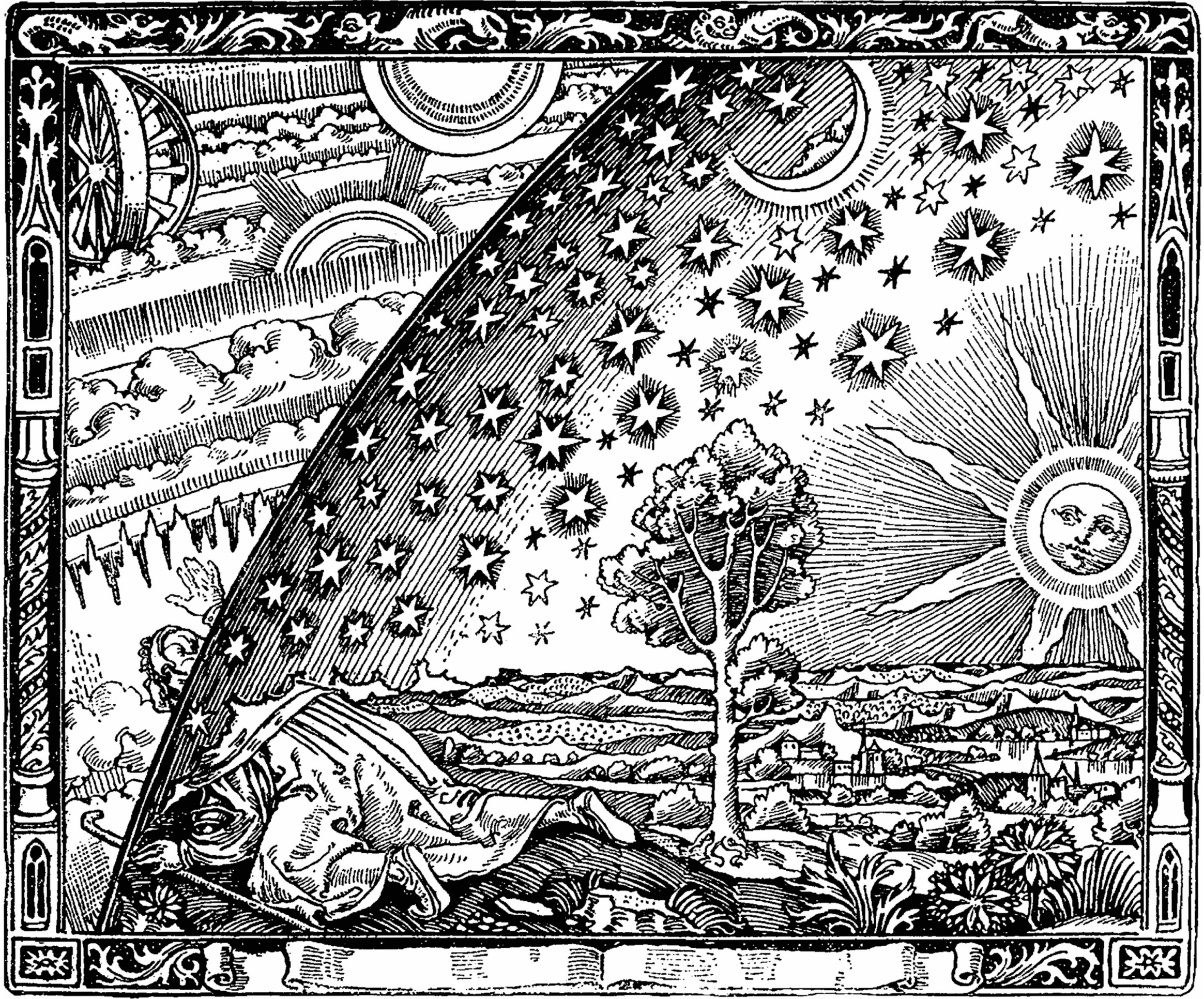


Challenges in (Federated) Computing in Astroparticle Physics

GÉANT May 2022

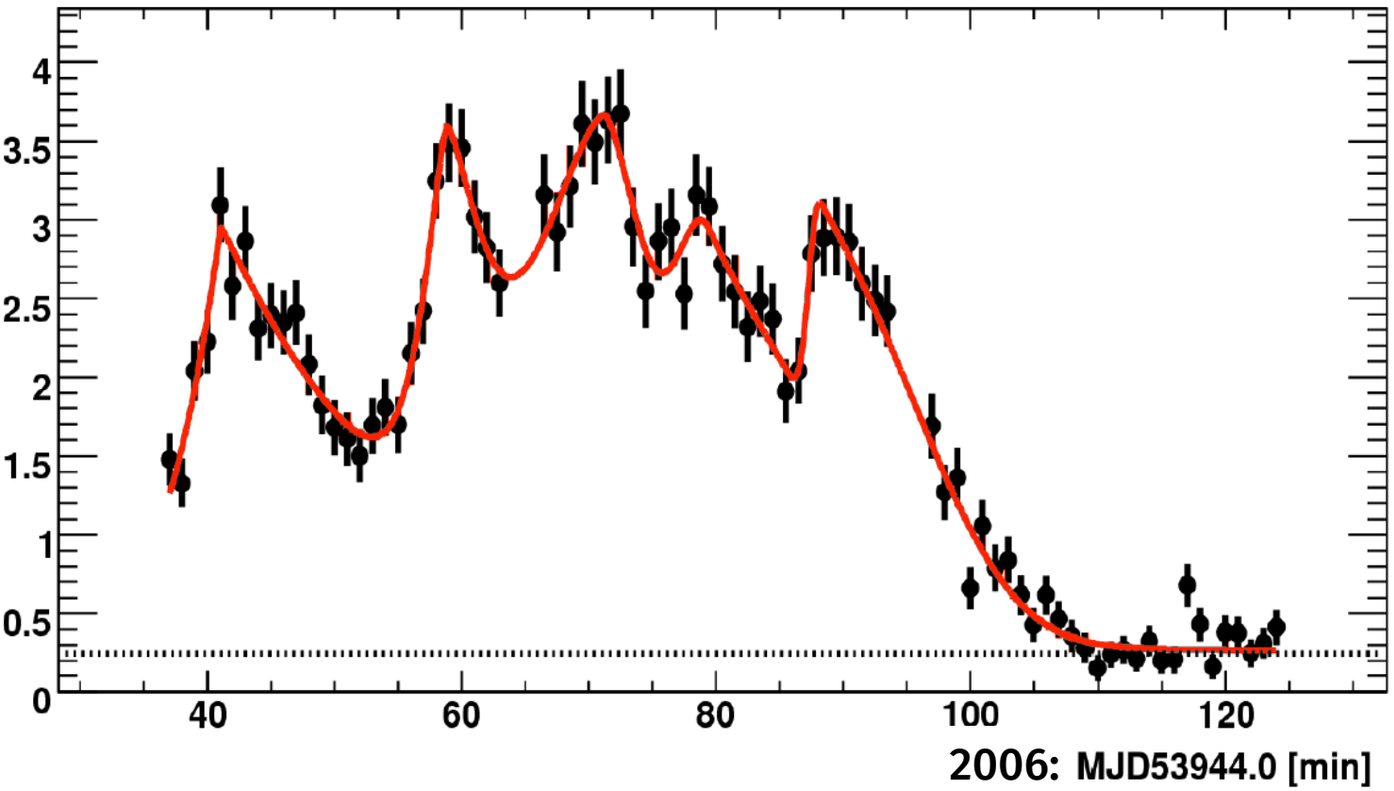
Canada & USA | Latin America | Europe | North Africa & Eastern [illegible] | West & Central Africa | Eastern & Southern Africa | [illegible] Asia | Asia-Pacific | Other R&E Networks

canarie ESnet INTERNET2 Red CLARA GÉANT EU4Digital SRENA CAREN TEIN

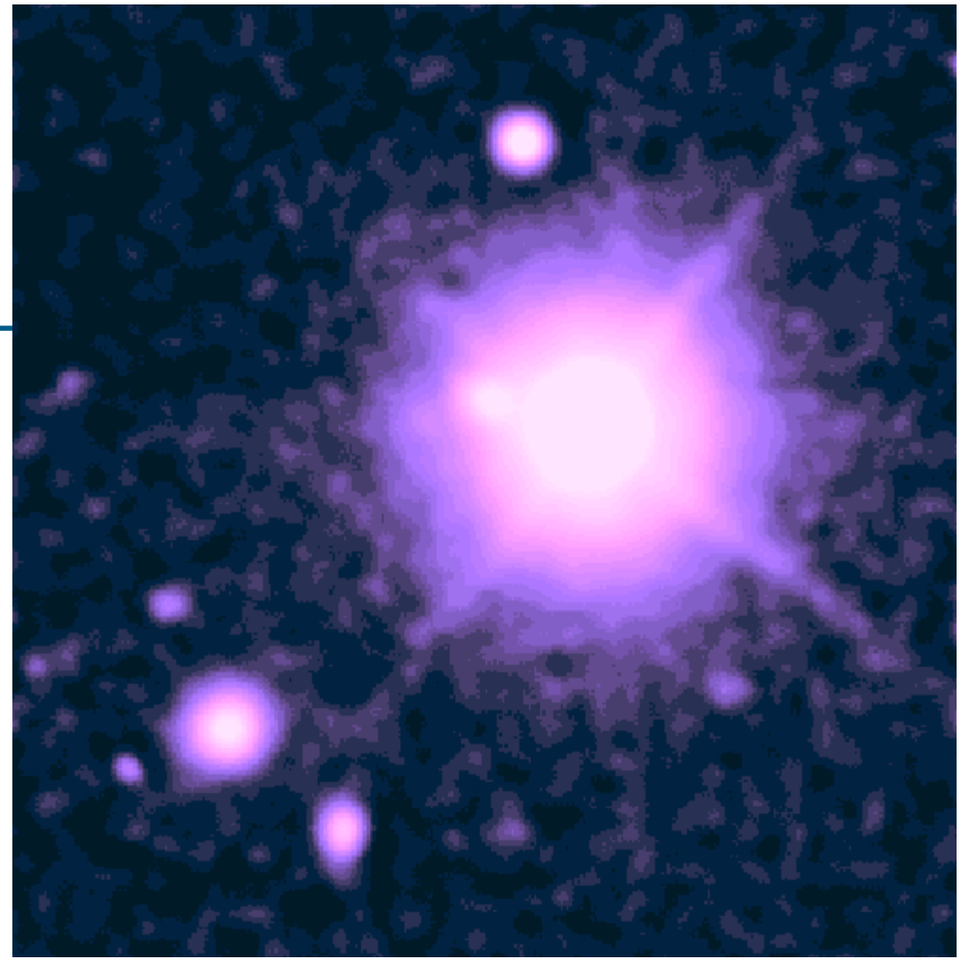
JENA Computing 2023

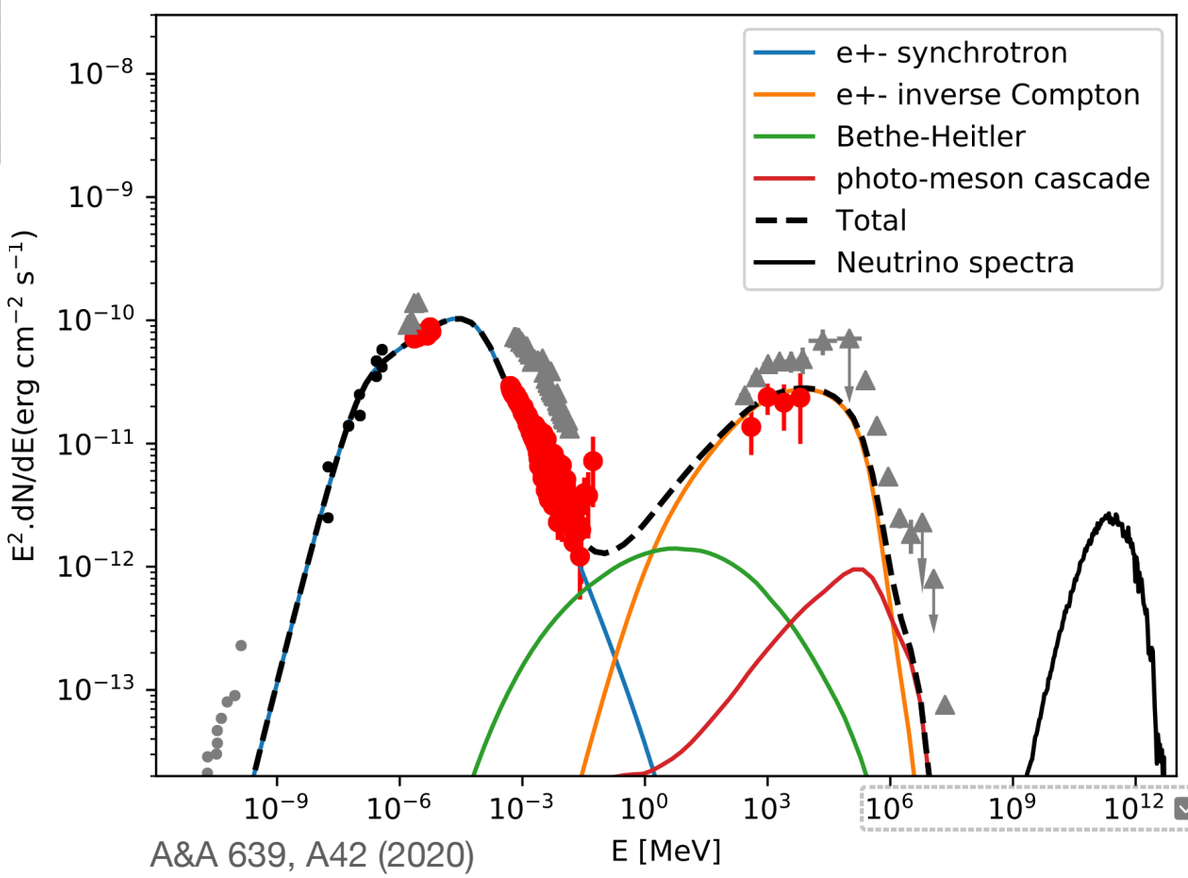
Blazar

PKS 2155



www.mpi-hd.mpg.de/hfm/HESS/pages/home/som/2010/07/

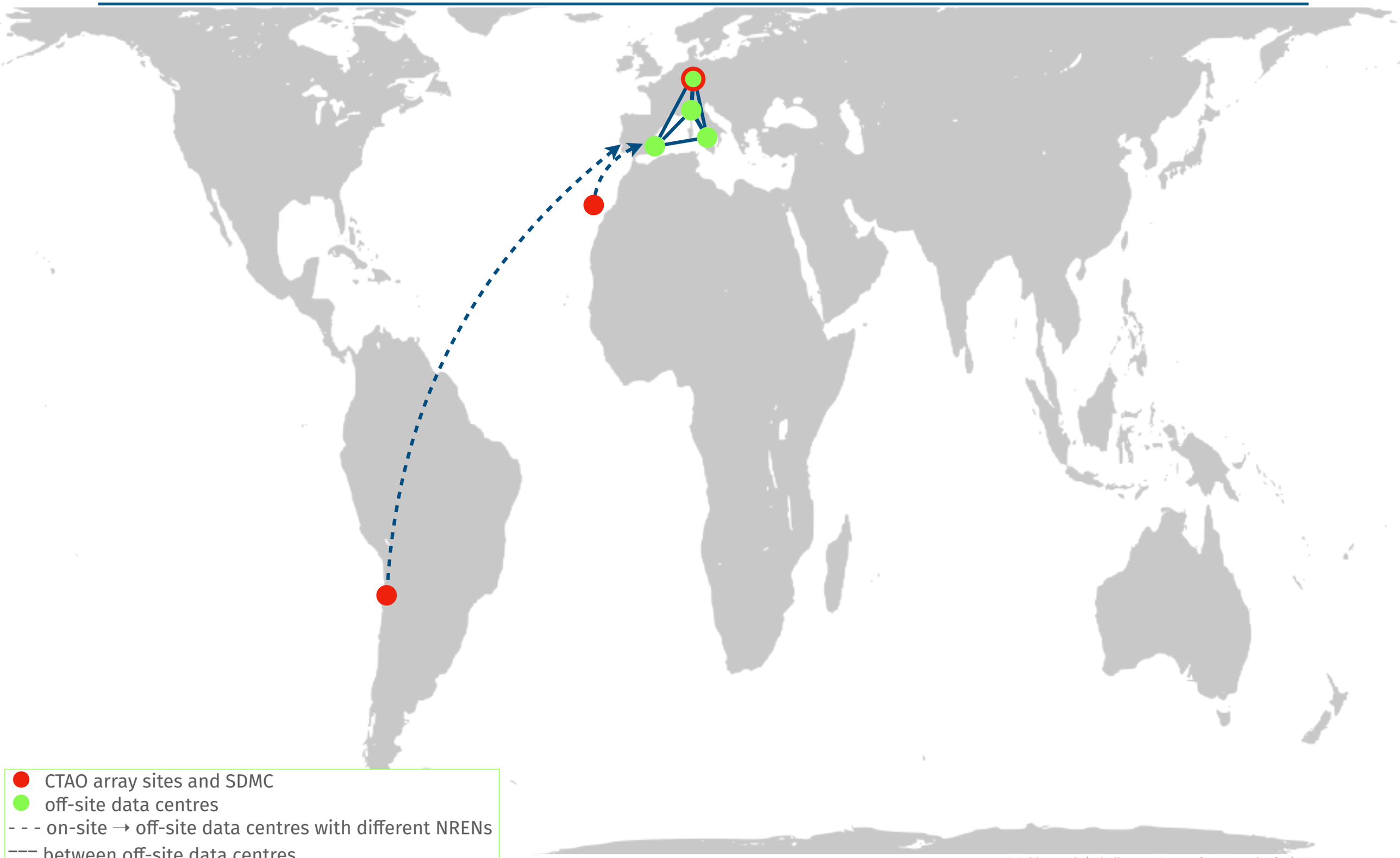




Dynamic of acceleration?

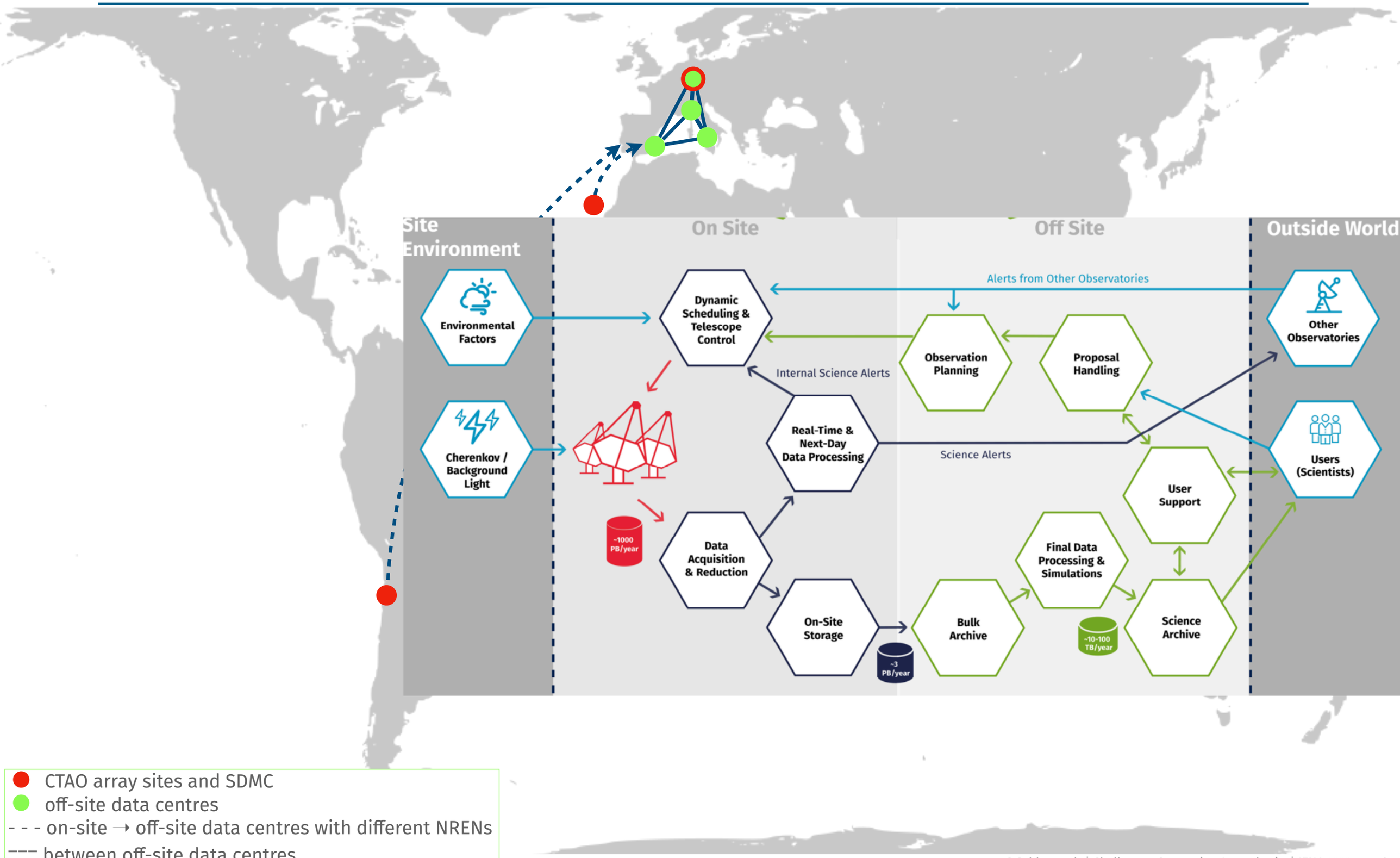
Cherenkov Telescope Array Observatory

From Site to Data Centres



Cherenkov Telescope Array Observatory

From Site to Data Centres



- CTAO array sites and SDMC
- off-site data centres
- - - on-site → off-site data centres with different NRENs
––– between off-site data centres

CTAO

Computing View

A distributed environment

- **Telescopes** and other array elements on Chile and La Palma **site**
  - experts/ maintenance teams, remote or on-site
- Operate array elements, collect and process data in two **on-site data centres** (DCs)
  - computing teams, operators and various on-site teams
- Transfer data → Europe → **four off-site DCs** → process data and preserve in bulk archive in those DCs
  - computing and SciOps teams → DCs, services, applications
- Process data and preserve in science archive in **two off-site DCs**
  - computing and SciOps teams → DCs, services, applications
- Monitoring and oversight in **Science Data Management Centre**
  - computing and SciOps teams
- **Science users** get access through a portal for observation proposals, and high-level data and software and status messages…

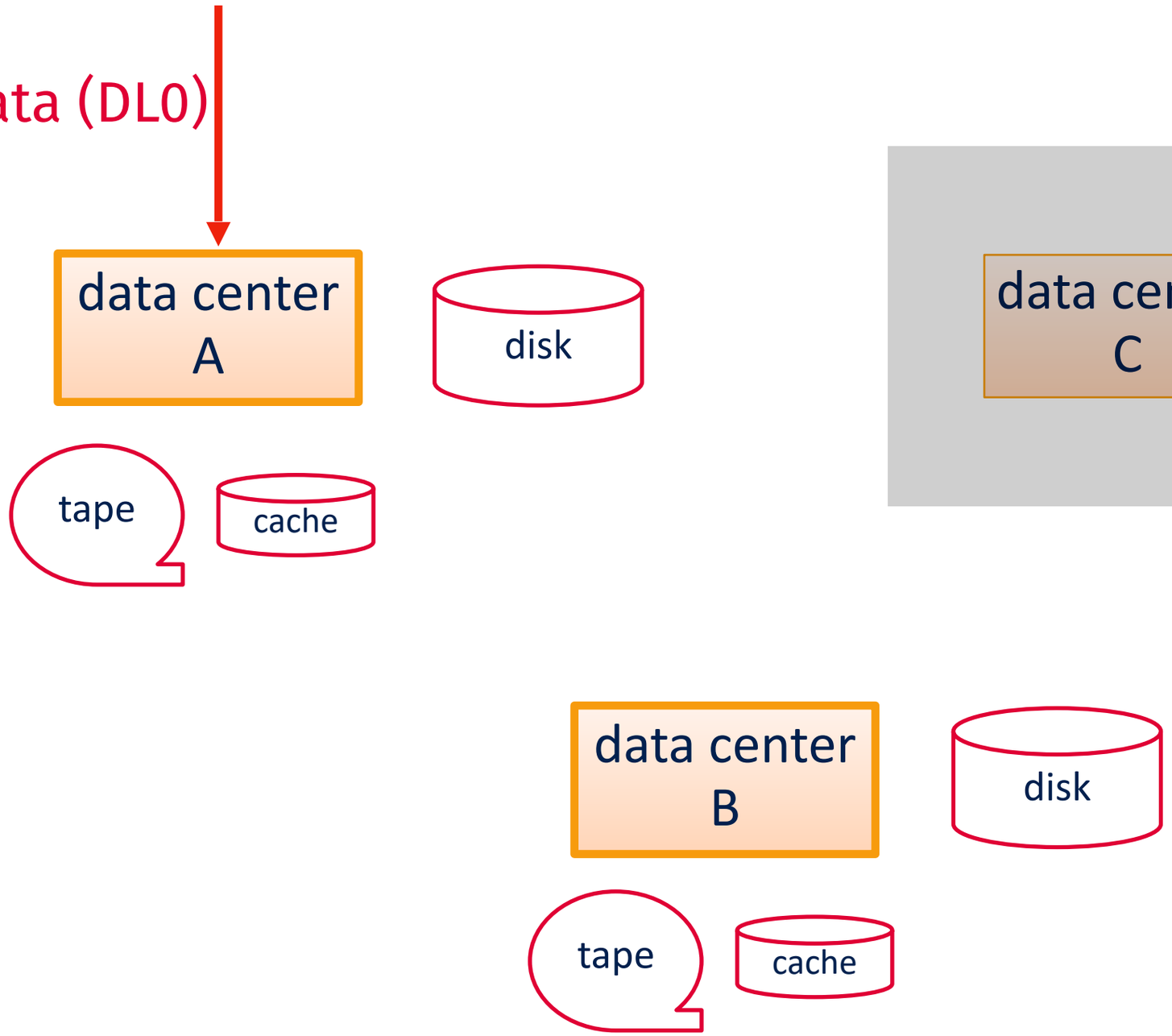
Off-site ICT

Data Flow for Raw Data Processing

simplified view

Observation night 1 from site 1

raw data (DL0)

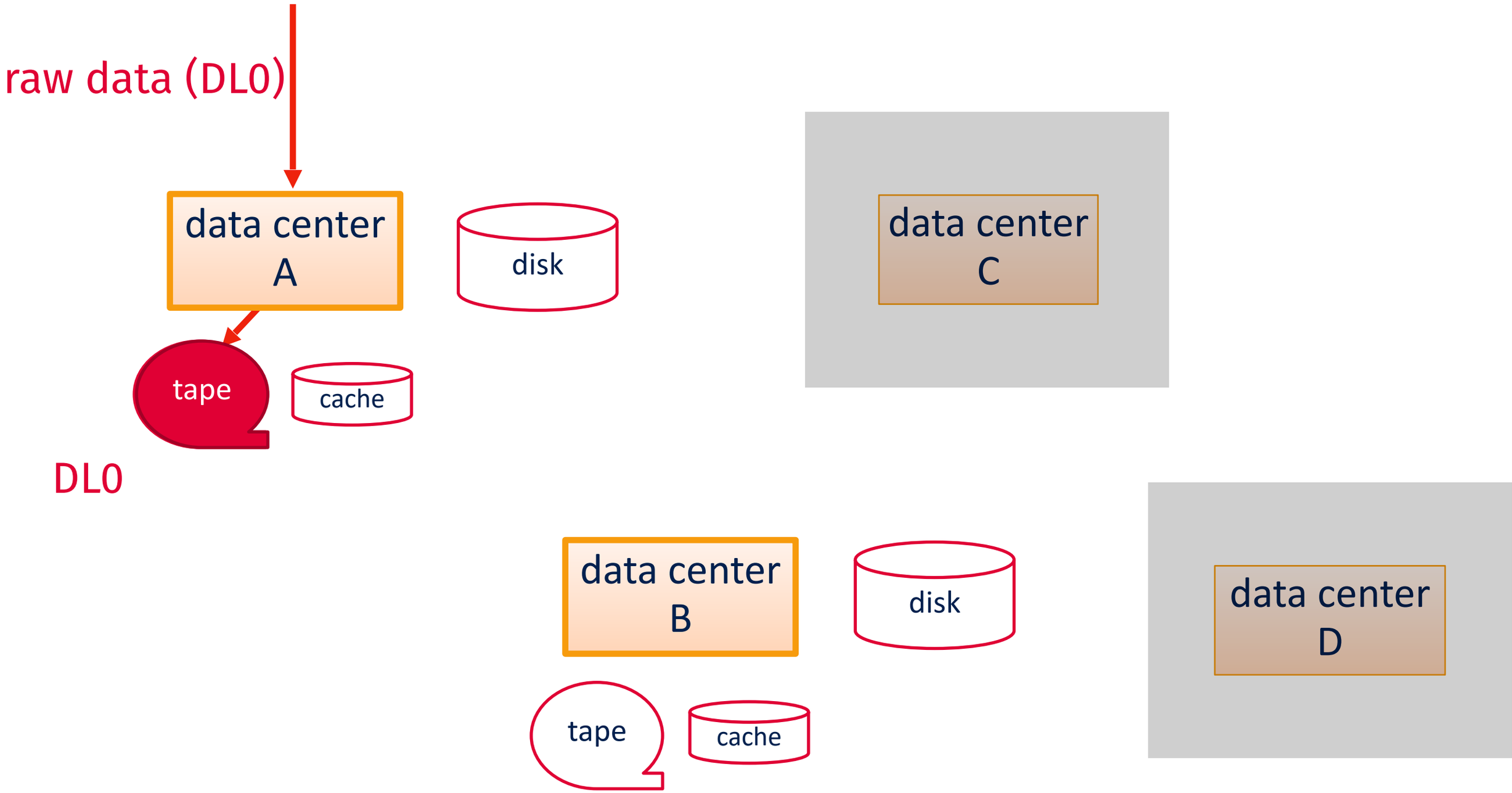


Off-site ICT

Data Flow for Raw Data Processing

simplified view

Observation night 1 from site 1

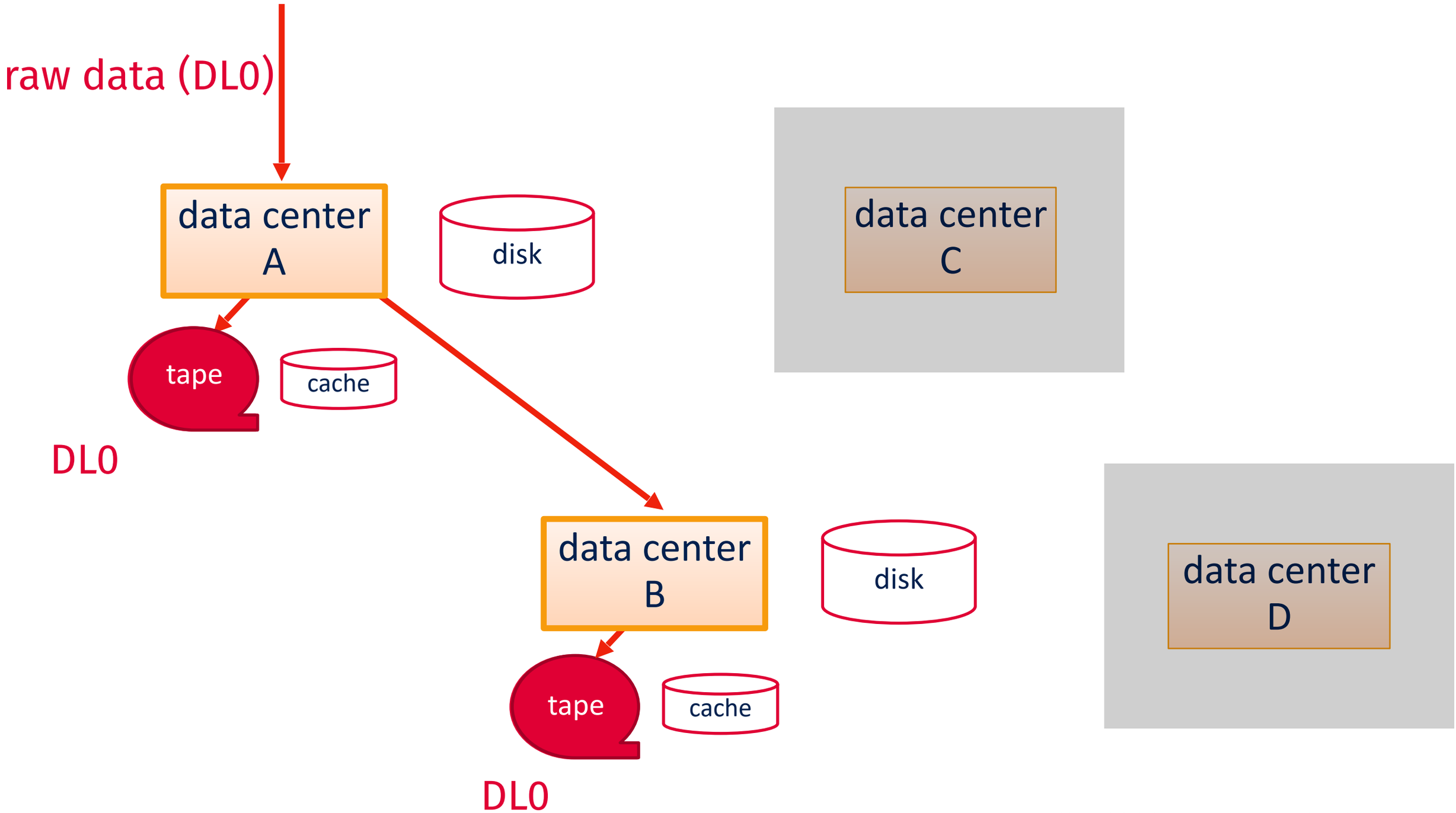


Off-site ICT

Data Flow for Raw Data Processing

simplified view

Observation night 1 from site 1

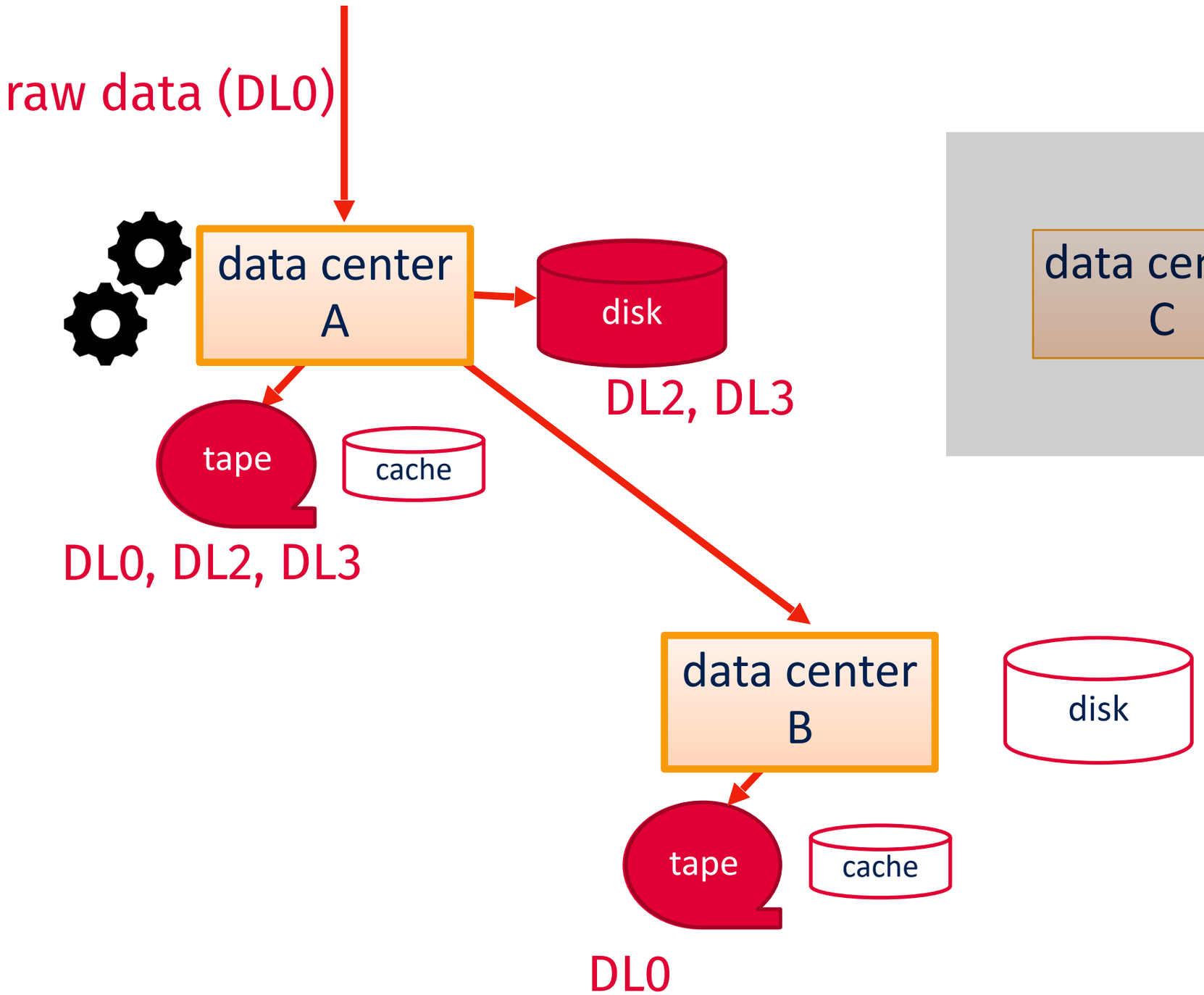


Off-site ICT

Data Flow for Raw Data Processing

simplified view

Observation night 1 from site 1

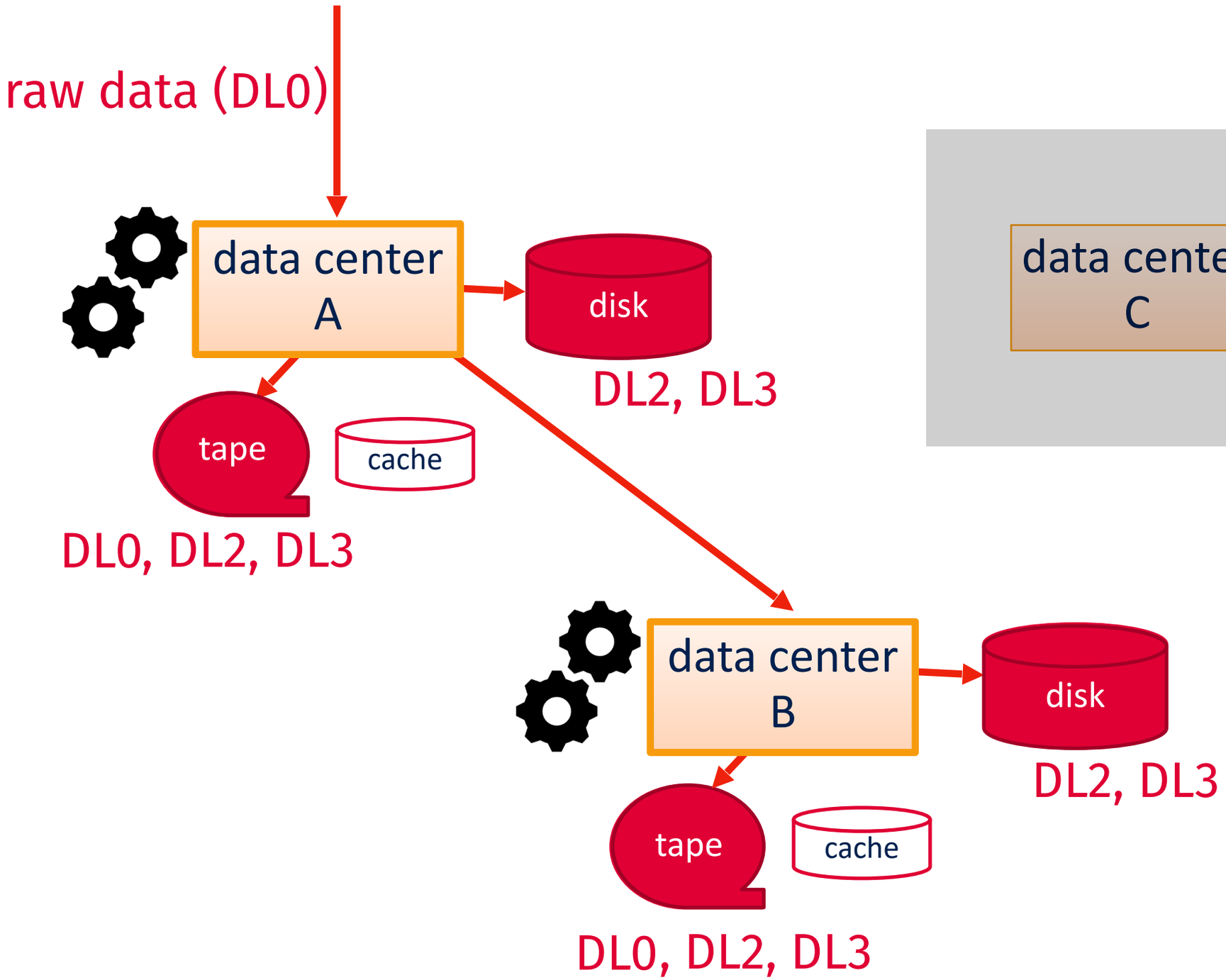


Off-site ICT

Data Flow for Raw Data Processing

simplified view

Observation night 1 from site 1

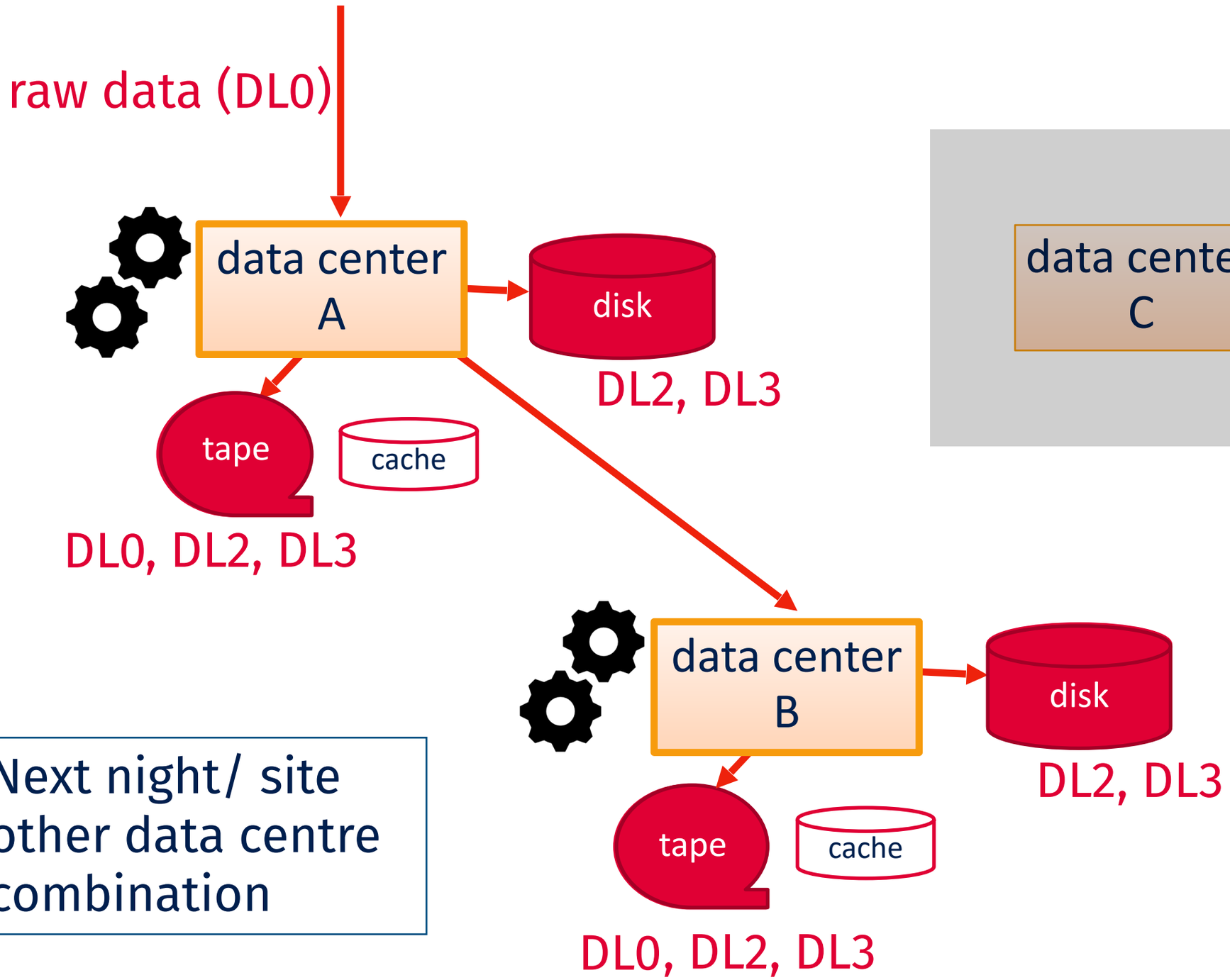


Off-site ICT

Data Flow for Raw Data Processing

simplified view

Observation night 1 from site 1

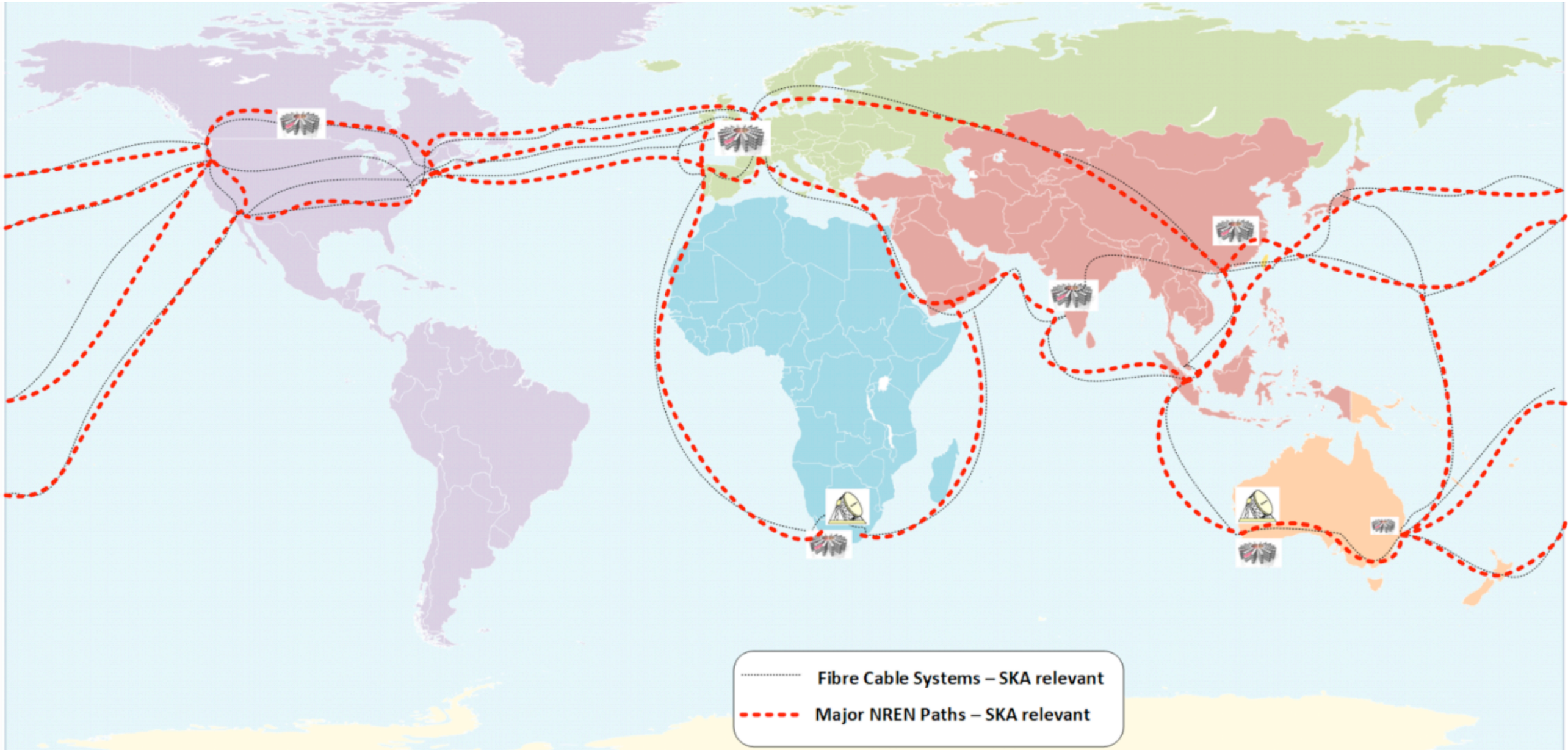


Next night/ site other data centre combination

Square Kilometre Array Observatory

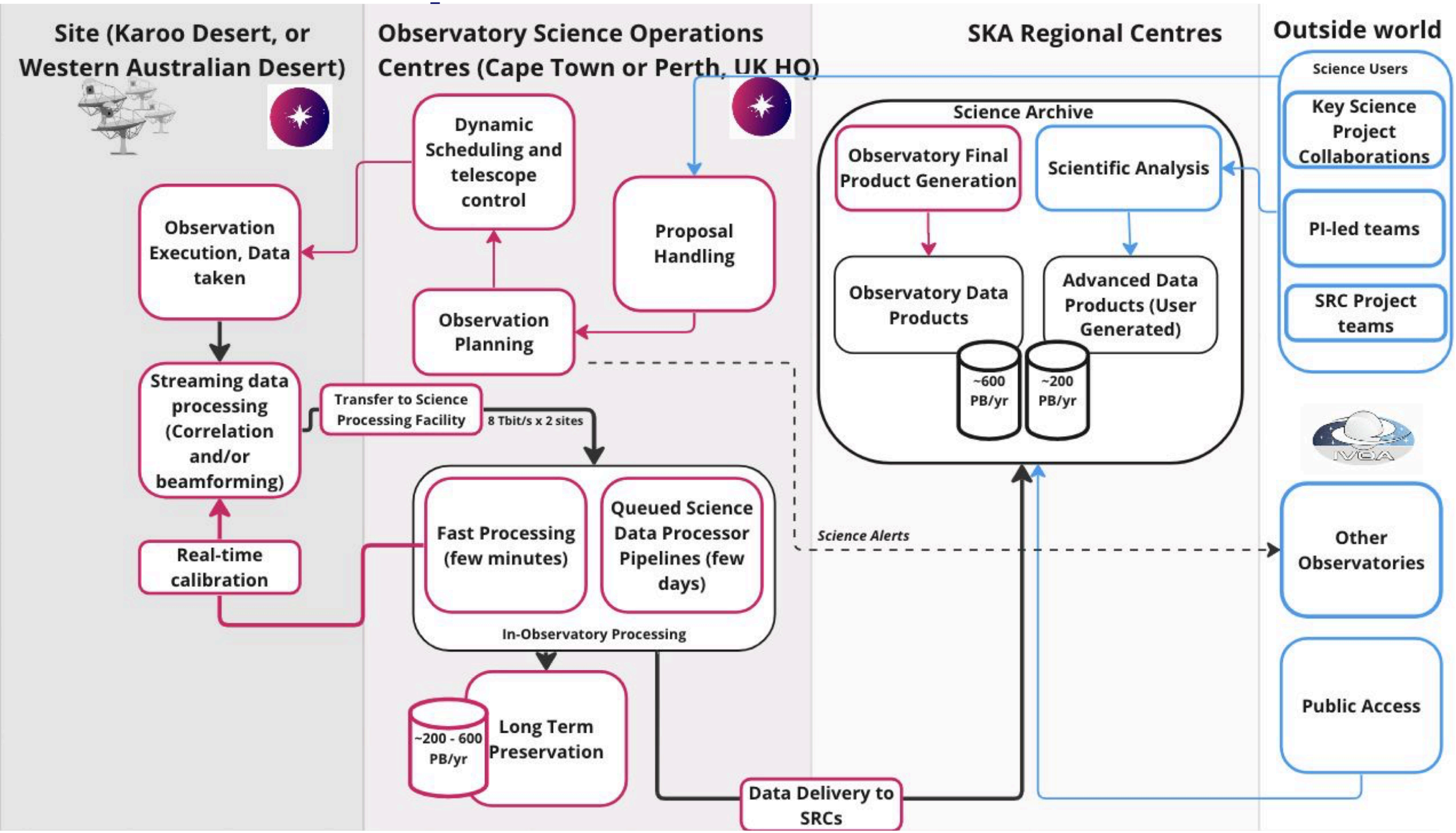
Worldwide Data Transport

SKAO Global Headquarters in the UK, the SKAO's two telescopes at radio-quiet sites in South Africa and Australia, and associated facilities to support the operations of the telescopes



SKAO

Science Operations Flow



R Bolton, CHEP23

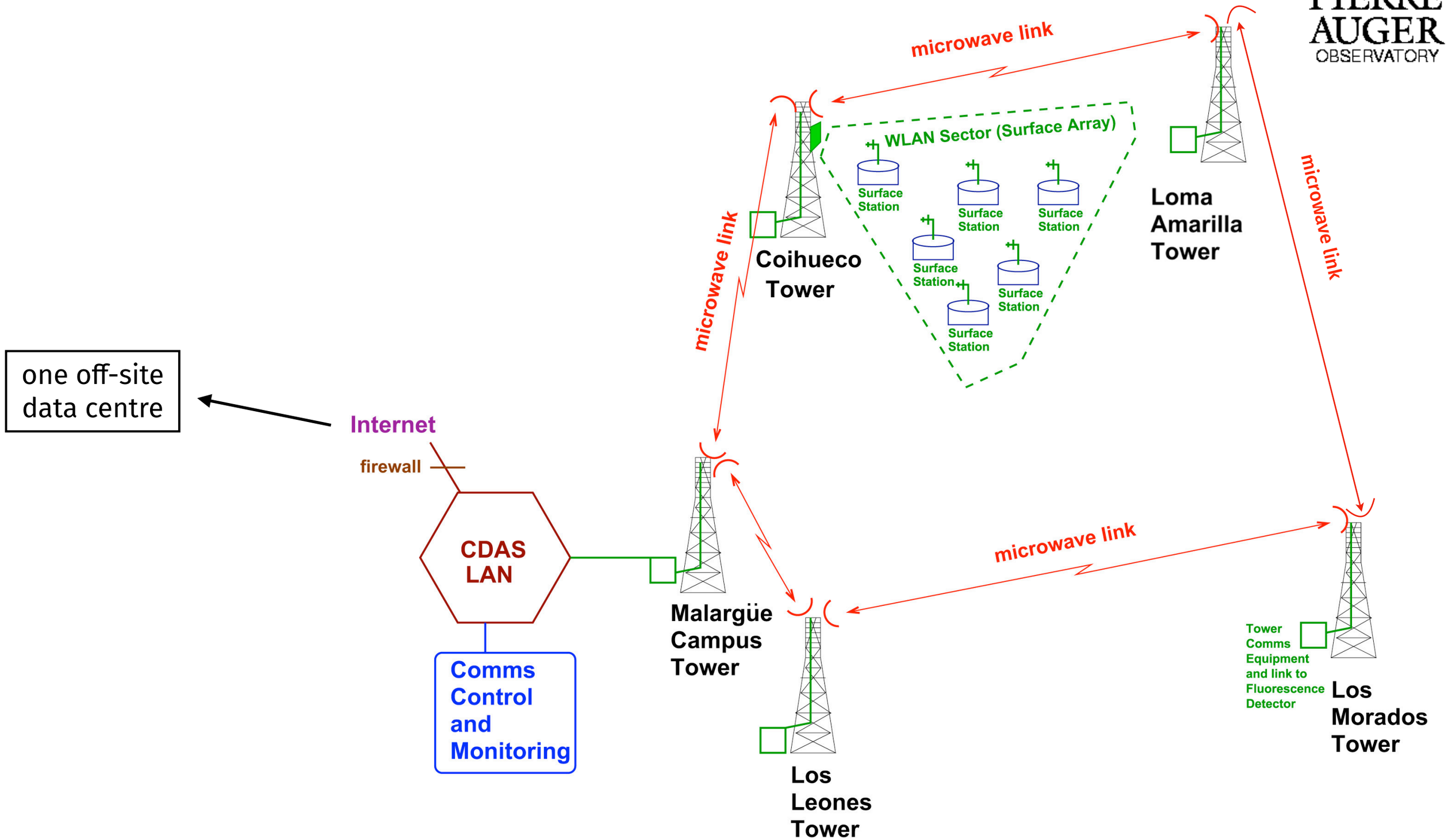
Data Processing

- Central Processing Facilities **on-site** with the telescopes
  - wide range of processing and support equipment
- **Science Processing Centres** - “close” to the sites
  - Use of supercomputers
- A global network of SKA **Regional Centres**
  - shared computational resources make data products available
  - ensure that scientists can access data products → make discoveries
  - SRCNet: a collaborative, interoperated ecosystem
    - support users (and their workflows) → sent them "to the data"
    - provide access to data products, platforms for advanced scientific analysis, and user support and training for astronomers
- Co-located data scenario
  - Plan data placement ahead of time
  - Users could be allocated to different SRCs based on data location
- Performed SKA Science Data Challenges to test the workflows

skao.int; R Bolton, CHEP23

Pierre Auger Observatory

Data Flow



E Santos, I Allekotte; astro-ph arXiv:1502.01323

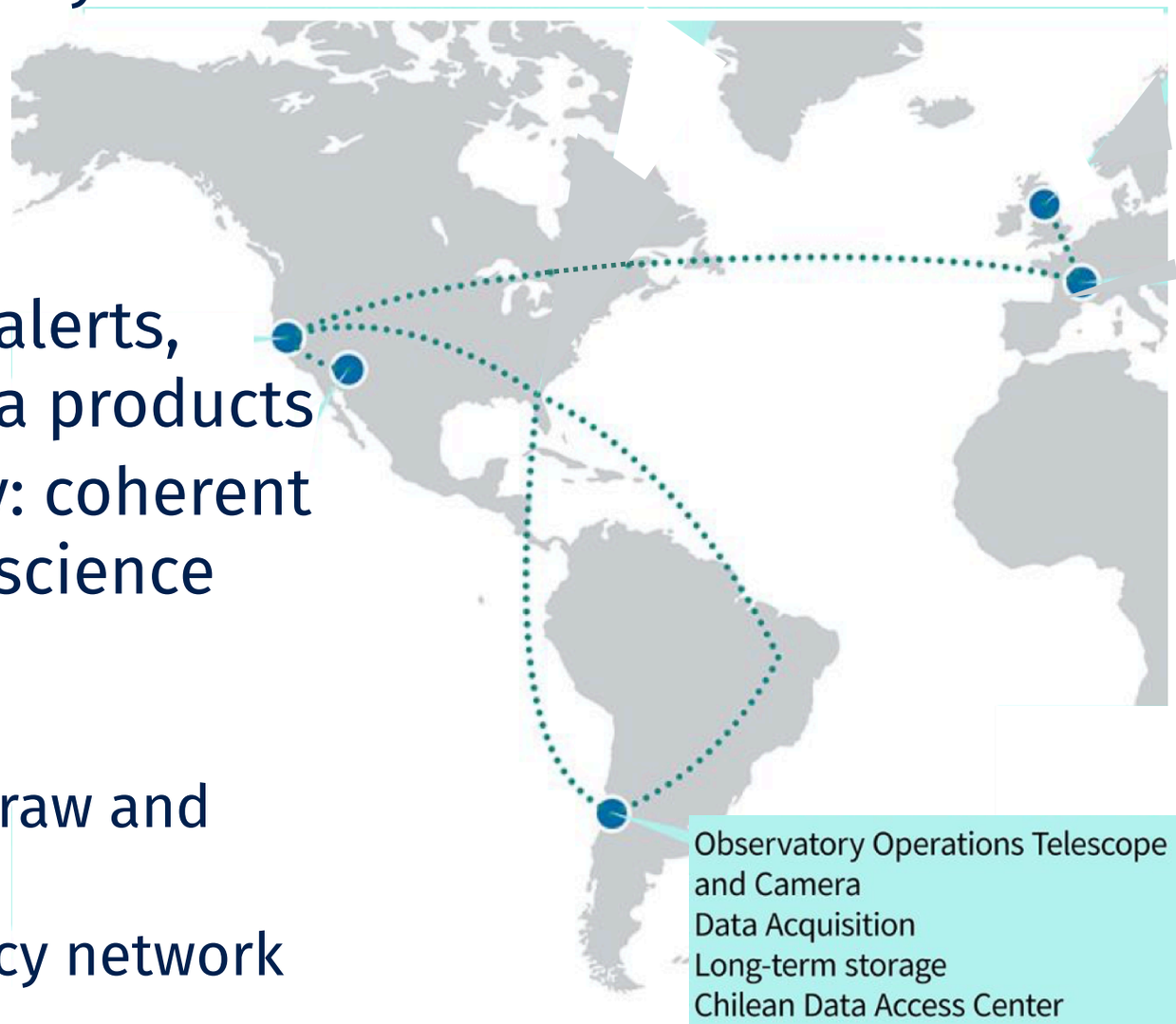
Vera C. Rubin Observatory

Data and Facilities

VERA C. RUBIN OBSERVATORY

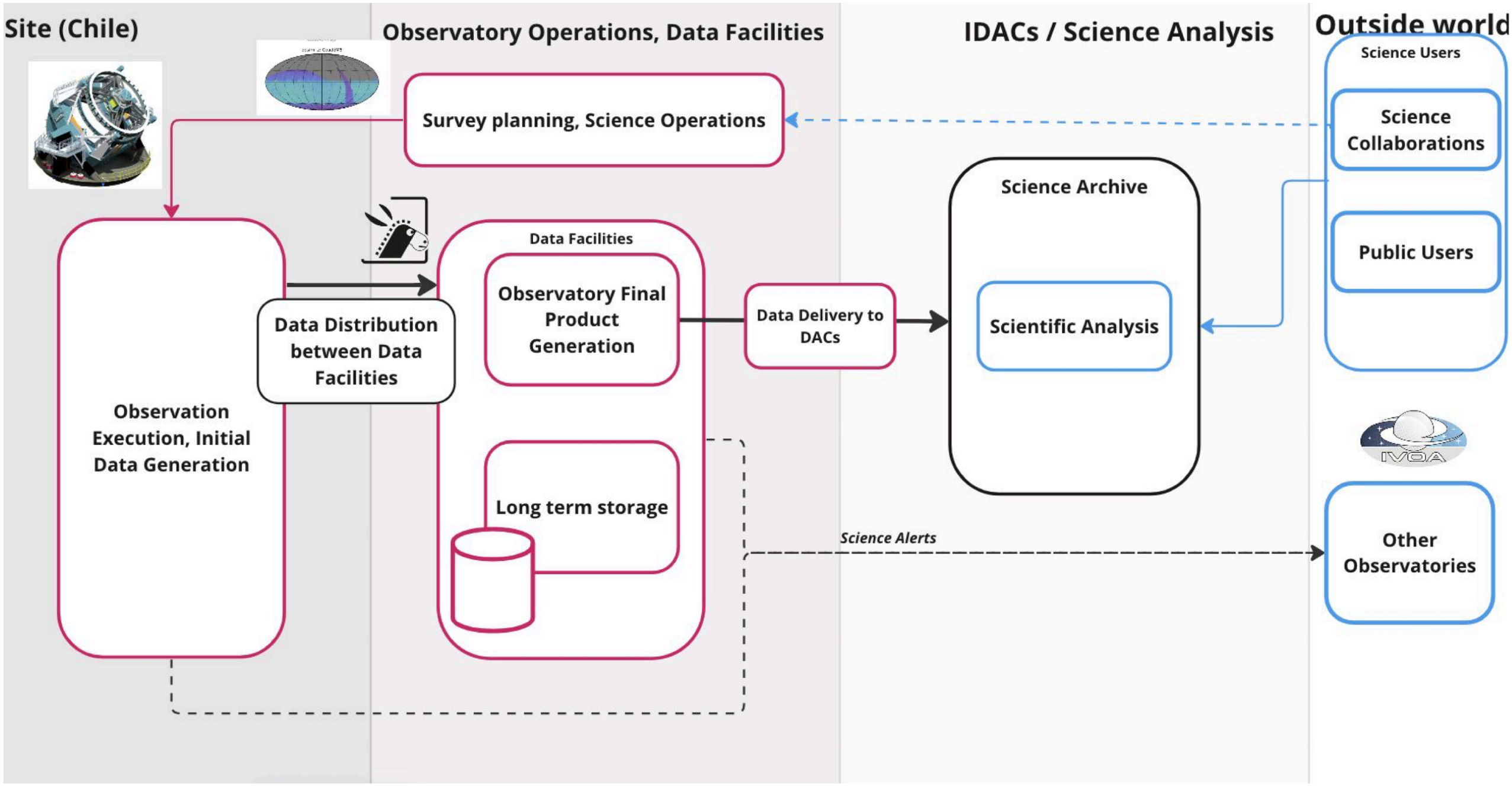
Previously Large Synoptic Survey Telescope

- Raw images → alerts & science-ready images & astronomical catalog → science collaborations
- Data Products:
  - Prompt data products: include alerts, source catalogs, and image data products
  - Release data products annually: coherent image processing of the entire science data set
    - performed at three data centres
    - one facility stores a full copy of raw and published data products
    - data replication over high-latency network
  - User Generated data products
- The Analysis Facility: catalog database and the Rubin Science Platform for interactive analysis



Vera C. Rubin Observatory

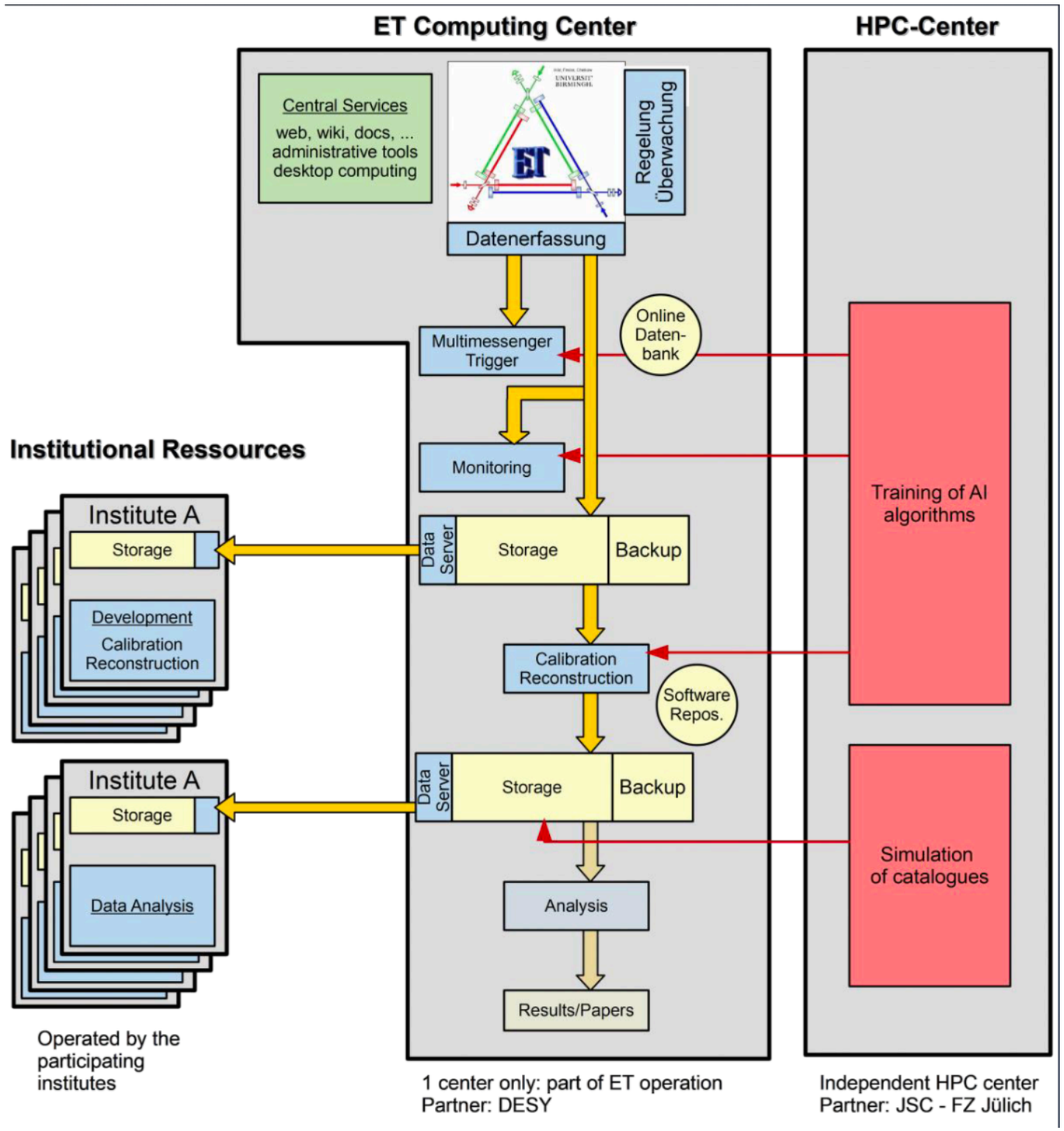
Science Operations Flow



R Bolton (F Hernandez), CHEP23

Einstein Telescope

Computing Concept



- Focus shifting from discovery physics towards observation and MM astronomy
- Online (site) center: Data acquisition and pre-processing, Instrument control
- Different options for offline analysis as distributed computing: Deep searches, parameter estimation etc
- Low-latency domain for Candidate search, Alert Generation services etc

S Bagnasco, 2021; A Stahl, 2022

The Astronet Science Vision & Infrastructure Roadmap 2022-2035

A few key recommendations for computing and data management

- Mission and facility planning should integrate plans for the production of science-ready data products and analysis tools
- Adopt and further develop a “tiered” approach for Data Infrastructure for all types of data including models, simulations, mocks
- Work towards a fully collaborative, open and synergistic view of the computing ecosystem, including data, software, analysis, simulations…
- Professional software engineering / computational skills base
- Engage with, and participate in, EOSC and other Open Science initiatives
- Promote, adopt principles of FAIR, Open data and open-source software
- Pursue long-term investment in state-of-the-art archives for storage, exploration and exploitation of data associated with missions/ facilities
- Beyond data access: concentrate on providing data analysis tools and resources on the site, where the data is stored

ISBN: 978-1-3999-5162-3

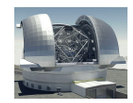
Interaction of Users with Observatories

Different Aspects

- User accounts (register, A&A)
- Allocation of observation time (proposals, KSPs, ToO, DDT, guaranteed…)
- Access and Data Access policies
  - access to raw data or to high-level data products
  - full data sets public/ open – or a fraction
  - Data Sharing: proprietary periods or (near-)immediate data release
  - Licences
- Standards e.g.
  - Data models, data formats, metadata
  - tools used by most experiments, be it programs or libraries
- Science Platforms
  - centralized or distributed computing
  - exchange of MM/MWL information (alerts, campaigns … catalogues)
- Science Analysis Tools and workflows (e.g. in notebooks)
- User Support and Collaboration Tools

Multimessenger Analysis

A **Multimessenger Ecosystem** will include a large number of existing and future facilities that will produce and consume alert triggers:

- Ground-based gravitational wave detector like ET EINSTEIN TELESCOPE
- Ground-based optical telescopes like the ELT in Chile 
- Large radiotelescope arrays like SKAO
- Facilities for cosmic-ray astronomy like the cta
- Cherenkov telescopes for highest-energy gamma-ray astronomy, in the Canary Islands and Chile like cta, and in Argentina Pierre Auger PIERRE AUGER OBSERVATORY
- Neutrino detectors like KM3NeT

Key components: Standards, Computational Infrastructure, Collaboration and Coordination, Data Sharing and Open Science

S Pardi, CHEP2023

Federated Computing

Federated computation moves the computation to the data and privately aggregates the results of those computations. [Informationweek]

A collection of realms (domains) that have established trust among themselves [...] typically includes authentication and may include authorization. [NIST]

The key difference between federated computing and traditional distributed computing lies in the data privacy and ownership aspects. [OpenAI]

Federation

In Astroparticle Physics

Use cases

- Access for scientists in different organisations (and the public) to data from different archives and bring the data together for MM studies
- Finding software and/or workflows
- Launching interactive analyses

Topics

- Federation of Authentication and Authorization Infrastructure (AAI)
- Federation of data lake
  - Data Infrastructure for Open Science (ESCAPE)

European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures

- Common APIs to batch computing
- Federation in ESFRI Science Analysis Platform (ESAP)
- Interoperability in International Virtual Observatory Alliance (IVOA)
- Federation in Open-source Scientific Software and Service Repository (OSSR) - an ESCAPE service

Open Science
FAIR Principles

Open science: unhindered access to scientific articles, access to data from public research, and collaborative research enabled by ICT tools – Research Institutions, Funding Agencies, Scientific Societies, Governments

Data Sharing (EOSC) – open and seamless access to services to

- analyze data
- re-use data and digital objects like software and publications

open-research-europe.ec.europa.eu;oecd.org; eosc.eu; go-fair.org

Open Science
FAIR Principles

Open science: unhindered access to scientific articles, access to data from public research, and collaborative research enabled by ICT tools – Research Institutions, Funding Agencies, Scientific Societies, Governments

Data Sharing (EOSC) – open and seamless access to services to

– analyze data

– re-use data and digital objects like software and publications

Findability, Accessibility, Interoperability, Reusability

- Metadata and data easy to find for both humans and computers
- Accessed of data for users - including authentication and authorisation
- Integrate data from several sources.
- The data need to interoperate with applications or workflows for analysis, storage, and processing
- Reuse of data: well-described metadata and data that they can be replicated and/or combined in different settings

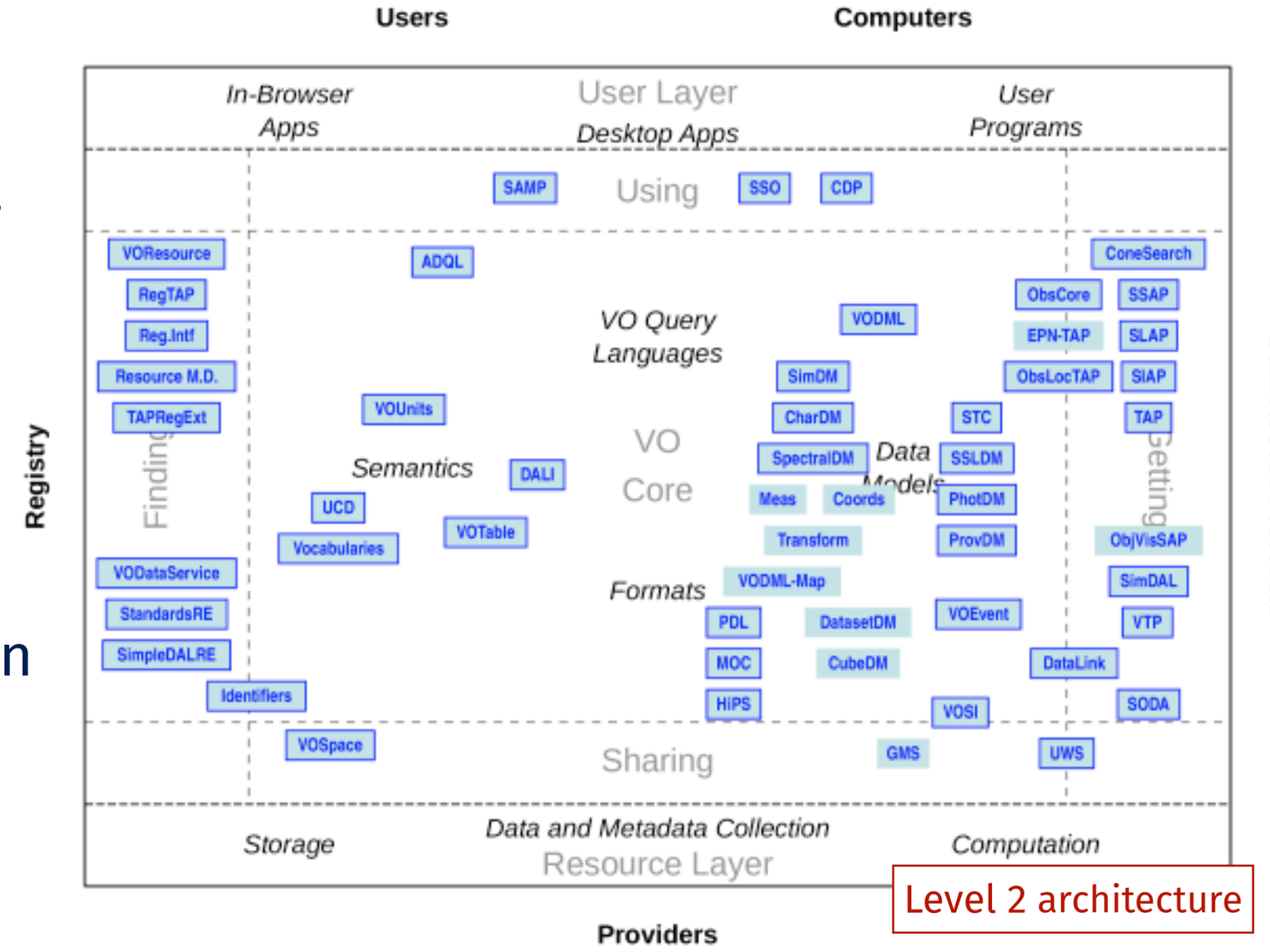
open-research-europe.ec.europa.eu;oecd.org; eosc.eu; go-fair.org

International Virtual Observatory Alliance

IVOA

Mission:"Facilitate the international coordination and collaboration for the development and deployment of the tools, systems and organizational structures necessary to enable the international utilization of astronomical archives as an integrated and interoperating virtual observatory.

- IVOA comprises 22 VO programs: countries, EU, ESA
- Develop and agree on key interoperability standards and technologies
- Federated Authentication
- VO FAIR standards
- Provide VO services and tools



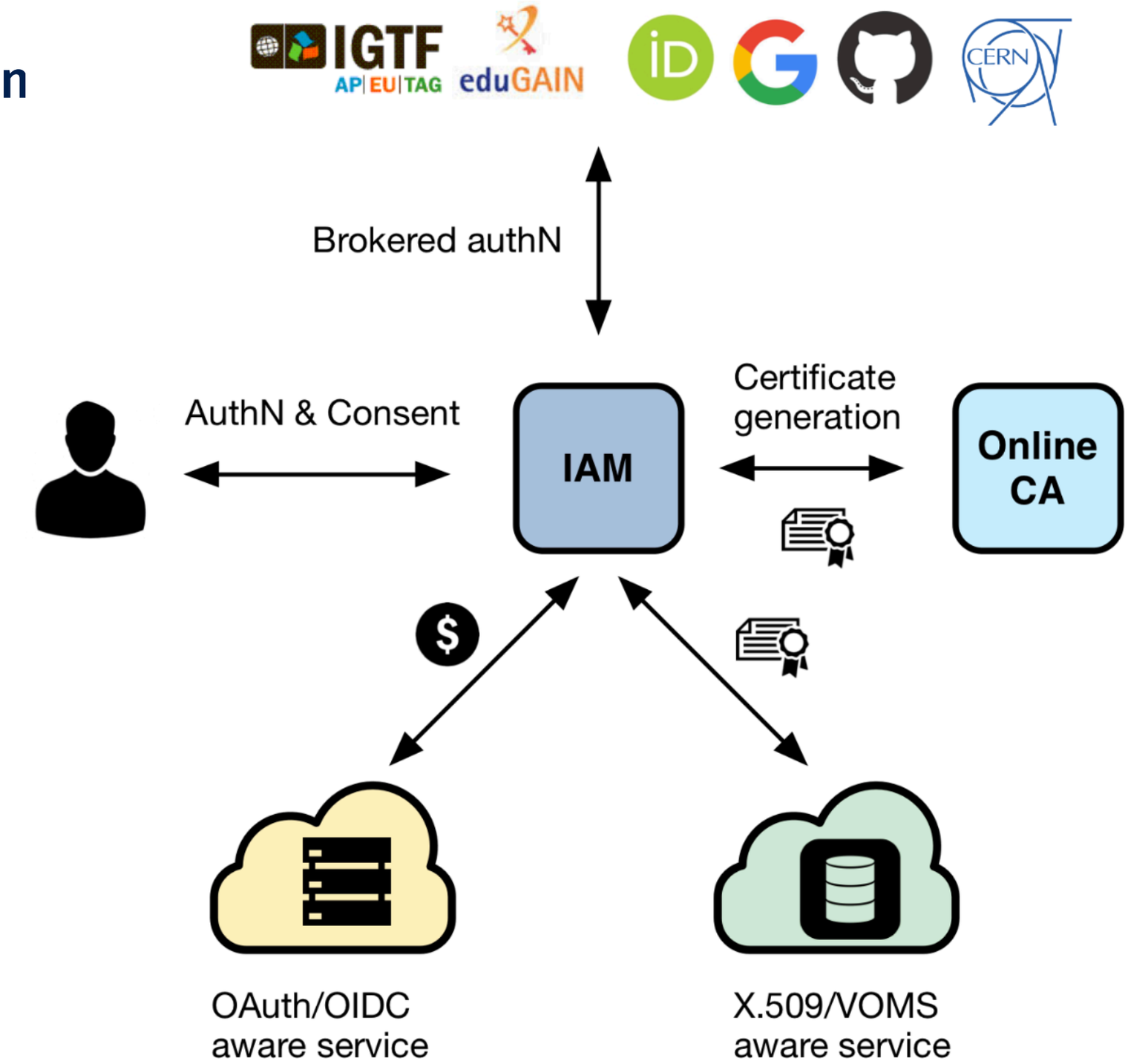
Level 2 architecture

Identity and Access Management Service

IAM

A VO-scoped authentication and authorization service that

- supports multiple authentication mechanisms
- can integrate existing Service aware of Virtual Organization Membership
- supports Web and non-Web access, delegation and token renewal
- manages Group membership



developed in the context of the INDIGO-Datacloud Horizon 2020 project

A Ceccanti, Jan 2021

ESFRI Science Analysis Platform

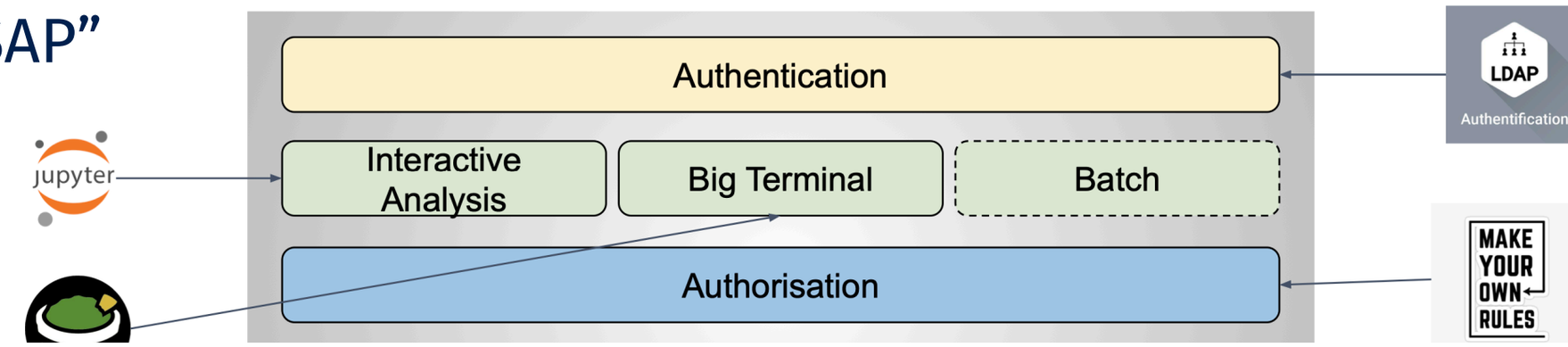
ESAP

Define and implement a flexible science platform for the analysis of open access data available through the EOSC environment

- Access and share your data and results
- Work with advanced analysis tools
- Integrate with a wide variety of other services

ESAP: a **focal point for integrating diverse services** which are drawn from other providers

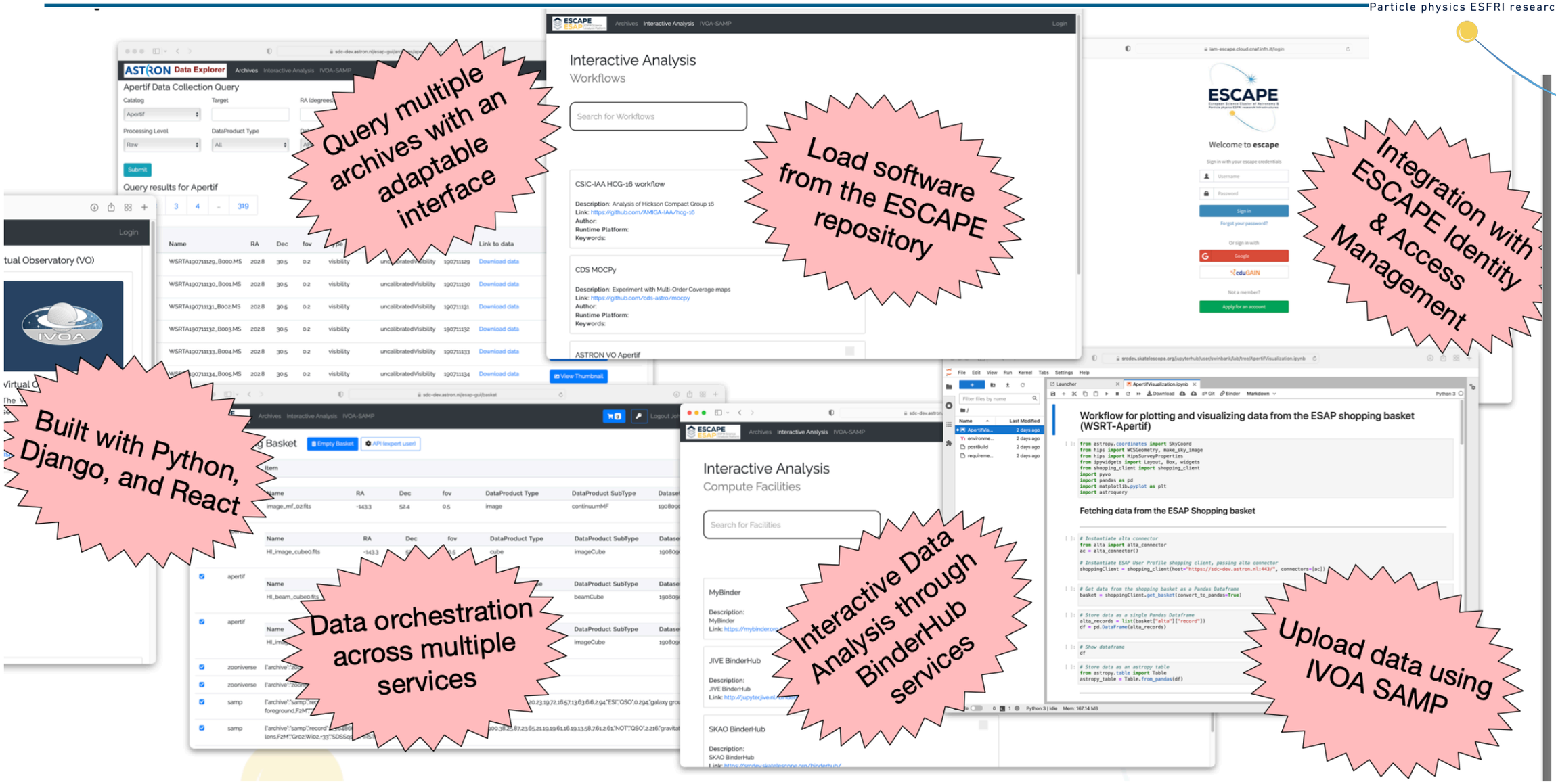
- Technically: web UI, and API gateway
- …with a range of pluggable, independent services
- A *toolkit* for building science platforms on different scales
  - "Centralized ESAP"
  - "Project ESAP"



projectescape.eu/esap; J Swinbank, Nov 2022, ESCAPE to the Future, Oct 2022; J Salgado, May 2023

ESAP

Example Capabilities



SKA (SRC) view

Users

Science Platform

Interface (API layer)

Open Source Scientific Software and Service Repository OSSR

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Software services for open science data-analysis of the ESFRI facilities

- A common repository for software across various ESCAPE infrastructures
- Designed both to *make software available* to practising scientists and to *preserve* software for future reproducibility.
- Built-in to common CI/CD systems (GitLab, Hub, etc)
- Built around the **Zenodo** system for long- term stability and preservation
- Programmatic access through the eOSSR library

Examples for such software: Auger OffLine and gammapy

Key outputs: A sustainable open-access repository, which will be dynamically enhanced and maintained by ESCAPE ESFRI projects and included in the EOSC catalogue

projectescape.eu/ossr; J Swinbank, Nov 2022

Federation

In Astroparticle Physics

Use cases

- Access for scientists in different organisations (and the public) to data from different archives and bring the data together for MM studies
- Finding software and/or workflows
- Launching interactive analyses

Topics

- Federation of Authentication and Authorization Infrastructure (AAI)
- Federation of data lake
  - Data Infrastructure for Open Science (ESCAPE)

European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures

- Common APIs to batch computing
- Federation in ESFRI Science Analysis Platform (ESAP)
- Interoperability in International Virtual Observatory Alliance (IVOA)
- Federation in Open-source Scientific Software and Service Repository (OSSR) - an ESCAPE service