NRC·CNRC

NRC.CANADA.CA

CANFAR and the IVOA

Brian Major

IVOA Interop, Bologna, May 2023

Canadian Advanced Network for Astronomical Research

- Inception in 2008
- Cloud platform for data and compute intensive astronomy
- Canadian Astronomy Data Centre (CADC)
- Digital Research Alliance Canada (DRAC)
- 1.0: Virtual Machines, Batch & Cloud Scheduler
- 2.0: Containers (2020 - Science Platform)

CADC | National Research Council Canada | Canadian Space Agency | University of Victoria | University of British Columbia | CANARIE | Compute Canada | Digital Research Alliance Canada

CANFAR Science Platform

General Purpose / Multi-wavelength
- ALMA, JWST, CFHT, ChimeFRB, CASTOR, Gemini, …
- Radio, Optical, Surveys, …
- High memory visualization, Machine Learning, catalogs, pipelines

Self-Service Model - Integrated Auth
- Groups (represent the projects)
- Images (containers)
- Project Data

Built around **Science Containers**

Tremendous v2.0 uptake (since 2020 operational start)

Science Portal

Active Sessions

carta1

Running

skaha/carta:3.0

started: **2023-05-08 12:16** UTC
expires: **2023-05-12 12:16** UTC

memory: **19M / 1G**
CPU cores: **0.001 / 1**

desktop1

Running

skaha/desktop:1.0.2

started: **2023-05-06 17:07** UTC
expires: **2023-05-10 17:07** UTC

memory: **145M / 1G**
CPU cores: **0.001 / 0.25**

notebook1

Running

skaha/astroml-notebook:23.04

started: **2023-05-08 12:17** UTC
expires: **2023-05-12 12:17** UTC

memory: **213M / 1G**
CPU cores: **0.812 / 1**

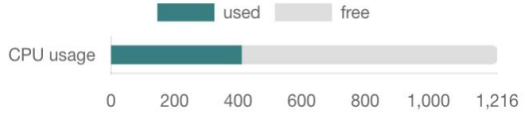
New Session

type	notebook
container image	arks/mcmc-notebook:v1.1
name	notebook1
memory	16
# cores	2

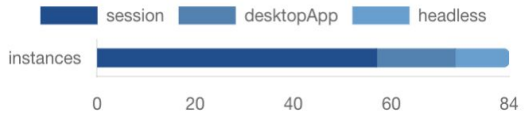
Launch Reset

Platform Load

Available CPUs: 805.75 / 1216



Running Instances: 84



*last update: **2023-5-8 12:22** UTC*

CANFAR Science Portal

Web Interface to the Interactive aspects of the CANFAR Science Platform

CANFAR - Project Workflow

1. Software specialist builds a container for your project
2. Publish the container to CANFAR image repository
3. Launch and use the container session

CANFAR - Trivial Workflow

~~1. Software specialist builds a container for your project~~

~~2. Publish the container to CANFAR image repository~~

1. Launch and use a container session

- Very easy to use: Few barriers to entry
- Open to all: Equity for astronomy

What things from the IVOA are working?

VOSpace - "cavern" built on of CephFS
- User and Project data seen in containers (POSIX) also available through VOSpace interface

Group Membership Service (GMS)
- Groups define projects
- Self-serve management extension to GMS invaluable
- Project owner and admins control access to:
  - data
  - images (containers)
  - platform access

Credential Delegation Protocol (CDP)
- Required for GMS and other inter-service call

What things from the IVOA are not working?

Authentication & Authorization items:

- GMS - Relationship to Groups coming from OpenID Connect needs to be stated
- SSO - How to work with "proxies" such as IAM and Keycloak
- Federated Identities

VOSpace*

- Intention of specification not well understood
- Clarification required in specification

(*VOSpace is an interface to distributed storage)

What does IVOA need to do next?

Observatory Networks Participation

The future is here: Big Telescope Data requires a network of cooperating and interoperating archives & platforms

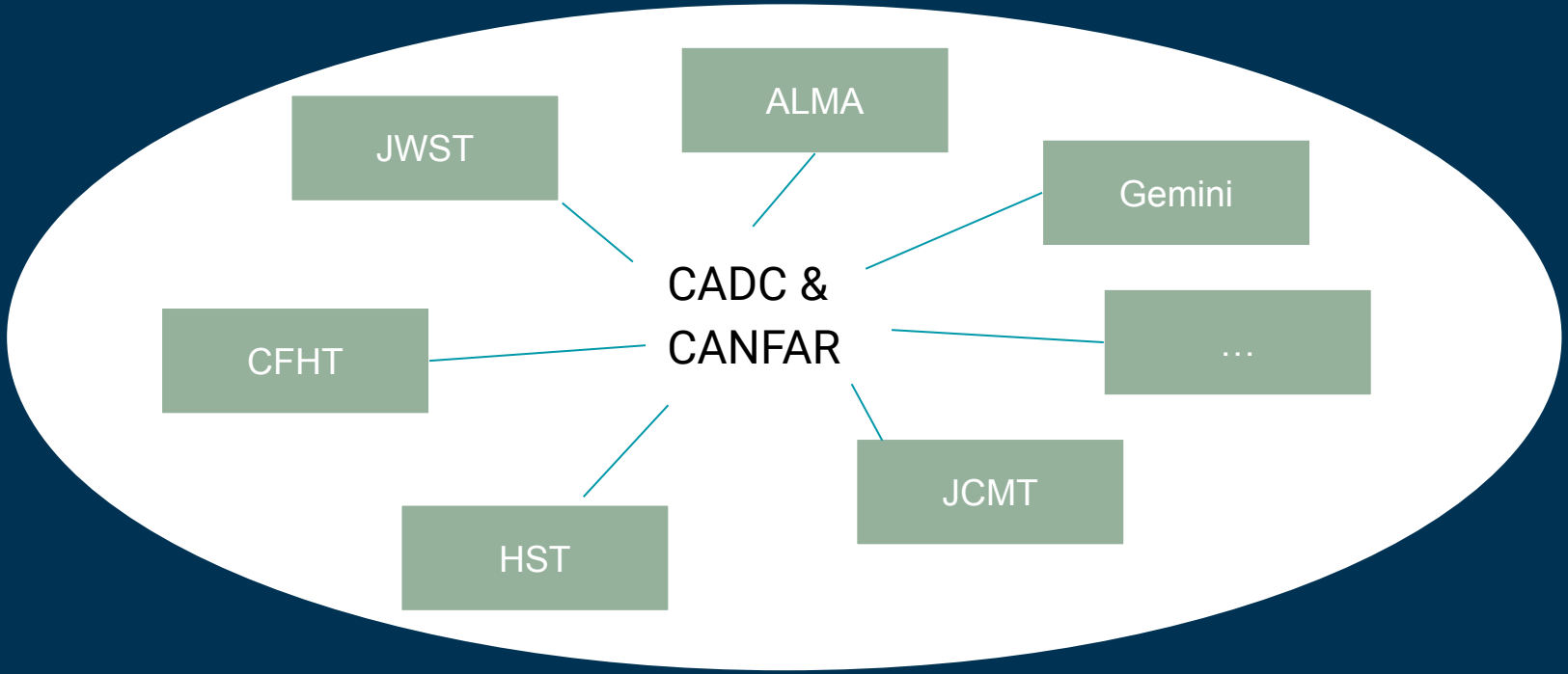
- SRCNet
- Rubin IDACs

CADC/CANFAR to absorb and expand, not duplicate

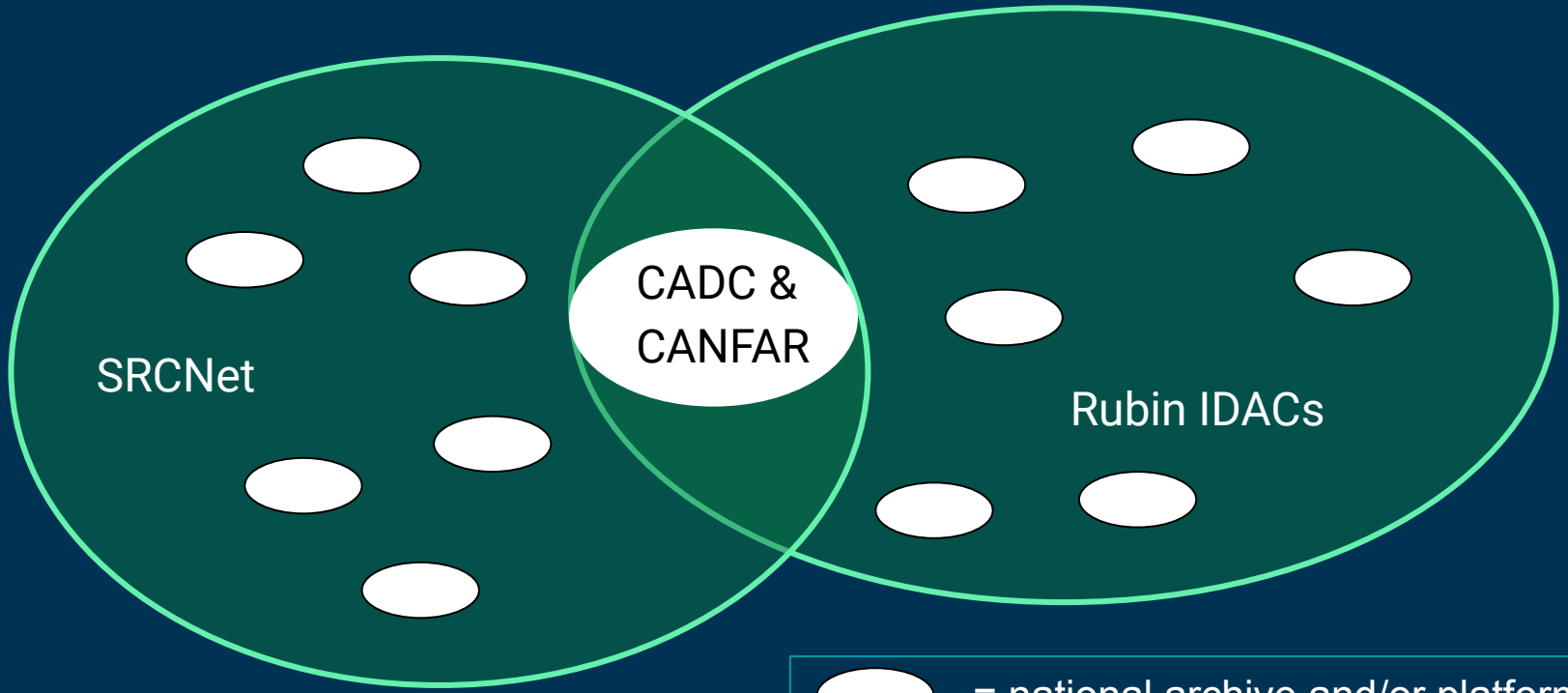
- General purpose enough to be agnostic to wavelength, data size, software requirements, etc…

The line between *Data Centres* and *Platforms* is blurring…

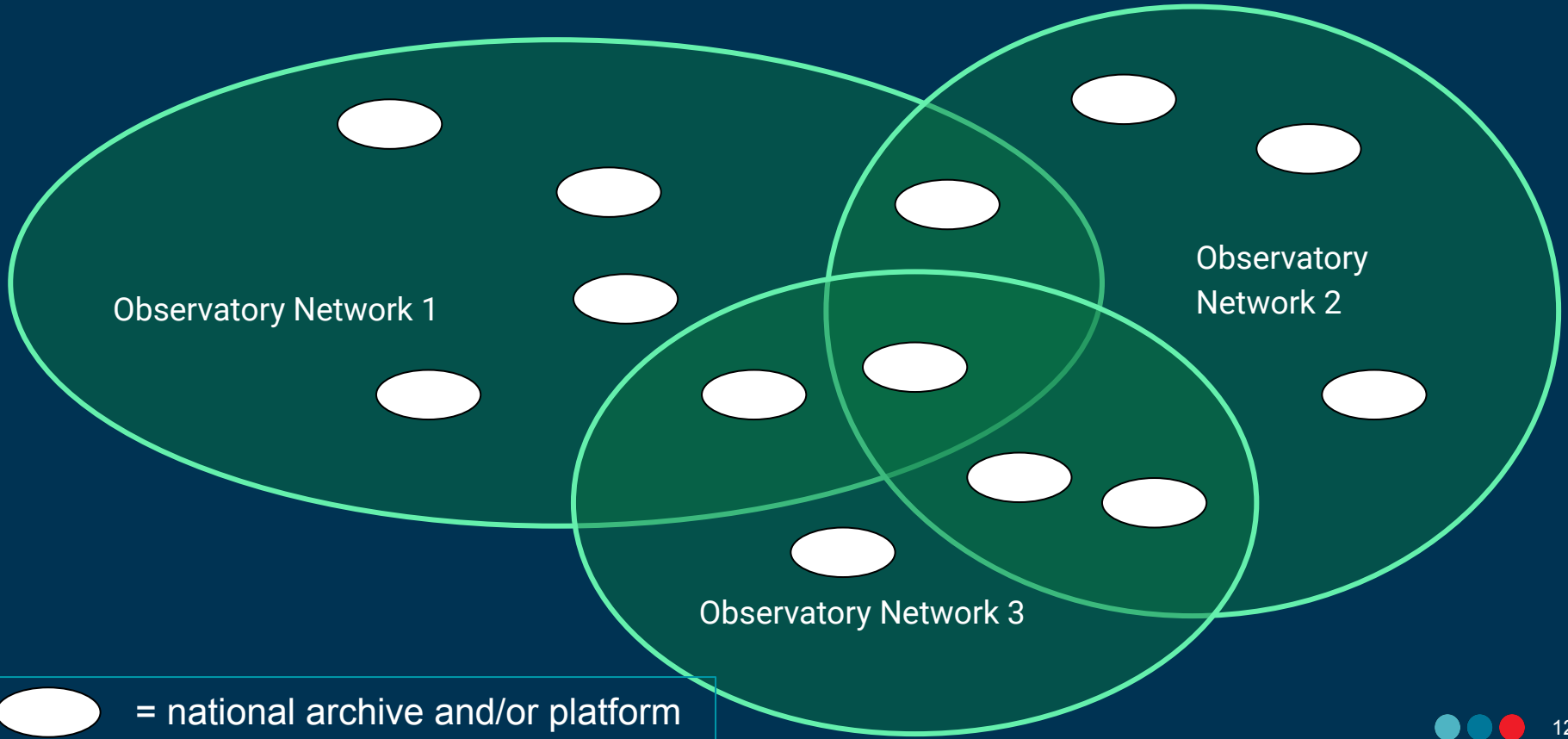
Multi-wavelength Archive & Platform



Observatory Network Participation



Observatory Network Participation

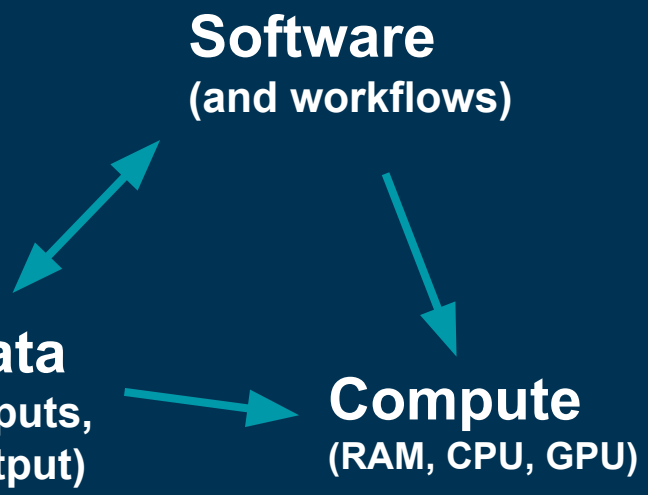


Observatory Network Participation

General characteristics of participation:

1. A given node may have some or all of the data
2. The data may be geographically distributed by science use cases
3. Data access is exclusively through the local platforms (no downloads)
4. Use of a node can come from any of the communities
5. Support for both proprietary and public data over time

Remote execution



Jobs run somewhere on the platform network and require

1. Software (portable)
2. Data (somewhat portable)
3. Compute (fixed)

Match all three to run a job (interactive or programmatic).
(See Execution Planner note)

- IVOA Historically Data Focussed.
- To support platforms IVOA must standardize on all three.
- Data models needed for software and compute.

Building and publishing an IVOA compatible container

Interactive Containers

- JupyterLab
- CARTA
- Visual Studio
- Pluto.jl
- Firefly (in progress)

- … and many more variants and project specific containers

- Remote Desktop
  - TOPCAT
  - CASA (v3..v6)
  - IRAF
  - Aladin Desktop
  - Starlink
  - Lenstool
  - … and many more

- Batch / programmatic

Standardize an IVOA Container

Containers that are eligible for a network of science platforms

Build, startup, and runtime requirements:

- Interactive: well-known port, must know own access path
- Must run as non-root user
- Auth disabled (leave that to the platform)
- Startup recipes, or defer to ENTRYPOINT, CMD?
- No hubs (leave that to the platform?)

Metadata Requirements

- To support discovery and execution planning
- Standard OCI metadata labels?

Auth "Challenges"

Currently

- ➔ Login per data center or platform (as per SSO 2.0)
- ➔ May share the same Identity (eg ORCID)

Ideal world

- ➔ One identity/login for all astronomy data and platform access

Must avoid

- ➔ Logins tied to specific data and access rights
  - ◆ Can't anticipate what science the user will do
  - ◆ Can't do multi-messenger science

What does the IVOA need to do next?

1. User Catalogs (TAP 1.2, "YouCat", Pat Dowler) - In Progress

2. Support multi-wavelength astronomy in observatory network nodes

3. Remote Execution
    a. Standard IVOA software (container?) characteristics and expectations
    b. Standards for Software and Compute Data Models

4. Auth: Federated Identity Protocols

IVOA is the perfect place for defining how these observatory networks and platforms (SRCNet, Rubin, and future!) can interoperate.

NRC·CNRC

NRC.CANADA.CA •

THANK YOU

Brian Major • CADC • brian.major@nrc-cnrc.gc.ca