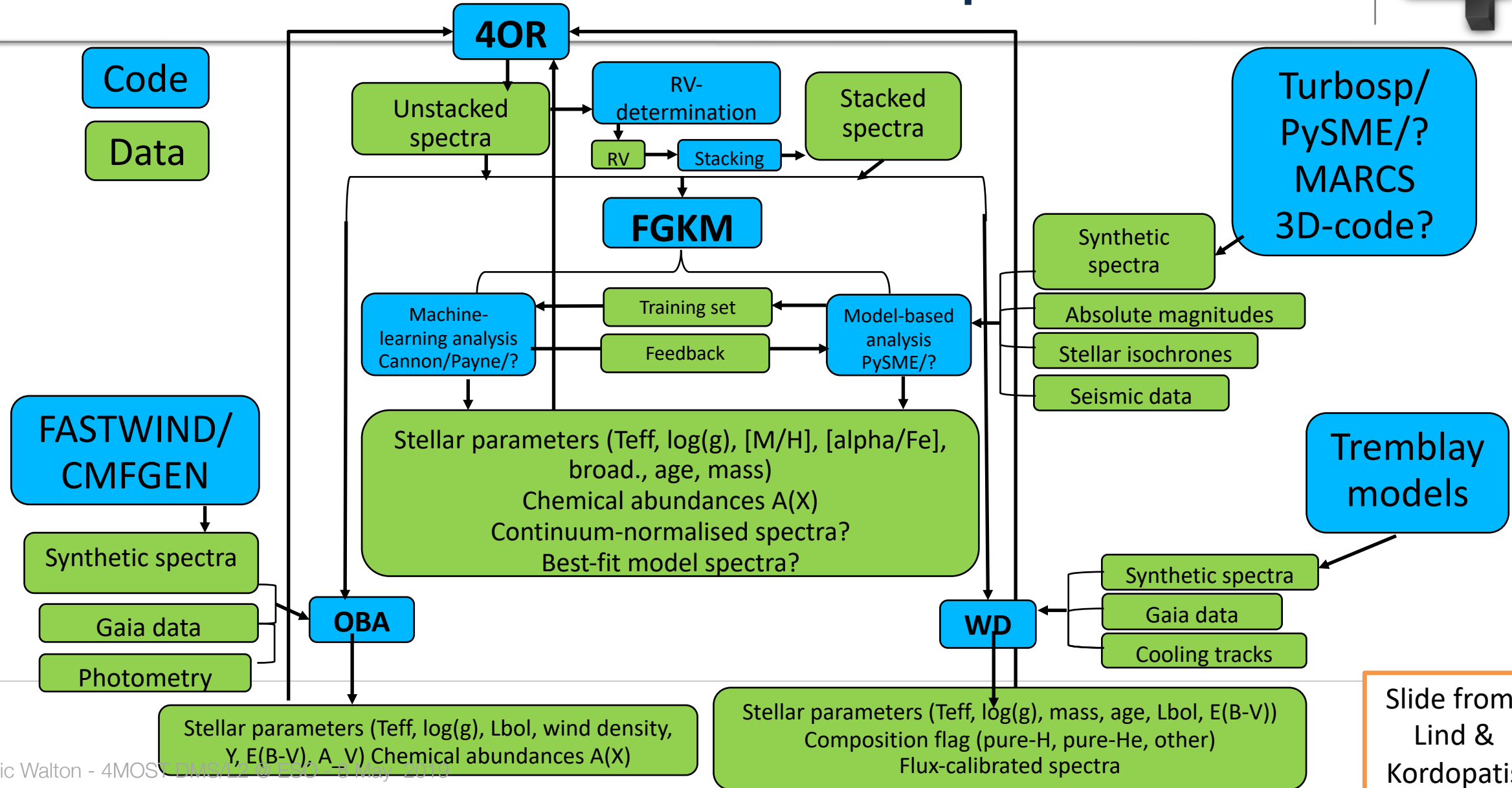
Slide from:  
Lind & Kordopatis



Survey	Type of target	S/N /Å	RV [km/s]	[Fe/H] dex	[X/Fe] Nr	[X/Fe] dex
<b>S1: LR halo</b>	RGB K-type stars	10-25	1-2	0.1-0.2	1 (alpha)	0.05-0.1
<b>S2: HR halo</b>	[Fe/H]<-0.5, TO+RGB, FGK	40-140	2	(0.1)	<20	<0.2-0.3
<b>S3: LR disk/bulge</b>	RGB K-type stars	10-30	1-2	(0.1)	~10	0.2
	White Dwarfs		5			
<b>S4: HR disk/bulge</b>	FGK stars	100	1	(0.05)	<20	<0.05
<b>S9: Magellanic Cloud</b>	OBAFGK stars	10-150	2	0.2	?	0.2

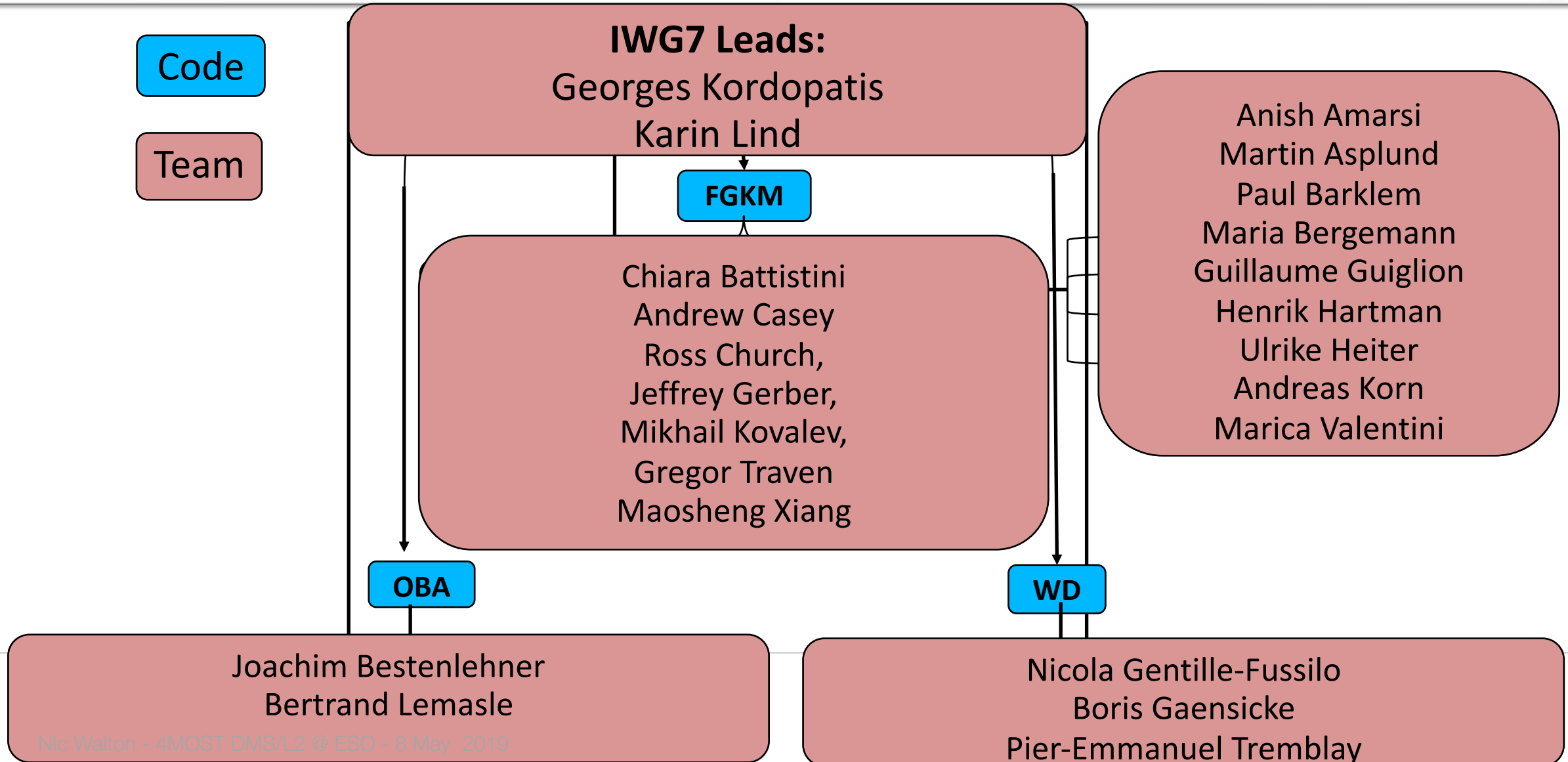


# 4GP: The 4MOST Galactic Pipeline

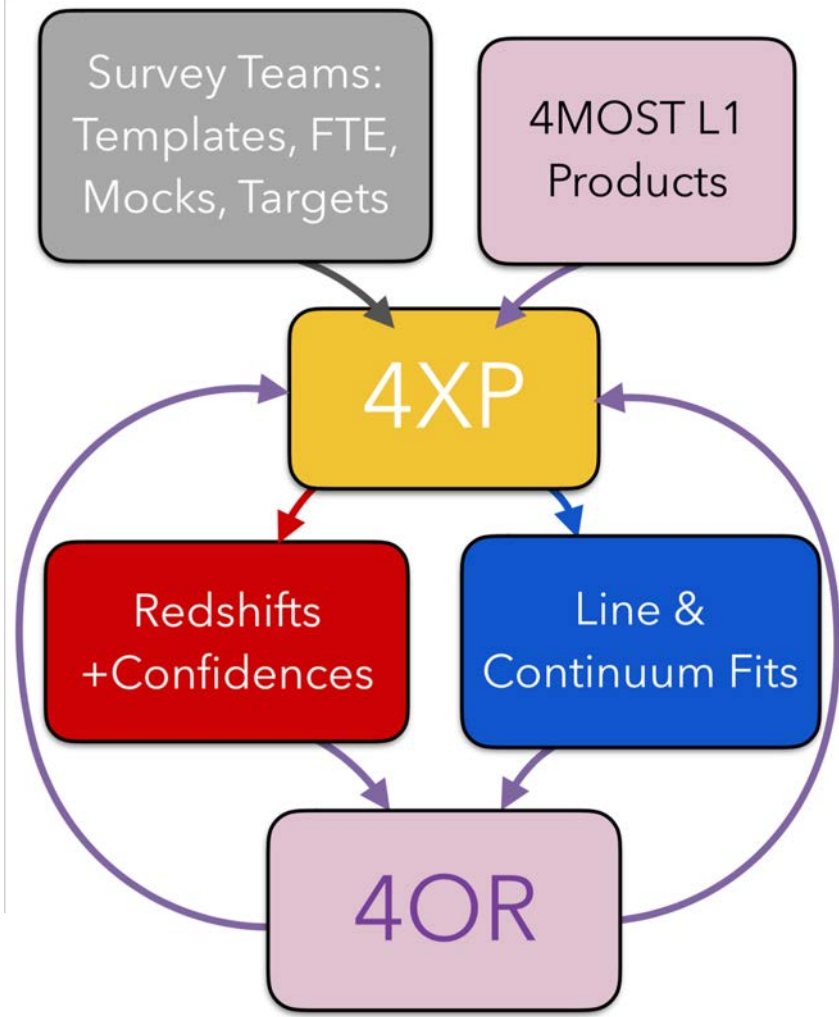


# 4GP: Key Analysis Team

Slide from:  
Lind & Kordopatis



# Extragalactic Analysis Pipeline: 4XP



## 4XP Overview

1. Takes **L1 data** + some investment from survey teams + **spectral templates** + **target catalogues** + **photo-z priors**
2. Turns them into **redshifts** + **line fits** + **continuum fits** + (+ **L3 derived products** - **SFR, metallicity, etc...?**)
3. Returns them to **4OR**

Slide from:  
Davies & Croom

Slide from:  
Davies & Croom

**4XP-Tool** - Interactive to provide GUI user interface to 4XP branches.



# 4XP: Status and Key Deliverables



## 4XP Status

- Test Versions of **4XP-Sim** (Lead: Davies) and **4XP-Z** (Lead: Davies & Croom) on GitHub
  - Internal IWG8 testing underway
- Initial **4XP-Fit** development will start soon (Lead: M. Owers – based on PPXF approach)
- Concept **4XP-Tool** (Lead: Davies) preliminary testing underway

## 4XP Key Deliverables

- Redshifts to 0.0005 ( $1+z$ ) accuracy (best and secondary)
- Redshift errors, probability of solution and best template
- Emission and absorption line measurements (EW, flux, width, continuum at line, SNR) for all lines specified in a pre-defined line list
- Additional lines identified which are not predicted from best-fit solution
- **Desirable**: Derived properties (D4000, H $\delta$  indices, luminosity weighted ages, mass-to-light ratio, BH mass,....)

Slide from:  
Davies & Croom

If you want to know more contact: [luke.j.davies@uwa.edu.au](mailto:luke.j.davies@uwa.edu.au)

# 4SP: 4MOST Selection Function

Slide from:  
Krumpe, Martell  
& Tempel



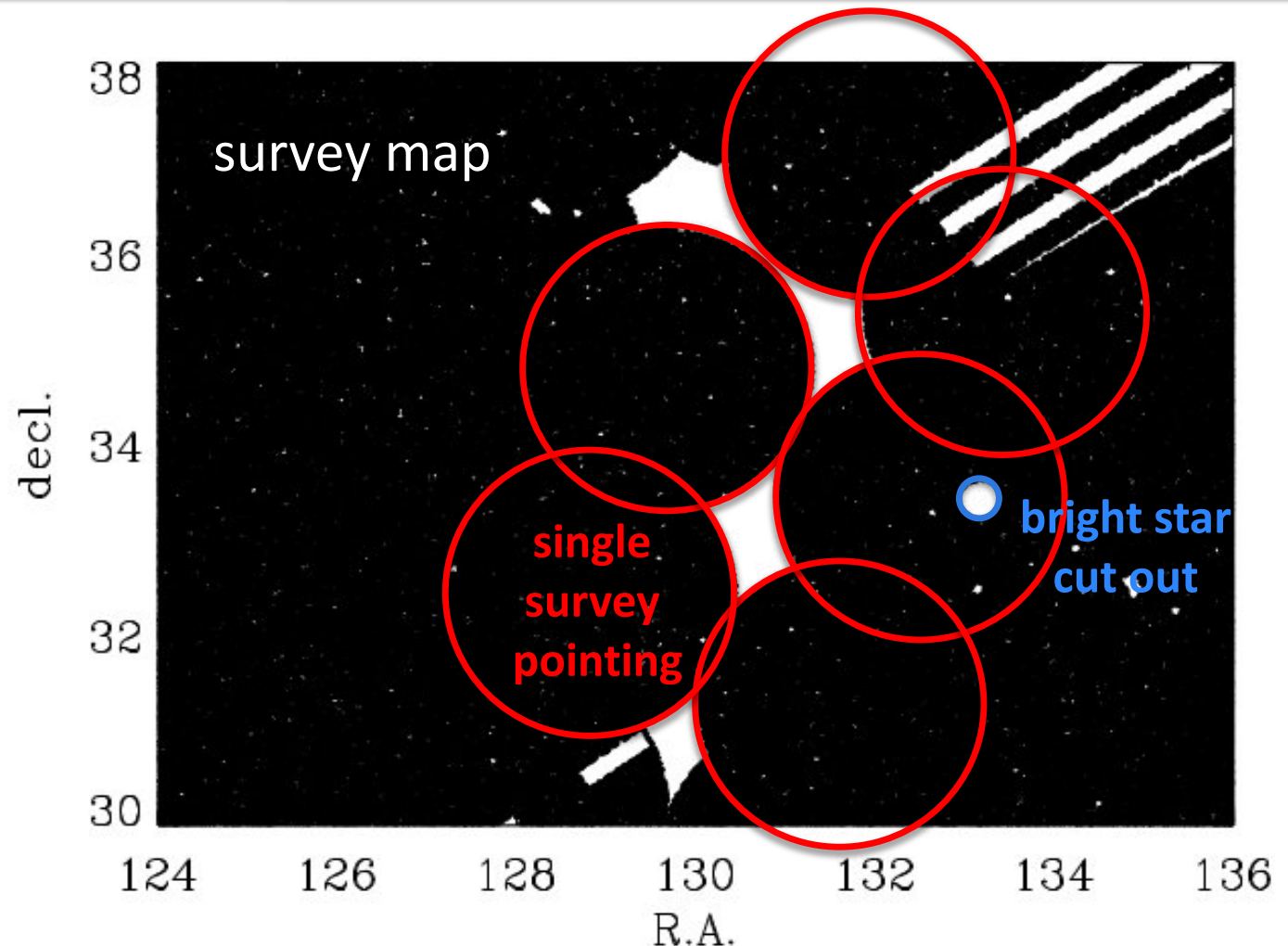
## 4SP: Quantifying Selection Biases

large samples of objects enable statistical measurements

but

what is an intrinsic property and what is due to survey imprint?

selection function allows you to unfold the data by the survey imprint and to recover the intrinsic properties





# 4MOST Selection Function – sky completeness



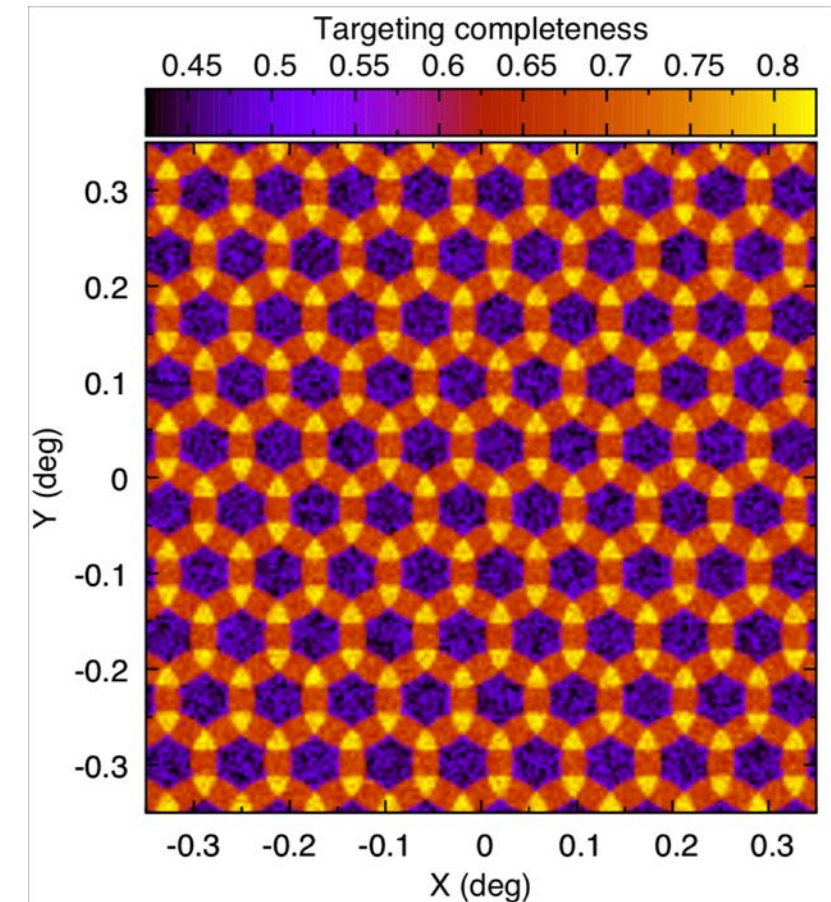
The instrument setup and observing strategy produce additional patterns in the sky.

Survey completeness depends on the following aspects:

- Number density of objects in **all** surveys – every survey is affecting all other surveys.
- Fixed fiber pattern in a focal plane (different for low- and high-resolution fibers).
- 4MOST targeting algorithm – fiber-to-target allocation is probabilistic.
- Other factors.

Slide from:  
Krumpe, Martell  
& Tempel

**Survey completeness is estimated in 4MOST pipeline**



**Pattern in the sky using random targets and random targeting**

# Object selection function

Slide from:  
Krumpe, Martell  
& Tempel



- The 4MOST object selection function quantifies how likely we are to:

## **Observe an object**

Controlled by input catalogue choices:

- Magnitude limits
- Parallax/parallax error limits
- Colour selection

## **Measure its features**

Controlled by the object:

- Feature strength
- Velocity dispersion
- Redshift

Controlled by the data:

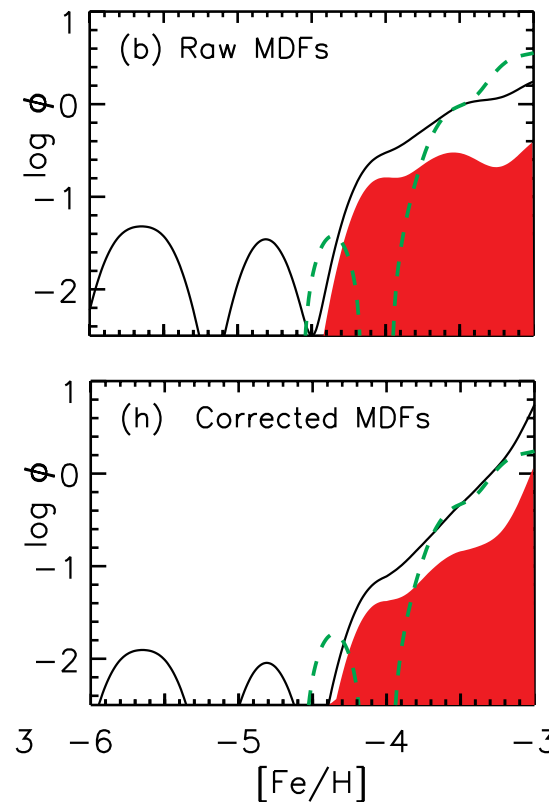
- Dispersion and resolution
- SNR
- Wavelength coverage

# Object selection function

Slide from:  
Krumpe, Martell  
& Tempel



- Any time we want to investigate the relative numbers of objects, we need to correct for the object selection function, e.g.:
  - Mass function
  - Metallicity distribution function
  - Space density of halo stars as a function of  $R_{GC}$



**As an example:** Surveys of metal-poor stars often have target selection biased toward metal-poor stars. To derive the true metallicity distribution function, the raw distribution needs to be corrected using the selection function.

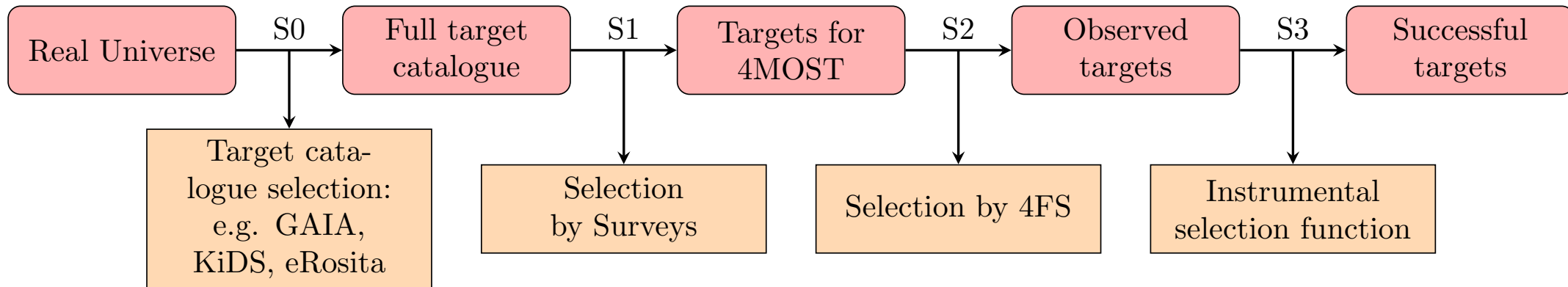
Figure from Yong et al. 2013, ApJ 762:27

# 4SP: Object Selection Function: Front End

## 4SP: Geometric and Observed SF: Back End



- 4SP: OSF optimizes input catalogues



- 4SP: GSF in HEALPix format - completeness information and achieved/limiting signal to noise ratio (SNR) for each HEALPix pixel in the coverage map
- 4SP: GOSF generates completeness functions specific to each survey and survey output parameter
  - E.g. S1: completeness as a function of parallax, proper motion, metallicity, etc

# 4CP: The L2 Classification Pipeline



- 4XP & 4GP template based, hence object classifications limited to the bounds of their training sets
- 4CP provides capability to widen scope of classification
- Observed target classification
  - Basic target classification (confirm input classifications)
- 4CP classifications of 4GP sources
  - Stellar binarity
  - Stellar type
- 4CP classifications of 4XP sources
  - Extragalactic galaxy sub-type: blazar, emission line galaxies etc
  - Quasar redshift



# 4MOST Operational Rehearsals

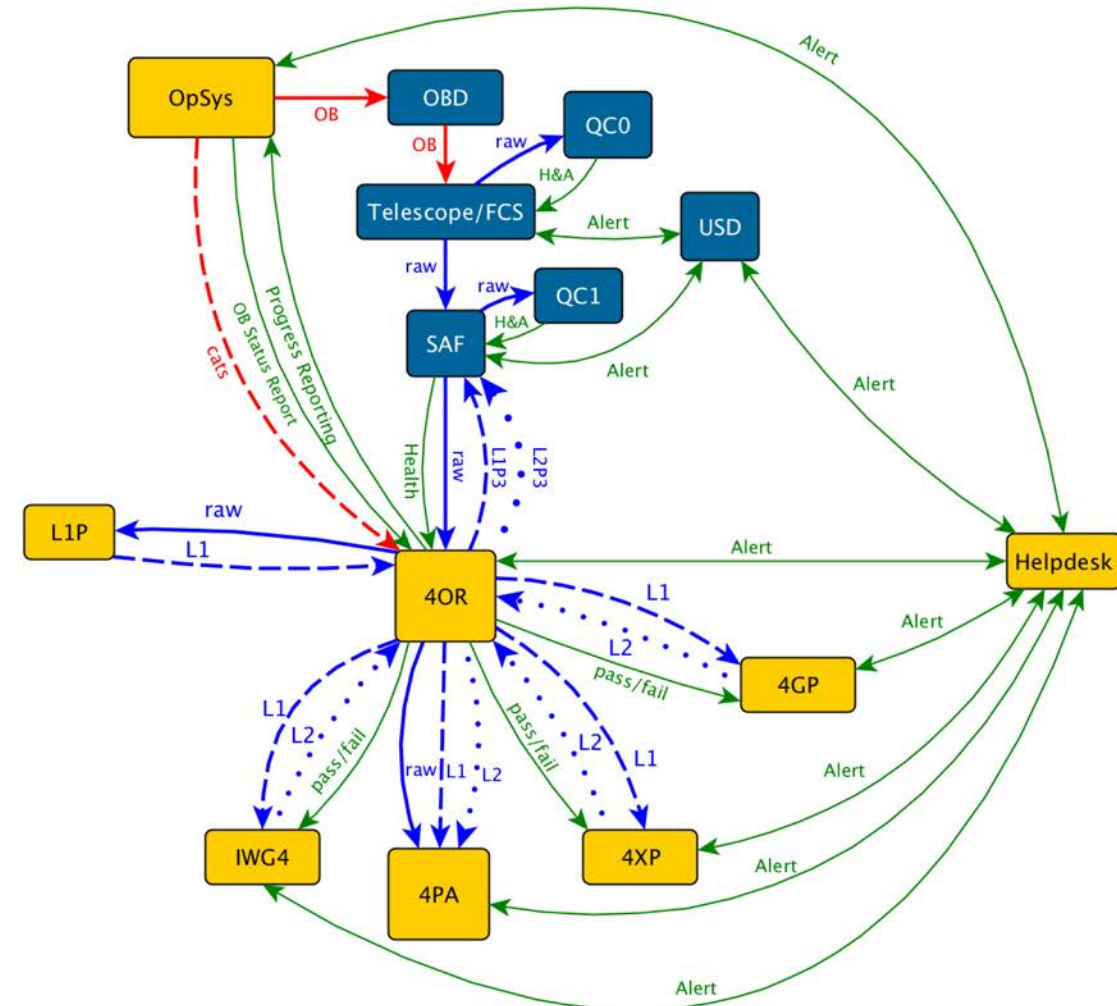
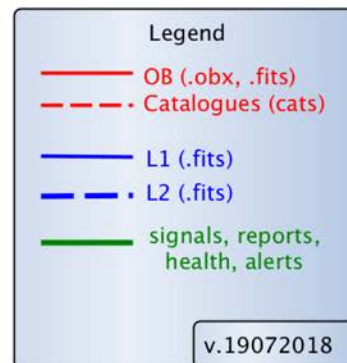
Early science access to the 4MOST (**simulated!**) survey



Series of increasingly complex rehearsals to test the 4MOST data flow system

4OpR1 tests the data interfaces including L2 pipelines

OpR-Secondary Data Flow



Data flow diagram shown here illustrative only



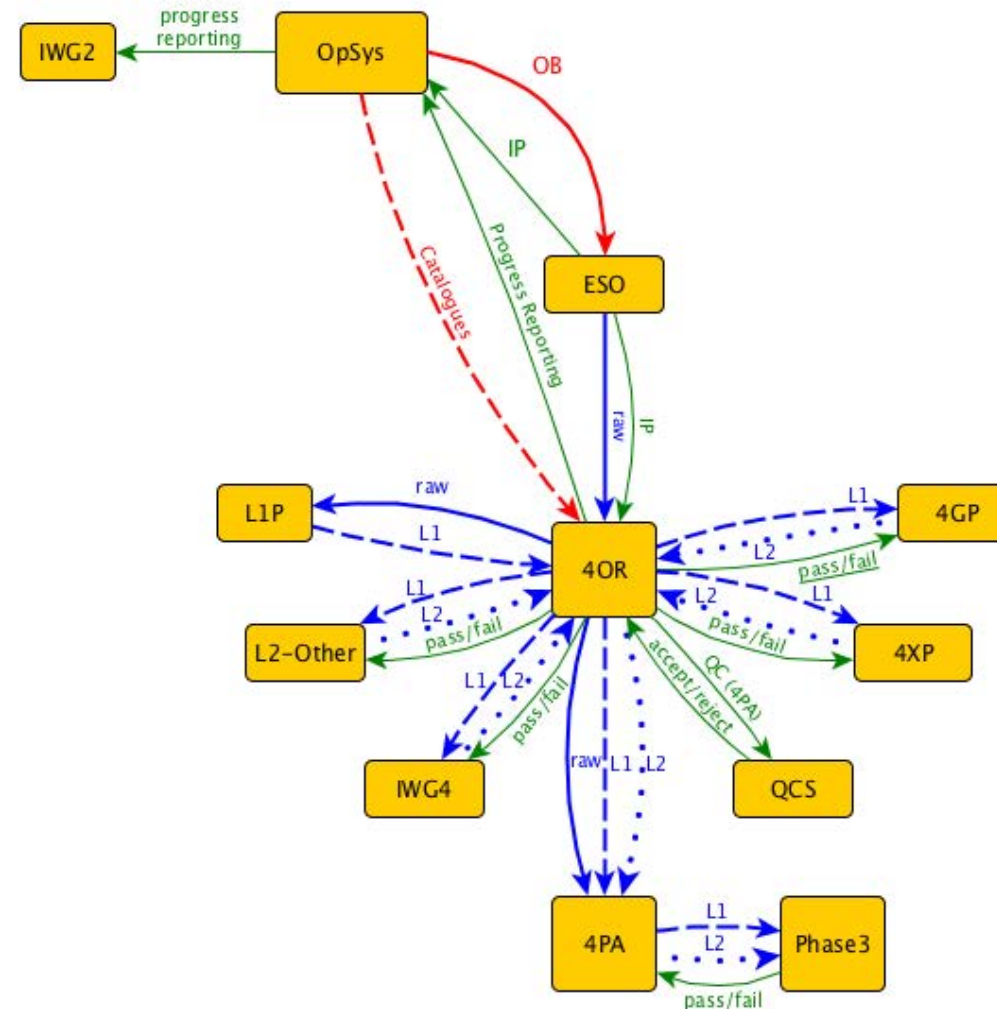
# 4MOST Operational Rehearsals

IWG and Science team participation



OpR2.5

4OpR2 tests the data products including those from the L2 pipelines



Legend

- OB (.obx,.fits)
- Catalogues
- raw (.fits)
- L2 (.fits)
- L1 (.fits)
- signalling
- health reporting
- progress reporting
- alerts

v.19022018

Data flow diagram shown here illustrative only

# OpR2+ Science Team Outputs QA

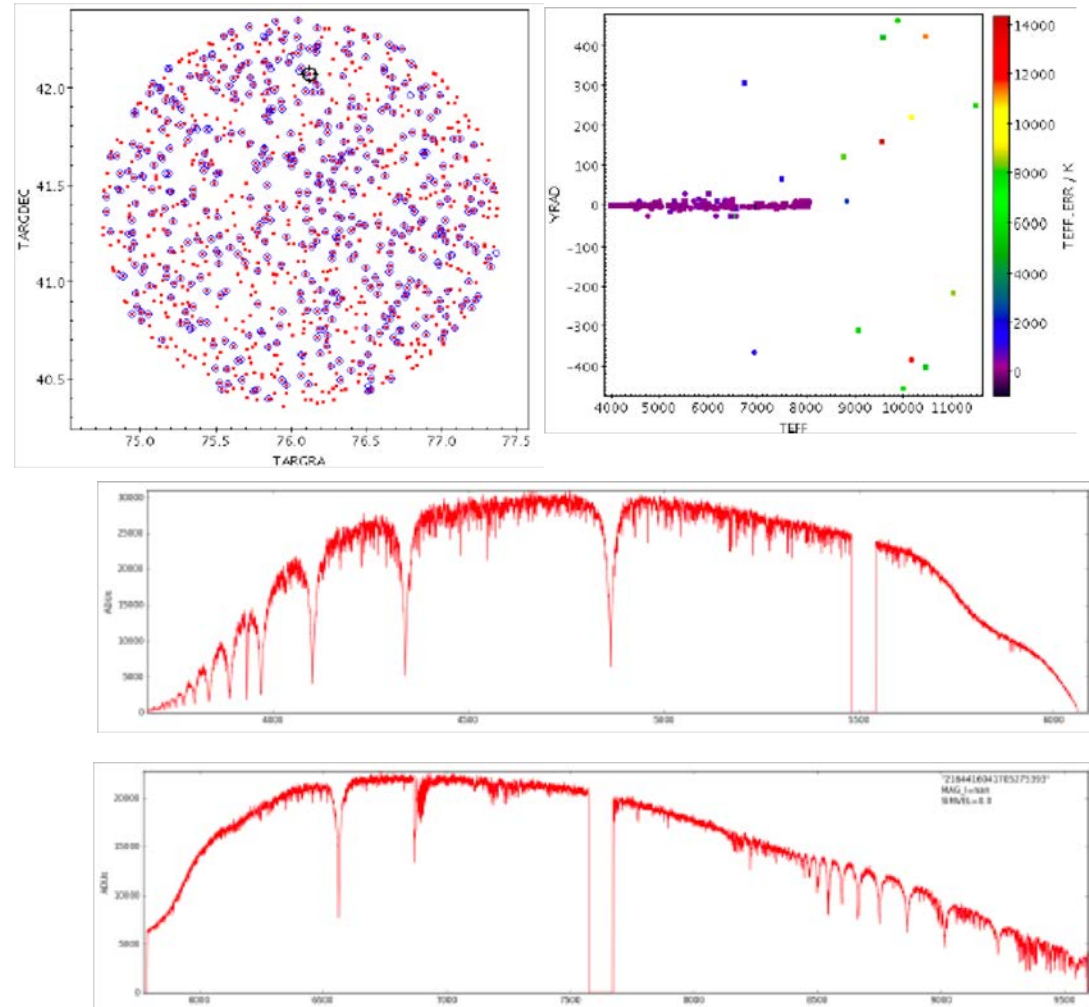
examples from WEAVE OpR2.5



QA of pipeline outputs,  
e.g.:

- Sky fibre distribution
- Sky background
- Spectral extraction
- Astrophysical parameters

Plots show some examples from a WEAVE LR halo field – OpR2.5 – fibres in a field, Vrad vs  $M_H$ , spectra extraction/classification (A1V star in this example)



# 4MOST Operational Rehearsals

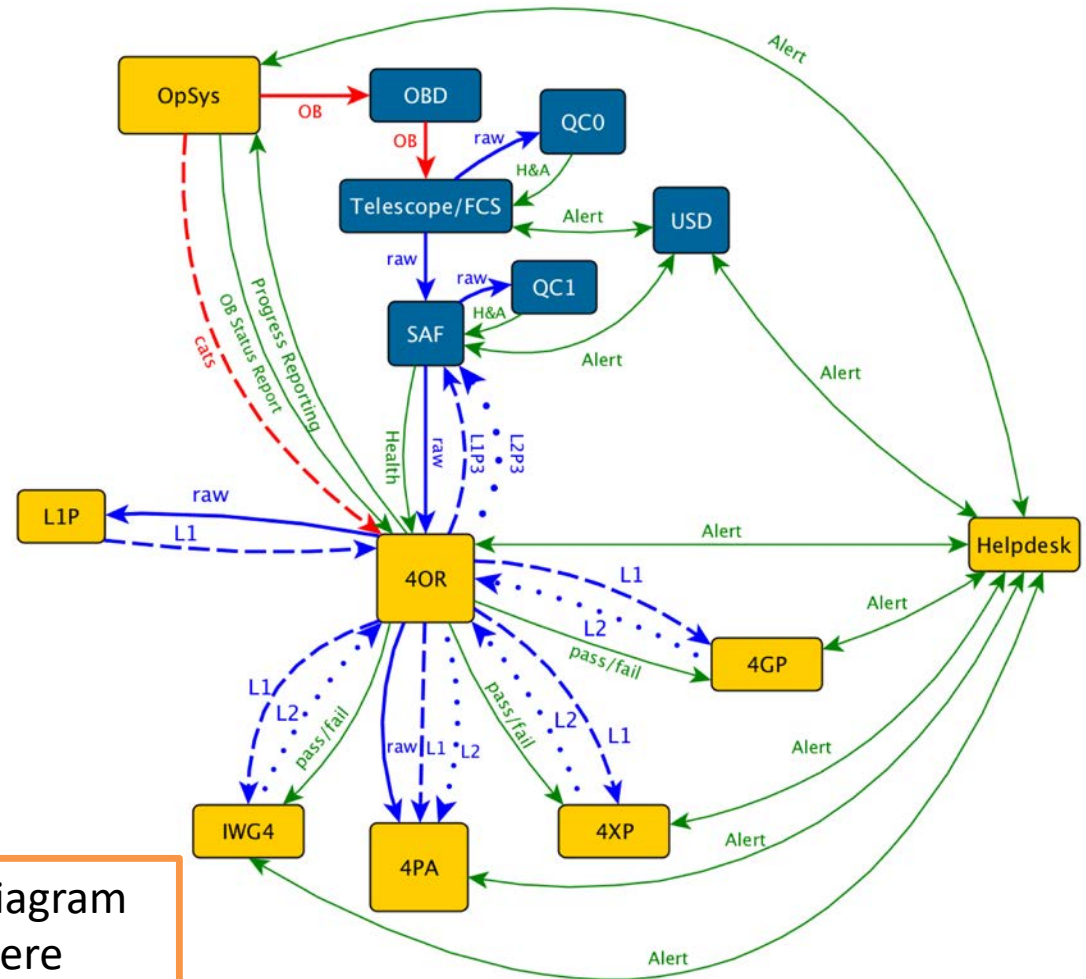
IWG and Science team participation



4OpR3 full simulation of multiple semester data runs testing complete processing system

Significant science team involvement in defining OBs and QA of output data products

Data flow diagram shown here illustrative only



# 4OpR Timeline

version May 2019

Milestone	Date
FDR	May 2018
OpR Plan Released	September 2018
‘Node’ Development Plans (provide what and by when)	Early 2019
OpR1 Readiness Review / 4OpR1 KO	19 June 2019
<b>OpR1</b>	<b>October 2019</b>
Local Acceptance HRS	July 2020
OpR2 Readiness Review	June 2020
<b>OpR2</b>	<b>September 2020</b>
Local Acceptance LRS (A,B)	March 2021
OpR3 Readiness Review	January 2021
<b>OpR3</b>	<b>April 2021</b>
TRR	August 2021
PAE	February 2022
System AIV commences	April 2022
<b>Commissioning (OpR4)</b>	<b>May 2022</b>
PAC	November 2022

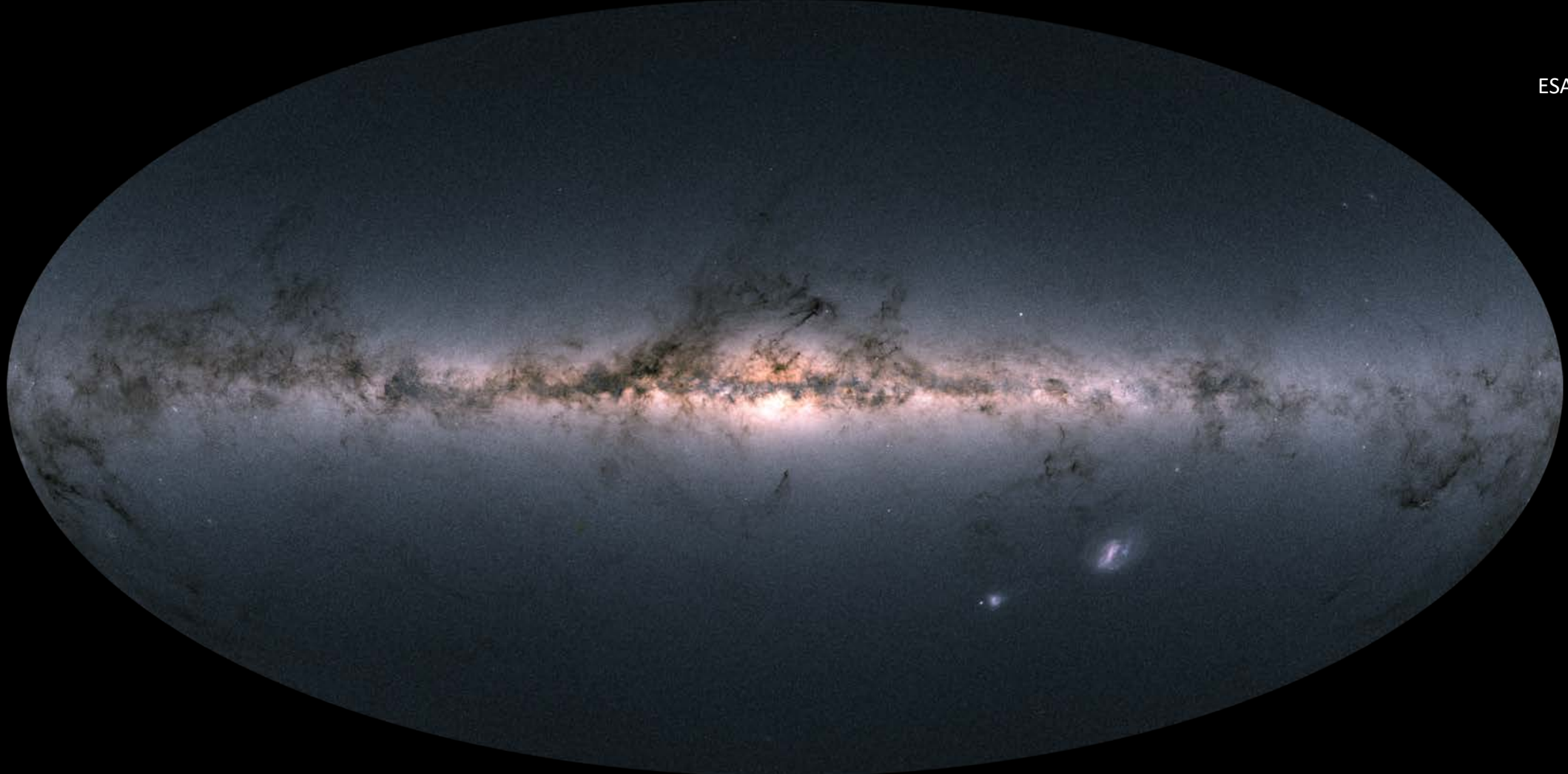


# Gaia DR2 Sky: all sky high resolution

image shows Gaia source colour



Credit:  
ESA/Gaia/DPAC



Targets for 4MOST/ Simplifying 4MOST calibration