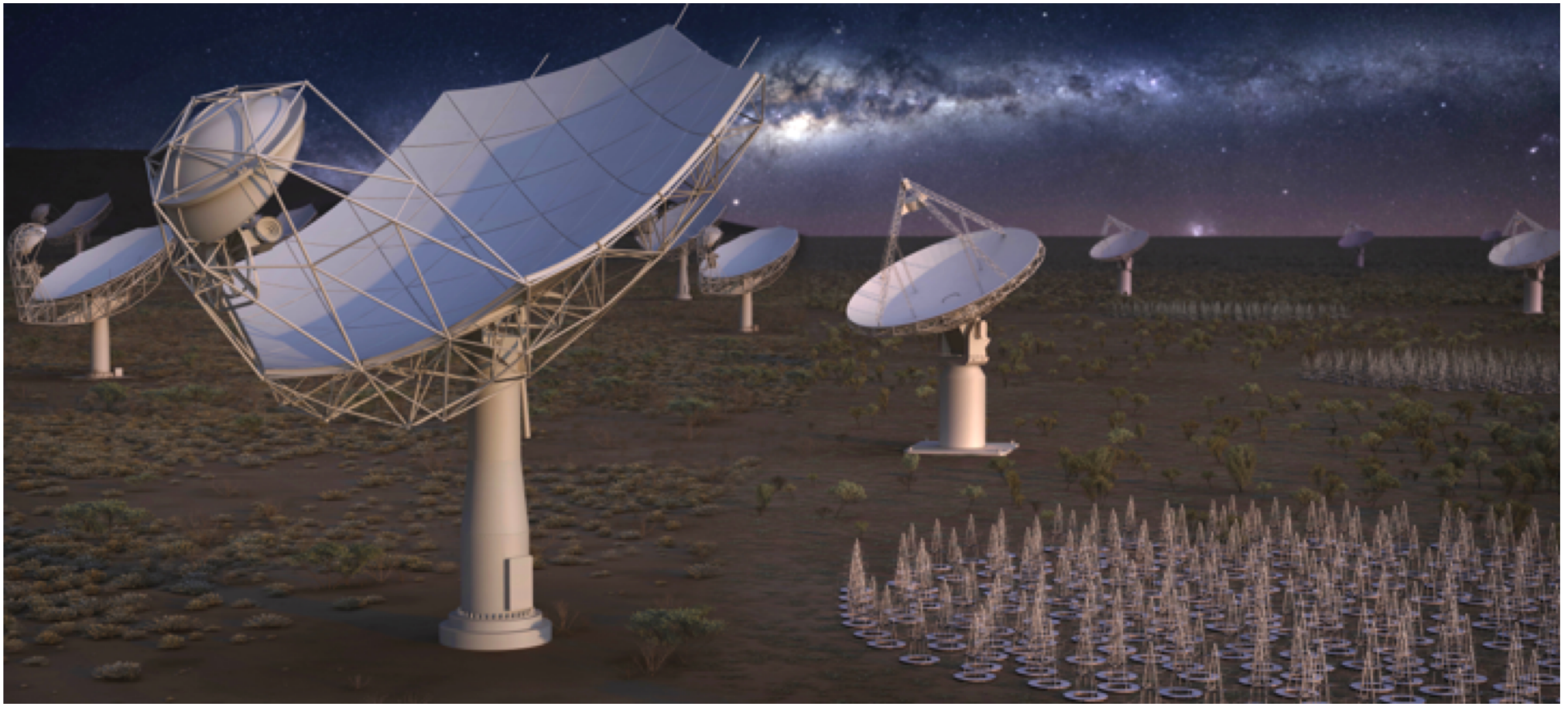
Square Kilometre Array

SKA & SRCCG update



Rosie Bolton

SQUARE KILOMETRE ARRAY

Exploring the Universe with the world's largest radio telescope

SKA Update (1)

Welcome to Spain (June ‘18) and France (July ‘18)

12 Members:
Australia
Canada
China
France
India
Italy
Netherlands
New Zealand
South Africa
Spain
Sweden
UK

50% European membership…

SKA Update (2)

- Convention has been initialed by 7 countries: IT, AU, ZA, SE, UK, CN and NL
  - Process of ministerial signing can now go ahead.

- It's Critical Design Review season!
  - Telescope Manager Passed
  - Central Signal Processor, overall review and sub-element reviews passed

- Science Data Processor consortium has a pre-CDR meeting in June and has now had its review readiness notice accepted for the main CDR – document deadline end October (so be patient with SDP members of AENEAS until then, after which be patient with the reviewers ahead of CDR meeting mid Jan 2019).

- SKA HQ – building has been handed over to SKA. A few offices occupied, remainder scheduled Oct/Nov 2018.

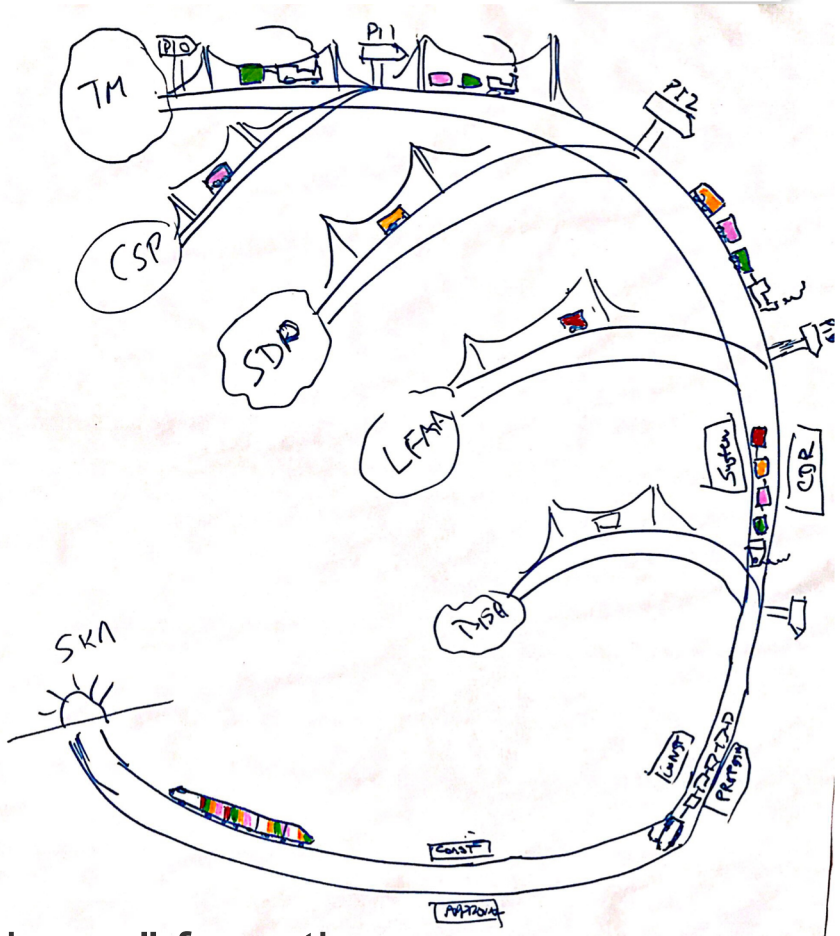
- SKA Science meeting in new HQ April 2019. https://indico.skatelescope.org/event/467/

Project Management

- CDRs signal the end of the formal consortium agreements
- As each CDR closes out, the work on that element of the design moves into the next phase:
  - Bridging: formally, this is the work needed to get the system design ready for system CDR
  - Specifically, the “Design Adoption Process” where SKAO takes ownership of the element designs after CRD and develops them into a complete observatory system design
  - Bridging is NOT construction: DOES NOT LEAD TO CONSTRUCTION CREDIT
  - Governed by individual MoUs between institutes providing in-kind effort and the SKAO
  - Gives the office access to the vital expertise that lies within many individuals

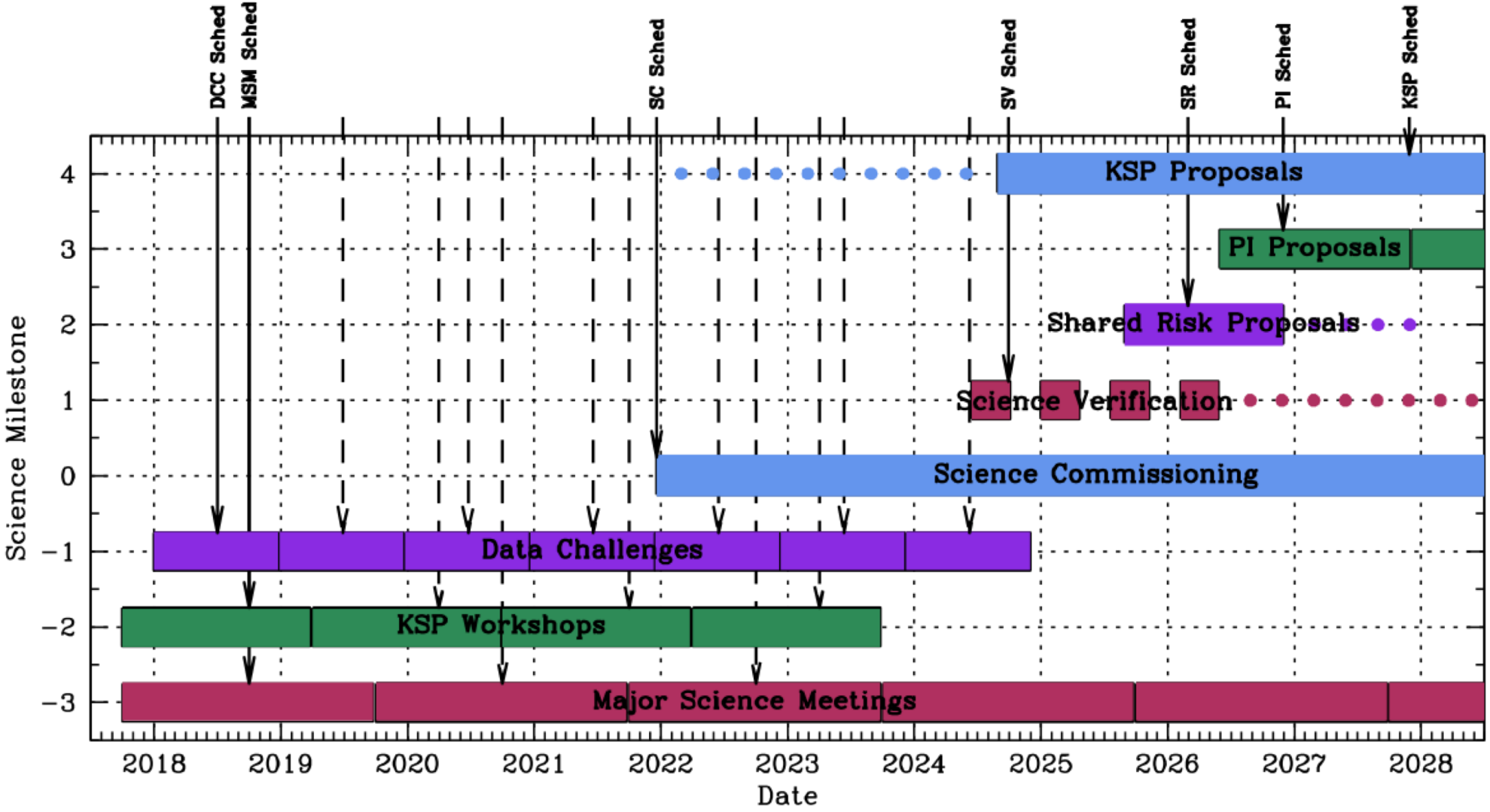
Scaled Agile Framework Project Management

- During the bridging period, SKAO will adopt an agile engineering approach to SOFTWARE DEVOLOPMENT
  - specifically SAFe – see https://www.scaledagileframework.com/essential-safe/
  - This is a pragmatic way of maximising output from a fixed-resource pool with continuous release
  - Works best with consistent FTE commitments from individuals



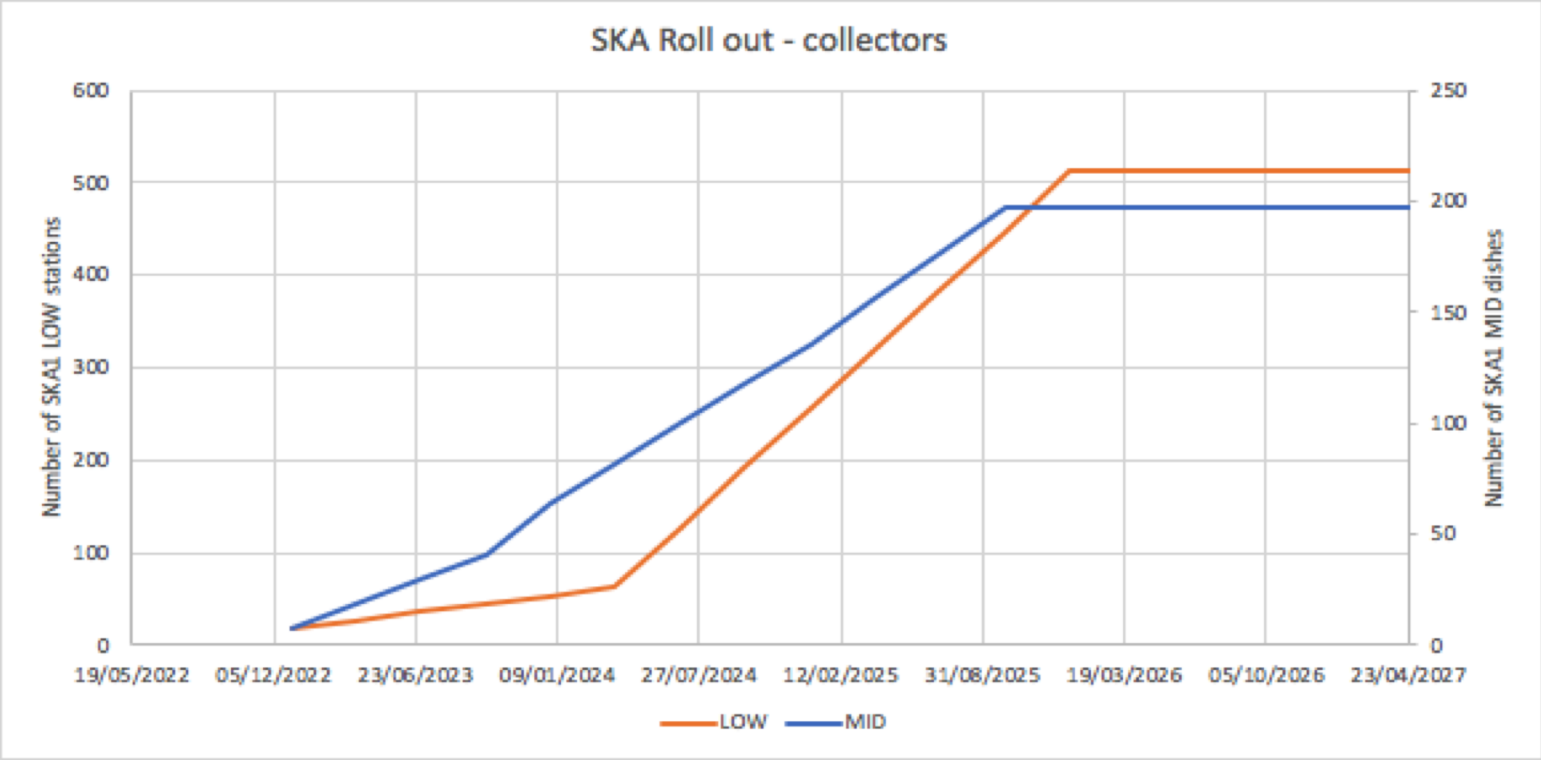
"In words, bridging is a railway where "carriages" from the pre-construction consortia "islands" cross bridges and join a single train running on the SKA track…As we progress the train gets longer and more streamlined until after bridging we are a streamlined bullet train aligned towards the goal of the SKA."
(https://confluence.skatelescope.org/display/SE/Bridging+Vision)

Schedule to SKA Science



Updating SRC requirements

- Quite complementary to the AENEAS WP3 work, which is HPSO based.
- Key Science Projects are not likely to be fully underway until 2028
- But, SRCs will need to be up and running smoothly before then
- Seek to develop Top level view of potential SRC work done in SKA's name before key science
- SKA tations



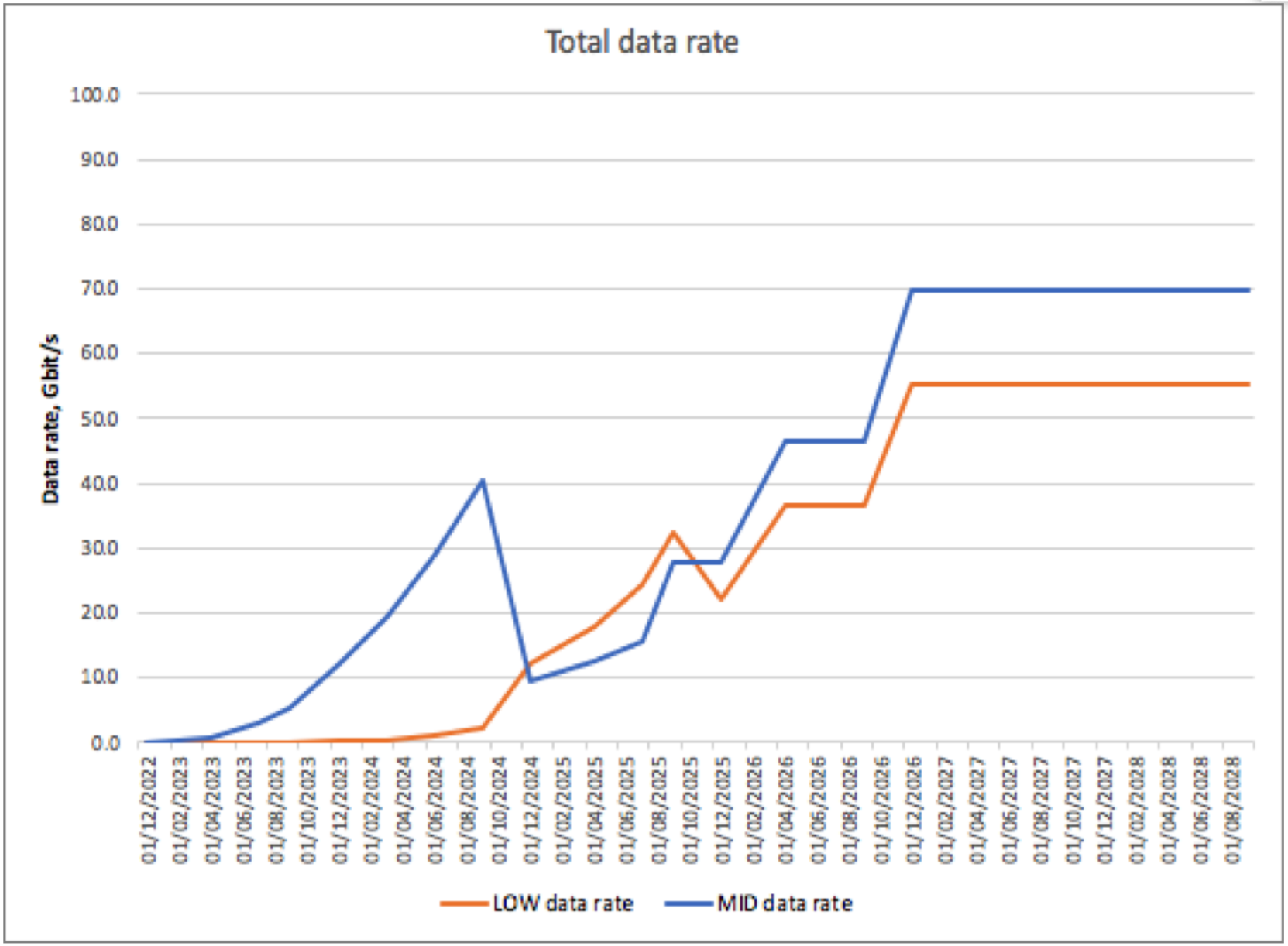
SKA roll out: How much data?

- Simple logic – but make lots of (clear) assumptions:
  - Take collectors and Correlator capability and calculate visibility data rate
  - Make assumptions about the fraction of time each array is used for imaging
    - Commissioning work in early years (assume 30%)
    - PI science once arrays complete (assume 75%)
  - During this imaging time assume:
    - Some (average) cadence for continuum images (1 per 4hrs), ~10Gbit/s total
    - Some (average) incidence of spectral line cubes, ~10 Gbit/s total

SKA roll out: How much data

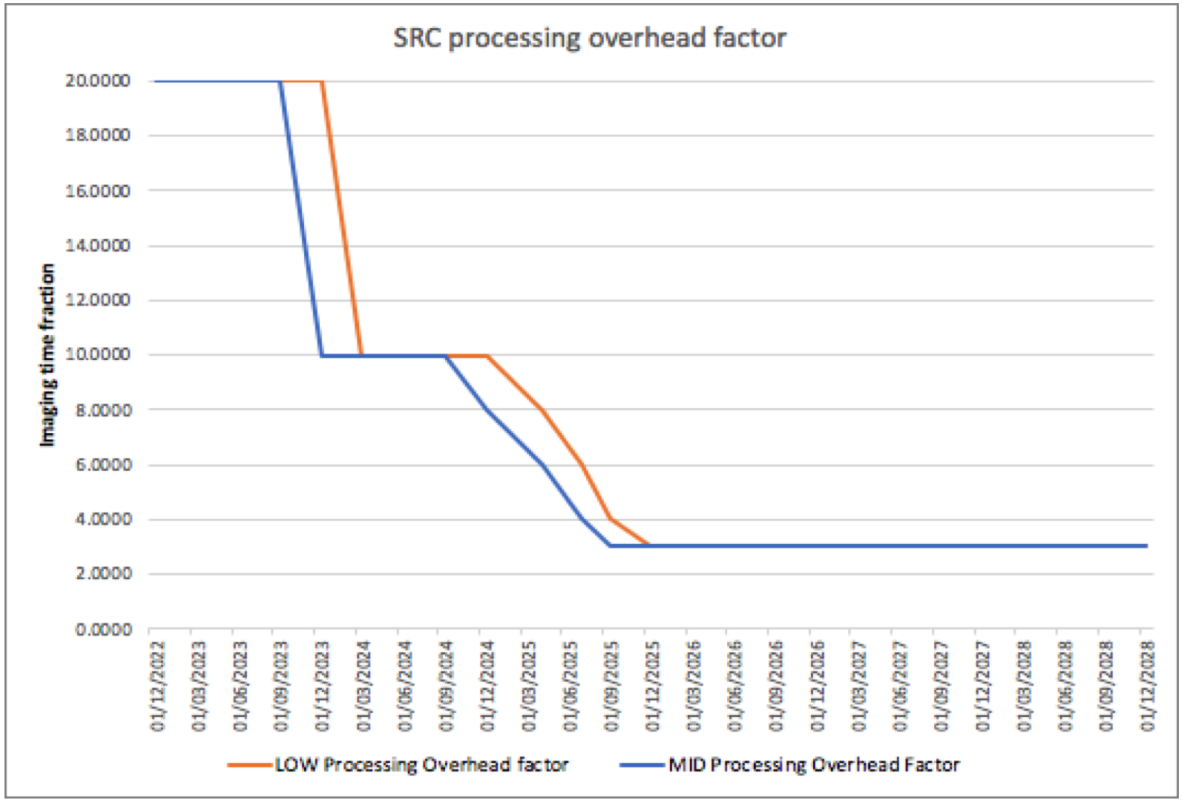
- IMPORTANTLY: Identify commissioning use case for SRCs
  - This is new to our thinking
  - SKA will have the capability to perform its own commissioning but it is recognized that engaging expertise from the comminuty can be very valuable, both for actually aiding commissioning tasks but also to develop TRUST and expertise in the community.
  - At full scale, even once fully commissioned, there will be occasions when successful SKA projects identify a need to use SKA Regional Centers to re-process
    - Parameter tuning / workflow selection for established SDP algorithms at a break-point in a large survey
    - Testing new SDP algorithms before porting them into the SDP (via SKAO)
    - BUT this does not lead to workflows that circumvent the SDP step as a general “but I’m a bit special” use case.
  - Export visibility data to SRCs a very small fraction of the time (2.5% once at full scale)

SKA Roll out: How much data?



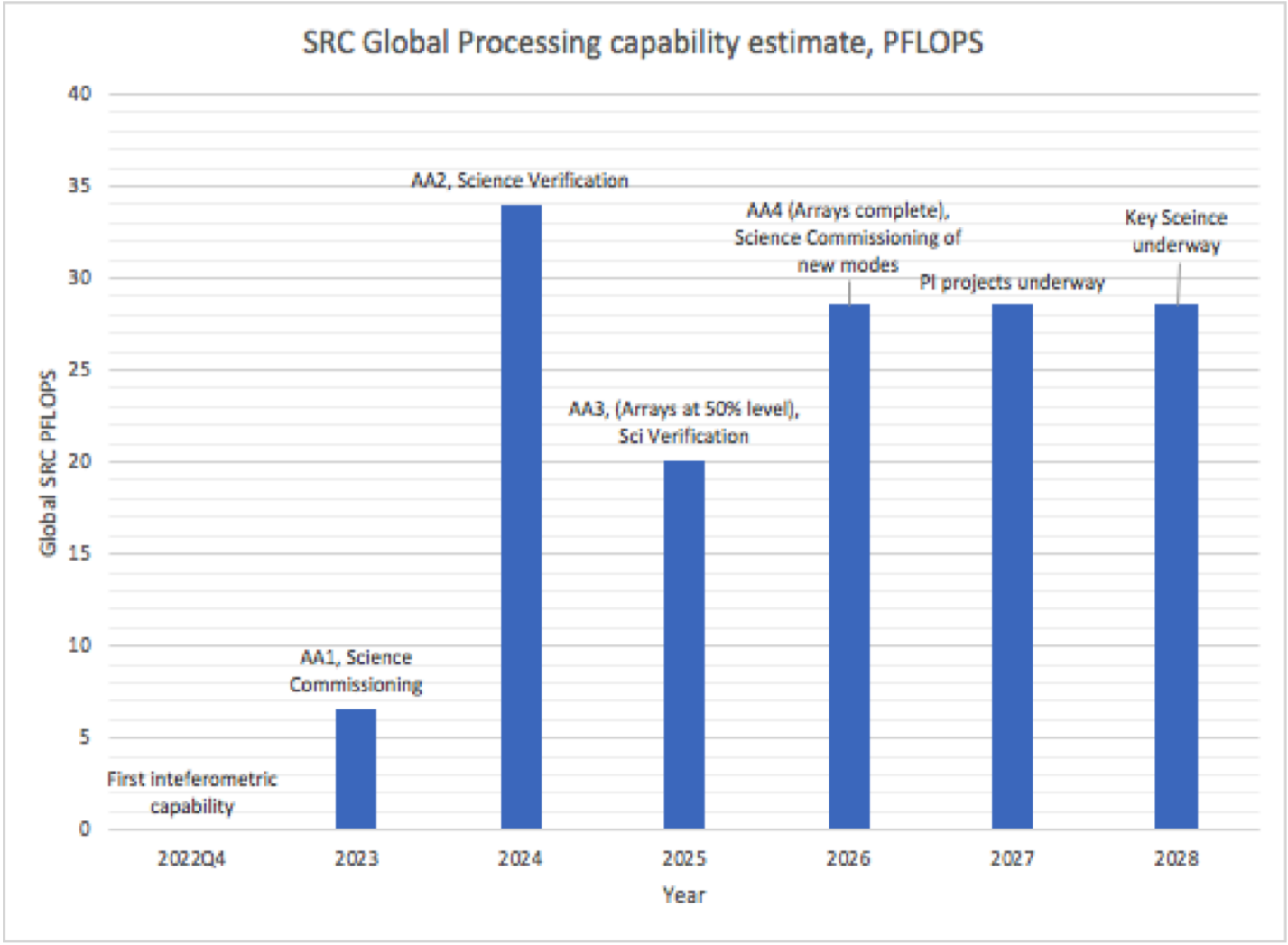
SKA Roll out: Data Processing

- ASSUME that the arithmetic intensity in the SDP is reflected in the SRC processing too
- Then assume that data products are processed multiple times within the SRCs – to allow users to determine the best workflow, for example
- Reduce this number over time to a factor of 3 at steady state.



SKA Roll out: Data Processing

This gives global SRC FLOPS estimates (excluding data challenges / simulatons work)



SKA Roll out: Data Storage

We assume that visibility data are deleted after 1 year.

We assume that image Observatory Data Products are used in the SRCs to generate Advanced Data Products which are a factor of 3 larger (on average) than the Image ODPs that generated them.

We assume that the visibility data products give rise to Advanced Data Products with volume 10% of the visibility volume.

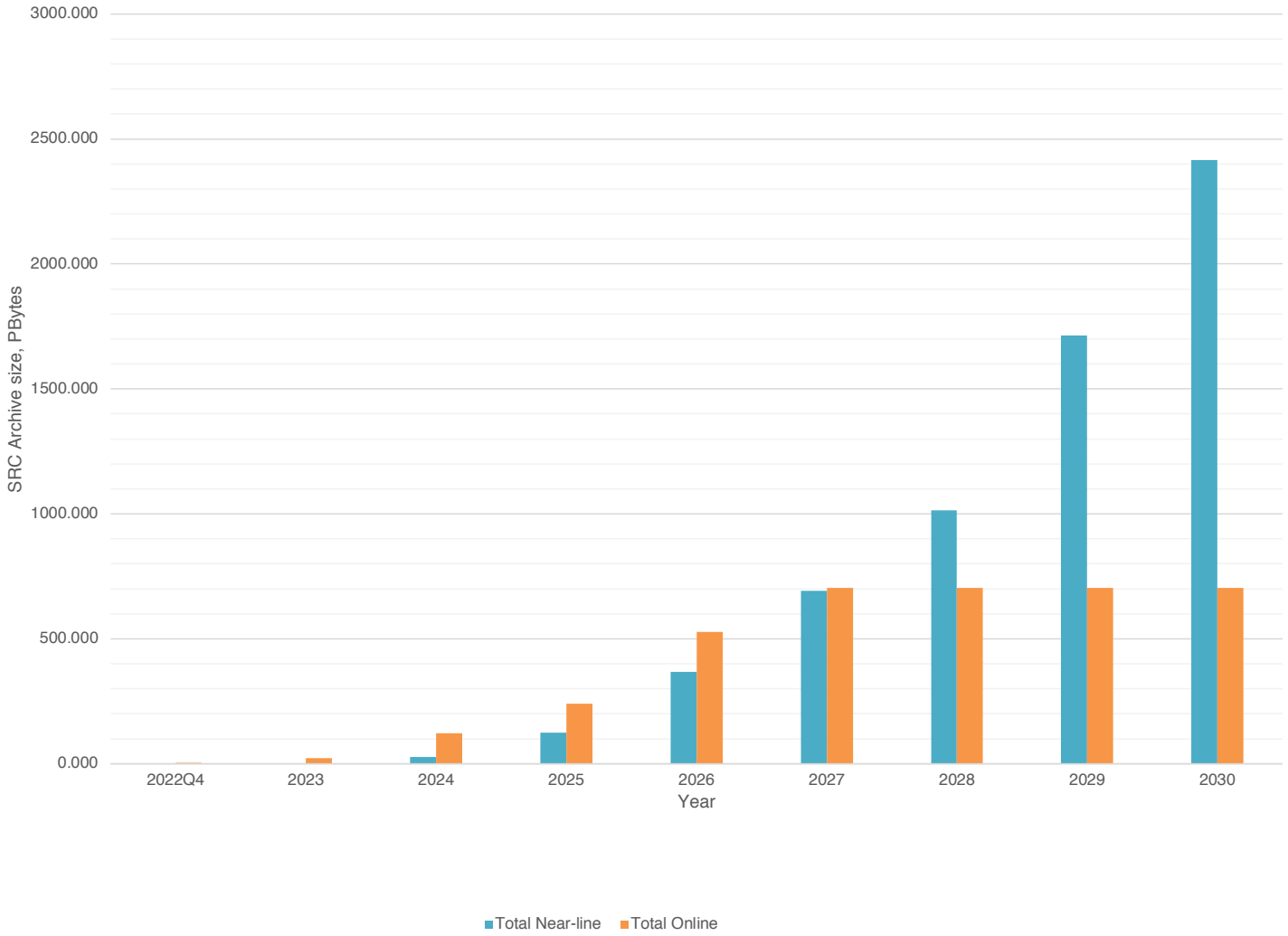
We assume that there are 2 copies of every data product within the global network of SRCs.

We assume that all data products are losslessly compressed with compression factors of 2.

We assume that there is sufficient “near-line” (medium-slow performance) storage available to store two copies of data products, including advanced data products

We further assume that there is sufficient “online” (high performance) storage to hold 1 year’s worth of data (this includes image observatory data products, advanced data products and the visibility data - see point 1).

SKA Roll out: Data Storage



Summary – bare minimum (ideally want headroom)

Year	Array Status	Archive size estimate and processing	Useful data set size	SRC Involvement
Pre-2022	Precursors and pathfinders	5 PB 50 TFLOPS	Up to 0.1 PByte scale	Data challenges with precursor, pathfinder and simulated data.
2022Q4	First dishes / stations on site	5 PBytes 60 TFLOPS	Up to 0.1 PBytes scale for data challenges, 50-200 GByte scale from SKA arrays	Voluntary acceptance of some principally useful visibility data sets, de-bug SDP pipelines and tune parameters. Science Verification work from 2024. Engagement of community experts from outside SKAO Interface testing
2023	Around 10-20% of total	24 PBytes, 7 PFLOPS	Up to 0.1 PBytes scale for data challenges, 1-30 TByte scale from SKA arrays	
2024	Up to around 50-70% of total	**150 PBytes, 35 PFLOPS**	0.02-1 PByte scale	
2025	Building to full scale	**365 PBytes, 35 PFLOPS**	0.1-5 PByte scale	A few Public data products available via SRCs
2026	Complete arrays, shared risk (early) PI science	**900 PBytes, 35 PFLOPS**	0.1-5 PByte scale	All easy modes enabled, commissioning of non-standard techniques Full scale SDP available (but HW may be at deployment baseline level) SRC role takes on project based user support for SRC processing and analysis of observatory products.
2027	PI science	**1.4 EBytes, 35 PFLOPS**	0.1-10 PByte scale	
2028	PI & KSP Science	**1.7 EBytes, 35 PFLOPS**	0.1-10 PByte scale	Full functionality SKA

Updated requirements

This is a very simple model but does lead to some important scalings:

Since the data rate out of the observatory is many times lower than the data rate into the SDPs, our previous estimate of ”2x bigger than the SDP” is rather hard to justify.

So, the global SRC processing scales not as 500 PFLOPS but closer to 35-80 PFLOPS. This looks like a fraction of a (design baseline) SDP

However, this analysis has not addressed the Key Science Projects, which commence around 2028

Personnel: FTE

- We have not attempted to estimate FTE costs in our analysis
- Expect significant regional (national) variation in these
- Variation in support needed for three different classes of SKA Regional Centre User:
  - PI programme scientists – potentially evolving to a "run of the mill" model with SRC work often involving re-use of existing workflows (e.g. simple image coaddition, source extraction)
  - Key Science Projects – likely to require challenging SRC workflows, simulations, potentially very large data sets but also could have identified compute expertise within their collaborations or funded from additional grants (?)
  - Archive users – support for these will need to be streamlined, but might involved significant "behind-the-scenes" SRC software development and data management.
- Can potentially learn how (if) support FTE cost scales with compute cost from other projects. (See WP3)
- We do not have a model for how FTE might be shared between SRCs, or whether all user support work could/would/should be shared or local.

End

(Note, discussion session on sizing 11:40-12:00)