SKA Science Data Processing and Observatory Data Products

Mark Ashdown

University of Cambridge

Outline

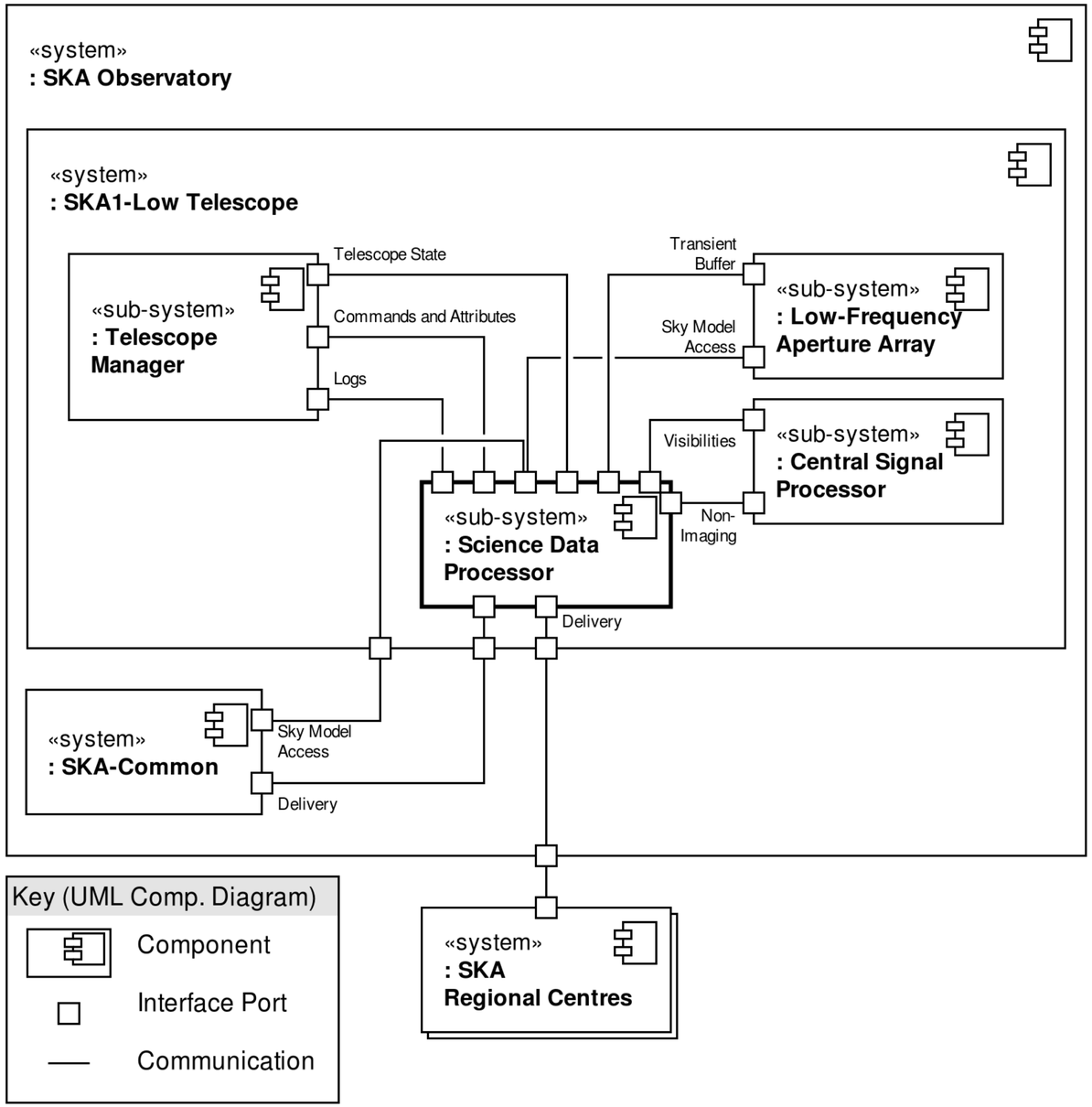
- Science Data Processor consortium
- Design: software and hardware architecture
- Computational requirements: parametric model
- Operational model
- Observatory data products

SDP Consortium

- Like other elements of the telescope, the SDP was designed by a consortium
- Started in 2013
- Led by the University of Cambridge with participation from 40 organisations around the globe
- SDP Critical Design Review took place November 2018 - March 2019
- Design was adopted by SKAO and has been taken forward into System CDR (happening now)

SDP Design

- Software and Hardware Architecture documented in Software Engineering Institute (SEI) style
- Made up of short(-ish) documents called “views” with a standard structure
  - Each one starts with a diagram
  - Describe the elements of the diagrams and their relationships
  - Reason about the properties they must have
- Different types of view
  - Component and connector views show the parts of the running system
  - Module views shows the elements of the software



SDP in context (C&C view)

Deployments:

- SKA1-Low (shown here)
- SKA1-Mid
- SKA Regional Centres

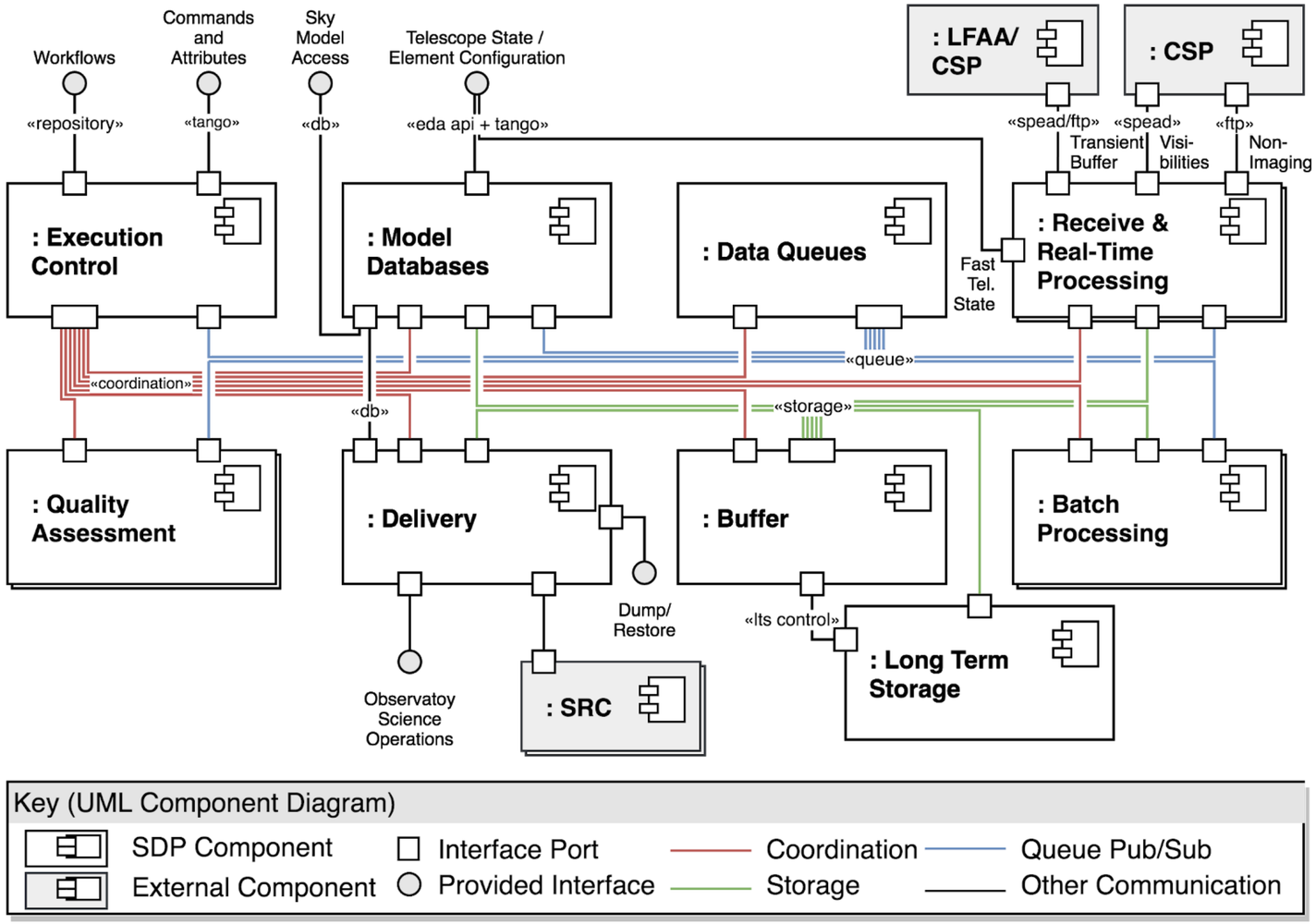
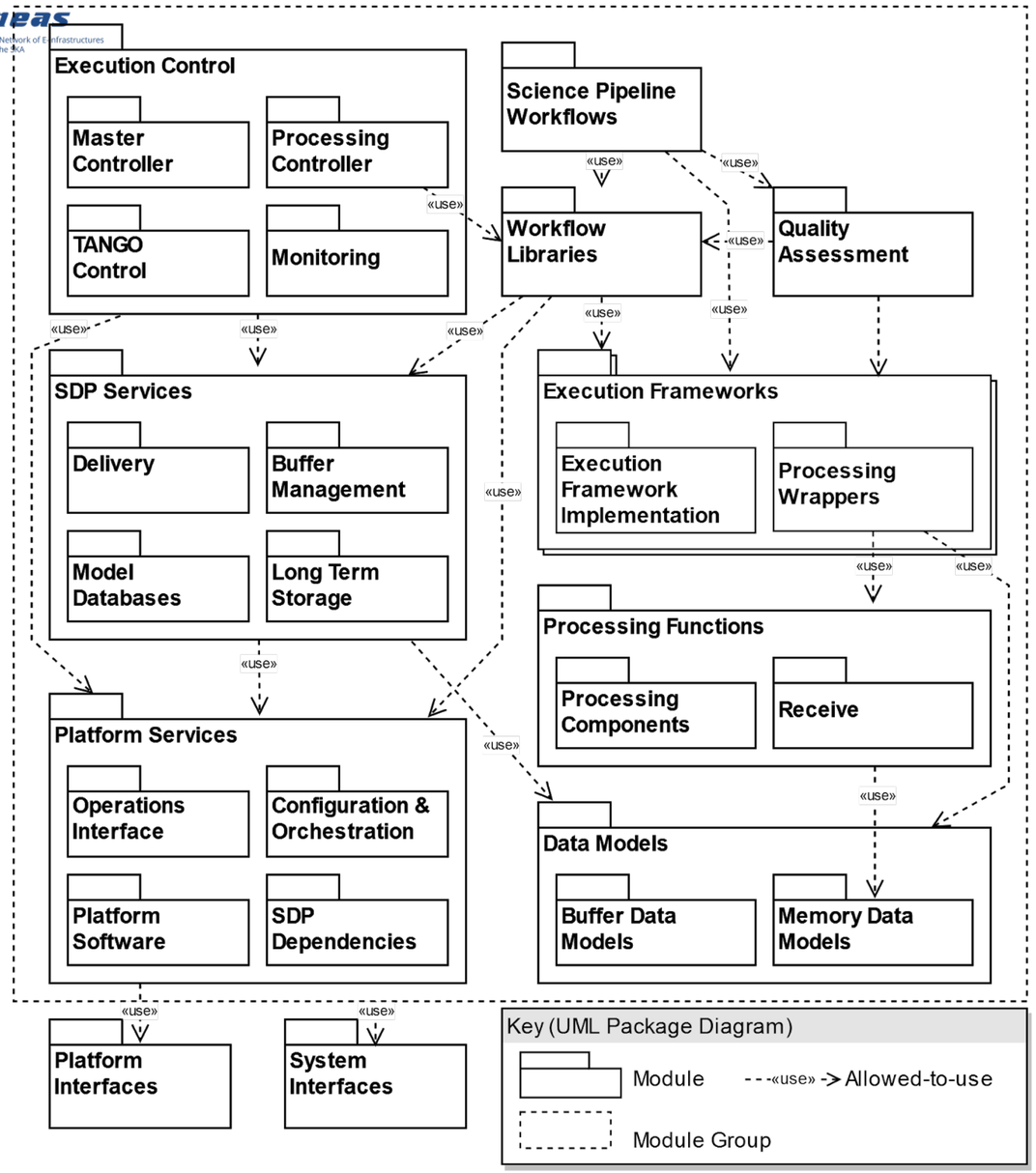


Figure 1: Science Data Processor (SDP) Component & Connector Primary Representation

SDP operational system C&C view

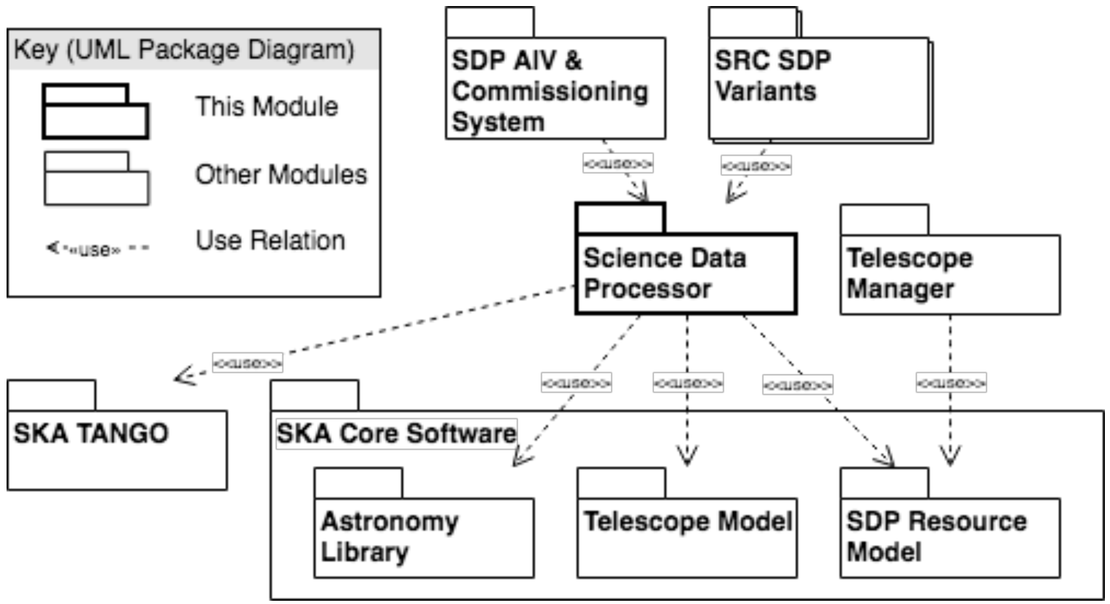
Shared Data & Publish/Subscribe:

- Buffer (Storage)
  high throughput
  high capacity
  requires locality!
- Data Queues
  low latency
  good throughput
  scales globally
- Coordination
  high reliability
  low latency (read)
  limited access



SDP Module views

- SDP system software (left)
- Relationships to SKA Core Software and SDP variants (below)



Computational Requirements

- The consortium estimated the size of the SDP at each of the two telescopes based on the high-priority science objectives (HPSOs)
- Mathematical model of the processing
  - Computes the number of flops required to process the data based on observational parameters (max. baseline, no. frequency channels, no. major cycles, …)
- This is the SDP parametric model
  - Implemented in Python with interactive Jupyter notebooks
  - http://gitlab.com/ska-telescope/sdp-par-model
- The model can be used with different input parameters
  - The HPSOs were the official “boundary conditions” for the CDR

Parametric Model Pipelines

- Real-time imaging pipelines
  - Ingest: receive and pre-process visibilities from CSP
  - RCAL: real-time calibration
  - FastImg: fast imaging for slow transient detection
- Batch imaging pipeline
  - ICAL: iterative self calibration (including direction-dependent calibration)
  - DPrepA: preparation of continuum image data products
  - DPrepB: preparation of coarse spectral image data products
  - DPrepC: preparation of fine spectral image data products
  - DPrepD: preparation of calibrated averaged visibilities data products (EoR projects)
- Non-imaging pipelines
  - Pulsar search and timing, single-pulse transient detection

SKA1-Low Compute and Data Rate

SKA1_Low:

HPSO	Time [%]	Tobs [h]	Ingest [Pflop/s]	RCAL [Pflop/s]	FastImg [Pflop/s]	ICAL [Pflop/s]	DPrepA [Pflop/s]	DPrepB [Pflop/s]	DPrepC [Pflop/s]	DPrepD [Pflop/s]	Total RT [Pflop/s]	Total Batch [Pflop/s]	Total [Pflop/s]
hpso01	15.6	5.0	0.63	0.75	0.38	6.88	2.35	2.50	5.12	0.30	1.76	17.16	18.92
hpso02a	15.6	5.0	0.63	0.75	0.38	4.01	2.35	2.50	5.12	0.30	1.76	14.29	16.05
hpso02b	15.6	5.0	0.63	0.75	0.38	4.01	2.35	2.50	5.12	0.30	1.76	14.29	16.05
hpso04a	39.8	0.7	0.63	0.22	0.12	-	-	-	-	-	0.96	0.00	0.96
hpso05a	13.4	0.7	0.63	0.22	0.12	-	-	-	-	-	0.96	0.00	0.96
Average	-	-	0.63	0.47	0.24	2.33	1.10	1.17	2.40	0.14	1.34	7.14	8.47

SKA1_Low:

HPSO	Time [%]	Tobs [h]	Npix (side)	Channels (DPrepB)	Channels (DPrepC)	Image size [GB]	Non-Vis Rate [Gbit/s]	Visibility Size [TB]	Visibility Rate [Gbit/s]	Total Rate [Gbit/s]
hpso01	15.6	5.00	18344	500	1500	2.7	8.5	205.8	91.4	99.9
hpso02a	15.6	5.00	18344	500	1500	2.7	8.5	205.8	91.4	99.9
hpso02b	15.6	5.00	18344	500	1500	2.7	8.5	205.8	91.4	99.9
hpso04a	39.8	0.67	-	-	-	-	0.7	-	-	0.7
hpso05a	13.4	0.67	-	-	-	-	2.6	-	-	2.6
Average	-	-	-	-	-	-	4.6	-	42.8	47.4

SKA1-Mid Compute

SKA1_Mid:

HPSO	Time [%]	Tobs [h]	Ingest [Pflop/s]	RCAL [Pflop/s]	FastImg [Pflop/s]	ICAL [Pflop/s]	DPrepA [Pflop/s]	DPrepB [Pflop/s]	DPrepC [Pflop/s]	DPrepD [Pflop/s]	Total RT [Pflop/s]	Total Batch [Pflop/s]	Total [Pflop/s]
hpso04b	1.0	0.2	0.60	0.94	0.36	-	-	-	-	-	1.91	0.00	1.91
hpso04c	3.1	0.2	0.60	0.56	0.23	-	-	-	-	-	1.39	0.00	1.39
hpso05b	2.1	0.2	0.60	0.95	0.55	-	-	-	-	-	2.11	0.00	2.11
hpso13	6.5	8.0	0.14	0.04	0.02	0.40	0.14	0.13	0.48	-	0.20	1.16	1.36
hpso14	2.6	8.0	0.15	0.03	0.01	0.15	0.09	0.09	0.31	-	0.19	0.63	0.83
hpso15	16.5	4.4	0.08	0.01	0.00	0.04	0.03	0.03	0.09	-	0.10	0.19	0.28
hpso18	13.1	0.0	0.60	0.94	0.36	-	-	-	-	-	1.91	0.00	1.91
hpso22	7.9	8.0	0.60	0.75	0.38	6.80	2.99	3.08	-	-	1.74	12.88	14.62
hpso27and33	13.1	0.1	0.19	0.09	0.05	0.26	0.31	0.47	-	-	0.33	1.04	1.37
hpso32	13.1	2.2	0.12	0.09	0.04	0.24	-	0.29	-	-	0.25	0.53	0.78
hpso37a	13.1	3.8	0.60	0.88	0.39	5.44	3.43	3.57	-	-	1.88	12.45	14.33
hpso37b	2.6	8.0	0.60	0.88	0.39	7.96	3.43	3.57	-	-	1.88	14.96	16.84
hpso37c	2.6	8.0	0.60	0.88	0.39	7.96	3.43	3.57	-	-	1.88	14.96	16.84
hpso38a	1.3	8.0	0.60	0.77	0.43	5.24	3.38	3.50	-	-	1.80	12.12	13.92
hpso38b	1.3	8.0	0.60	0.77	0.43	6.93	3.38	3.50	-	-	1.80	13.81	15.61
Average	-	-	0.36	0.44	0.20	1.92	1.01	1.10	0.06	0.00	1.00	4.09	5.09

SKA1-Mid Data Rate

SKA1_Mid:

HPSO	Time [%]	Tobs [h]	Npix (side)	Channels (DPrepB)	Channels (DPrepC)	Image size [GB]	Non-Vis Rate [Gbit/s]	Visibility Size [TB]	Visibility Rate [Gbit/s]	Total Rate [Gbit/s]
hpso04b	1.0	0.17	-	-	-	-	2.3	-	-	2.3
hpso04c	3.1	0.17	-	-	-	-	2.3	-	-	2.3
hpso05b	2.1	0.25	-	-	-	-	6.9	-	-	6.9
hpso13	6.5	8.00	25339	160	3200	5.1	4.2	-	-	4.2
hpso14	2.6	8.00	18814	300	5000	2.8	2.8	-	-	2.8
hpso15	16.5	4.40	10837	260	2500	0.9	0.8	-	-	0.8
hpso18	13.1	0.02	-	-	-	-	0.1	-	-	0.1
hpso22	7.9	8.00	110601	1000	0	97.9	48.1	-	-	48.1
hpso27and33	13.1	0.12	23549	700	0	4.4	99.3	-	-	99.3
hpso32	13.1	2.20	-	-	-	-	1.3	-	-	1.3
hpso37a	13.1	3.80	94195	700	0	71.0	60.6	-	-	60.6
hpso37b	2.6	8.00	94195	700	0	71.0	28.8	-	-	28.8
hpso37c	2.6	8.00	94195	700	0	71.0	28.8	-	-	28.8
hpso38a	1.3	8.00	113204	1000	0	102.5	50.4	-	-	50.4
hpso38b	1.3	8.00	113204	1000	0	102.5	50.4	-	-	50.4
Average	-	-	-	-	-	-	28.4	-	0.0	28.4

Operational Model

- How will telescope users interact with SDP?
- The answer is: not directly!
- At the observation design stage, users will work with SKA Observatory staff to choose the pipelines to be run on their data (and the parameters of the pipeline)
  - Use a model of the computational cost to see how feasible the desired processing is
- SDP pipelines will be available to users at SRCs
  - Test out the pipelines with simulations
  - Possibility of receiving a small but representative visibility data set to tune the pipeline parameters for your project

Observatory Data Products

- Image data products
  - Image cubes
  - Gridded visibilities
- Non-image data products
  - **Science Data Model** - more on this in a moment
  - Calibrated averaged visibilities (designed for EoR projects)
  - Transient source catalogue
  - Pulsar timing solutions
  - Sieved pulsar and transient candidates
  - Transient buffer data

Science Data Model

- The Science Data Model is very import for users of the image and gridded visibility data products
- It encapsulates all of the information about
  - Telescope configuration and state during the observation
  - Sky model (input and output)
  - Processing carried out on the data, for example calibration solutions (and their history)
  - Quality assessment of the data products
  - etc.
- The purpose of the SDM is to provide as much information as possible (in the absence of the raw visibilities) to understand the image data products