ICRAR

International Centre for Radio Astronomy Research

The Need for Regional Centres

in Data Intensive Research

Markus Dolensky

Techical Leader

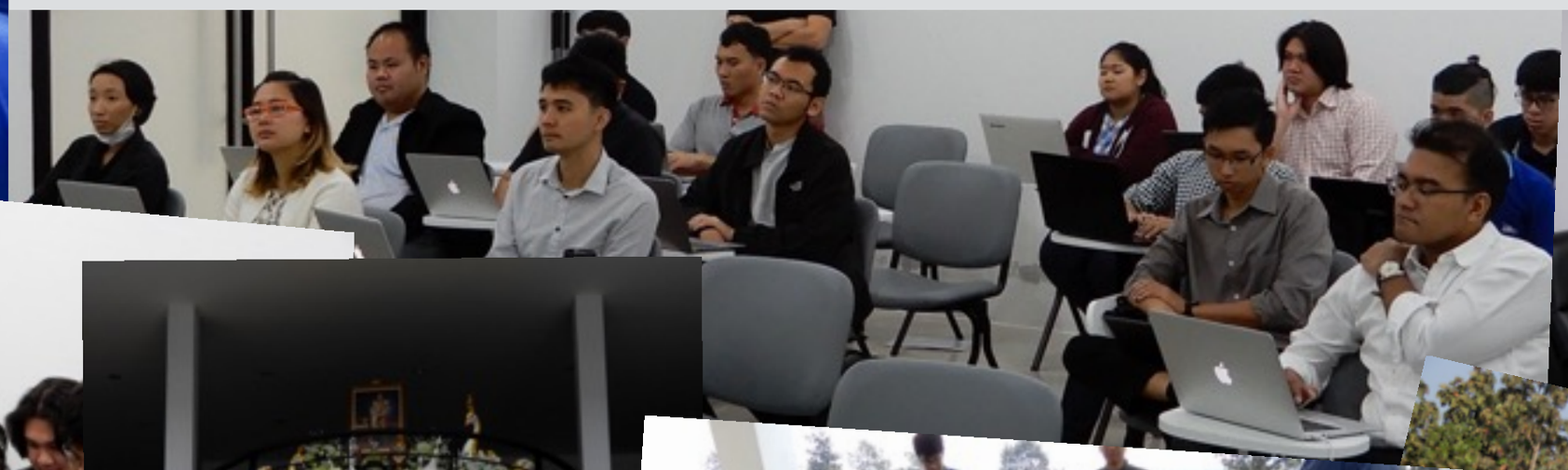
credits: Peter Quinn, Slava Kitaeff and Data Intensive Astronomy Team

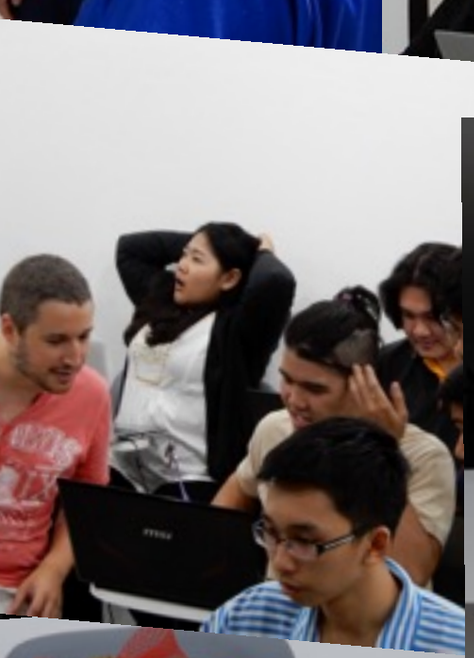
Curtin University

The Government of Western Australia

The University of Western Australia

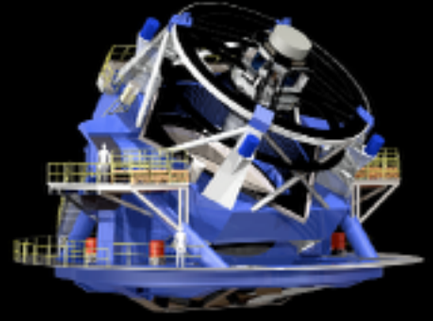
NARIT-ICRAR Workshop on Data Intensive Astronomy
Australia-ASEAN Council Grant Program 2016-2018





Data Rates and Volumes

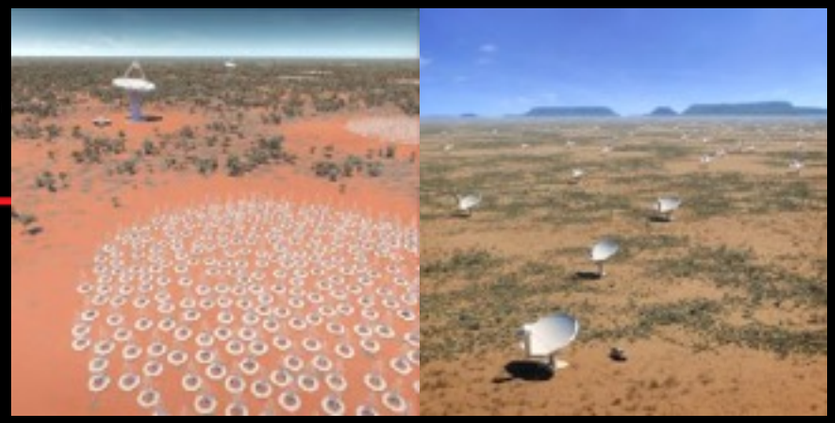
ALMA Science Data Products ~ 1 PB/year



LSST Science Data Products 4-5 PB/year

LHC High Luminosity LHC 2023
Science Data Products ~ 400 PB/year

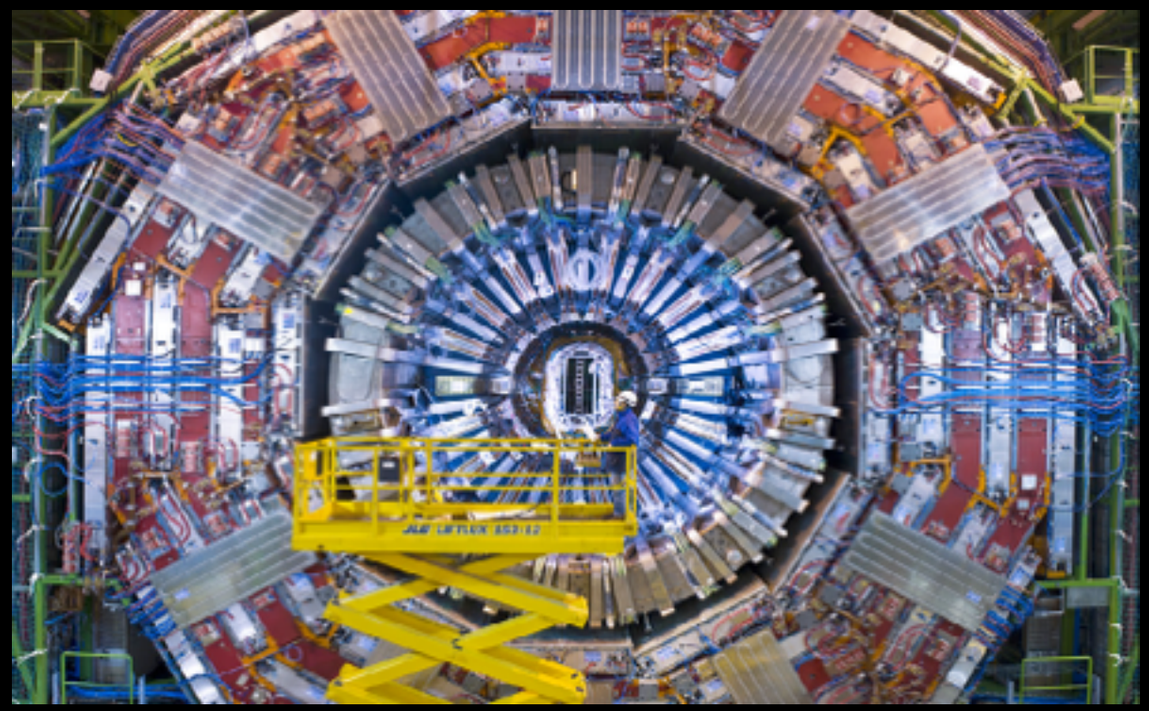
SKA1 Data Volume



Preservation System	LOW	MID	
Long term storage volume	110	340	PB
Medium Performance Buffer	30	100	PB
Science Data Volume (HPSOs)	55	170	PB
8 TB drives LTS (0.11+0.34 EB)	13750	42500	Unit
26 TB drives LTS	4200	13000	Unit

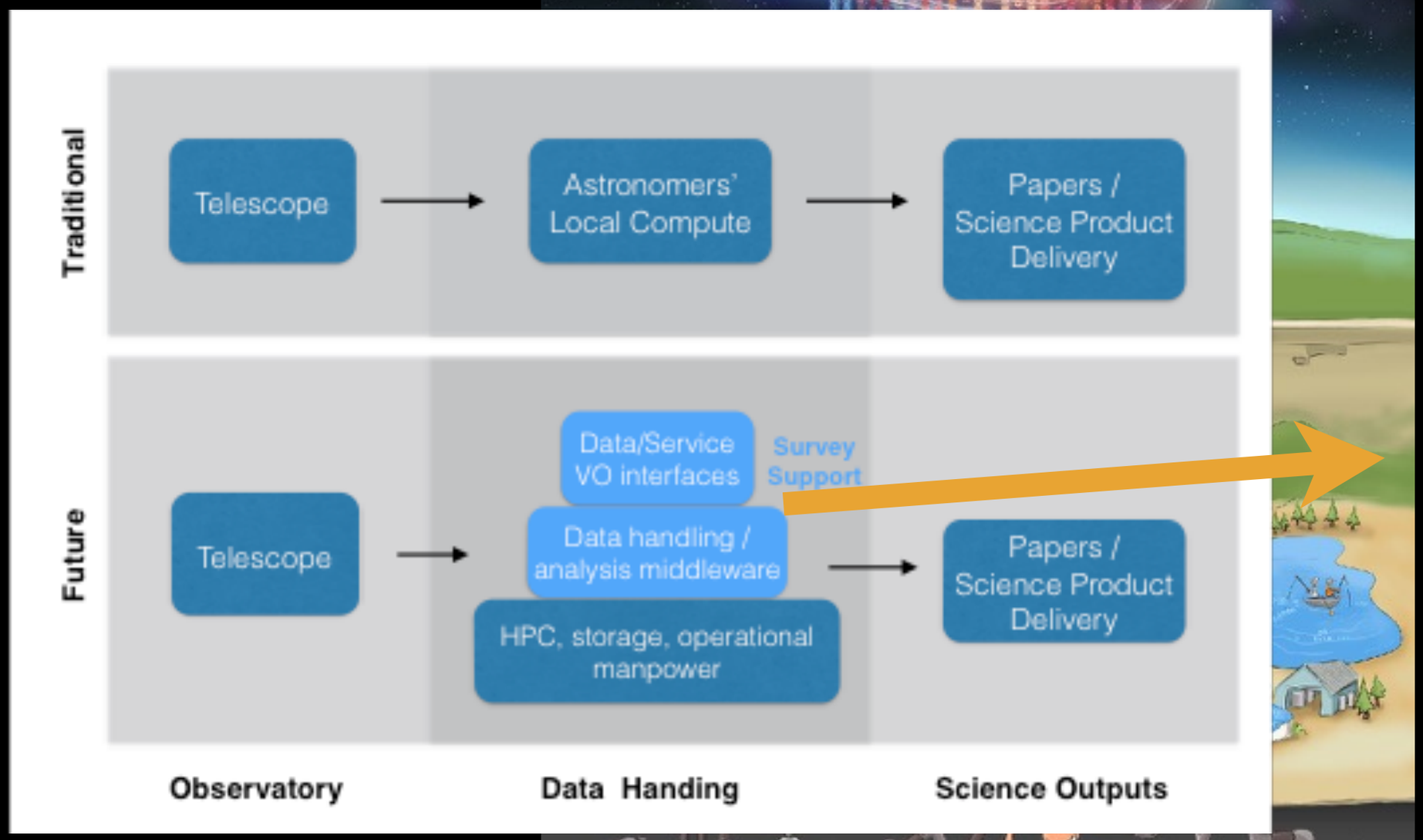
The LHC decision

- LHC produced 70 PBytes in its first 3 years (expected 15PB/year)
- High Luminosity LHC (2023) 400 PB/ year and multi Exabyte data sets for CMS and ATLAS on SKA timescale



- **2001 LHC realized compute/store cost exceeded accelerator and detector costs**
- **Created the Worldwide LHC Computing Grid (the “grid”)**
  - separately funded, reviewed and governed but fused with LHC facility/ CERN, MOUs from contributing organisations and consortia
  - Diversify funding source opportunities
  - More completely engage the community

A change in approach…



The next 5-10 years

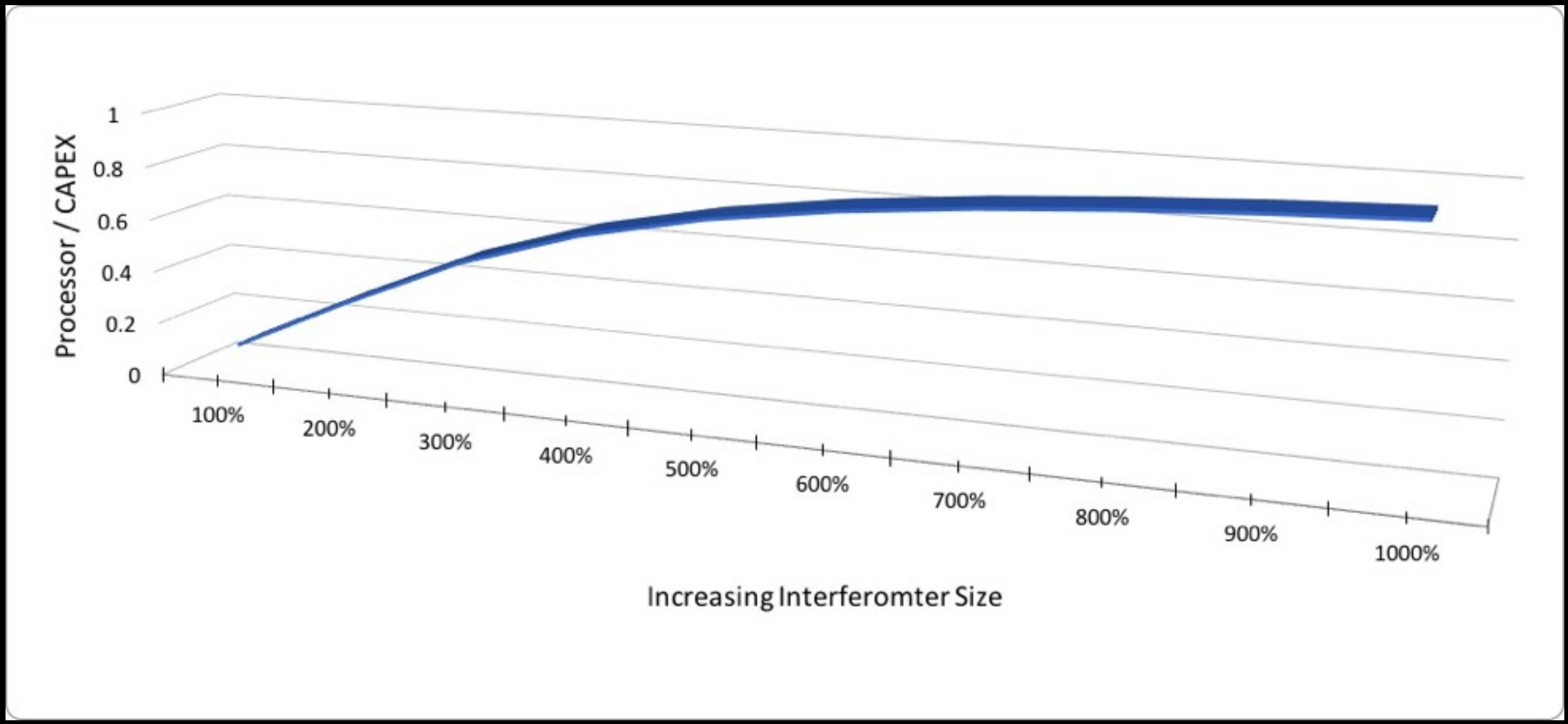
CHILES → x10 → **ngVLA**

Not possible without the formation of SKA Regional Centres and Alliances

x10 → **FAST/**

x10 → **SKA**

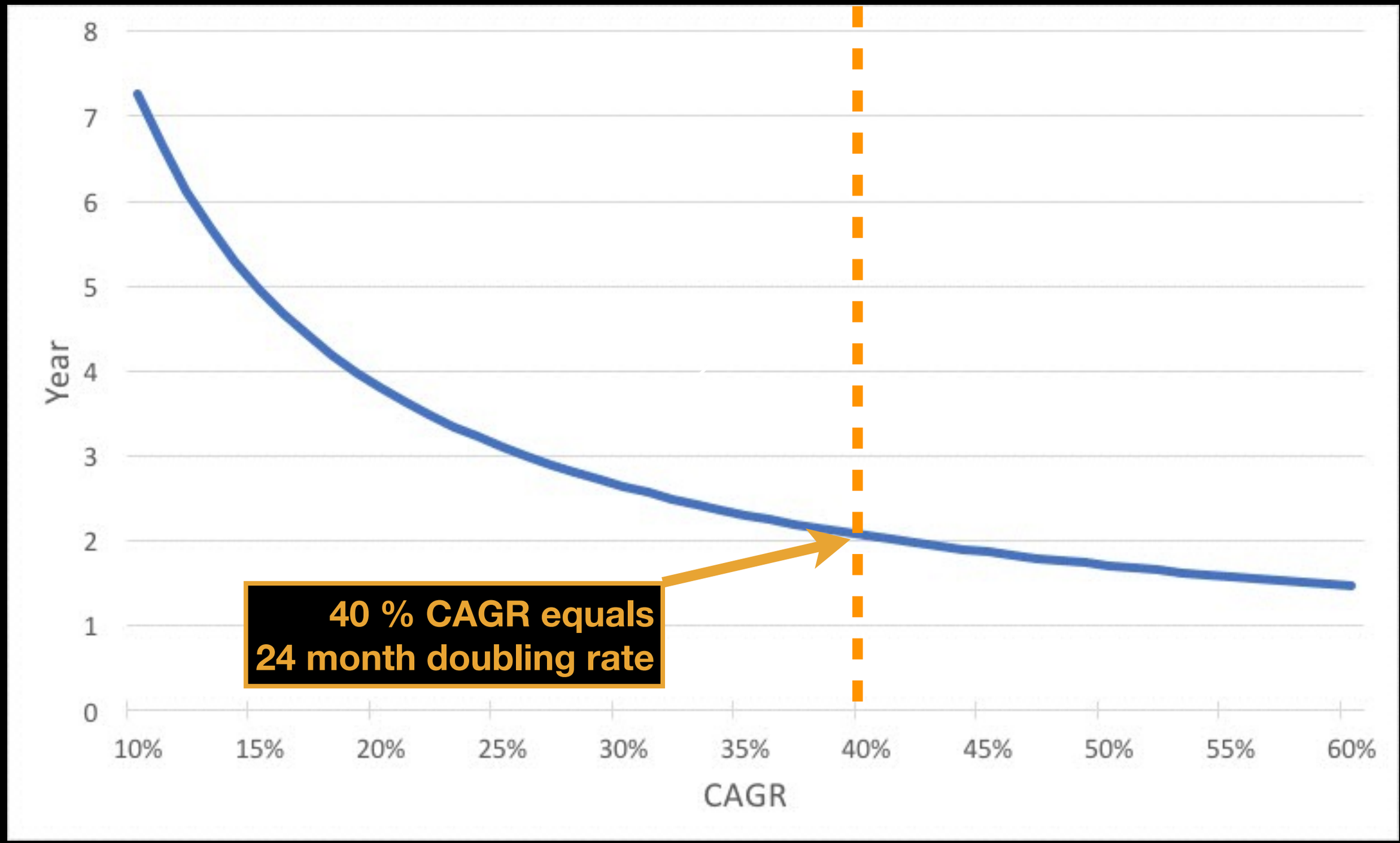
Imaging goes with $O(n^3)$



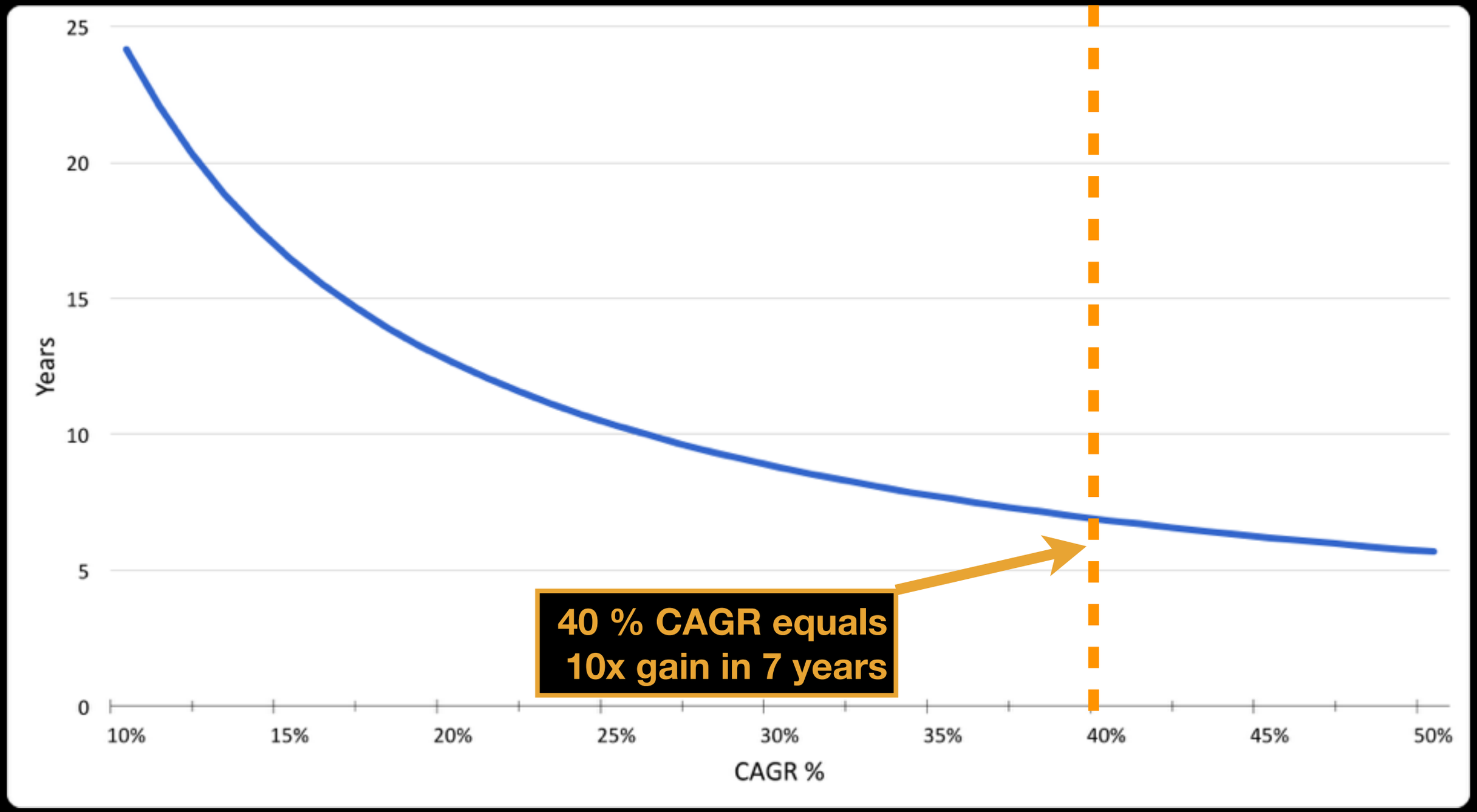
Doubling Interferometer means 8x Processing - $O(n^3)$: Memo 49, T. Cornwell

Processing System quickly dominates CAPEx.

CAGR and Doubling Time



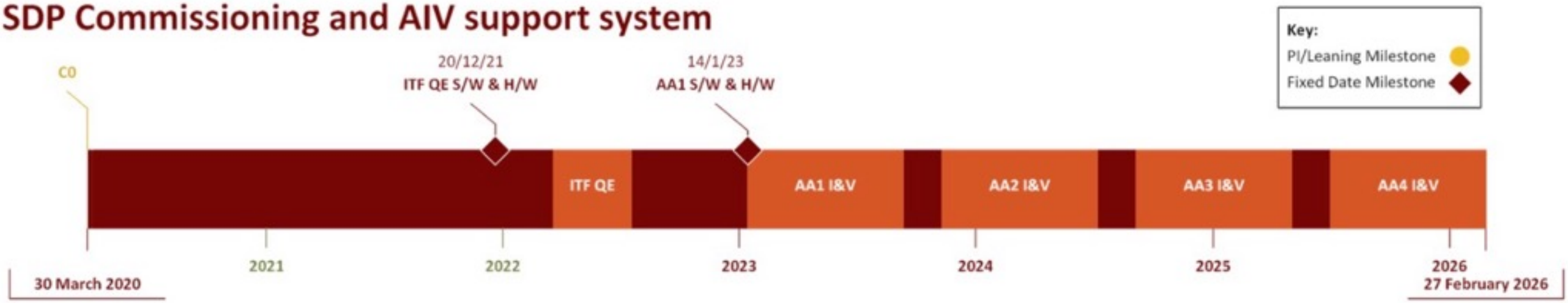
Regaining or Loosing a Factor of 10?



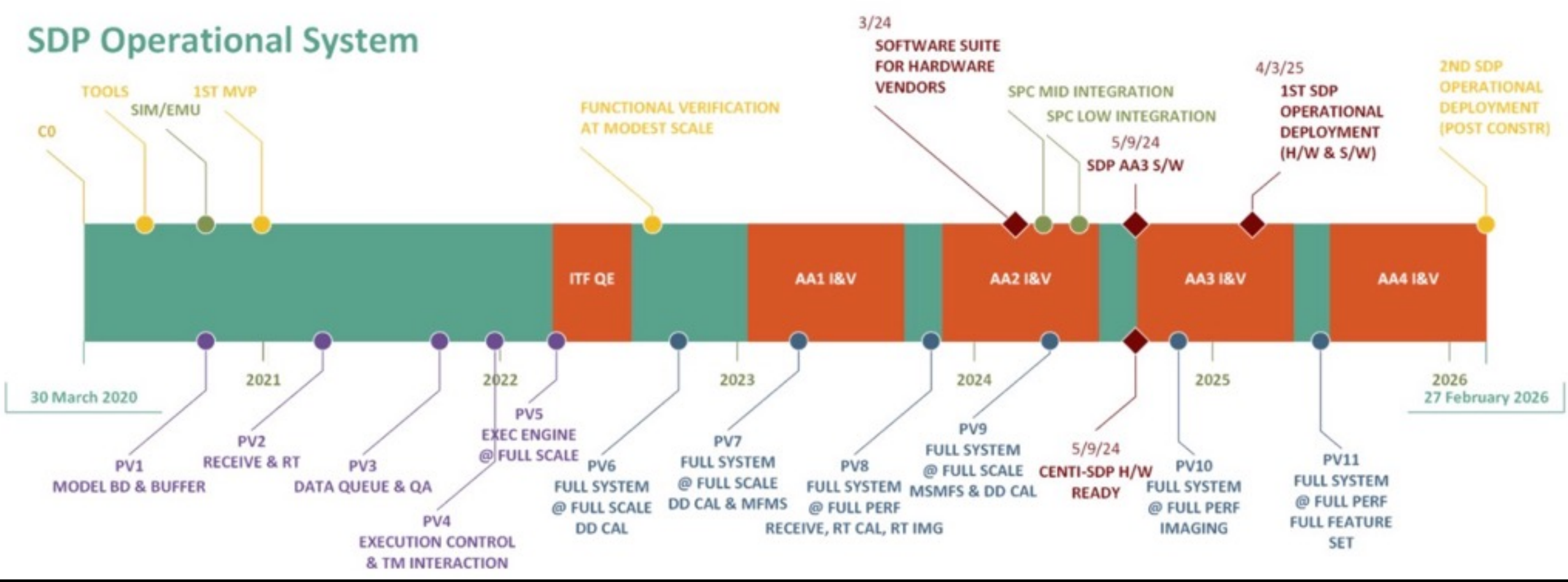
The rebaselining downsized by a factor of 10.
To regain a factor of 10 takes many years depending on CAGR.

Rollout Plan 2020 - 2026

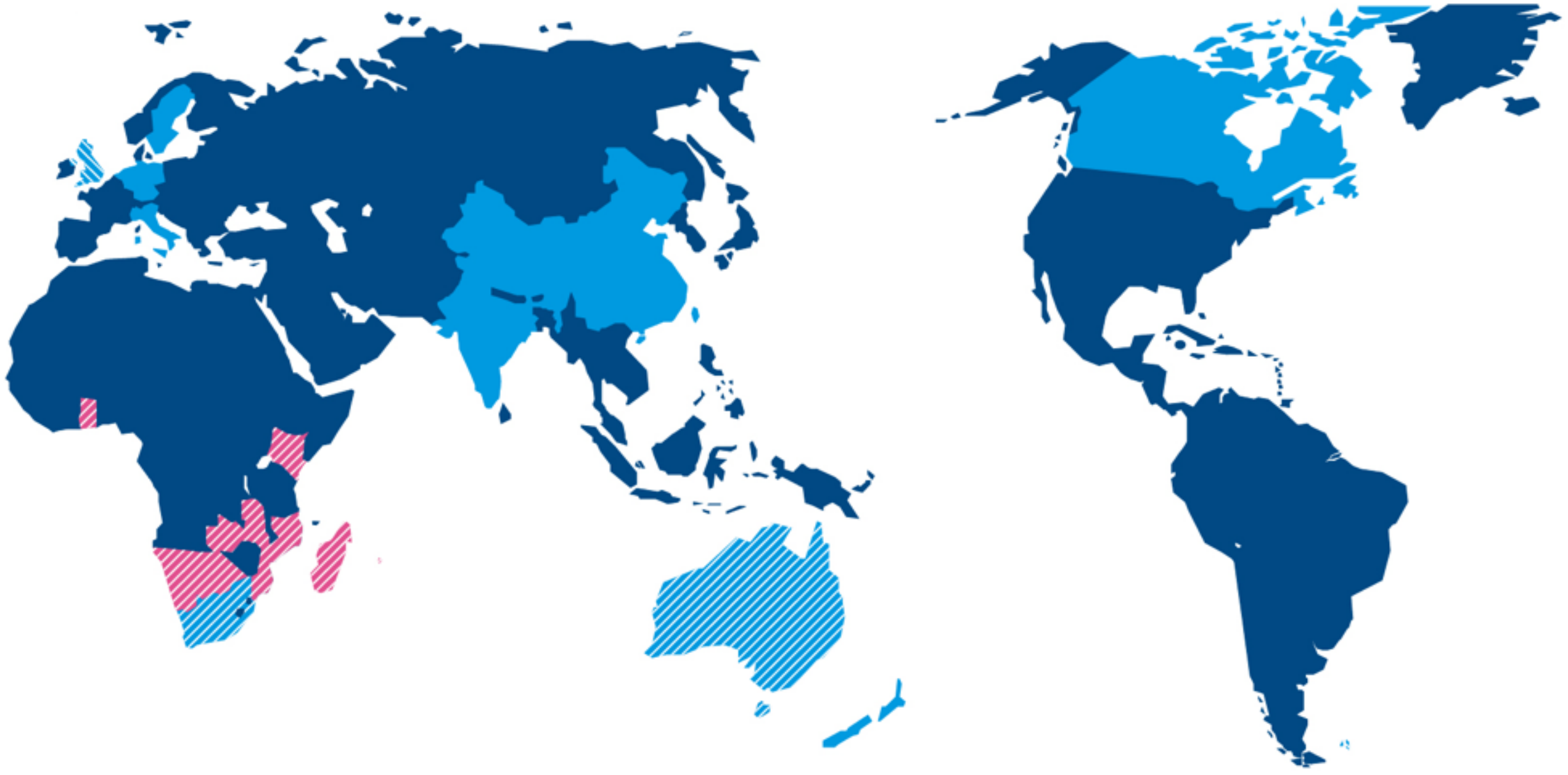
SDP Commissioning and AIV support system



SDP Operational System

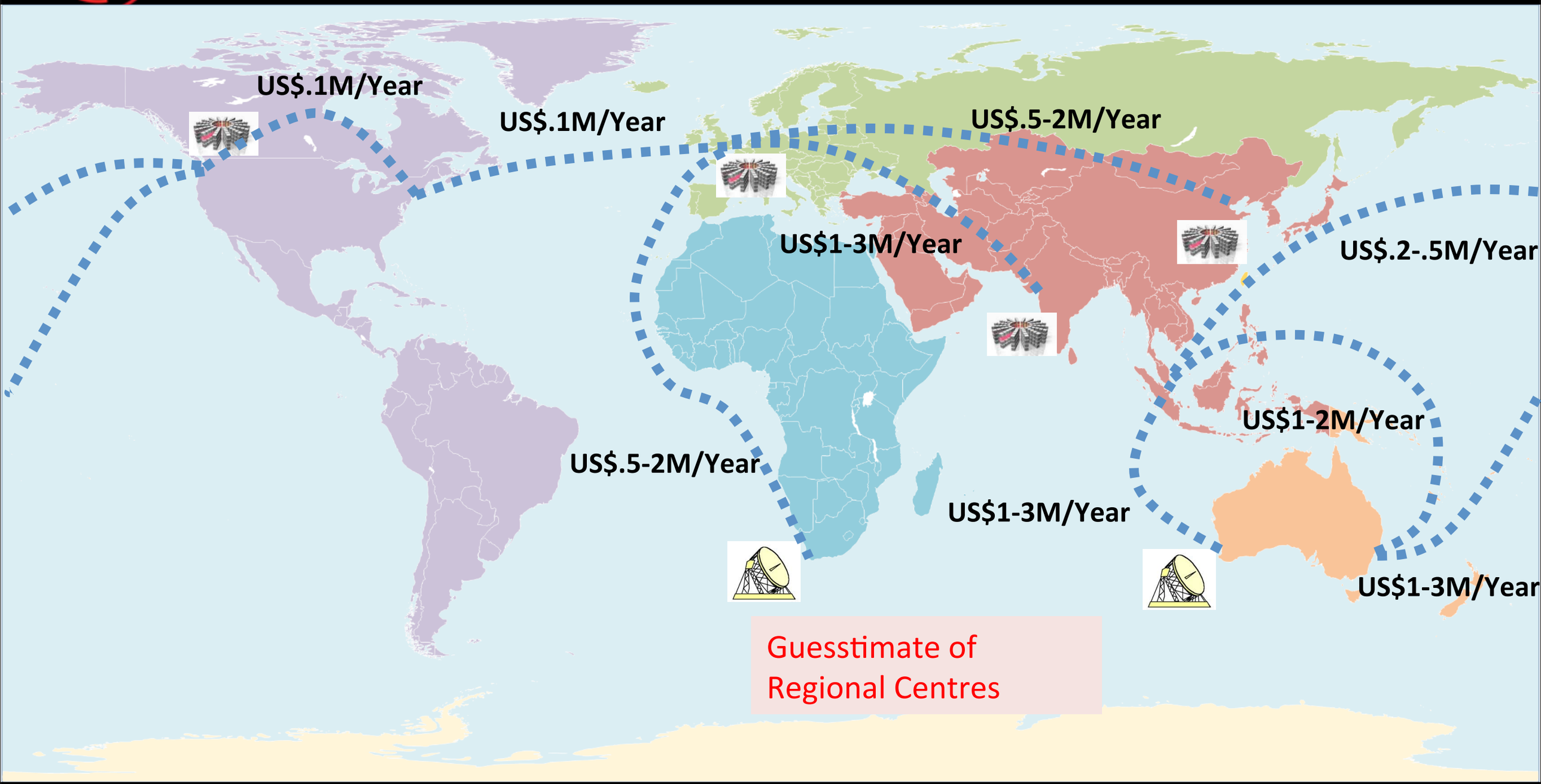


Regional Networking



Where does post processing/analysis happen?
Data access and maximizing scientific return

Estimated SDP to World Costs



- Guesstimate of Regional Centre locations
- 10 year IRU per 100 Gbit circuit 2020-2030

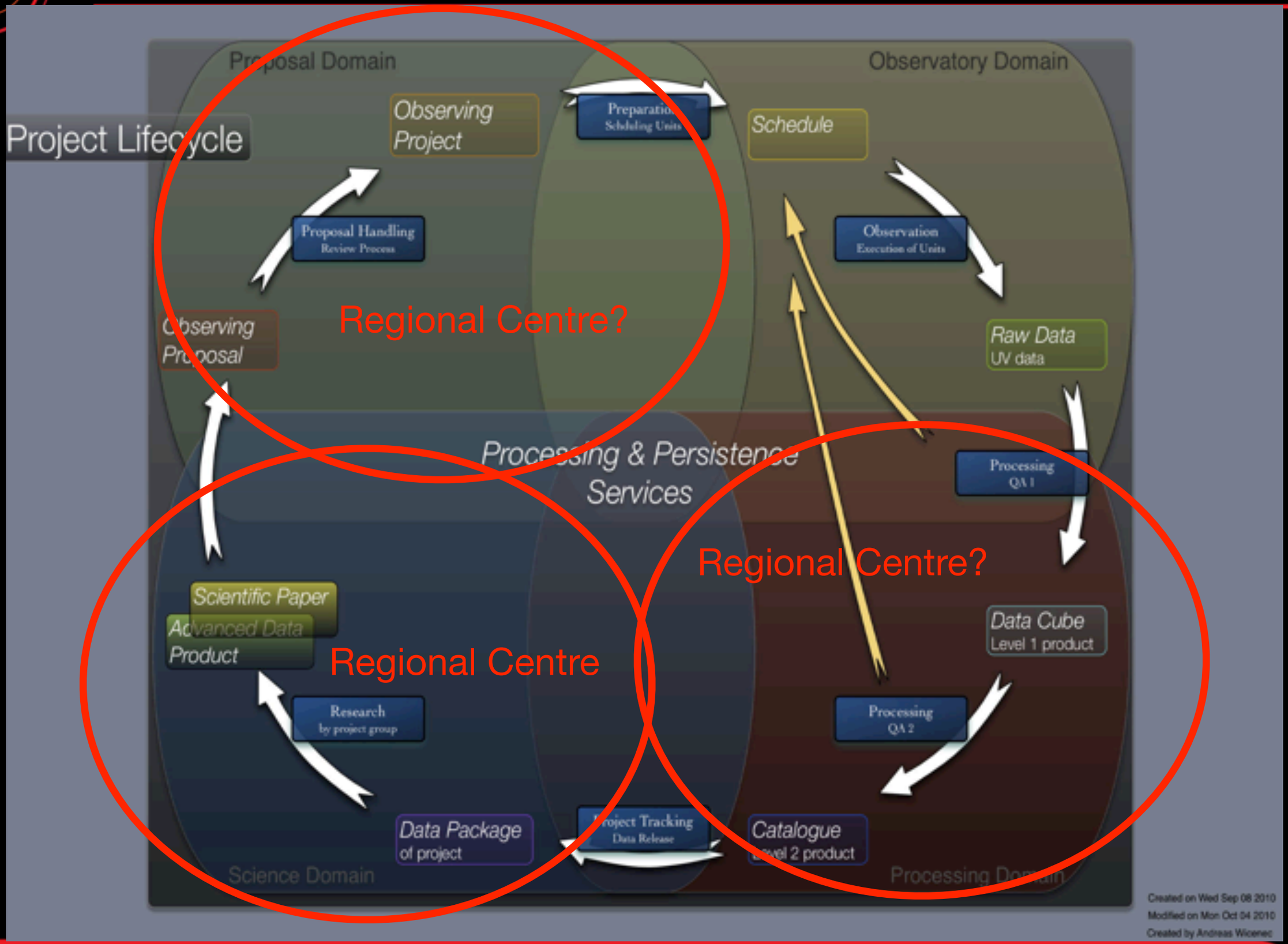
Existing Regional Networks

- CERN LHC
  - Tier 0: CERN
  - Tier 1: large computing centres
  - Tier 2: analysis centres
- ALMA
  - regional centres
  - regional centre nodes
- EUMETSAT
  - national meteorological bureaus
  - regional (implementation) centres

SKA Data Product Levels

Level	Definition	Responsibility
7	Enhanced data products e.g. Source identification and association	
6	Validated science data products (released by Science Teams)	
5	Calibrated data, images and catalogues	SDP
4	Visibility data	CSP
3	Correlator output	CSP
2	Beam-former output	LFAA
1	ADC outputs	LFAA

What Observatories Do



SRCCG

The SKA Regional Centre Coordination Group (SRCCG) is instructed to:

- define a minimum set of requirements for individual SRCs;
- develop a set of requirements for the SRC network as a whole, including plans for the physical network connecting SRCs, both together and with the SKAO sites;
- develop the requirements for implementing an SKA data archive;
- identify the requirements for, and coordinate the development of, technical interfaces between the SRCs and the SKA;
- draft the MoU to be agreed between the SRCs and the SKA Observatory that accept the stated minimum requirements and SKA data access policies;
- develop a set of data challenges to test and verify the capabilities and resources of each SRC, as well as the SRC network;
- develop an accreditation process for the SRCs, including branding and appearance policies;
- develop a process to ensure that software tools are interoperable across the SRC network;
- develop a policy for the inclusion of user-generated advanced data products into the SRC archive;
- develop SKA data heritage and access policies to that heritage data set for approval by the SKAO; and
- investigate models for the future governance of the collaborative network of SRCs.

ASRCWG

1. To define the Australian SKA Regional Centre scope, opportunities, requirements and potential funding opportunities, in line with SKAO Operational planning and the SKAO Board's response to the recommendations of the SKA Data Flow Advisory Committee
2. To explore opportunities to expand the Australian SKA Regional Centre activities to include collaboration with similar activities in China, New Zealand and the broader Asia-Pacific Region with a particular focus on business case development and precursor enabled technological and scientific programs
3. To initiate a detailed study (ERIDANUS) of the data and processing requirements and costs within an Australian SKA Regional Centre in support of Australian Survey Science with the SKA precursors and SKA-1

Data Canals and Rivers

FUNDED: 15 July 2017

Advanced European Network of E-infrastructures for Astronomy with the SKA

Exascale **R**esearch **I**nfrastructure for **D**ata in **A**sian-Pacific astro**N**onomy **U**sing the **S**KA

Launched April 10 2017

eridanus.net.au

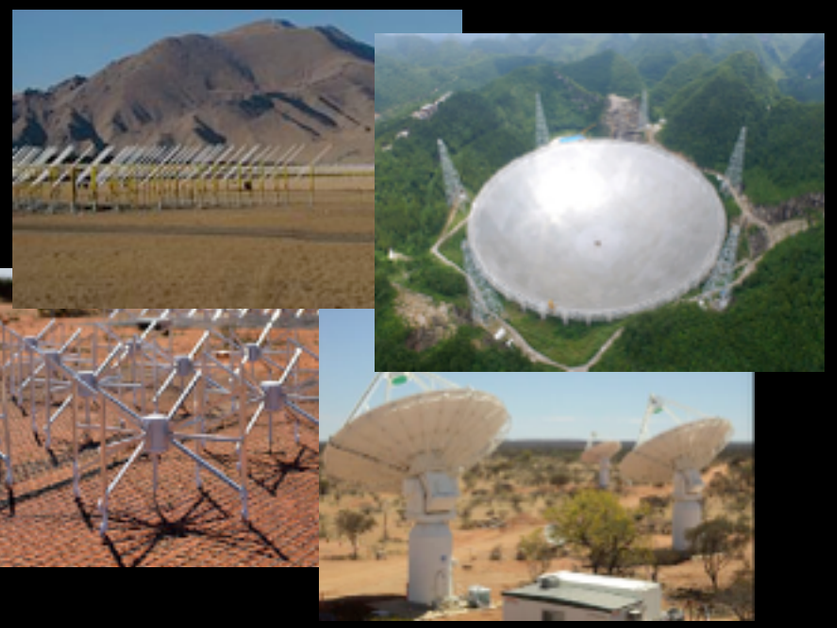
ERIDANUS

The ERIDANUS Project:

- A three year, **bottom-up**, design study commencing in 2017, aimed at deploying prototype data intensive research infrastructure and middleware, between and within Australia and China, capable of addressing SKA-class data and processing challenges.
- Will respond to challenges identified by the SRCCG and will work in a coordinated and collaborative manner with the European Aeneas Project.

SKA-1 Scale

Technology Development Projects



Precursor Science Projects

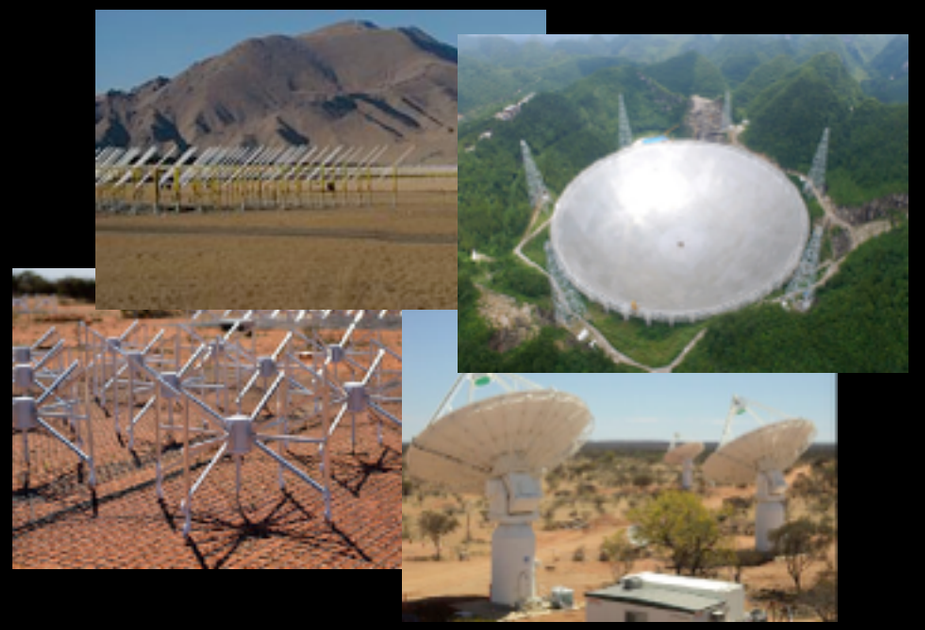
ERIDANUS

The ERIDANUS Project will engage with:

- The astronomical communities within Australia, China, …
- Industrial partners
- The providers of networking and computing research infrastructure
- National and international committees coordinating efforts on SKA regional centre developments

SKA-1 Scale

Technology Development Projects



Precursor Science Projects

ERIDANUS

The ERIDANUS Project:

- Founded by ICRAR and SHAO in November 2016 - now includes CSIRO and interest from CADC -
- Focus:
  - Data flow architectures
  - Execution frameworks
  - Strategies and systems to monitor and control cost
  - Networking optimisation
  - Cloud computing options
- **OPEN INVITATION FOR ALL TO JOIN!**

SKA-1 Scale

Technology Development Projects

Precursor Science Projects

Australia-China SKA Big Data Meeting

2020/21

- Funded Regional Centre initiatives, at least in China and Australia
- ASKAP, MWA and FAST in full science operations using components from ERIDANUS developments
- First science results from Australia/China precursor science teams using ERIDANUS infrastructure
- SKA Regional Centre Alliance formed
- First ERIDANUS and AENEAS success in SKA data challenges
- SKA-1 construction started
- Joint proposal to governments of Australia, China and further partners for an Asia/Pacific Regional Centre network with SKA1-class capability by 2025

SKA in Australia

SKAO

SKA Observatory

RSA

AUS

SKA Community

AUS

MRO

Pawsey

Geraldton

SKA Operations

Scheduling
Execution
QC
Calibration

Pipelines
Data Packages
Site
Maintenance

Perth

SKA Regional Centre

Community Support
Survey Science Team Support
High level data products
Science, Engineering, DIA innovation
Education/Outreach

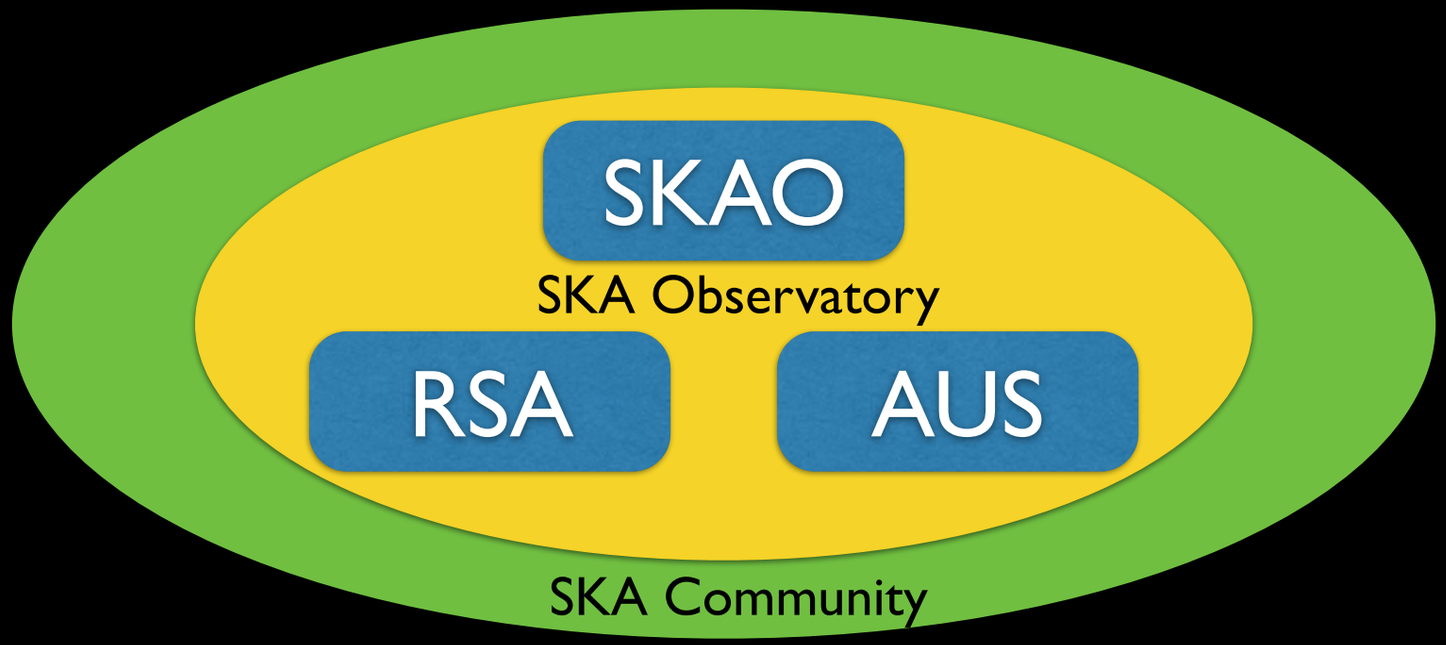
Node 3

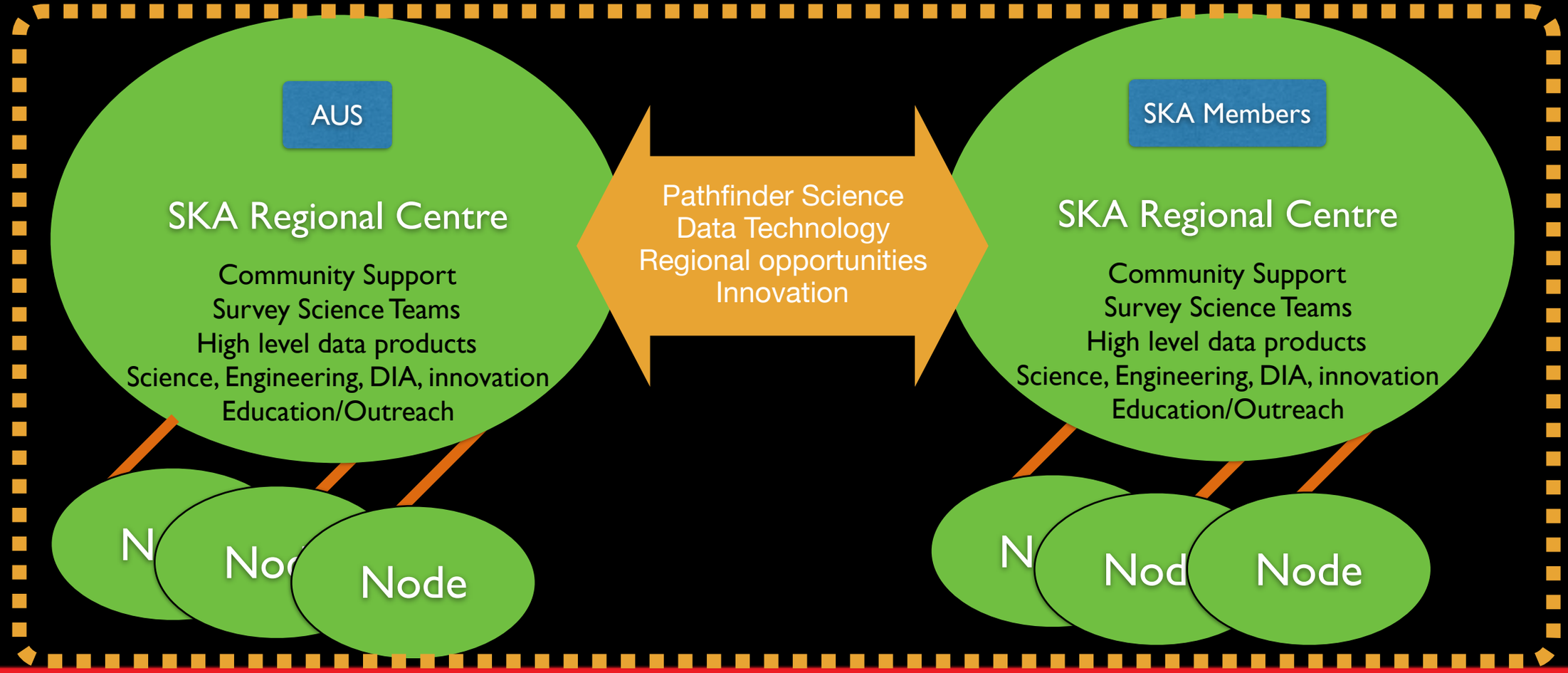
Node 2

Node 1

+ Community

Asia-Pacific SKA Regional Centre Network





Murchison Radio Astronomy Observatory

Video Link:

https://www.youtube.com/watch?v=Nk5_Uwh-P-U

Thank You