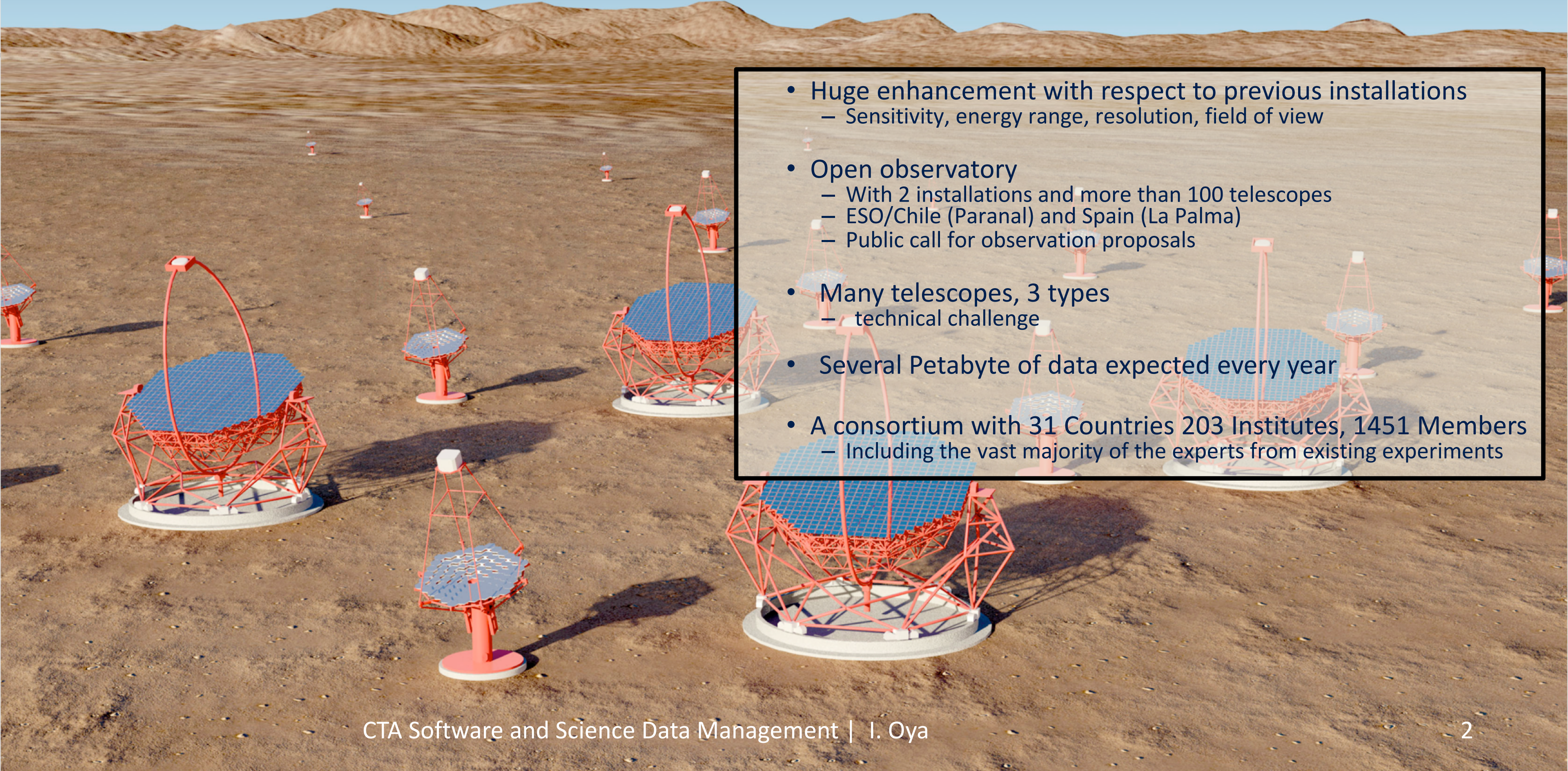
cherenkov telescope array

CTA Software and Science Data Management

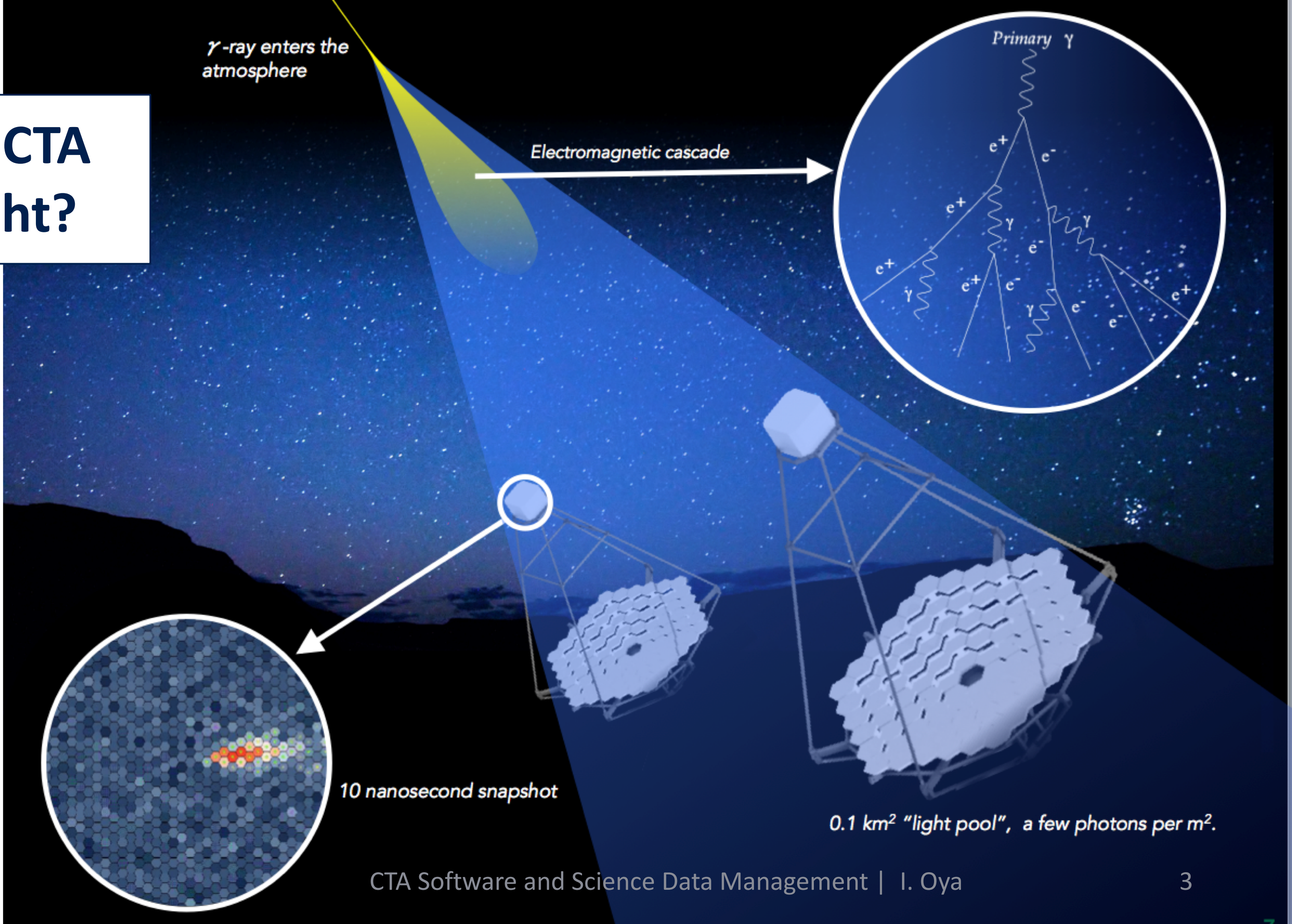
I. Oya, M. Füßling, J. Hinton
K. Kosack.
CTA Observatory

The Cherenkov Telescope Array



- Huge enhancement with respect to previous installations
  - Sensitivity, energy range, resolution, field of view
- Open observatory
  - With 2 installations and more than 100 telescopes
  - ESO/Chile (Paranal) and Spain (La Palma)
  - Public call for observation proposals
- Many telescopes, 3 types
  - technical challenge
- Several Petabyte of data expected every year
- A consortium with 31 Countries 203 Institutes, 1451 Members
  - Including the vast majority of the experts from existing experiments

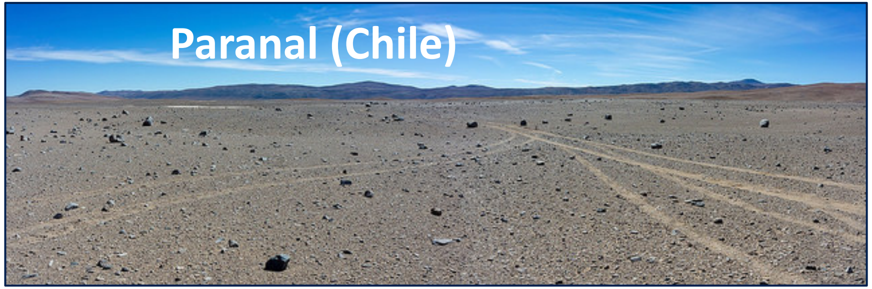
How will CTA detect light?

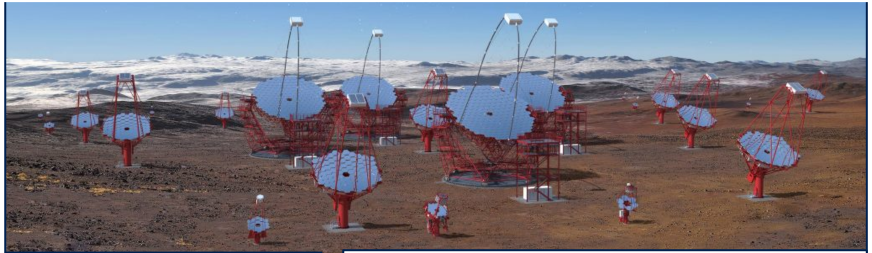


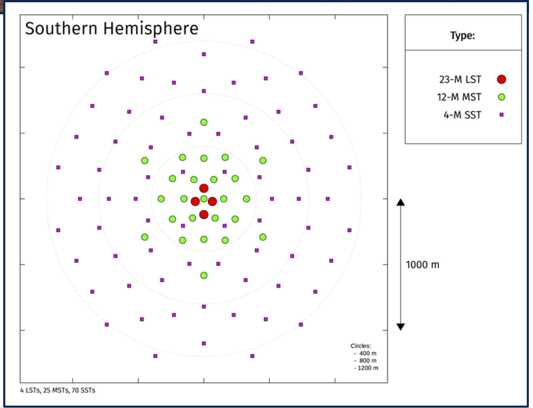
CTA Sites

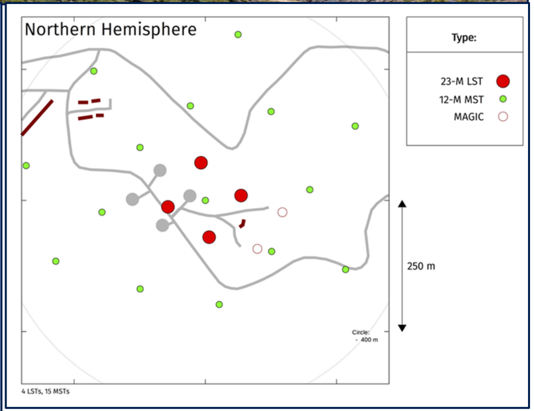

Zeuthen
DESY
Spain
La Palma
Bologna
INAF
northern hemisphere
southern hemisphere
Chile
Paranal
ap: ©2015 Google, Inst. Geogr. Nacional
: ©2015 Google
Array Sites
Headquarters
Science Management Centre

CTA at Paranal & La Palma









CTAO Legal Entity

- In 2014, the CTA Observatory gGmbH was founded as interim legal entity, located at Heidelberg, under German law, to prepare the CTA implementation:
  - Select and prepare two array sites + HQ + Science Data Management Centre (SDMC)
  - CTAO Staff: 30 persons (Jan 2019)
  - Hiring, check CTAO webpage for positions

- The final legal entity for full construction, a *European Research Infrastructure Consortium* (ERIC), is being set up under European Union law (early 2020?)

The CTA consortium (CTAC)

31 Countries
203 Institutes
1451 Members

CTA Observatory and Consortium

Input to the CTA design (CTA Consortium)

Consortium

CTAC Key Science Projects

- Gal. plane survey
- Gal. center survey
- Extragal. survey
- LMC survey
- …

In-Kind contributions to construction (CTAC institutes)

Contributions to operation (CTAC institutes)

Observatory

Cash for construction (Agencies)

Operation of the full Observatory

Operations funding (Agencies)

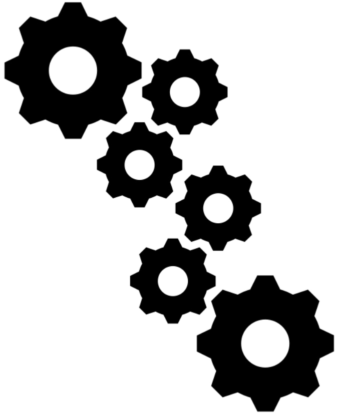
Design and Construction of the full Observatory

Observatory

Proposal-driven open user programme

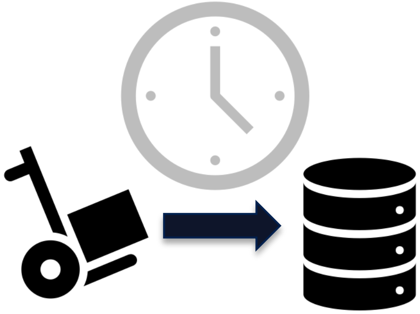
CTA Observatory Operations

- An Open Observatory / User Facility
  - For the first time in this waveband
  - Annual AoO, TAC ranking, long-term schedule
  - Proposal preparation support, tracking, helpdesk, …
  - Public science data archive – After proprietary period
- Two Telescope Arrays & HQ & SDMC – one Observatory
  - Inter-site coordination
  - Uniform approach to science operations
- Main Challenges
  - Sub-array operation, wide field of view, instrument response generation, background modelling, rapid alert generation and response, data volume, science operations during construction
- A Software Instrument
  - Software plays a critical role in all steps of the Observatory

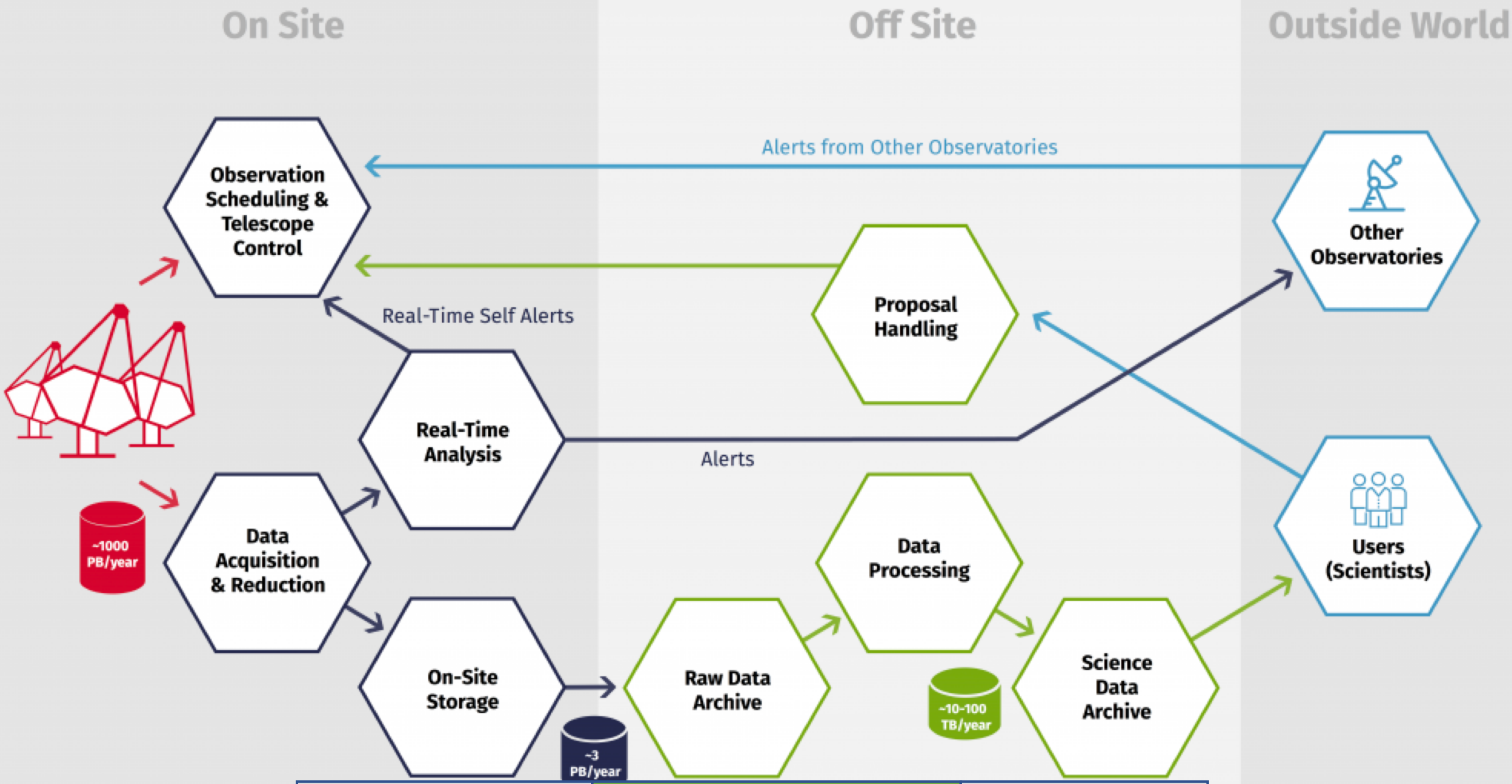


CTA Observatory Operations – Commitment

- “Data” is the final product of CTAO.
  - Everything else is “just support to get data”.
- Operating CTA as an Observatory impacts many areas
  - Supplier – Customer relationship
    - CTAO = Supplier, Science User = Customer
  - “Observatory” means “Commitment”
    - To deliver data of defined scope, quality and within time
    - To treat science users equally and fair
    - To keep the observatory at the forefront of state-of-the-art research
  - 30 years life time
    - Significant maintenance effort
  - Limited operations budget
    - Build, operate and maintain simple and robust CTA system



CTAO OPERATIONS

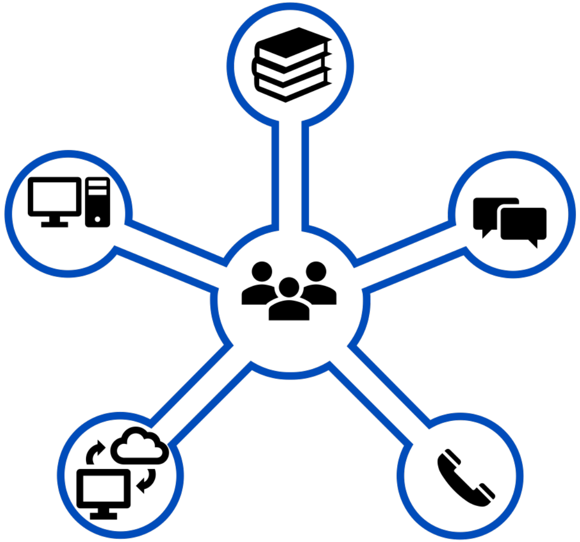


Observation Execution	Data Processing and Preservation	Science User Support

Initial calib./reduction →Transmission from site →Bulk data archive →Science data archive

Science Operations – Implications

- Science Operations = Data processing + User Support Services + Computing Facilities Support
- Main elements for efficient and sustainable science operations include
  - Processing software
  - Computing facilities and storage
  - Expert people to run
    - Processing software
    - MC simulations
- Prepare and deliver science data to users
- User Support Services

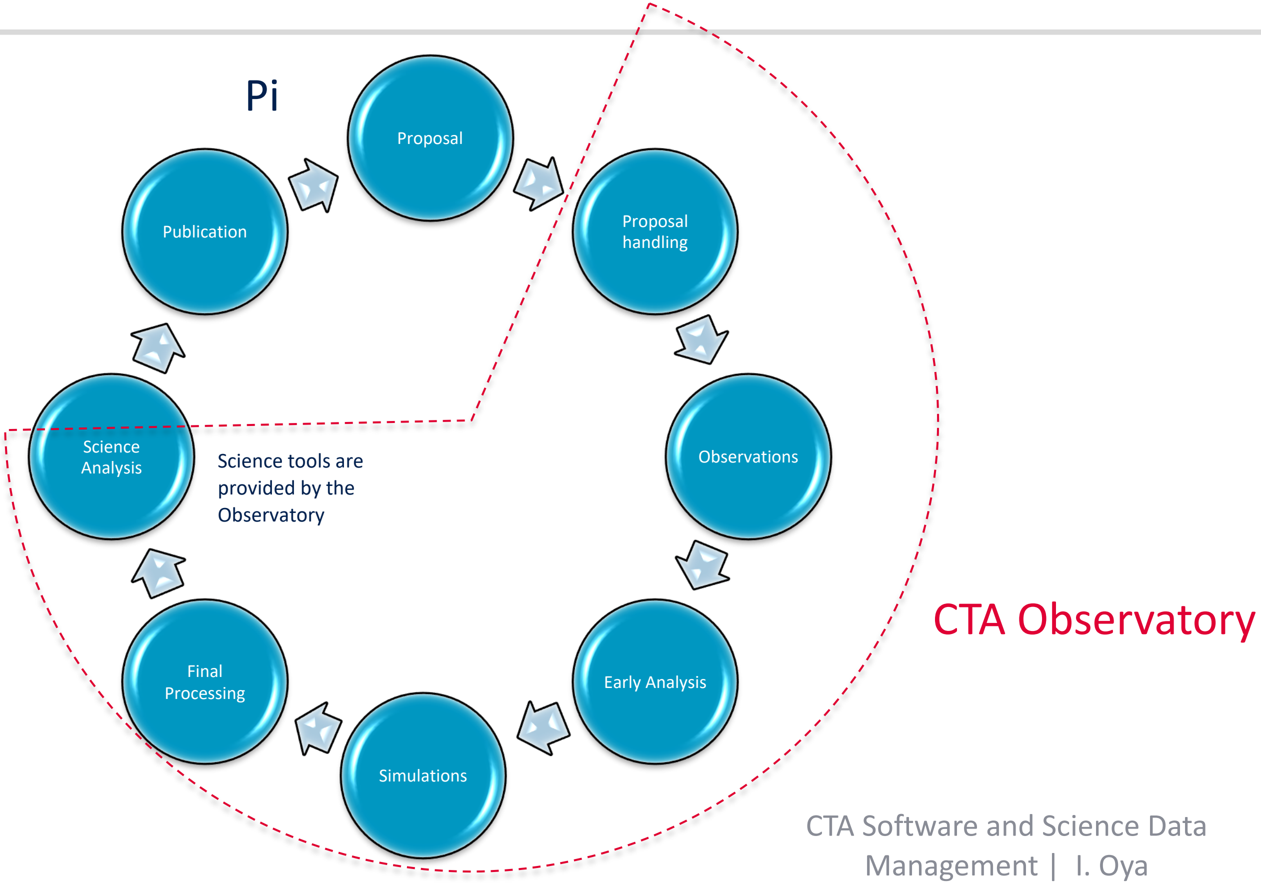


Science Operations – Implications

- Science Operations = Data processing + User Support Services + Computing Facilities Support
- Main elements for efficient and sustainable science operations include
  - Processing software
  - Computing facilities and storage
  - Expert people to run
    - Processing software
    - MC simulations
- Prepare and deliver science data to users
- User Support Services

- Call for Proposals
- Proposal Handling
- User support services
- Science Portal, FAQ, Newsletter
- PI interactions, incl. science alerts
- Provide tools for proposal preparation, data analysis
- Science planning
- Data processing and data quality monitoring
- Data archiving
- MC simulations
- Provide instrument response functions
- Science Analysis tools
- …

CTA Observatory and PI



Computing Challenges

- Reducing On-site Data Volume
  - 1000 PB/y directly coming out from cameras (mostly noise)
  - Impossible to transfer via internet link (see later)
- Storing and Processing Big Science Data:
  - Computing resource requirements imply distributed computing model
  - Data volume is too big to separate storage from computing
- Simulating CTA
  - Development of extensive air showers
  - Propagation of Cherenkov light
  - Cherenkov photon ray-tracing through the telescope optics to Camera
  - Photosensors simulation and camera electronics
- Developing a Data Processing Pipeline
  - PBs of data
  - Distributed team of developers.

(+many more computing challenges for e.g. control and supervision of 100 telescopes, sub-array operation, scheduling)

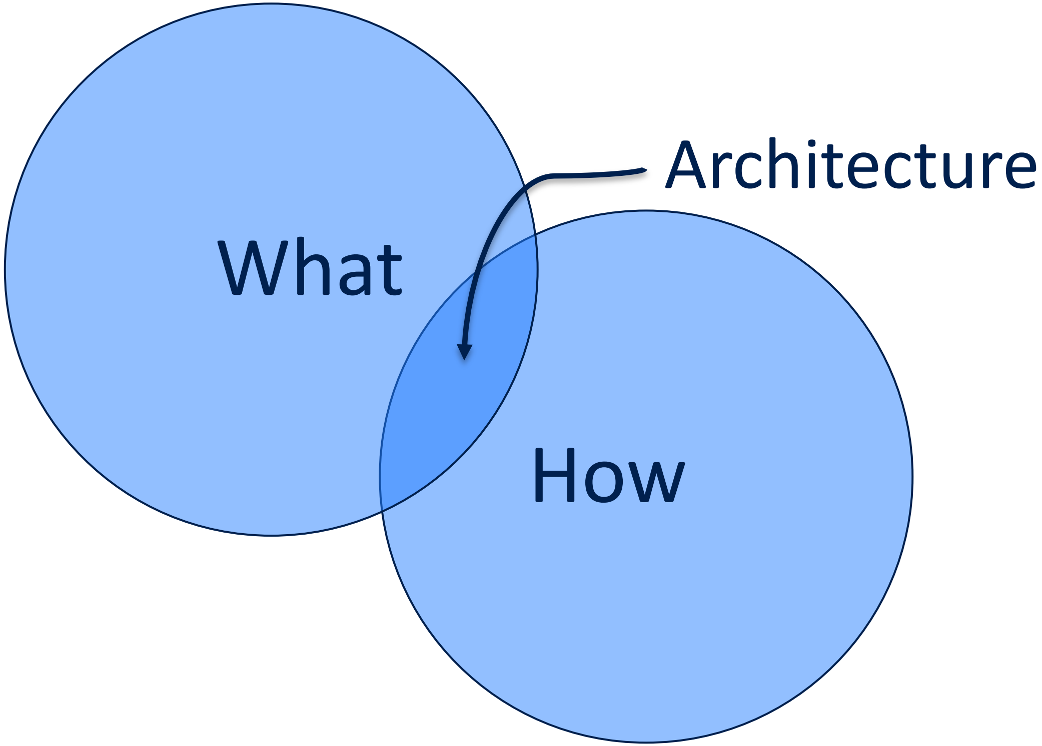
Computing Challenges

- Reducing On-site Data Volume
  - 1000 PB/y directly coming out from cameras (mostly noise)
  - Impossible to transfer via internet link (see later)
- Storing and Processing Big Science Data:
  - Computing resource requirements imply distributed computing model
  - Data volume is too big to separate storage from computing
- Simulating CTA
  - Development of extensive air showers
  - Propagation of Cherenkov light
  - Cherenkov photon ray-tracing through the telescope optics to Camera
  - Photosensors simulation and camera electronics
- Developing a Data Processing Pipeline
  - PBs of data
  - Distributed team of developers.

(+many more computing challenges for e.g. control and supervision of 100 telescopes, sub-array operation, scheduling)

A CTA SW architecture needed

What is Software Architecture?



Main CTA Systems

Science Operations

- Observation Execution System (OES)
- Data Processing and Preservation System (DPPS)
- Science User Support System (SUSS)
- “The HUB”

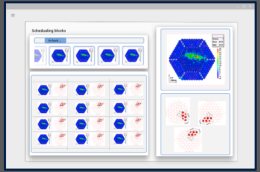
Technical Infrastructure

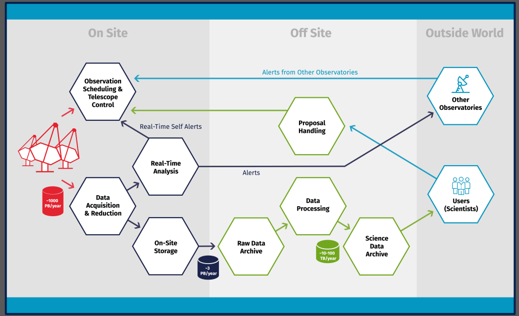
- Telescope (TEL) [*]
- Safety System
- Auxiliary Instruments (AUX)
- Array Infrastructure Elements (AIE)

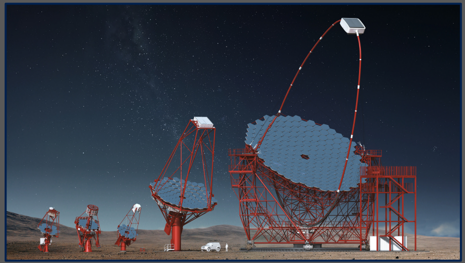
Observatory Operations and Administration

- Operations Support System (OSS)
- Management and Administrative System (MAS)

[*] 1 to N telescope systems, several types



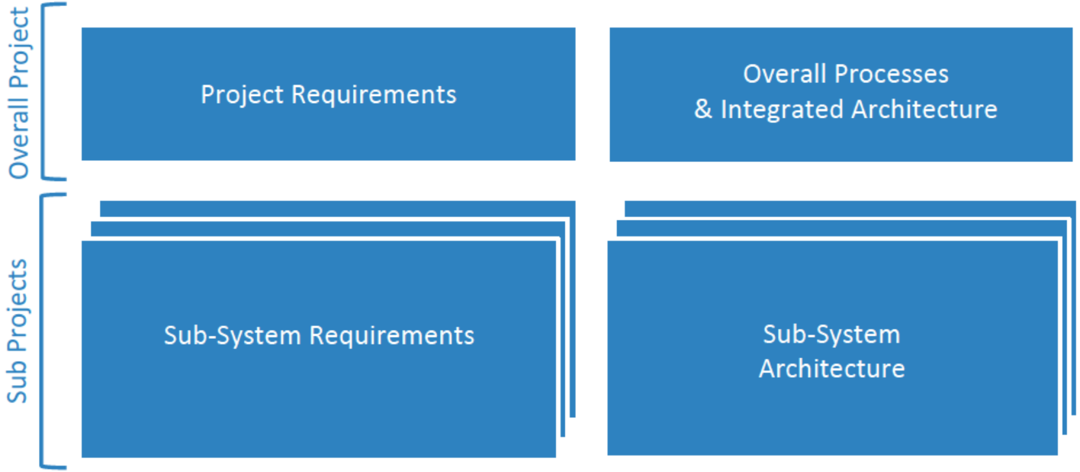




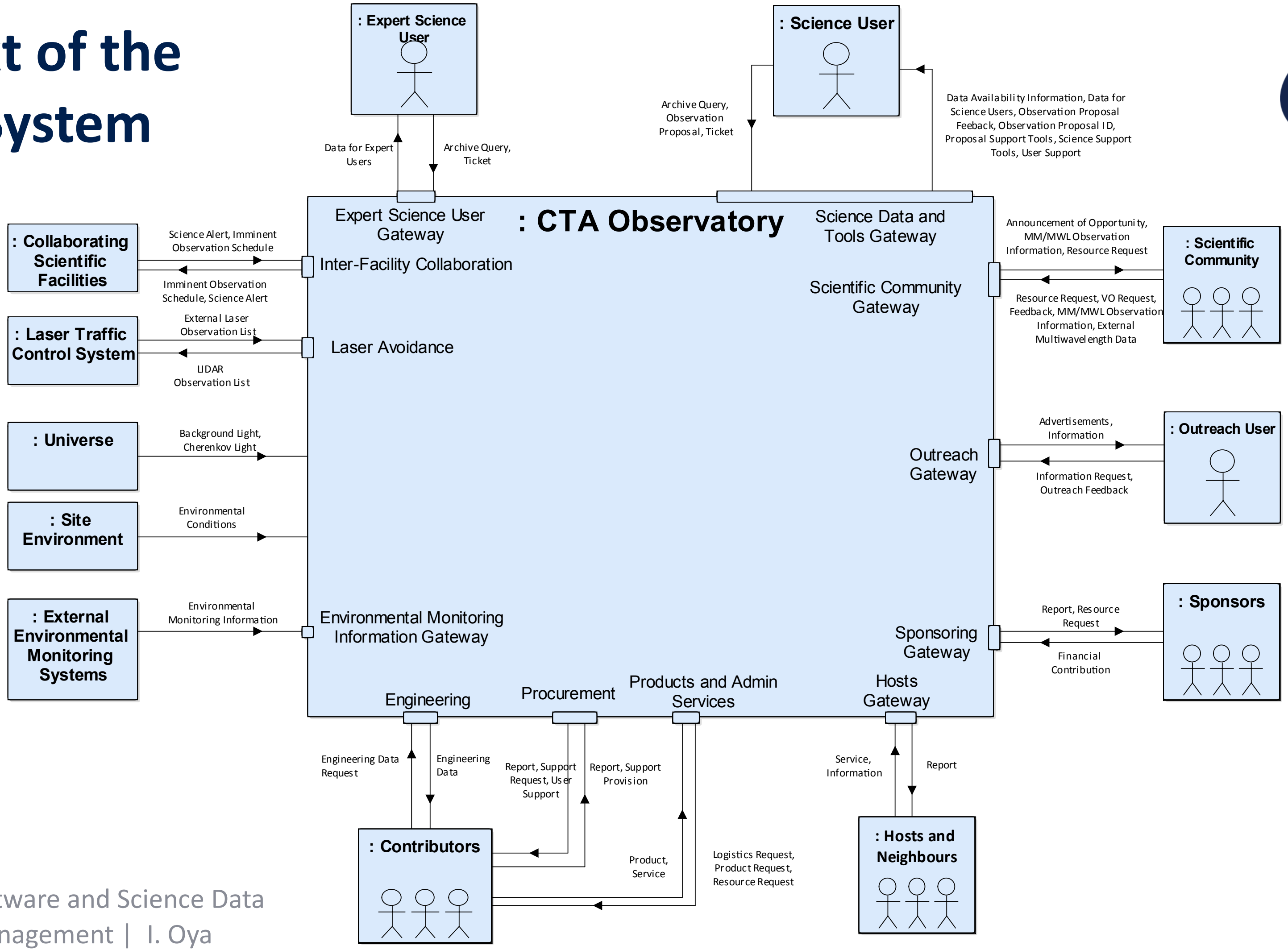
The CTA Observatory System Architecture

CTAO + Fraunhofer IESE

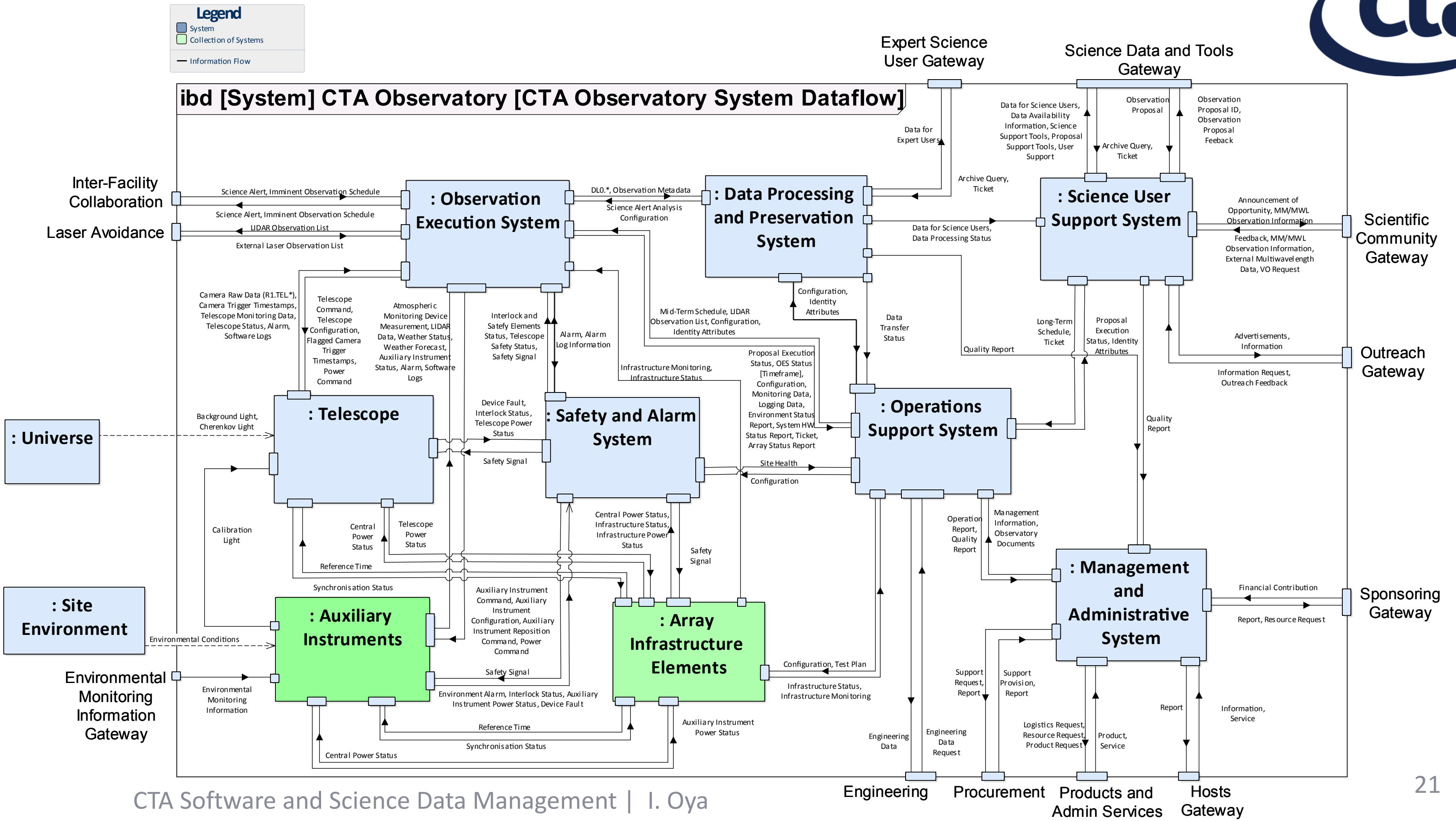
- Architecture summarizes all decisions taken in the project, with views representing the resulting system structures
- **Complete scope**: the whole CTAO System
- **Integration architecture**: Integrates all CTA (sub-)systems
- **Includes everything**: systems, stakeholders, processes, data, interfaces…
- **Model-based and formal approach**:
  - SysML/UML notation
  - Implemented in Enterprise Architect



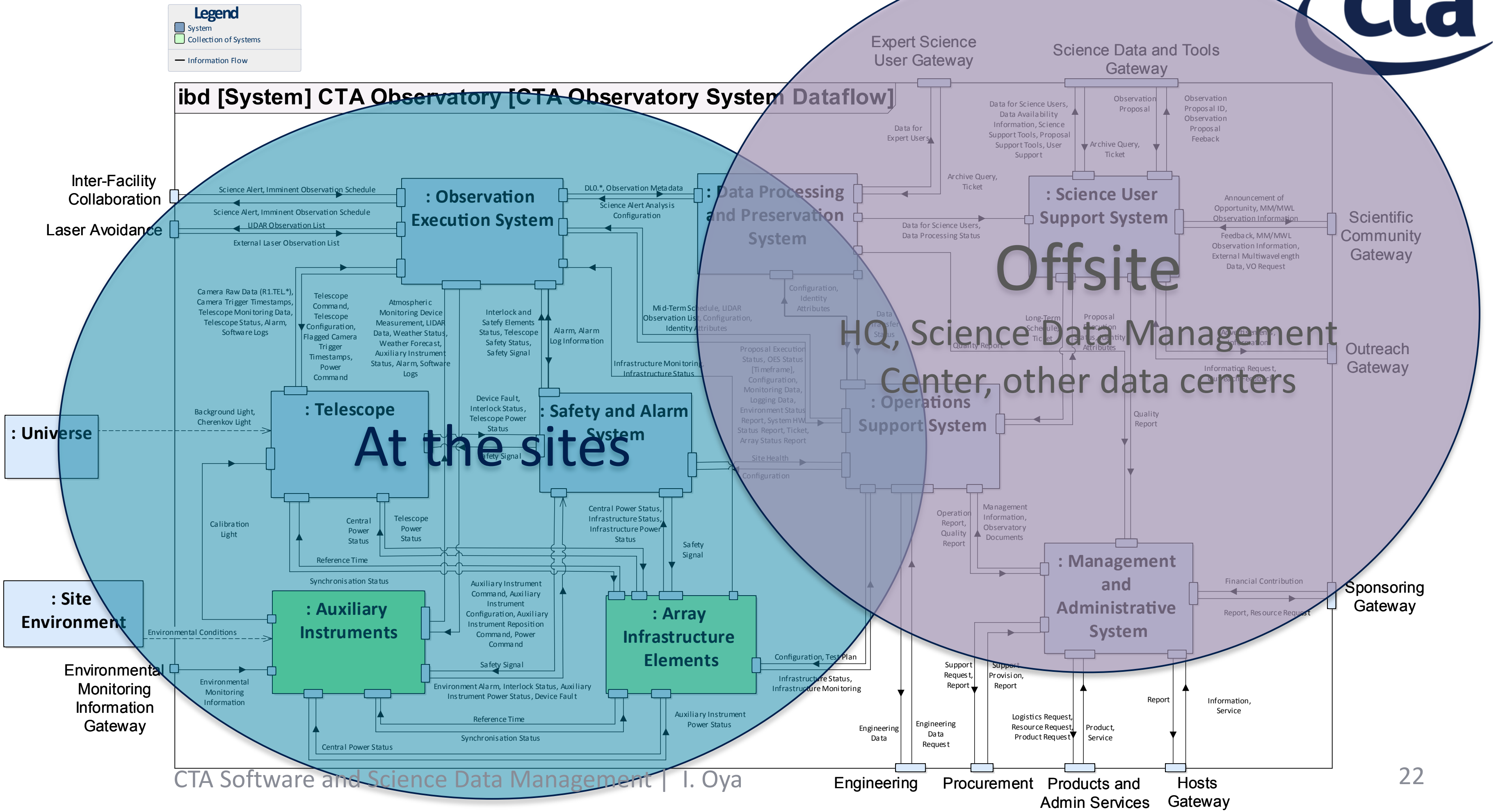
Context of the CTAO System



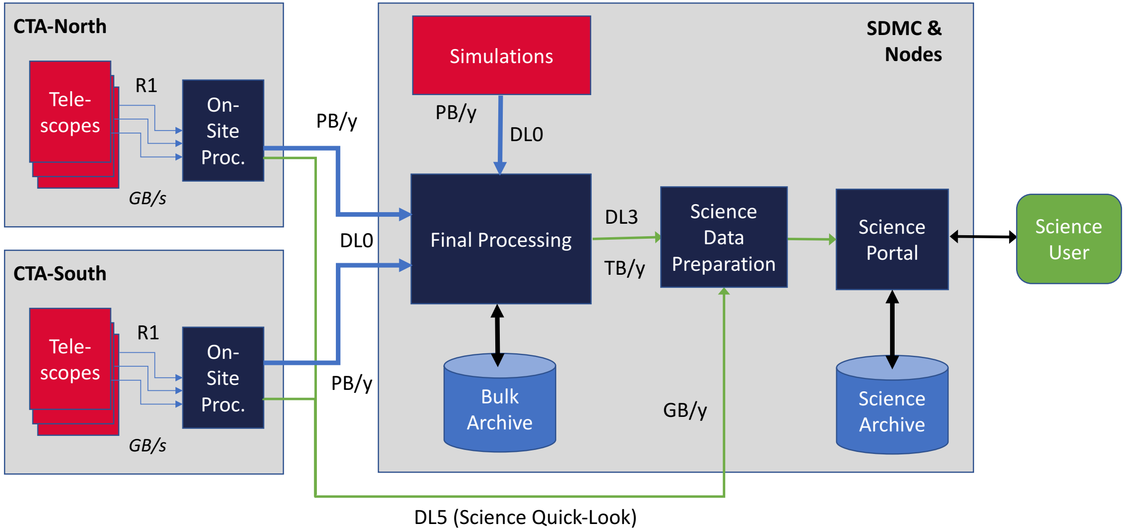
The CTA System Structure



The CTA System Structure

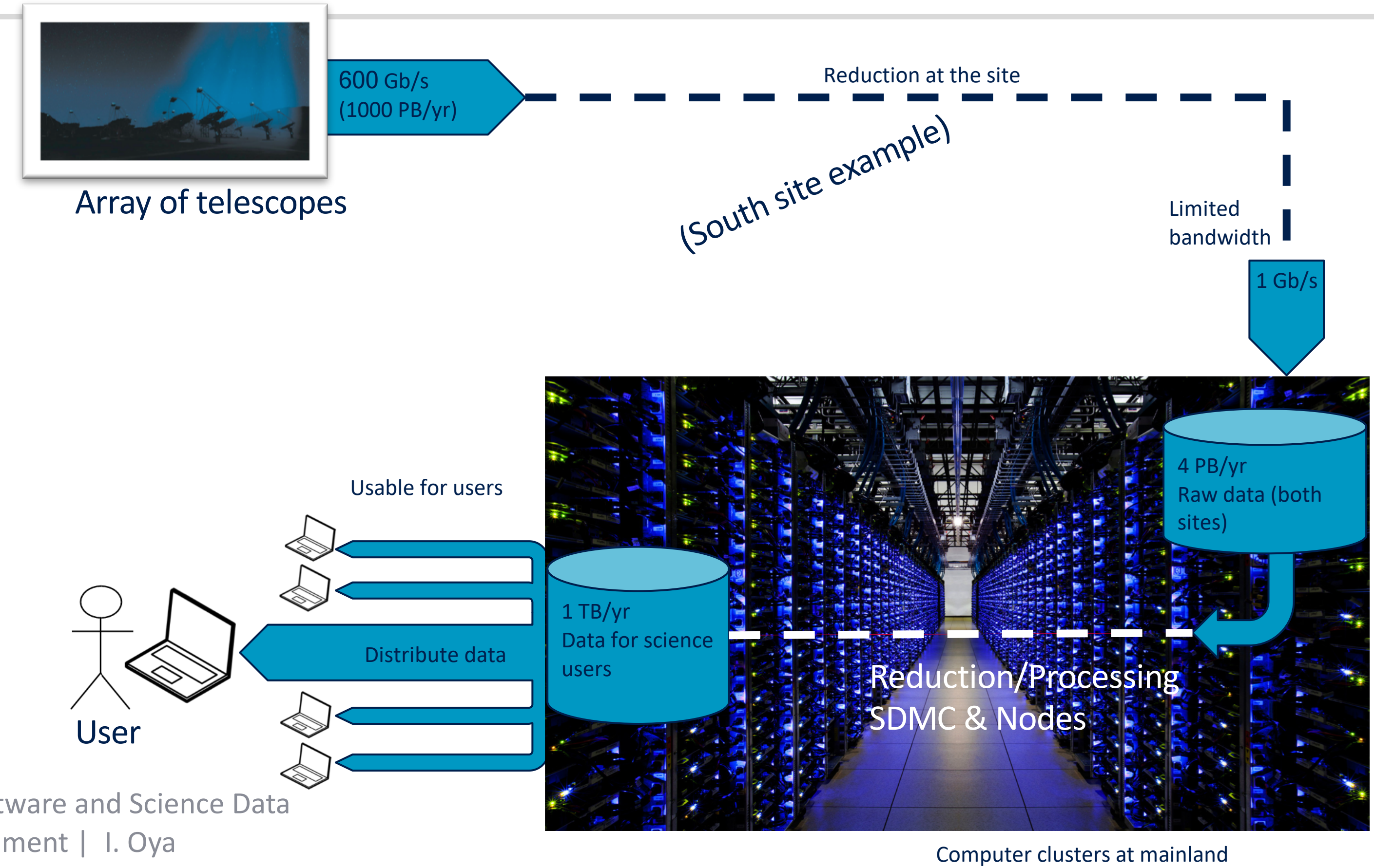


Data Flow & Analysis Categories

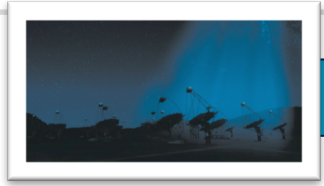


- **Three analysis categories**
  - *Cat A*: at the sites, real time, robust, generate internal science alerts, online data quality
  - *Cat B*: At the sites, offline (e.g. next day), tailored version of "C" → Higher sensitivity than Cat. A, at next morning
  - *Cat C*: offsite, final analysis results with maximum sensitivity, much more computing (and time) required
- **Strong data reduction**
  - Data will be reduced art several levels, from the photon-detector data up to the data delivered to the science users.

Data Flow – From the Cameras to the User



Solving the Data Volume Problem (I)



Array of telescopes

600 Gb/s
(1000 PB/yr)

Reduction at the site

Limited bandwidth

1 Gb/s

4 PB/yr
Raw data (both sites)

Note: larger bandwidths may become eventually available

Reduction at the sites

(South site example)

Central trigger

130 PB/y

3% full waveform signal, remaining signal integrated

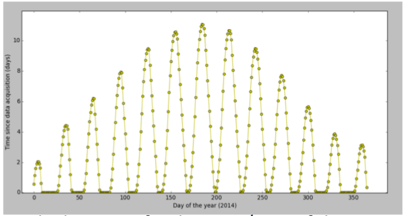
21 PB/y

Compression, event selection…

~ 2 PB/y

Still need factor 10 reduction due to available bandwidth

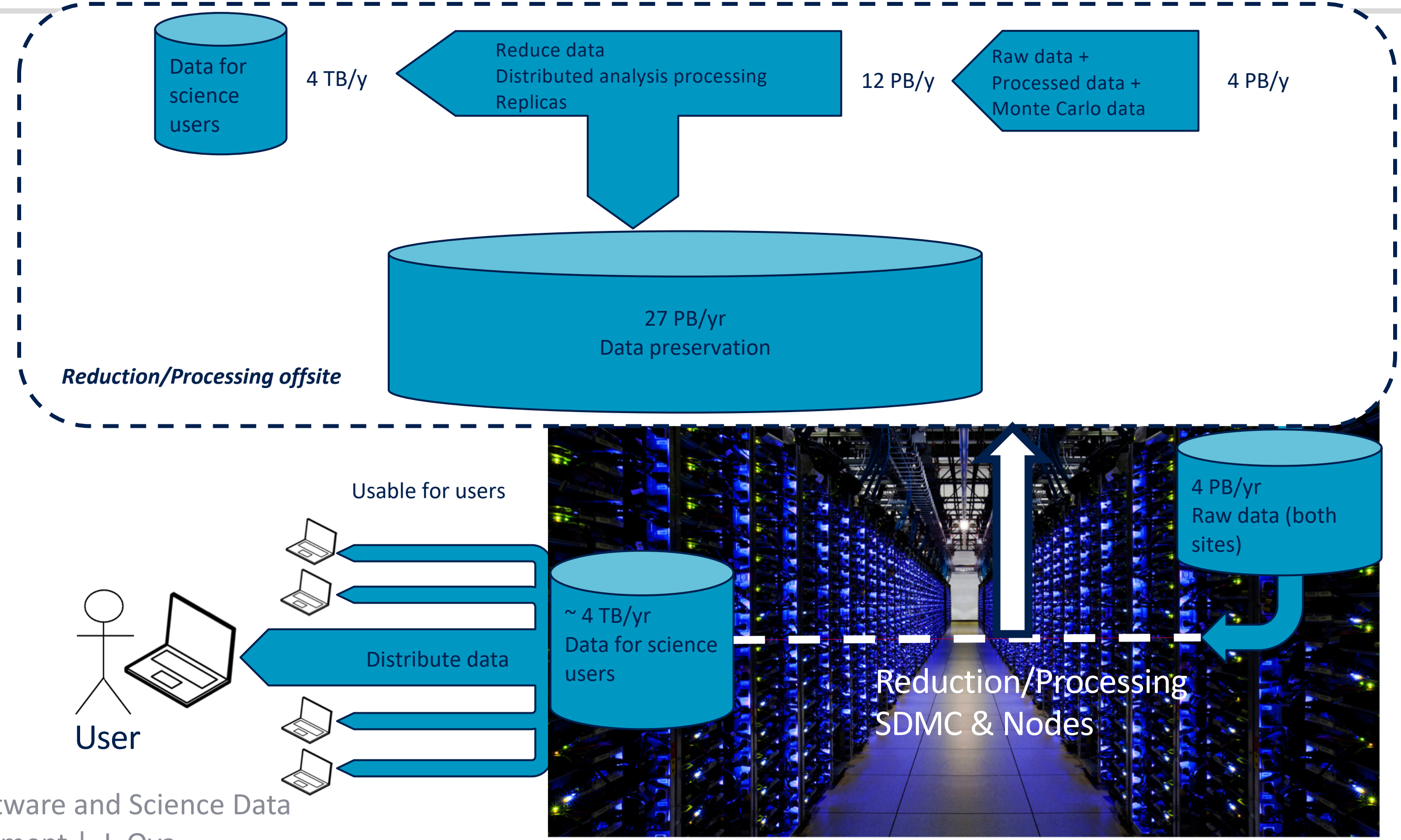
Time since Data Acquisition(days)



Daily data transfer duration/ Day of the year

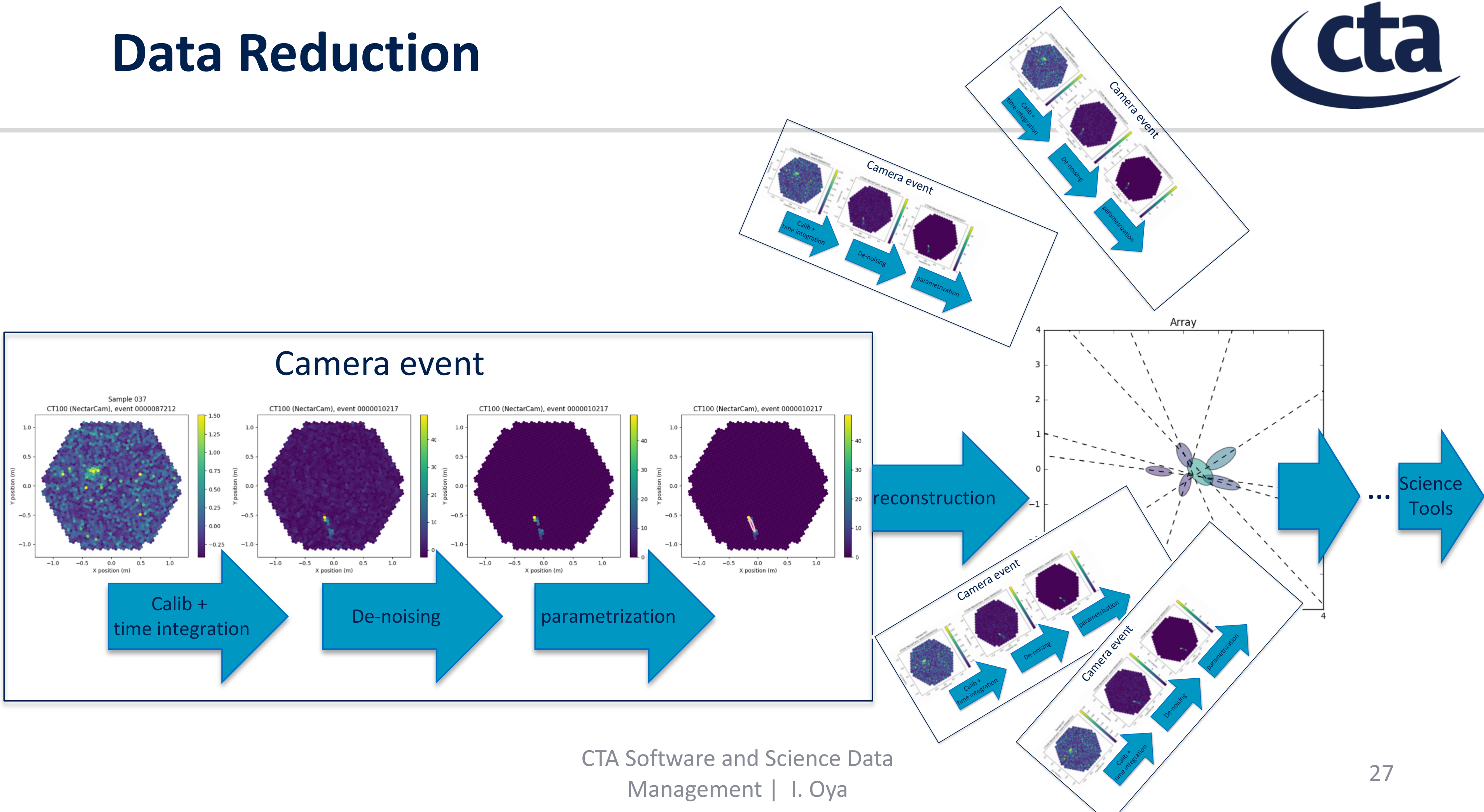
Take into account variations in data volume acquire/day of the year…

Solving the Data Volume Problem (II)

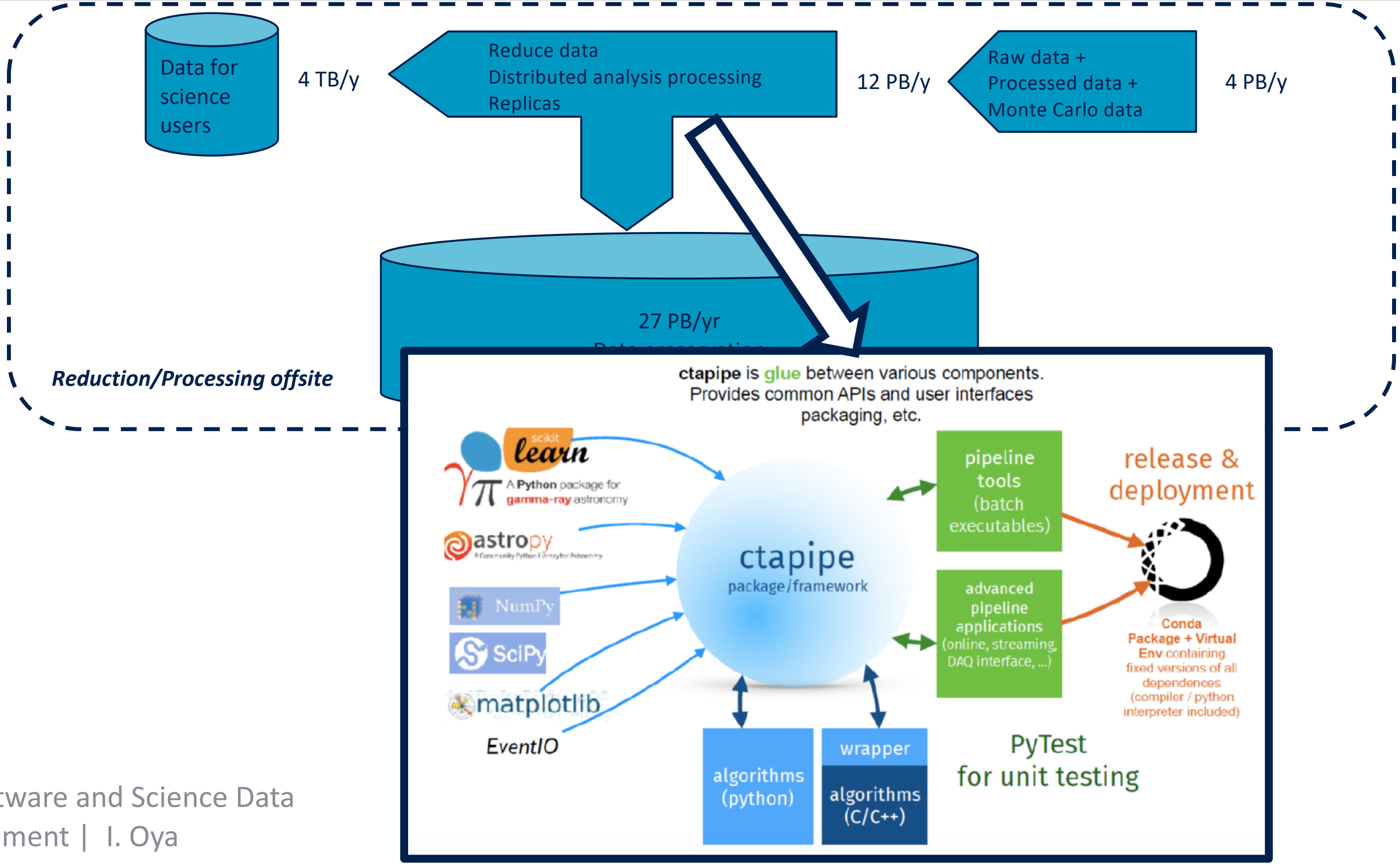


Computer clusters at mainland

Data Reduction



CTA Pipelines: Common Core Package



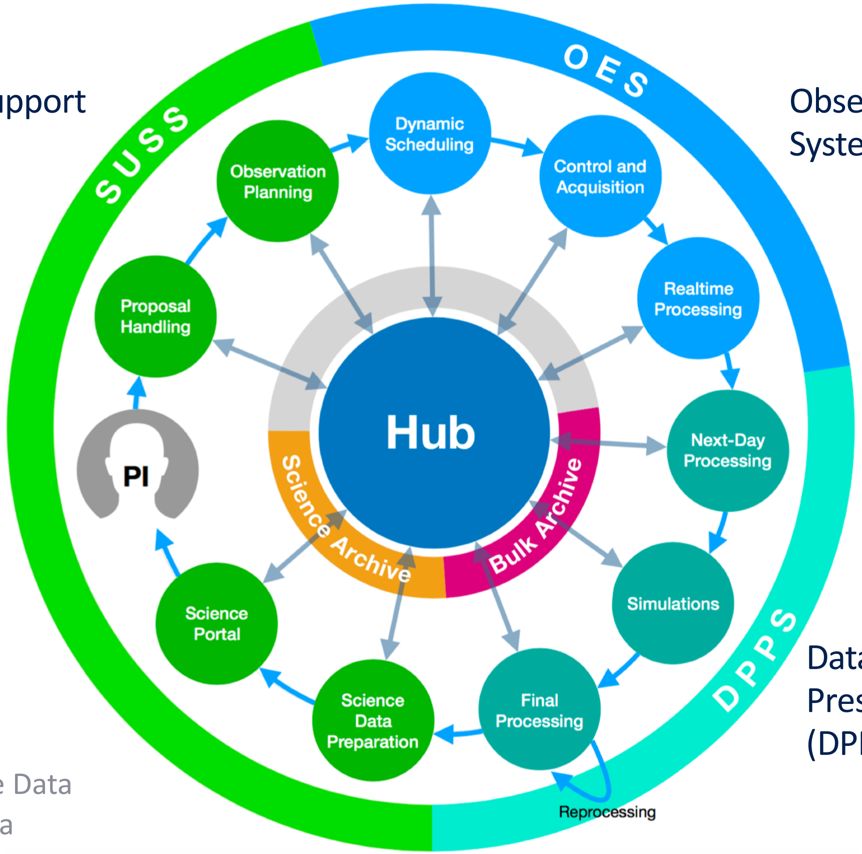
Exploring technologies for the DPPS data centers

File Transfer	CERN FTS, DIRAC-DMS mirroring
Bulk Archive	CTA Archive Prototype (Xtreme DataCloud), DIRAC-DMS/SMS, CTA Data Model
Simulation	Corsika/SimTelArray, GROptics, ROBAST, ...
Pipelines	ctapipe, astripipe
Computing and workflow management	CTA-DIRAC, OPUS
User Interface	Lightweight Science Gateway, OPUS web interface, Dirac web interface.

CTA Science Operation Information Synchronization – “The Hub”

Science User Support System (SUSS)

Observation Execution System (OES)



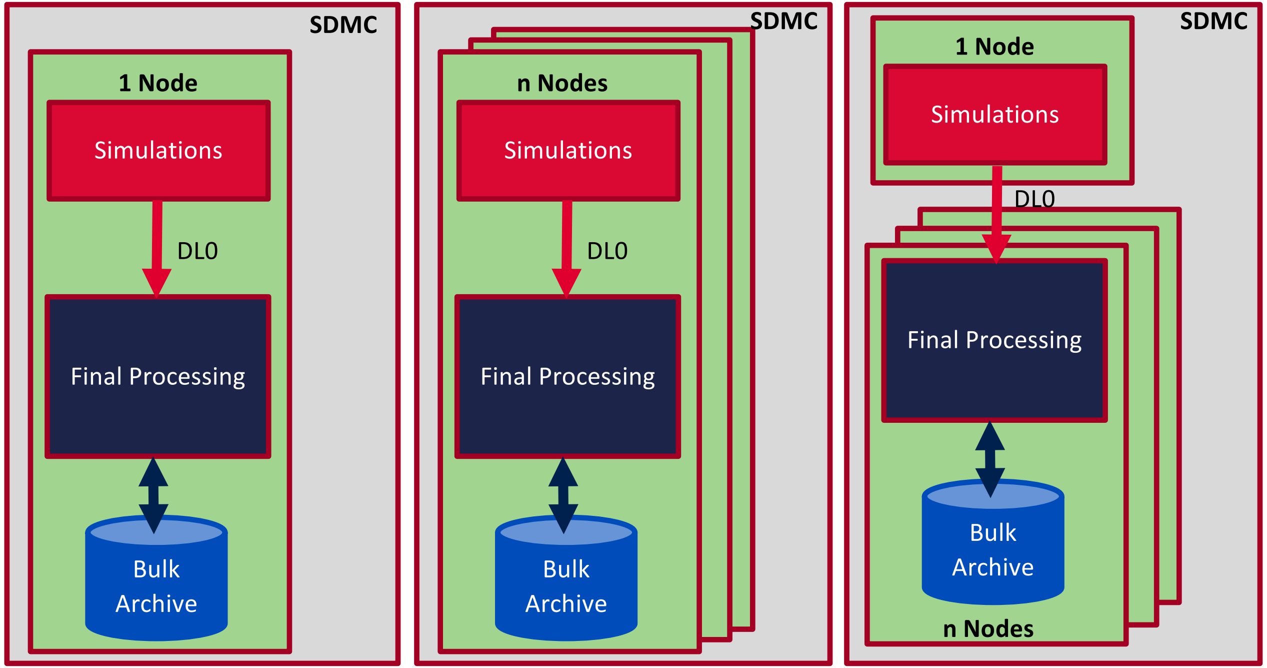
Data Processing and Preservation System (DPPS)

Role of the CTAO Science Data Management Centre at Zeuthen

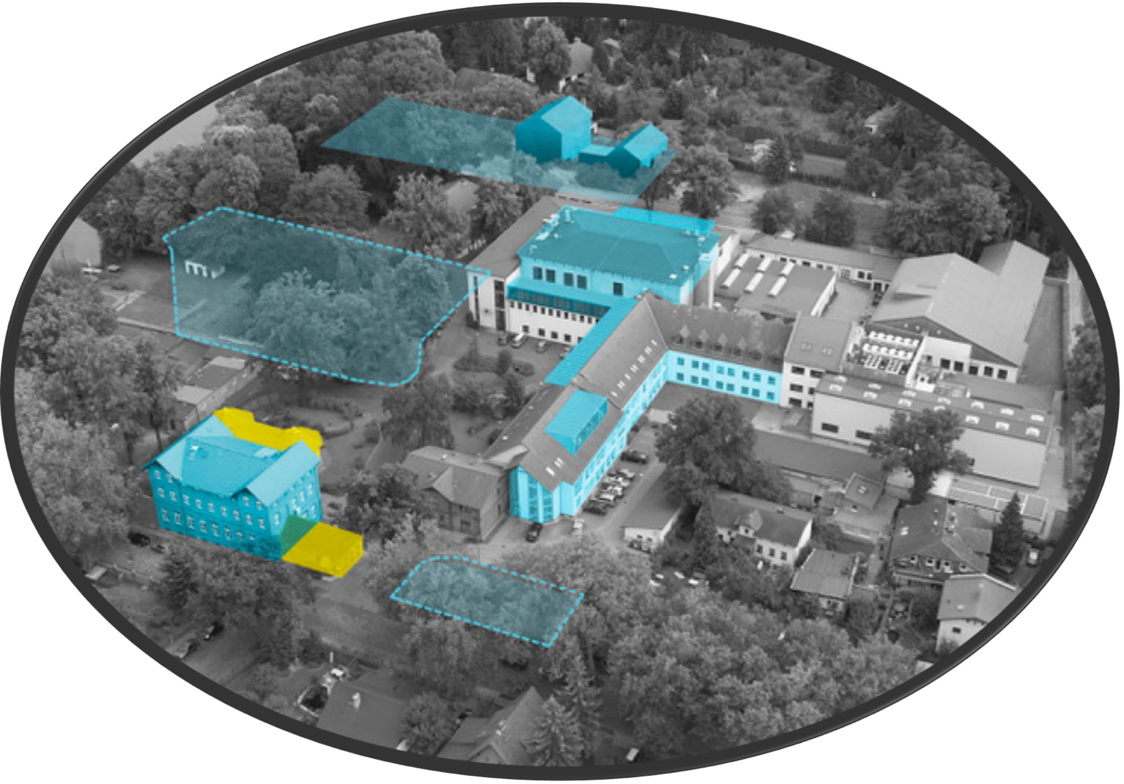
- In charge of science operations and making CTA's science products available to the worldwide community
- Manage CTAO offsite processing software. In particular:
  - Data Processing and Preservation System (DPPS)
  - Science User Support System (SUSS)
  - Master node of "the HUB"

Models for Data Processing as part of Science Operations

- Different computing models under investigation for CTAO
- SDMC computing resources: local + CTAO Contributing institutes and (+ cloud?)

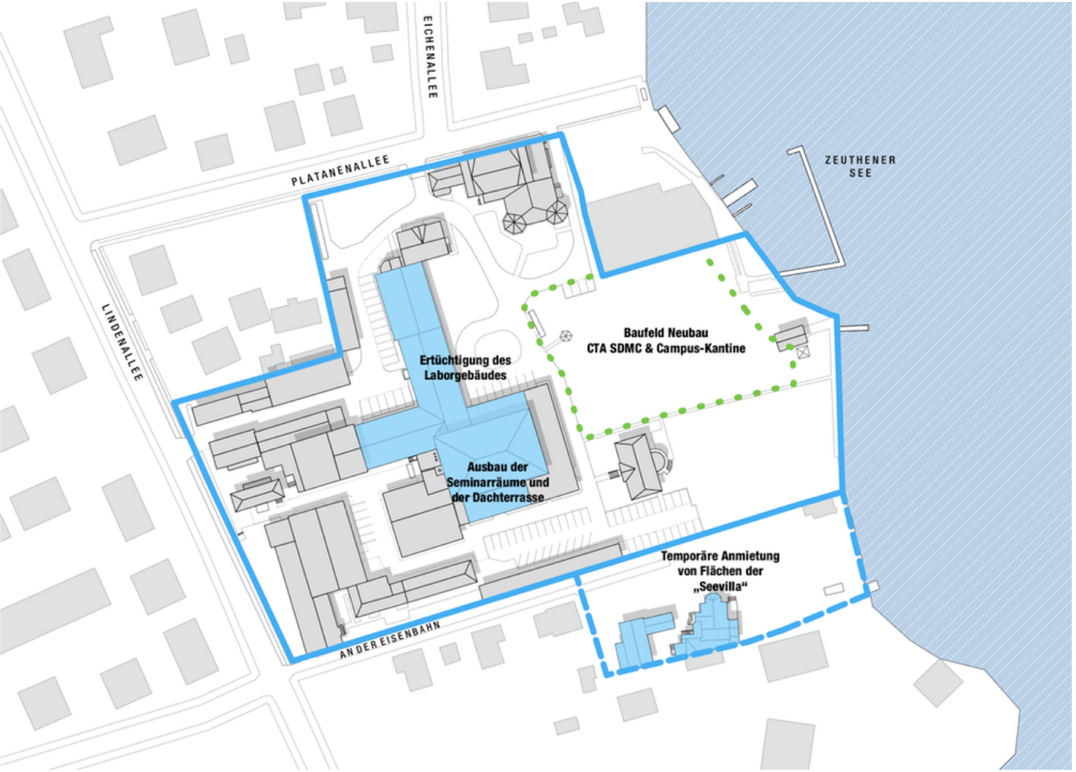


SDMC Building at the DESY Campus in Zeuthen



- New SDMC building as part of a DESY Zeuthen campus master plan
- Competition to design and construct a new building started
  - 1,200 s square metres of useable space
  - canteen
  - education centre
- Final decision on the winning design is planned for early March 2019

SDMC Building at the DESY Campus in Zeuthen

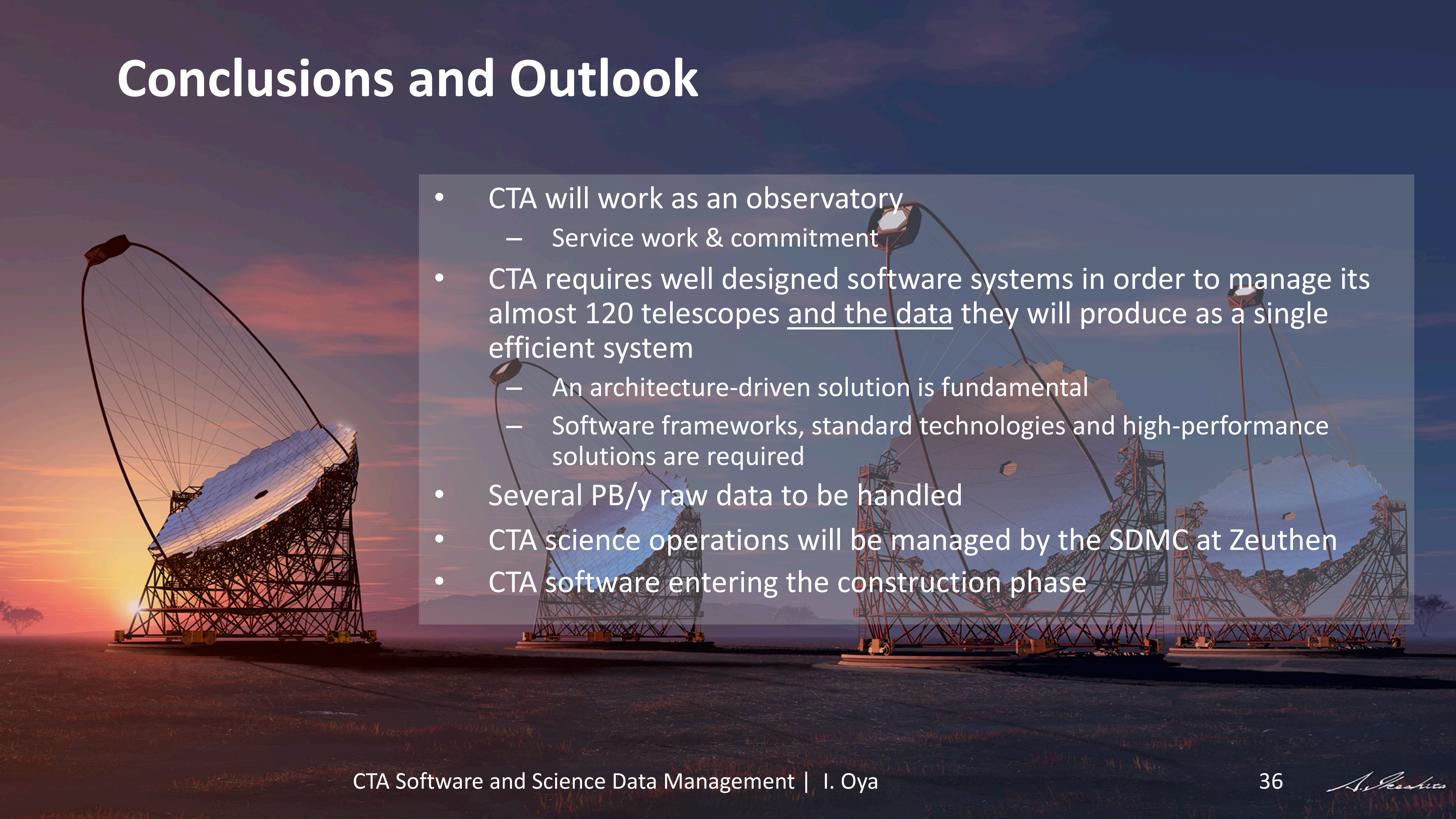


- New SDMC building as part of a DESY Zeuthen campus master plan
- Competition to design and construct a new building started
  - 1,200 s square metres of useable space
  - canteen
  - education centre
- Final decision on the winning design is planned for early March 2019

SDMC – Now

- 2 CTAO staff already allocated in an existing building on campus
  - Me and MF – SW teams coordination
- Room for adding immediately 6 more
  - will grow to support the construction of the CTA software products and to organize science operations
- Already significant activity in our meeting room

Conclusions and Outlook



- CTA will work as an observatory
  - Service work & commitment
- CTA requires well designed software systems in order to manage its almost 120 telescopes <u>and the data</u> they will produce as a single efficient system
  - An architecture-driven solution is fundamental
  - Software frameworks, standard technologies and high-performance solutions are required
- Several PB/y raw data to be handled
- CTA science operations will be managed by the SDMC at Zeuthen
- CTA software entering the construction phase

Register now

#ctasymposium2019

WHEN/

6-9 MAY 2019

WHERE/

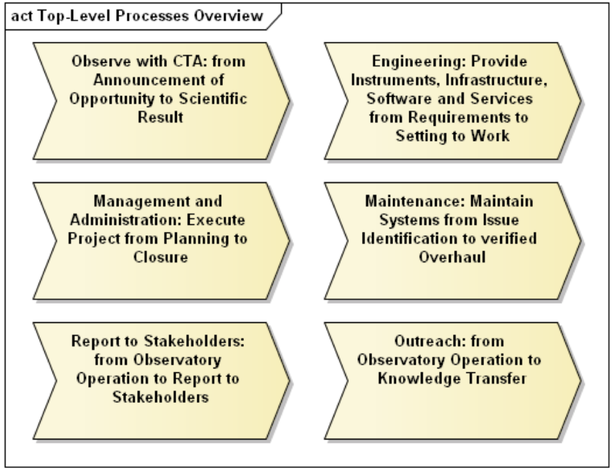
Teatro Duse

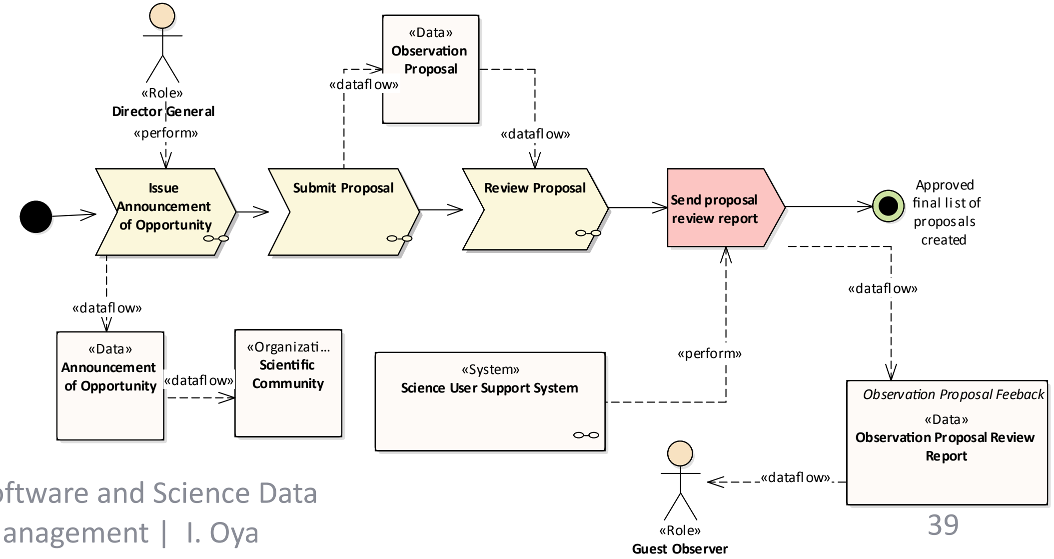
Via Cartoleria 42, Bologna

Backup

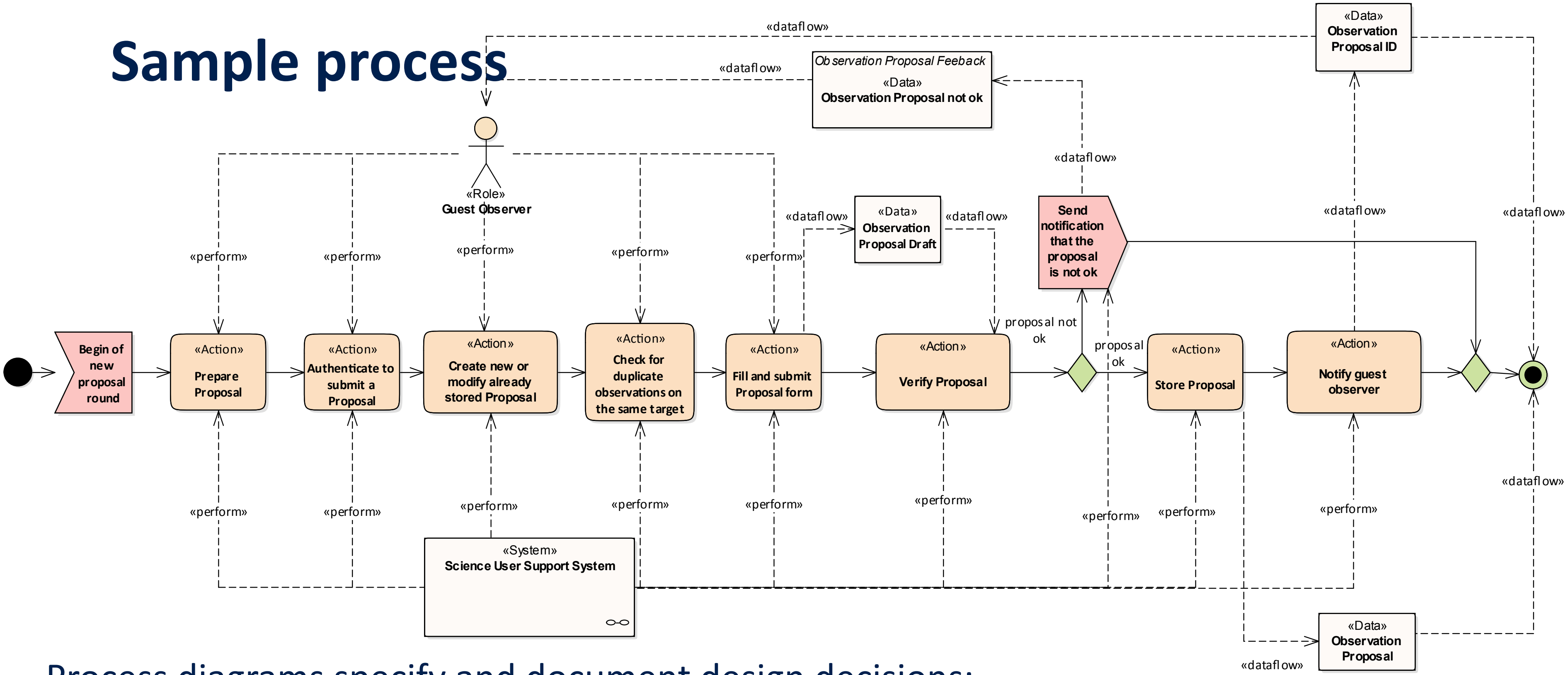
Processes of the CTA Observatory

- Process-based analysis
- Covering full scope of CTA
- 150+ processes at different levels of abstraction
- Modeled with tailored activity diagrams





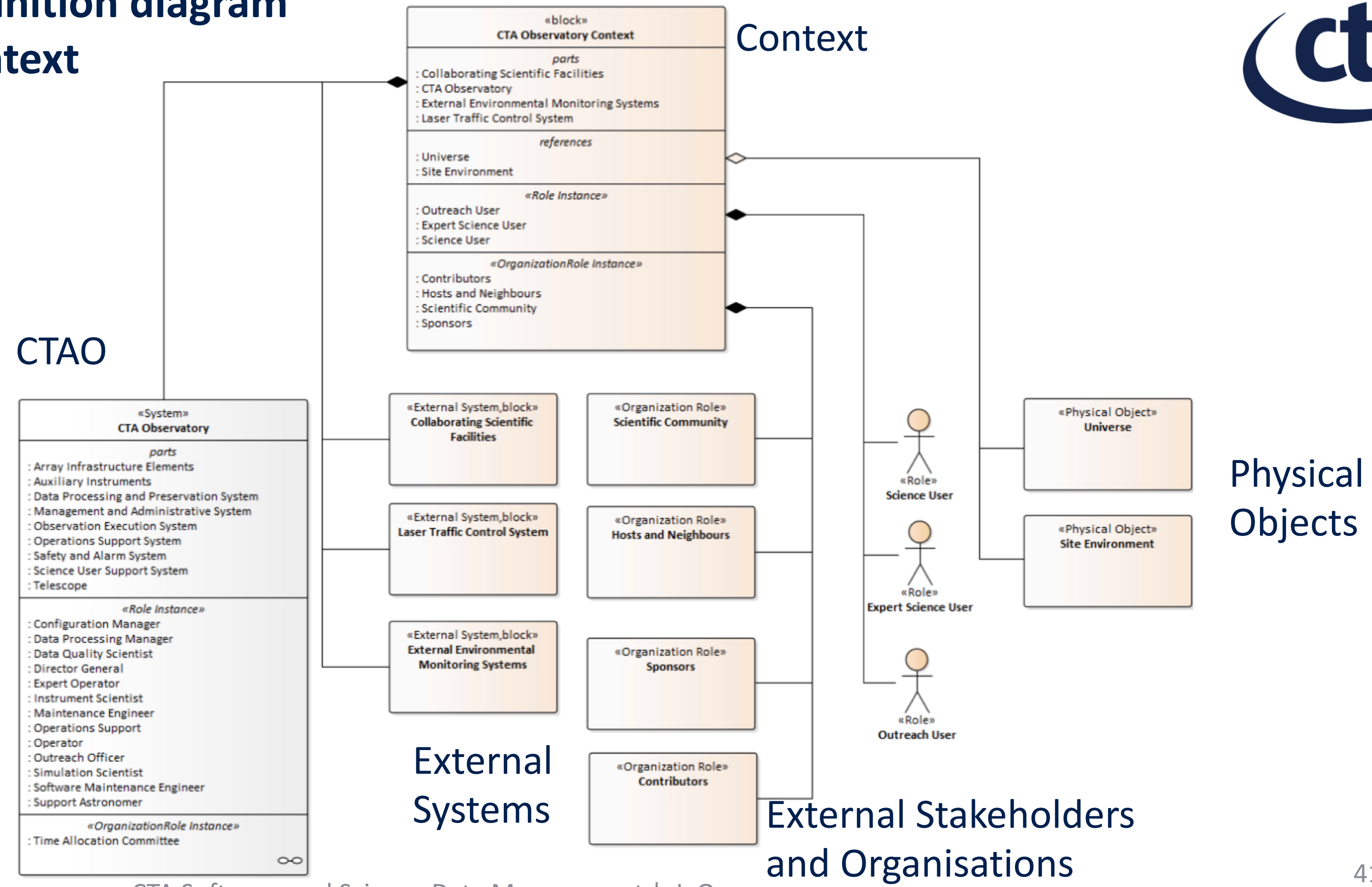
Sample process



Process diagrams specify and document design decisions:

- Tailored activity diagrams
- Conceptual data flow
- Performing system and/or human actor (“stakeholder”)

Block definition diagram
CTAO Context



Architecture Reference Documentation

- Reference Documentation for each individual system
- Input for detailed (sub-)system architecture and product-level requirements
- Includes scope, context, functionality and interfaces
- Diagrams and accompanying specifications and notes

4.3.6.6 Users

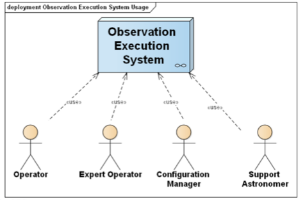


Figure 31: Users of the OES.

4.3.6.4 Main Processes

The main processes to which the OES contributes are (only top two levels listed):

1. Observe with CTA: from Announcement of Opportunity to Scientific Result
   a. *Schedule Refinement: from Long-Term Schedule to Short-Term Schedule*
   b. *Observation Execution: from Short-Term Schedule to Collected Data*
   c. *React to Target of Opportunities: from Received Science Alert to Updated Short-Term Schedule*
   d. *Science Alert Generation: from Collected Data to Generated Science Alert*
   e. *Process Data: from Collected Data to Archived Science Products*
2. Maintenance: Maintain Systems from Issue Identification to verified Overhaul
   a. *Calibration: Perform, Check and Refine Calibration and Calibration Instruments*
   b. *Issue Identification: from Collected Data to Identified Issue*
   c. *Maintenance and Upkeep: from Identified Issue to Verified Maintenance, Calibration or other Action*

SUMMARY	
	Summary table
NAME AND ACRONYM	
	Name and acronym of the system used throughout the section.
SCOPE	
	Summary of the main responsibilities and functionalities of the system under discussion.
CONTEXT	
	Environment of the system under discussion showing the main interactions and information flow with the surrounding internal and external systems and stakeholders.
USERS/STAKEHOLDERS	
-	**EXTERNAL USERS**
	External users of the system which interact with the system and are supported by the system to execute certain actions. The users can be external or internal to the CTAO.
-	**INTERNAL USERS**
	Internal users of the system which interact with the system and are supported by the system to execute certain actions. The users can be external or internal to the CTAO.
MAIN DATA ELEMENTS	
	Main conceptual data elements that are exchanged with the users and other systems (either as produced or consumed data) and are handled and/or stored within the system under discussion.
FUNCTIONALITY	
-	**USER-SUPPORTED PROCESSES AND ACTIONS**
	Main actions and related processes in which the system supports an internal or external stakeholder.
-	**AUTOMATIC PROCESSES AND ACTIONS**
	Main actions and related processes that are performed by the system.
INTERFACES	
	List of all the main interfaces with their purpose.
DEPLOYMENT	
	Number of instances of the system currently envisaged to run simultaneously (single or multiple instances) and its deployment location (either on site, off site or both).
NOTES	
	Additional notes.