



SKA Regional Centre Data Management Update

Rucio Community Workshop 11-11-22

Rohini Joshi, on behalf of the Data Operations Group

Operations Data Scientist, SKAO



Overview

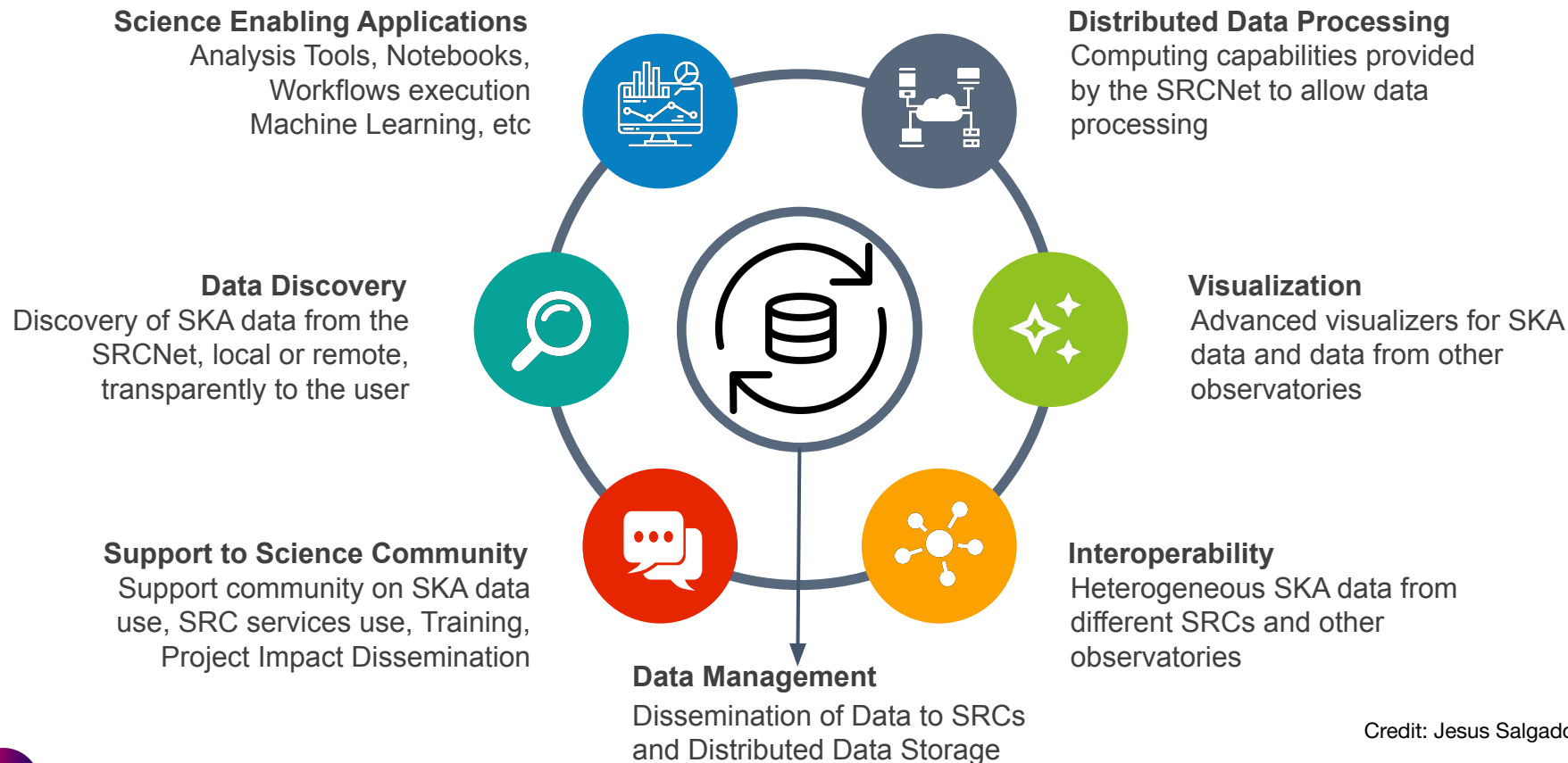
- SKA Regional Centres (SRCs) landscape
- SKAO Rucio prototype: Current status
 - Tokens
 - Metadata
 - Science Data Challenges
- What's next



SKAO data product delivery



SKA Regional Centre Capabilities



Credit: Jesus Salgado

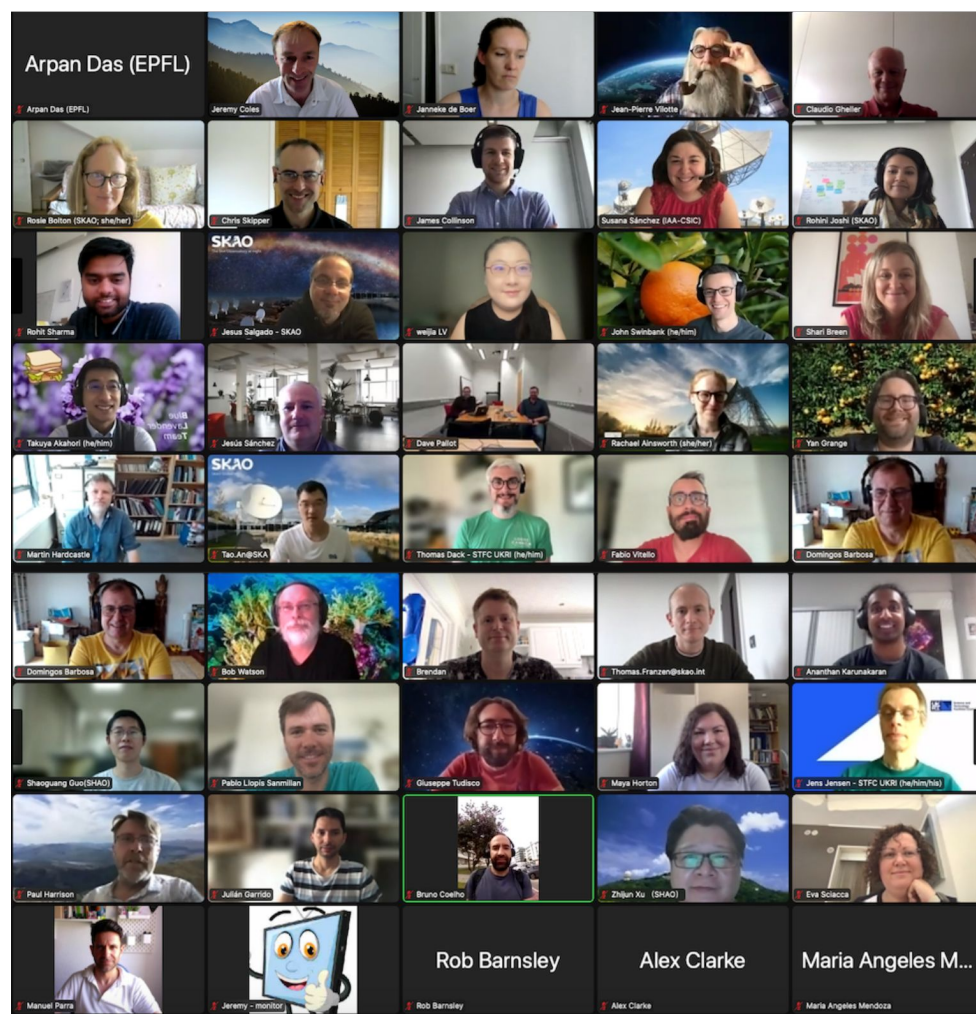


SRC Prototyping

- Prototyping started June 2022
- Team members from 12 countries plus SKAO
- about 1000 developer-days per 3-months
- Anticipate growing as national funding to develop SRC nodes is available

See Rosie's talk at the WLCG workshop for broader context!

<https://indico.cern.ch/event/1162261/contributions/5083367/>



SKAO Rucio Prototype

- Recap of the deployment
 - See [last year's talk](#)
 - <https://gitlab.com/ska-telescope/src/ska-rucio-prototype/>
- Headlines
 - More sites
 - 1.29, End to end token-based auth
 - Metadata development
 - Functional testing every [15 min](#)
 - [Rucio task manager](#) - modular & easily extensible framework for automating common maintenance, functionality tests and health assessment on a Rucio datalake (we welcome contributions!)
 - Monitoring dashboard





General / Rucio Events ☆ 🔗



now-90d to now-10d



Source RSE

All

Destination RSE

All

Protocols

All

Scopes

testing_functional

Filename glob

*

Binning

auto

Min file size for averaged throughput calculations

1000000

2022-08-13 09:42:13

to

2022-11-01 09:42:13

Local browser time

Data Discovery

Overview

TRANSFERS OK

DELETIONS OK

Transfers from source RSE by geographical location



successful transfers

2.90 Mil

failed transfers

719 K

successful deletions

4.01 Mil

failed deletions

113 K

volume transferred

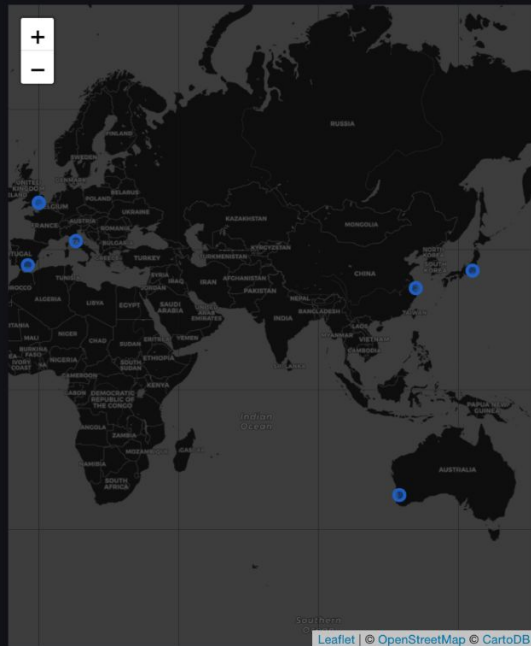
117 GB

average throughput

253 kb/s

volume deleted

148 GB





Events by type over time



Transfers



Transfer success site matrix

Src\Dst	SPSRC_STORM	JPSRC_STORM	IMPERIAL	CNSRC_STORM	CNAF	AUSSRC_STORM	STFC_STORM
STFC_STORM	97%	100%	100%	100%	100%	99%	NO DATA
SPSRC_STORM	NO DATA	100%	100%	100%	100%	99%	27%
JPSRC_STORM	100%	NO DATA	100%	100%	100%	99%	99%
IMPERIAL	98%	100%	NO DATA	99%	100%	99%	39%
CNSRC_STORM	99%	100%	100%	NO DATA	97%	98%	100%
CNAF	98%	100%	100%	99%	NO DATA	99%	34%
AUSSRC_STORM	86%	100%	100%	99%	100%	NO DATA	37%



Tokens

- A bit more treacherous than expected
- Issues with conveyor submitter, poller, reaper were discovered as we went along. All fine now*



Metadata

- Put in some dev effort into improving metadata functionality
- Metadata -> data discoverability -> with a view to IVOA 'integration', see talks from yesterday

Science Data Challenges

- Extended Rucio client for managing structured SDC data with Rucio
- Hope to use it for distributed SDC3 data to partner facilities



What's next

- IVOA metadata
- Permissions
 - Fine grained auth is crucial
 - View permissions are not a given!
- Integration with Compute

Housekeeping tasks

- Migrating the Rucio deployment to new HA k8s infrastructure
- Migrate from ESCAPE Indigo IAM -> SRC Indigo IAM
- More Conveyor daemons



The TL;DR is now...



Summary

- SRC prototyping activities ongoing, bootstrapped by ESCAPE
 - SKAO DM prototype giving Rucio new eyes and exposure
- Engagement from the SRC community will continue to shape the direction of the DM efforts as we dig deeper and look more closely at compute integration, permissions.
 - Engagement and activities from the Rucio community will also play a crucial role.

Misc suggestions

- +1 for persistent slack history
- Schematics capturing the different Rucio setups (maybe via dev roundtables)
- Faces at dev meetings :)



Questions?

*We recognise and acknowledge the
Indigenous peoples and cultures that have
traditionally lived on the lands on which
our facilities are located.*



www.skao.int