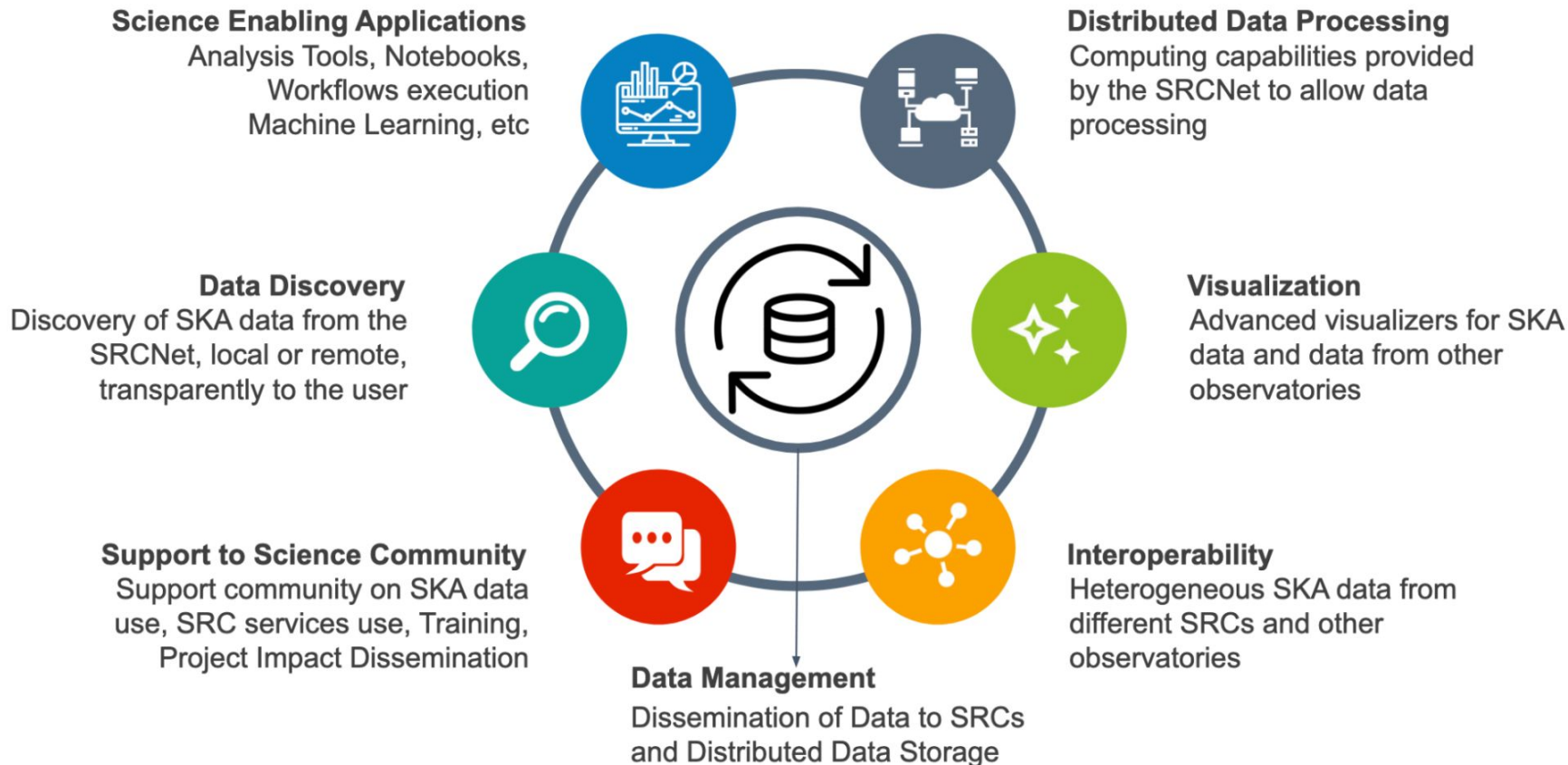SRCNet Coral update

SRCNet Infrastructure updates for CHSRC

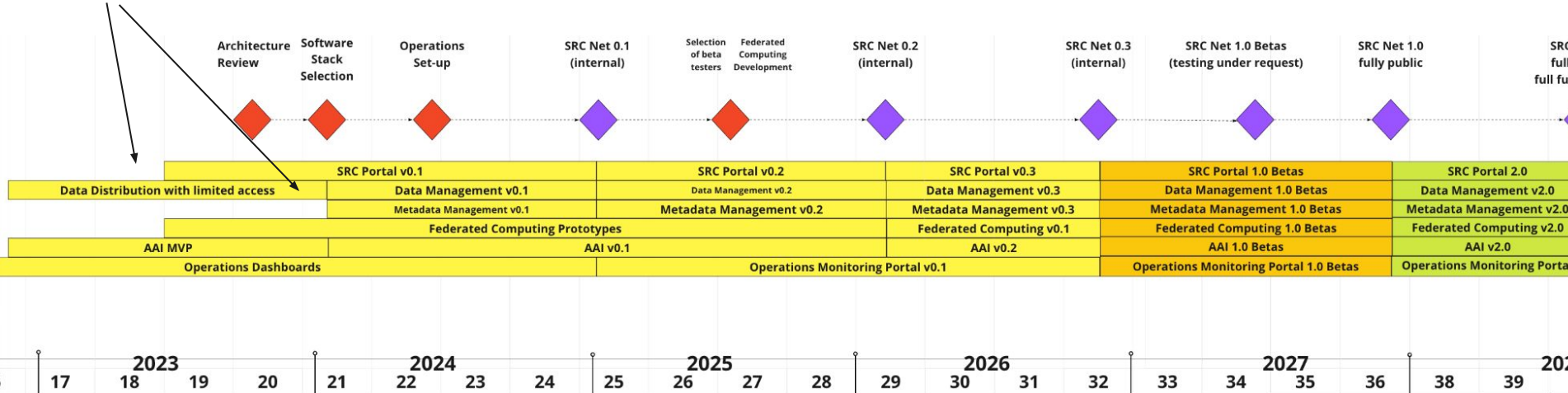Swiss SKA Days - September 6th, 2023
Carolina Lindqvist and Pablo Llopis

SRCNet Objectives

Science Enabling Applications
Analysis Tools, Notebooks, Workflows execution Machine Learning, etc

Distributed Data Processing
Computing capabilities provided by the SRCNet to allow data processing

Data Discovery
Discovery of SKA data from the SRCNet, local or remote, transparently to the user

Visualization
Advanced visualizers for SKA data and data from other observatories

Support to Science Community
Support community on SKA data use, SRC services use, Training, Project Impact Dissemination

Interoperability
Heterogeneous SKA data from different SRCs and other observatories

Data Management
Dissemination of Data to SRCs and Distributed Data Storage

SRCNet Roadmap

Focus on Data Management Solutions (Data and Metadata)

Coral Team - Collaborating Institutions

Coral Team - Purpose and Goals
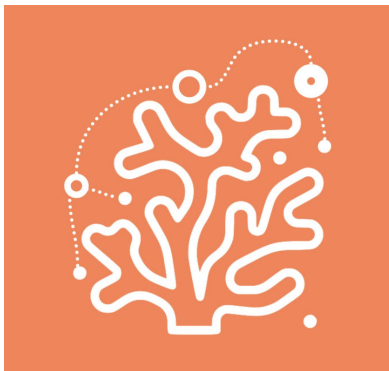
🔍 **Identifying** computing **architectures** and **resources** available in SRC sites

🧑‍💻 **Prototyping**

- ○ **Data** management testbed
- ○ **Auth**entication and Authorisation Infrastructure
- ○ **Science Platform**
- ○ Assessing **performance** in different SRC sites and architectures

🚀 **Impact**: gather and share **knowledge**

🎓 Provide **early feedback** on the adoption of proposed solutions

SRCNet high level topics for PI19 (Coral Team)

- **Data management solutions**

- Building blocks for science platforms

- Collection of workflows

SRCNet high level topics for PI19 (Coral Team)

- **Data management solutions**

  - Supporting Rucio Data Lake operations

  - Storage Inventory and Metadata OpenCADC services

- Building blocks for science platforms

- Collection of workflows

SRCNet high level topics for PI19 (Coral Team)

- Data management solutions

- **Building blocks for science platforms**

- Collection of workflows

SRCNet high level topics for PI19 (Coral Team)

- Data management solutions

- **Building blocks for science platforms**

  - Jupyterhub

  - Dask

  - Data lake integrations

- Collection of workflows
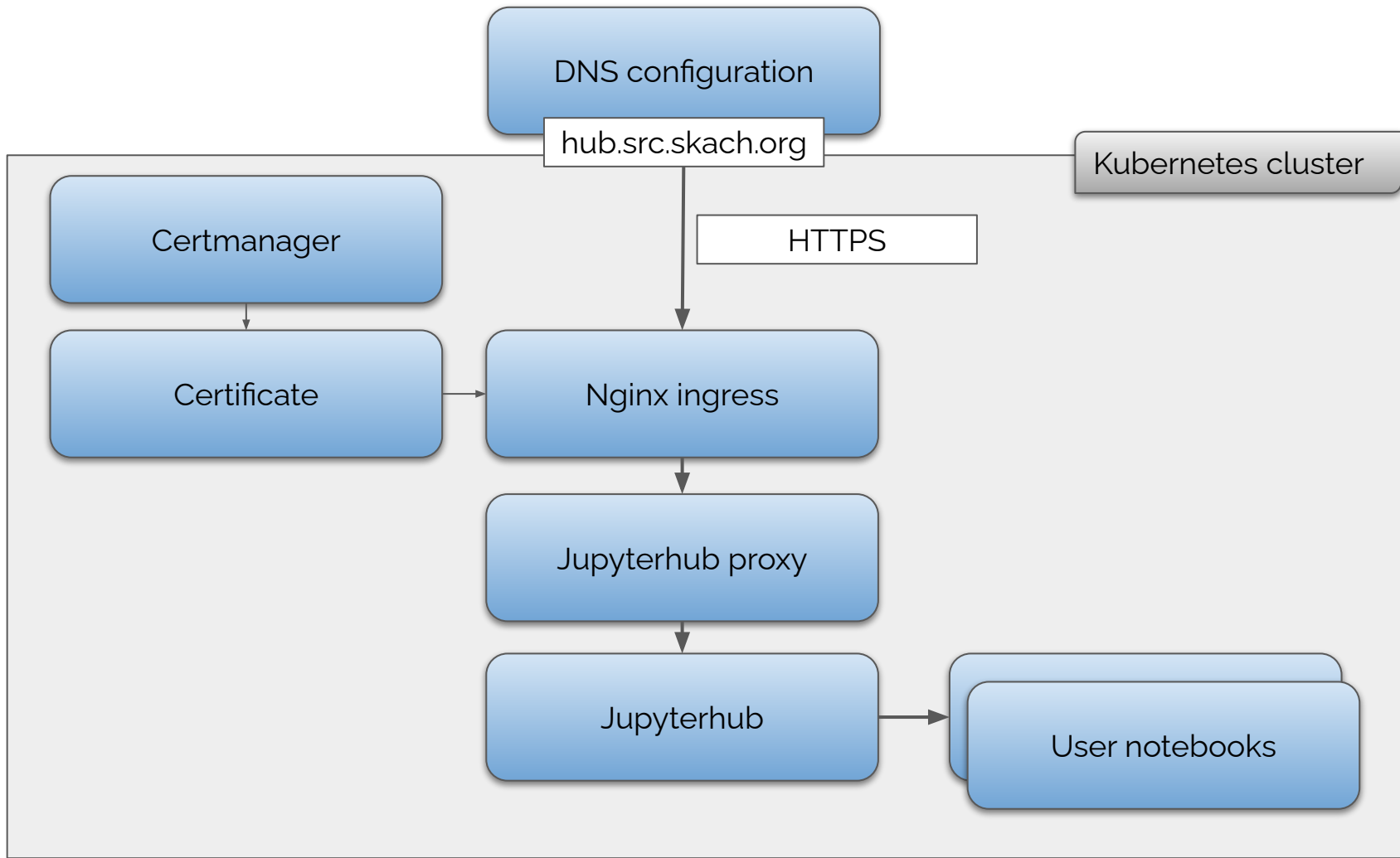
SRCNet high level topics for PI19 (Coral Team)

- Data management solutions

- Building blocks for science platforms

- **Collection of workflows**

SRCNet high level topics for PI19 (Coral Team)

- Data management solutions
- Building blocks for science platforms
- **Collection of workflows**
  - Offer a "shopping list" of reference workloads, organised by category.
  - Integrate with CI/CD to run workloads as a pipeline
    - SRCNet system validation
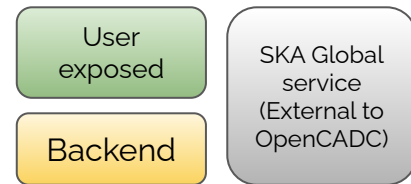   - Performance and regression testing
  - https://gitlab.com/ska-telescope/src/src-workloads

SKA SDC JupyterHub prototype deployment

- Based on a [Helm chart](#) provided by Magenta team

- DNS (domain name) configured at external service provider

- Appropriate configuration for SKACH infrastructure

    - All services running in a Kubernetes cluster

    - Separated by Kubernetes namespaces

    - Enabling HTTPS with [Let's encrypt](#) certificate and [Certmanager](#) deployment

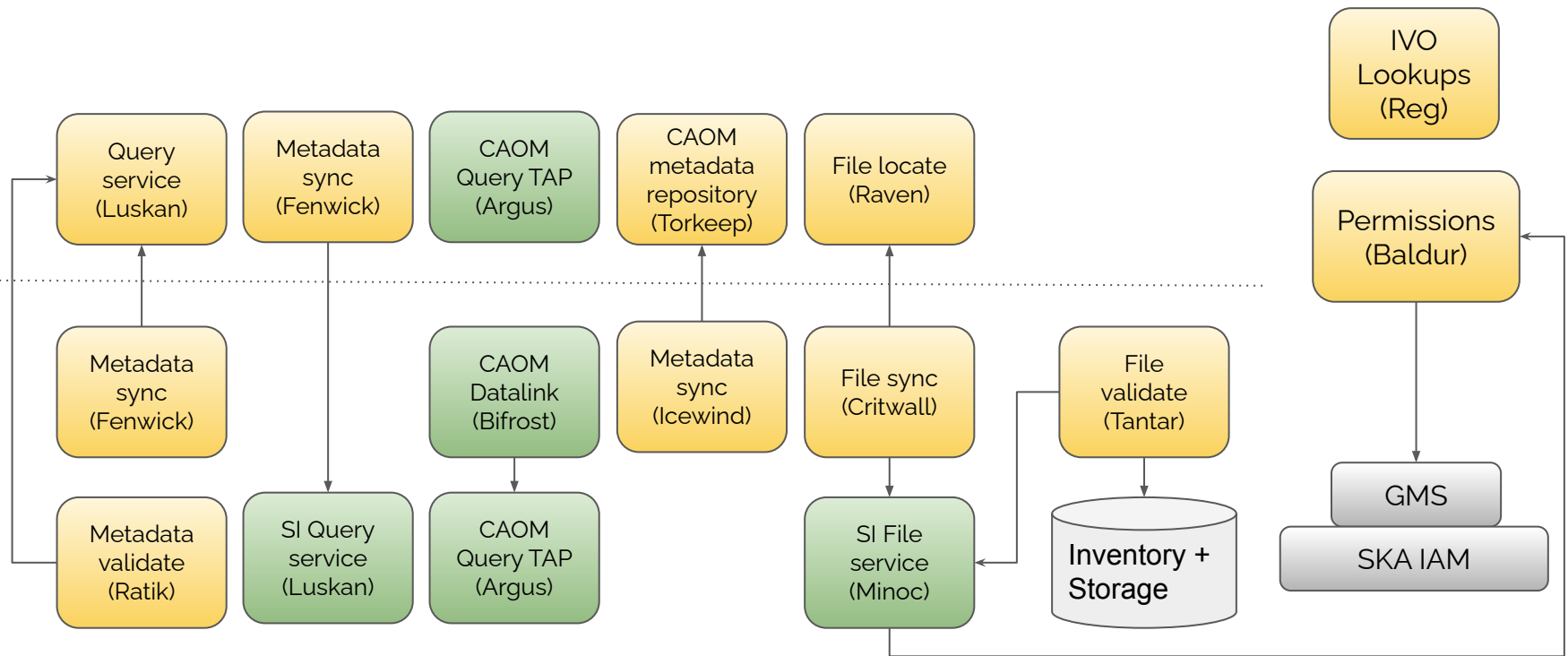   - [Ingress Nginx](#) deployment and configuration

OpenCADC Storage Inventory (SI)
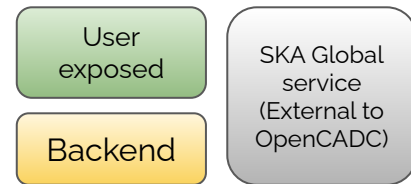# Mini-SRCNet prototype: A bird's eye view (All)

User exposed

Backend
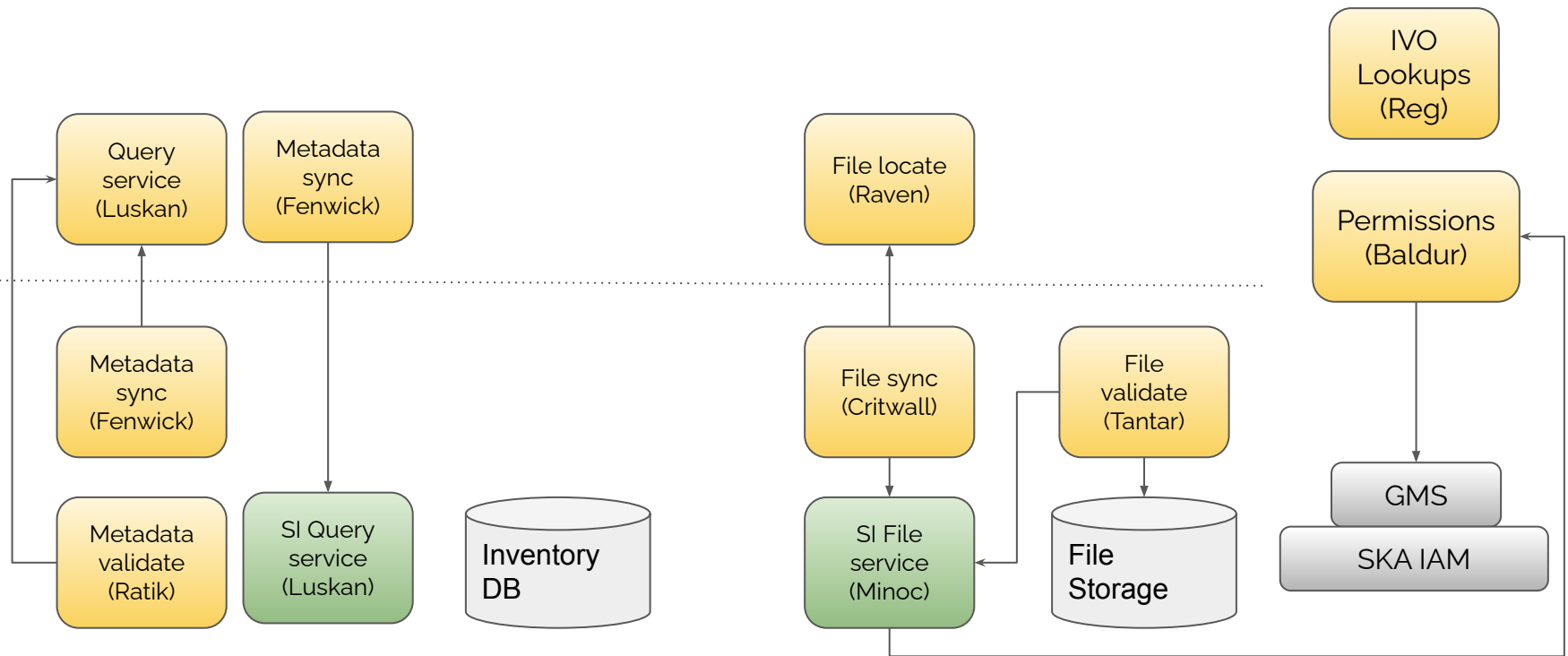
SKA Global service (External to OpenCADC)

Storage Inventory Metadata

Storage Inventory Data

Global IVO Abstr.

Global Site

Local Site

Query service (Luskan)

Metadata sync (Fenwick)

CAOM Query TAP (Argus)

CAOM metadata repository (Torkeep)

File locate (Raven)

IVO Lookups (Reg)

Permissions (Baldur)

Metadata sync (Fenwick)

Metadata validate (Ratik)

SI Query service (Luskan)

CAOM Datalink (Bifrost)

CAOM Query TAP (Argus)

Metadata sync (Icewind)

File sync (Critwall)

File validate (Tantar)

SI File service (Minoc)

Inventory + Storage

GMS

SKA IAM

OpenCADC Storage Inventory (SI)
Mini-SRCNet prototype: A bird's eye view (SI only)

User exposed

Backend

SKA Global service (External to OpenCADC)

Storage Inventory Metadata

Storage Inventory Data

Global IVO Abstr.

Global Site

Query service (Luskan)

Metadata sync (Fenwick)

File locate (Raven)

IVO Lookups (Reg)

Permissions (Baldur)

Local Site

Metadata sync (Fenwick)

Metadata validate (Ratik)

SI Query service (Luskan)

Inventory DB

File sync (Critwall)

File validate (Tantar)

SI File service (Minoc)

File Storage

GMS

SKA IAM

storage site: local metadata and data
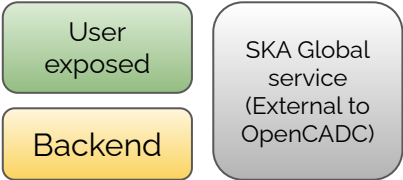
Image credit: https://github.com/opencadc/storage-inventory

OpenCADC Storage Inventory (SI)
Mini-SRCNet prototype: A bird's eye view (CAOM only)

User exposed

Backend

SKA Global service (External to OpenCADC)

Storage Inventory Metadata

Storage Inventory Data

Global IVO Abstr.

Global Site

CAOM2 TAP Query (Argus)

CAOM2 Metadata repository (Torkeep)

File locate (Raven)

IVO Lookups (Reg)

Permissions (Baldur)

Local Site

CAOM Datalink (Bifrost)

Inventory

CAOM Query TAP (Argus)

Metadata sync (Icewind)

GMS

SKA IAM
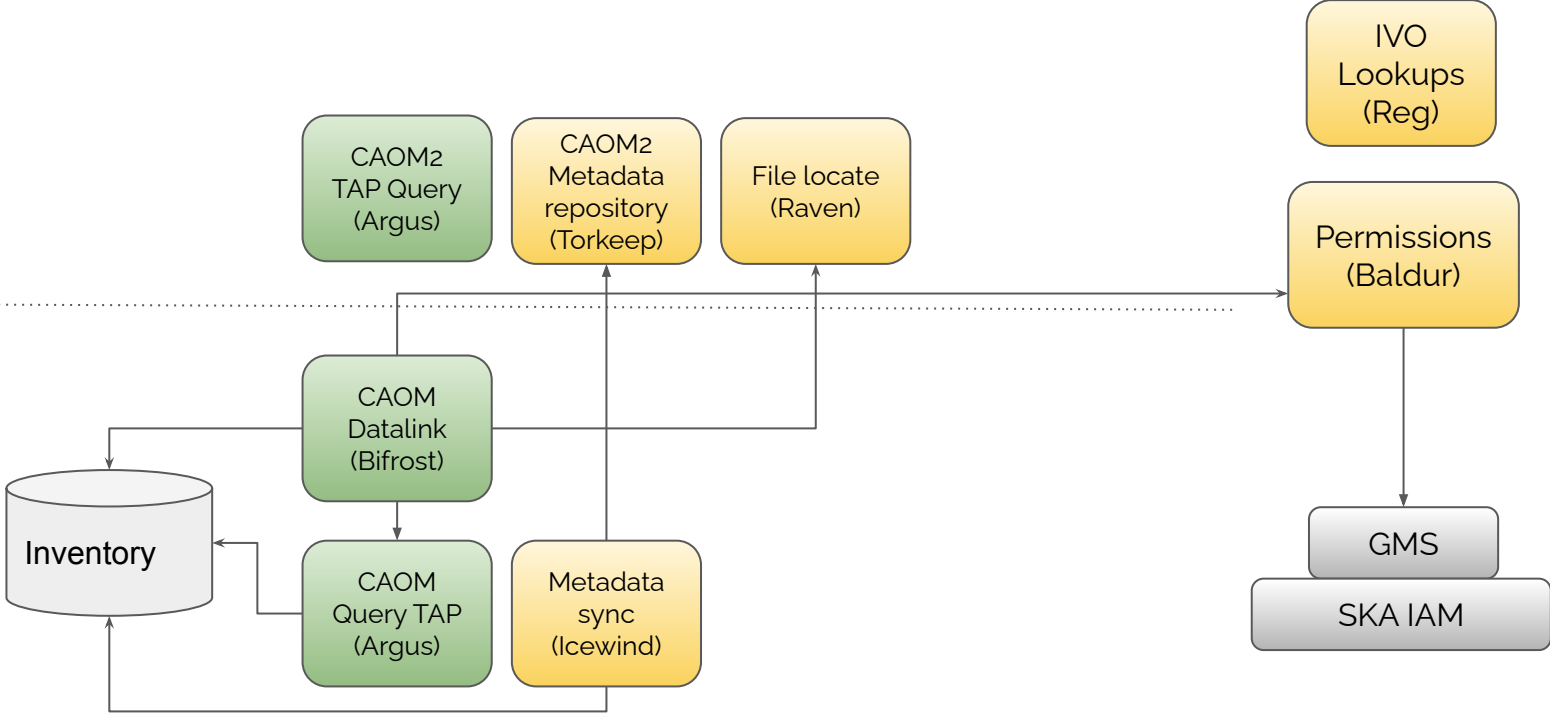
global: metadata curation

«repository»
torkeep

«query»
«TAP»
argus

«database»
postgresql

<<storage>>
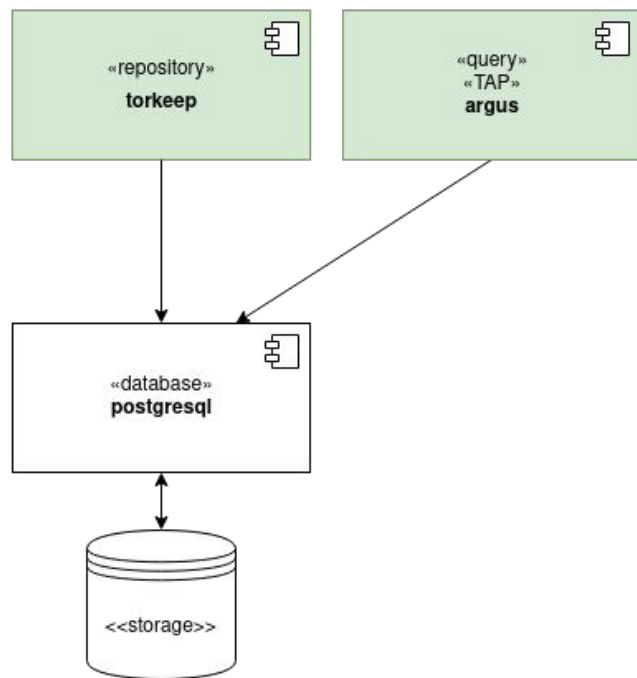
node: metadata and access services

«query»
«TAP»
argus

«datalink»
bifrost

«metadata-sync»
icewind

«metadata-validate»
**TBD**

«database»
postgresql

<<storage>>
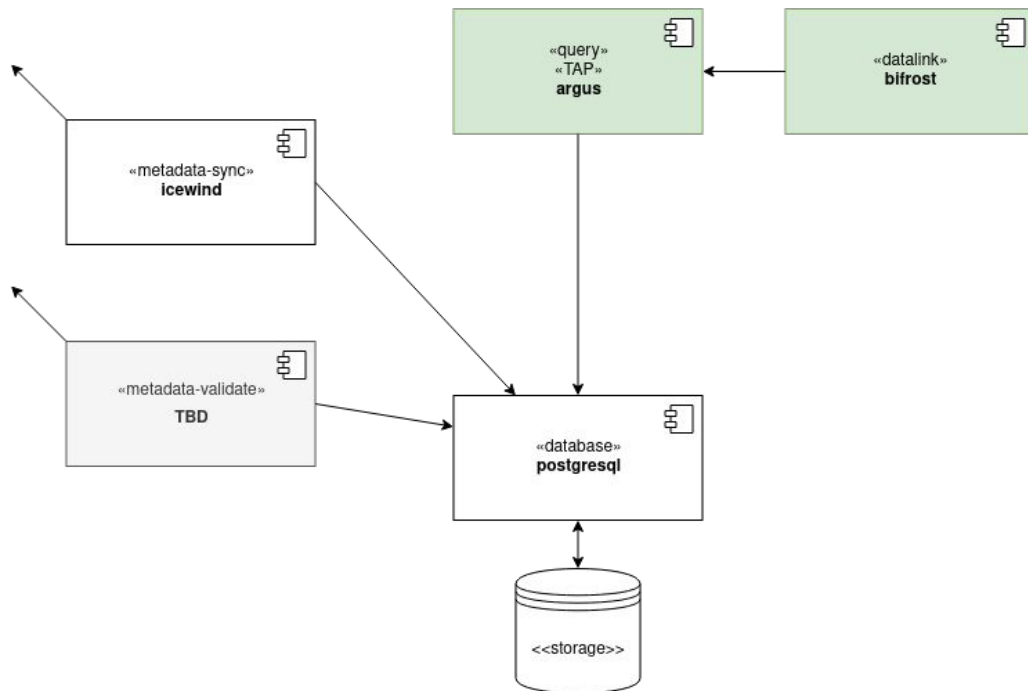
# OpenCADC Storage Inventory (SI) File Synchronization example

1. User PUTs a file to the site1.minoc service, either directly or via negotiation with a global raven service.
2. global.fenwick.site1 discovers the new inventory metadata for the file by querying site1.luskan.
3. site2.fenwick.global discovers the new inventory metadata for the file by querying global.luskan.
4. site2.critwall finds the locations of the new file via global.raven -- this returns a list of URLs from which the file can be downloaded.
5. site2.critwall downloads the file from site1.minoc.

Image credit: https://www.opencadc.org/storage-inventory/ops/
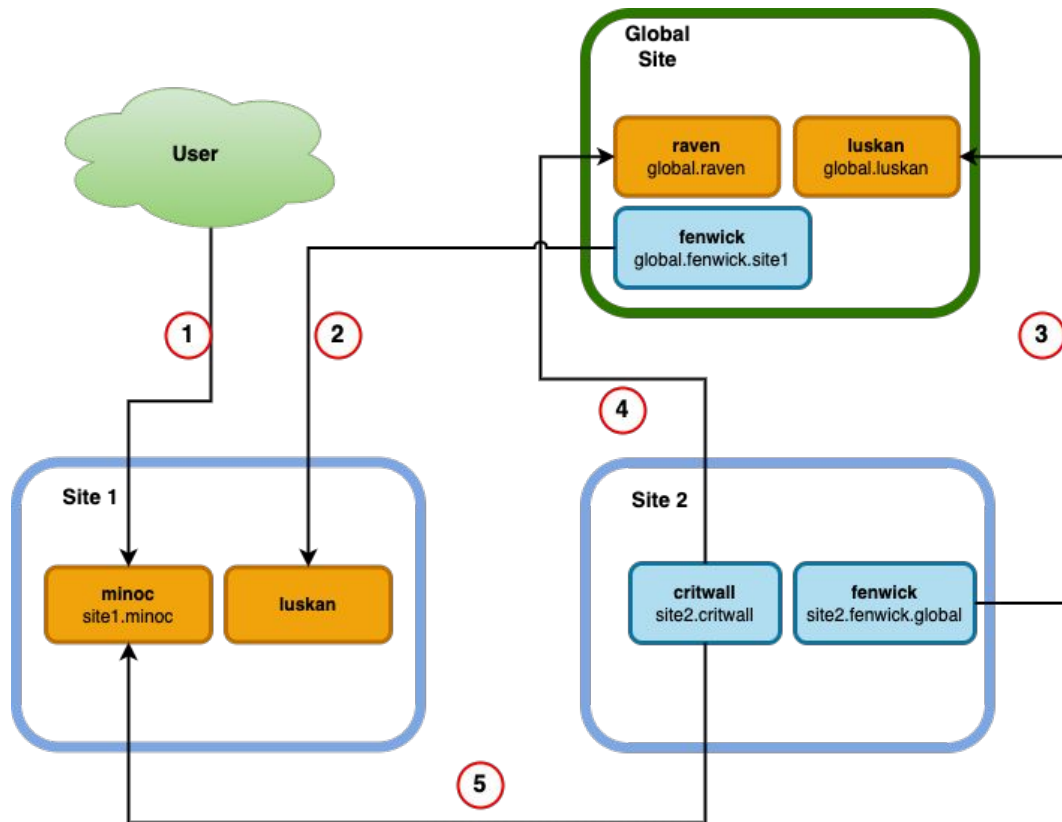
DEMO 1
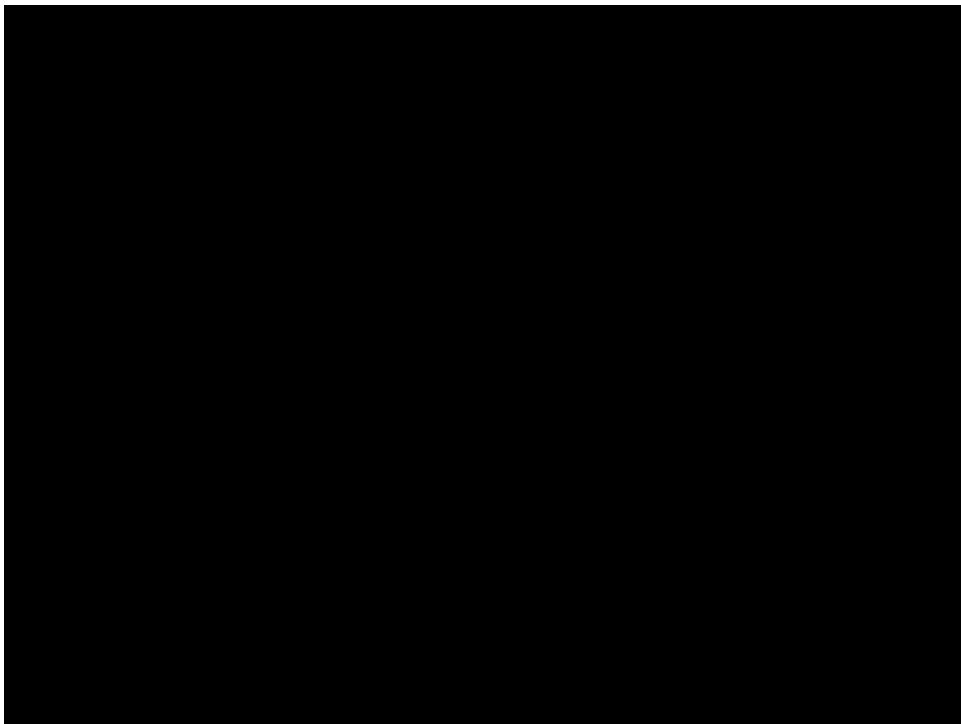
DEMO 2
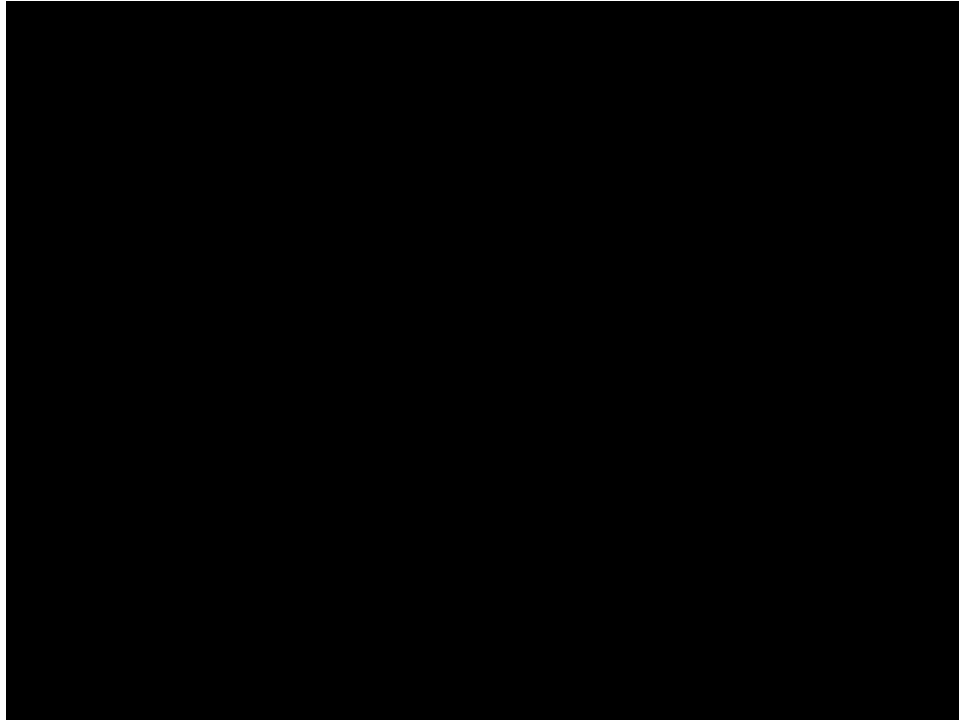
PI20 features

- Mini-SRCNet Science Platform deployment

- Continuation of work for SRCNet workloads repository

- Migration to new CSCS Kubernetes cluster

  - Rucio + Storage Inventory service migration

  - Data migration for Rucio + Storage Inventory

  - JupyterHub, Dask

  - Internal services (monitoring, Secret management system, Registry)

- [Explore Workload Management System solutions](#)

- [Secure Authentication for SRCNet Nodes in Shared HPC Environments](#)