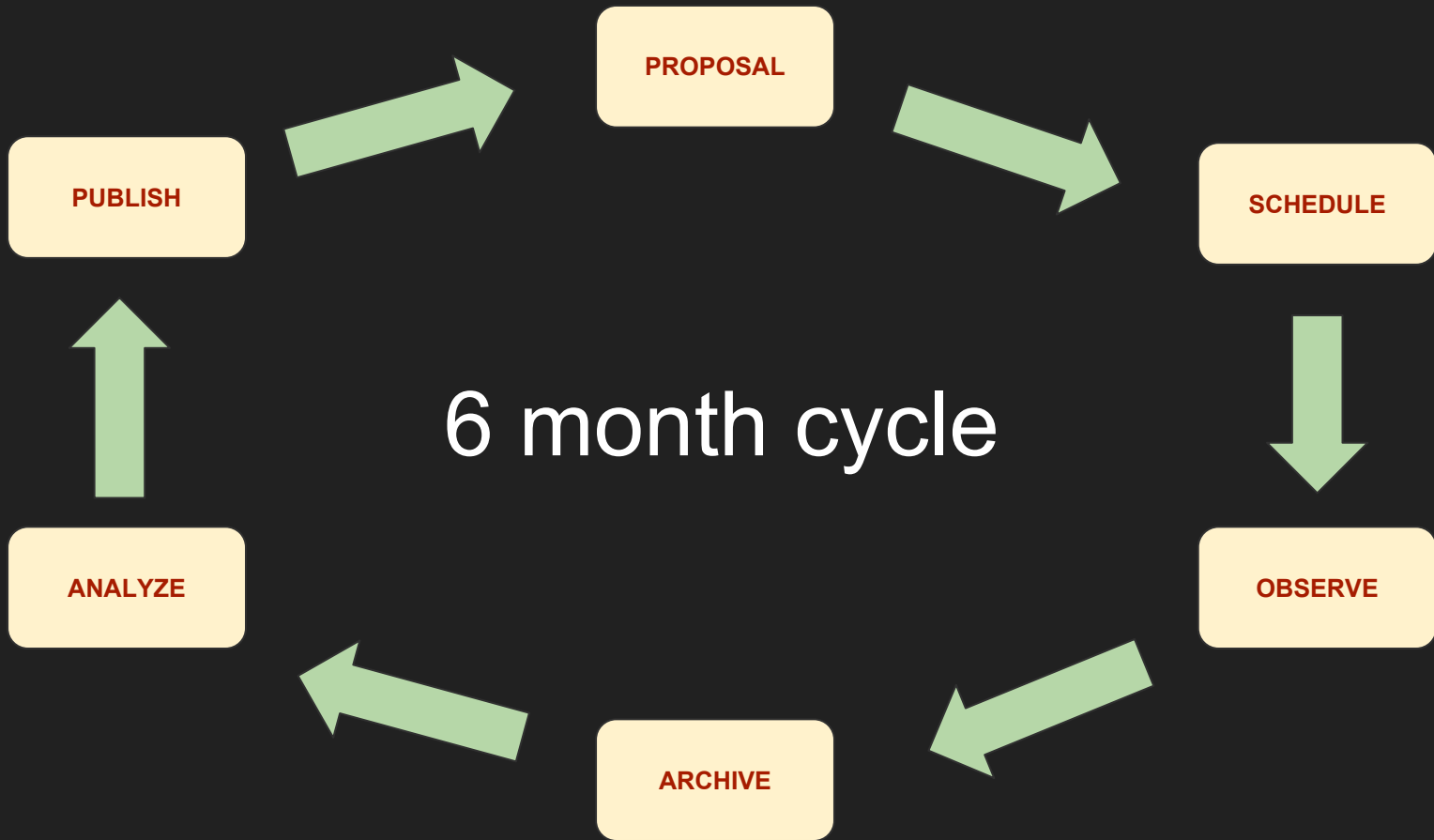


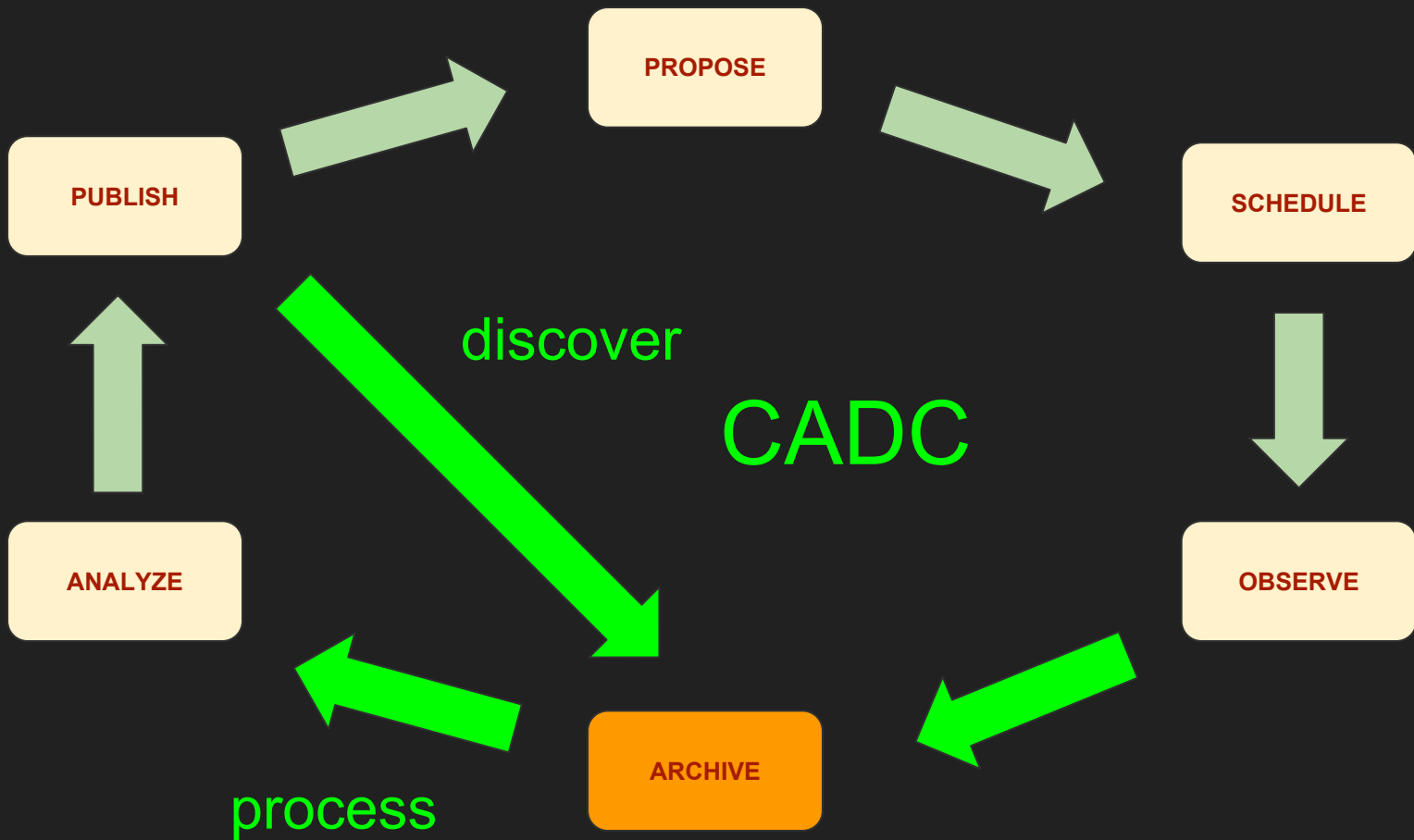
Astronomy Data Delivery and Processing Services with CADC and CANFAR

Sébastien Fabbro
for the Canadian Astronomy Data Centre

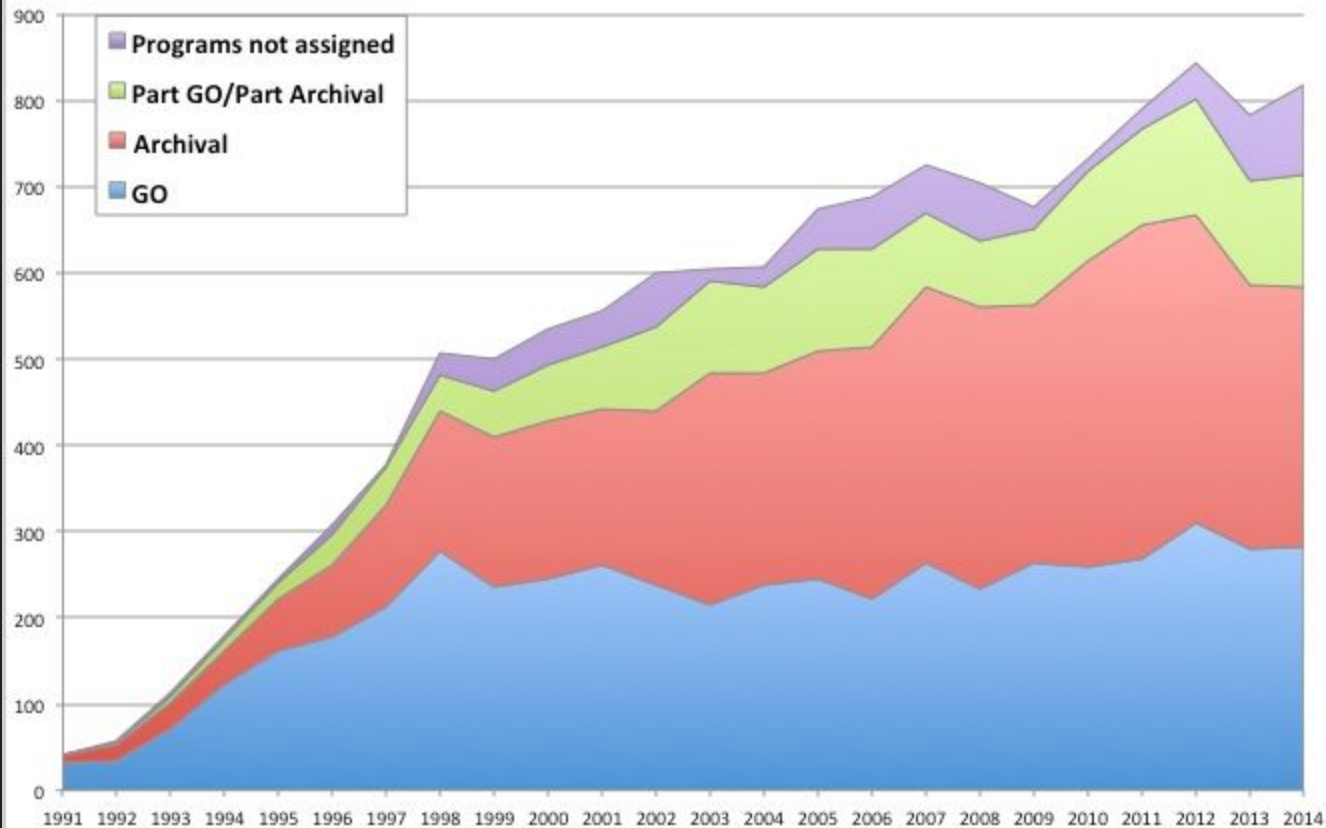
CADC users

- astronomy collaborations
- telescopes
- individual telescope observer PI
- other astronomy data centres





HST Publication Rate

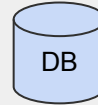
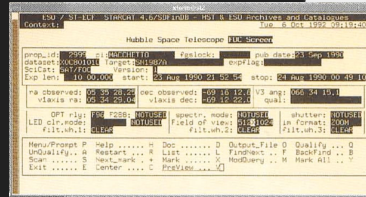


from
archives

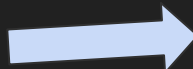
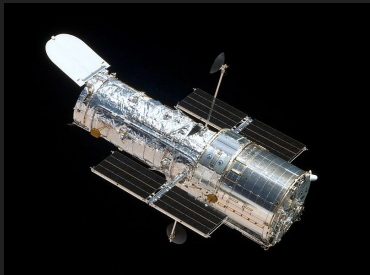


data
archiving

1986



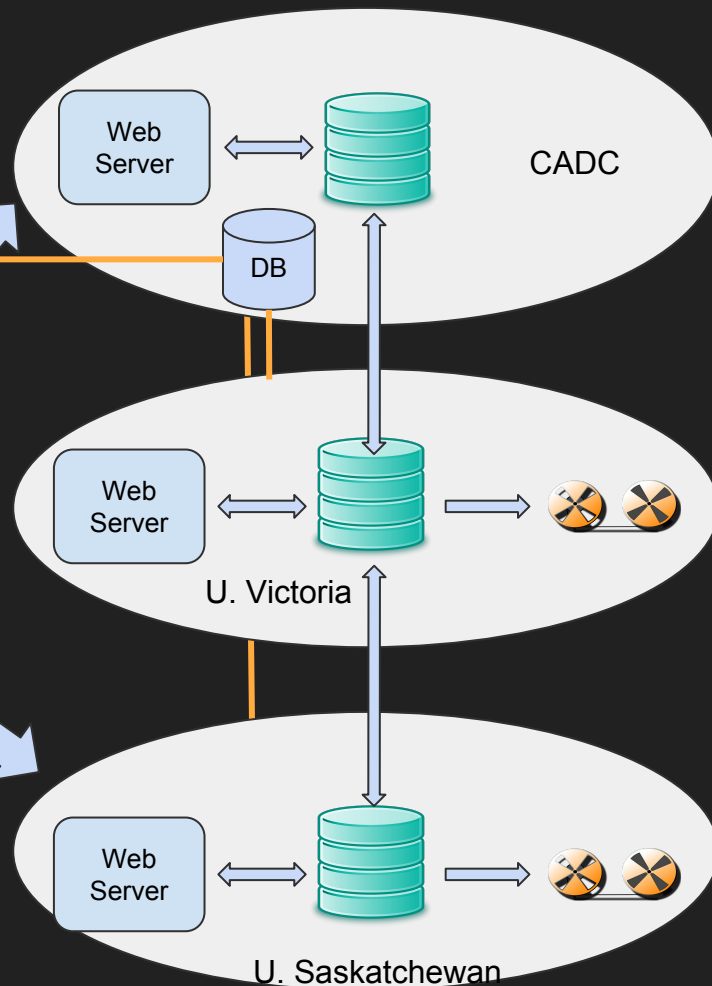
CADC



2016



PUT
GET
DELETE
INFO





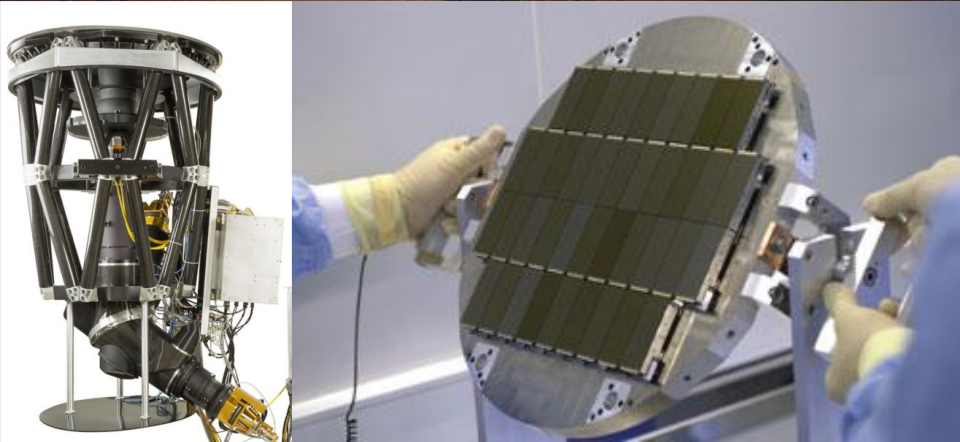
Hubble Space Telescope

- since 1986
- in-house processing
- UV, optical, infrared, spectra



Canada France Hawaii Telescope

- since 1991
- PI distribution
- large file sizes
- megaprime post-processing





James Clerk Maxwell Telescope

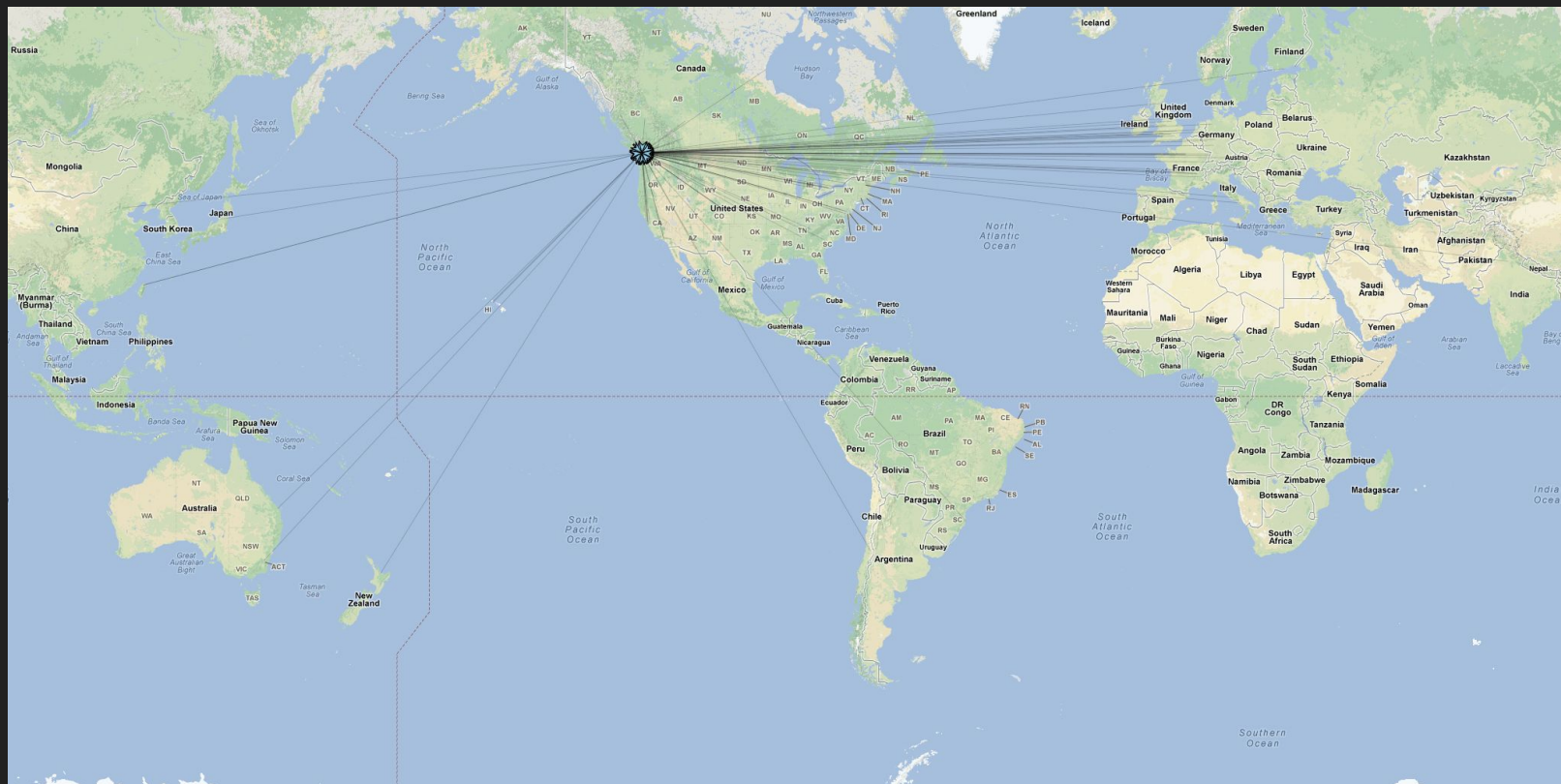
- since 1993
- many small files
- using all CADC web services
- today largest usage at CADC



196 User Storage Spaces

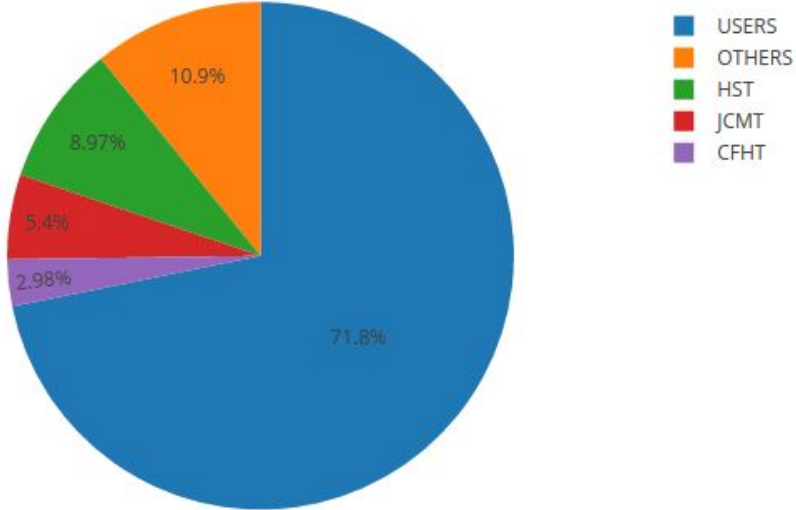
- since 2011
- VOSpace IVOA standard
- browser UI, cli and FUSE
- X.509 based

geography of user PUTs

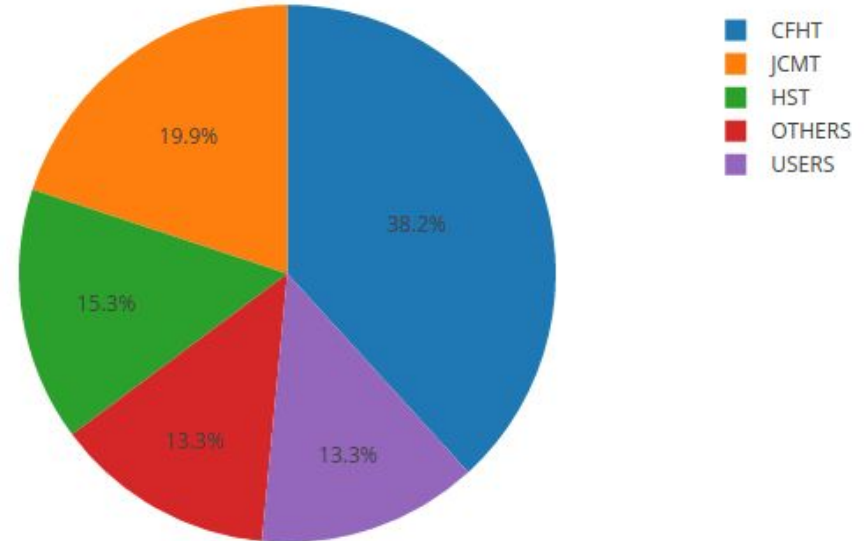


data stored per replica in 2016

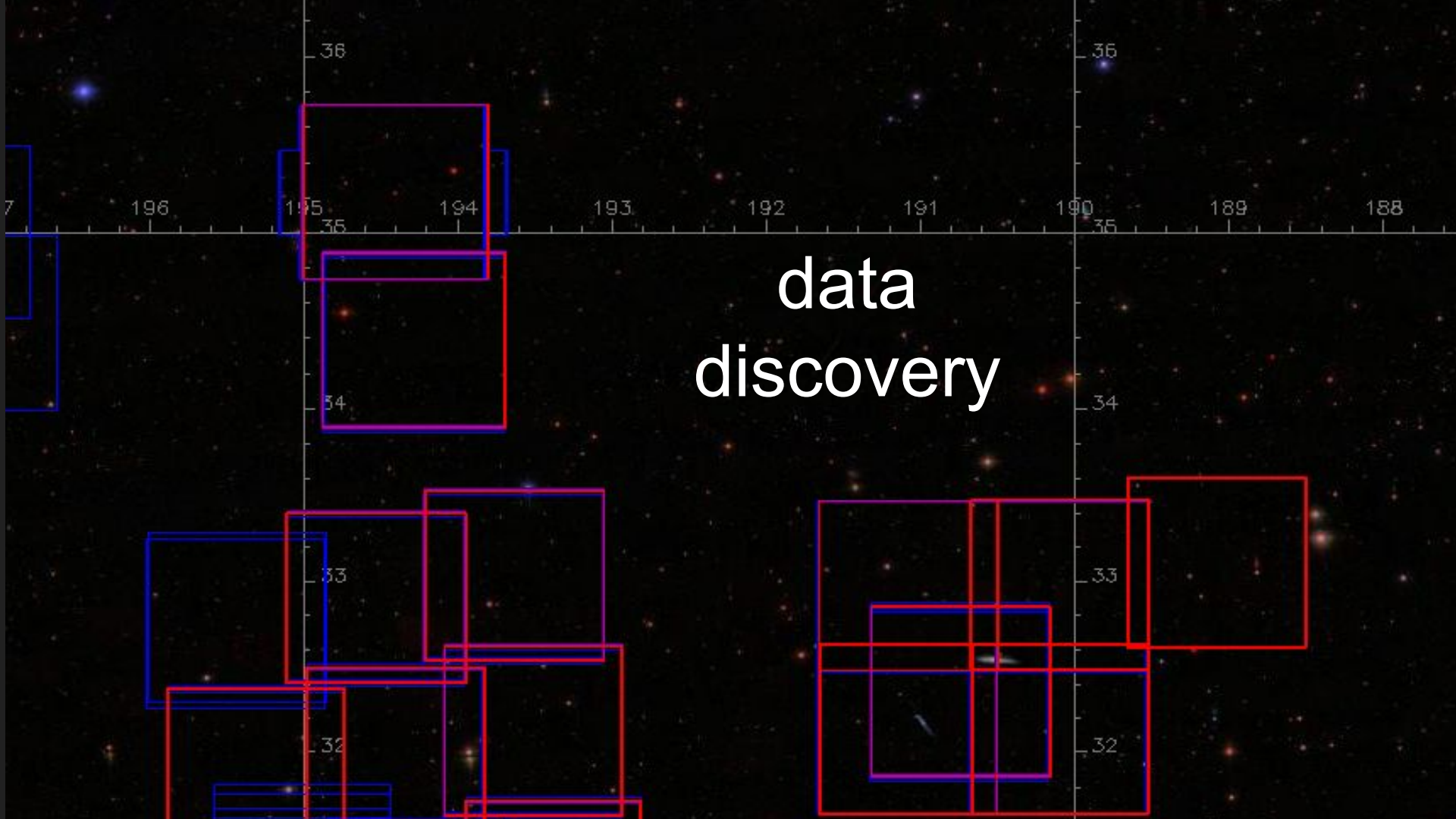
244,964,334 Files



1.22 PB

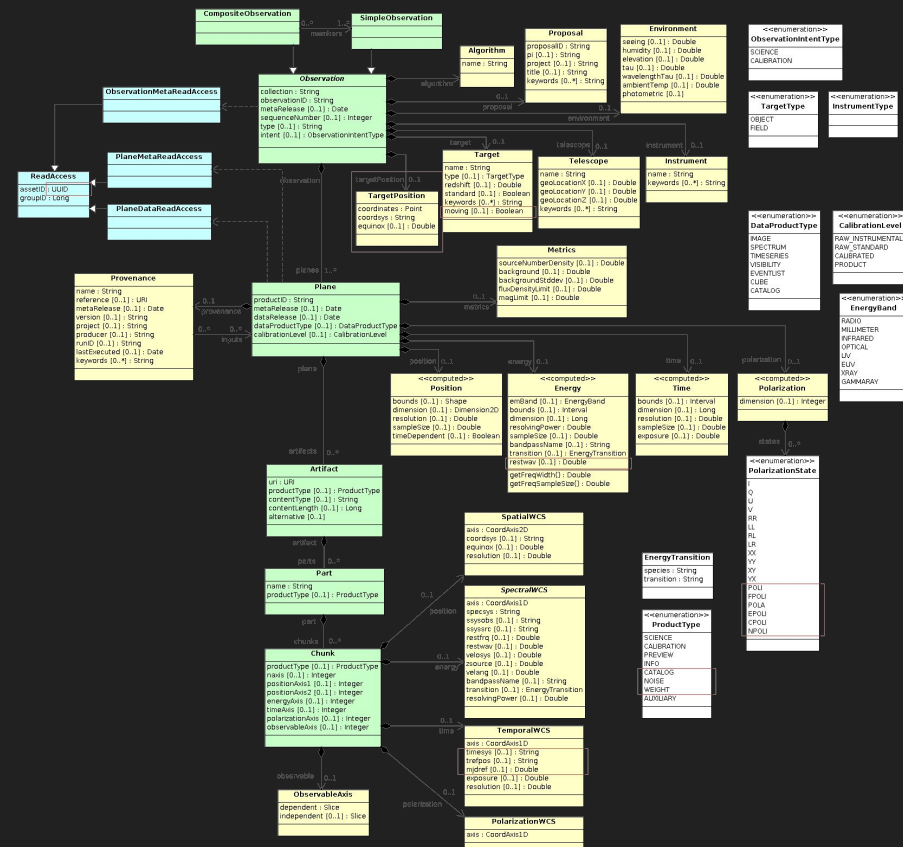
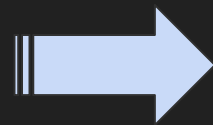


data
discovery



Common Archive Observation Model

- location
- time
- energy
- polarization



current archive

- 23 ingested collections
- 115 instruments
- 4 calibration levels
- delivery: web services

Telescope Data Products - Advanced Data Products - Services - Documentation - Advanced Search - Login

CADC Home > Advanced Search

Advanced Search

Search Results Error ADQL Help

Search Reset

Click on ? for explanations

Observation Constraints

- ▶ Observation ID ?
- ▶ P.I. Name ?
- ▶ Proposal ID ?
- ▶ Proposal Title ?
- ▶ Proposal Keywords ?
- ▶ Data Release Date ?

Science and Calibration data ▼

Spatial Constraints

▼ Target (m87) ?

Resolve with (if applicable)

Automatic (recommended) ▼

m87 ✓

- Or -

Choose file No file chosen

▼ Pixel Scale ?

Do Spatial Cutoff

Temporal Constraints

- ▶ Observation Date ?
- ▶ Integration Time ?
- ▼ Time Span (= 3650.0 days) ?

10y

Spectral Constraints

▼ Spectral Coverage ?

600nm-700nm

- ▶ Spectral Sampling ?
- ▶ Resolving Power ?
- ▶ Bandpass Width ?
- ▶ Rest-frame Energy ?
- ☐ Do Spectral Cutoff ?

Additional Constraints

Band	Collection	Instrument	Filter
All (7)	All (21)	All (115)	All (2149)
Gamma-ray	CFHT	ACS	0.35MB
Infrared	CFHTMEGAPIPE	Apogee USB/Net	0.35um
Millimeter	CFHTTERAPIX	COS	0.45MB
Optical	CFHTWIRWOLF	CPADIR	0.45um
Radio	HST	Cassegrain Spectrograph	0.75um
UV	HSTHLA	Cassegrain Spectropola	0.85um
Unknown	GEMINI	Direct image	1.083 um
	JCMT	ESPaDOnS	1.210 um
	JCMTLS	F2	1.282 um
	DAO	FTS2-SCUBA-2	1.3um
	DAOPlates	Fabry image	1.4um

Target

m87

Service

NED(ned.ipac.caltech.edu)

Coordinates (RA Dec)

187 70593 12.39112

ICRS

Object Type

G

Morphology Type

Time (ms)

324

(1) Raw Standard

(0) Raw Instrumental

Unknown

Data Type

All (6)

catalog

image

Other

spectrum

timeseries

Observation Type

All (57)

ACQUIRE

ALIGN

ARC

ASTAR

BIAS

CAL

CALIB

COMPARISON

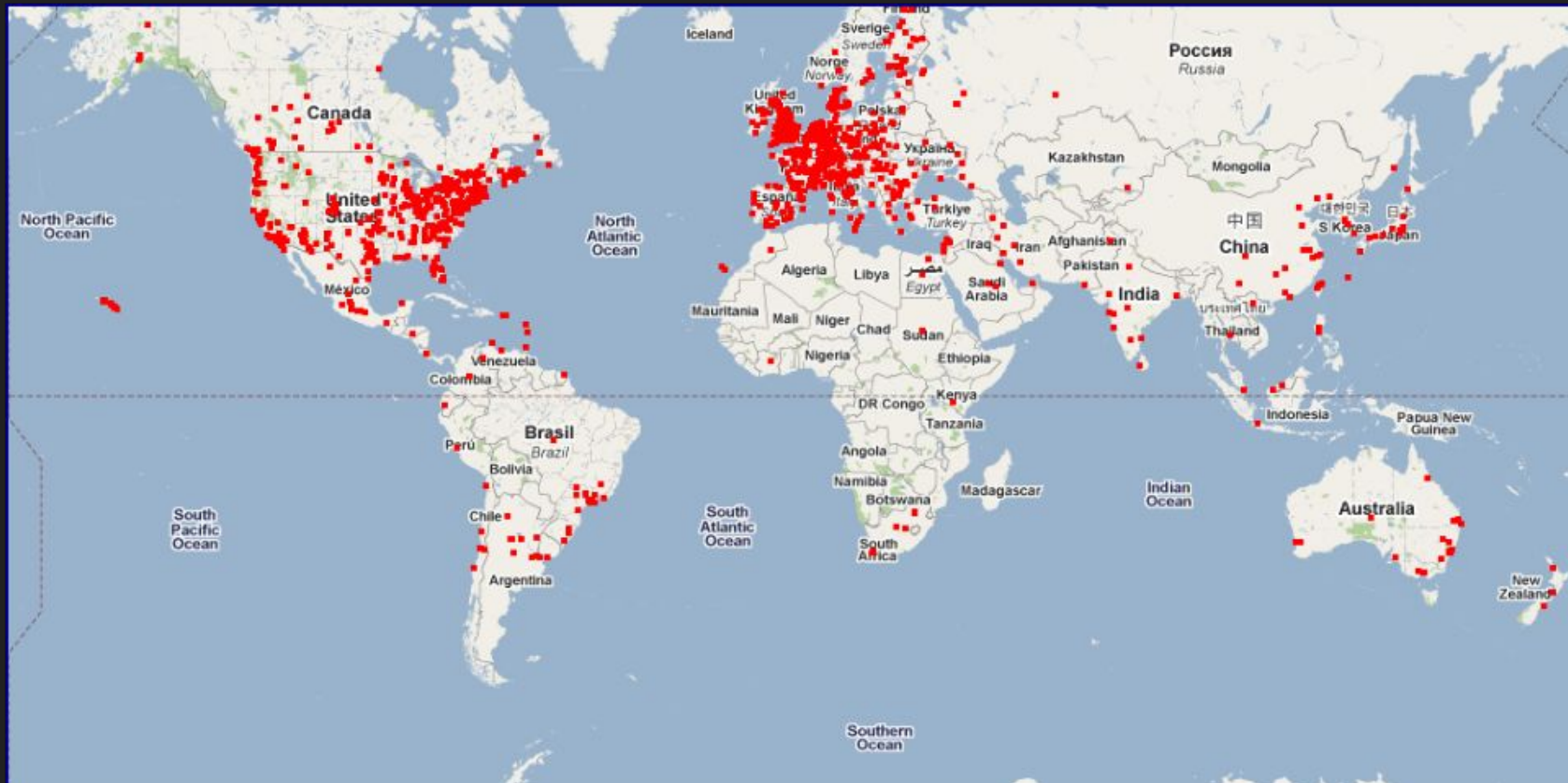
DARK

DIM

DOME_FLAT

Search Reset

deliveries in 2015: 91M files, 1.1PB, 5k users



discovery challenges

- automated discovery with processing
- low user uptake

data expertise + data model complexity ?



data
processing

multi-cloud processing

- 2 OpenStack clouds
- Astronomy-aaS
- 2015: 4.4M jobs

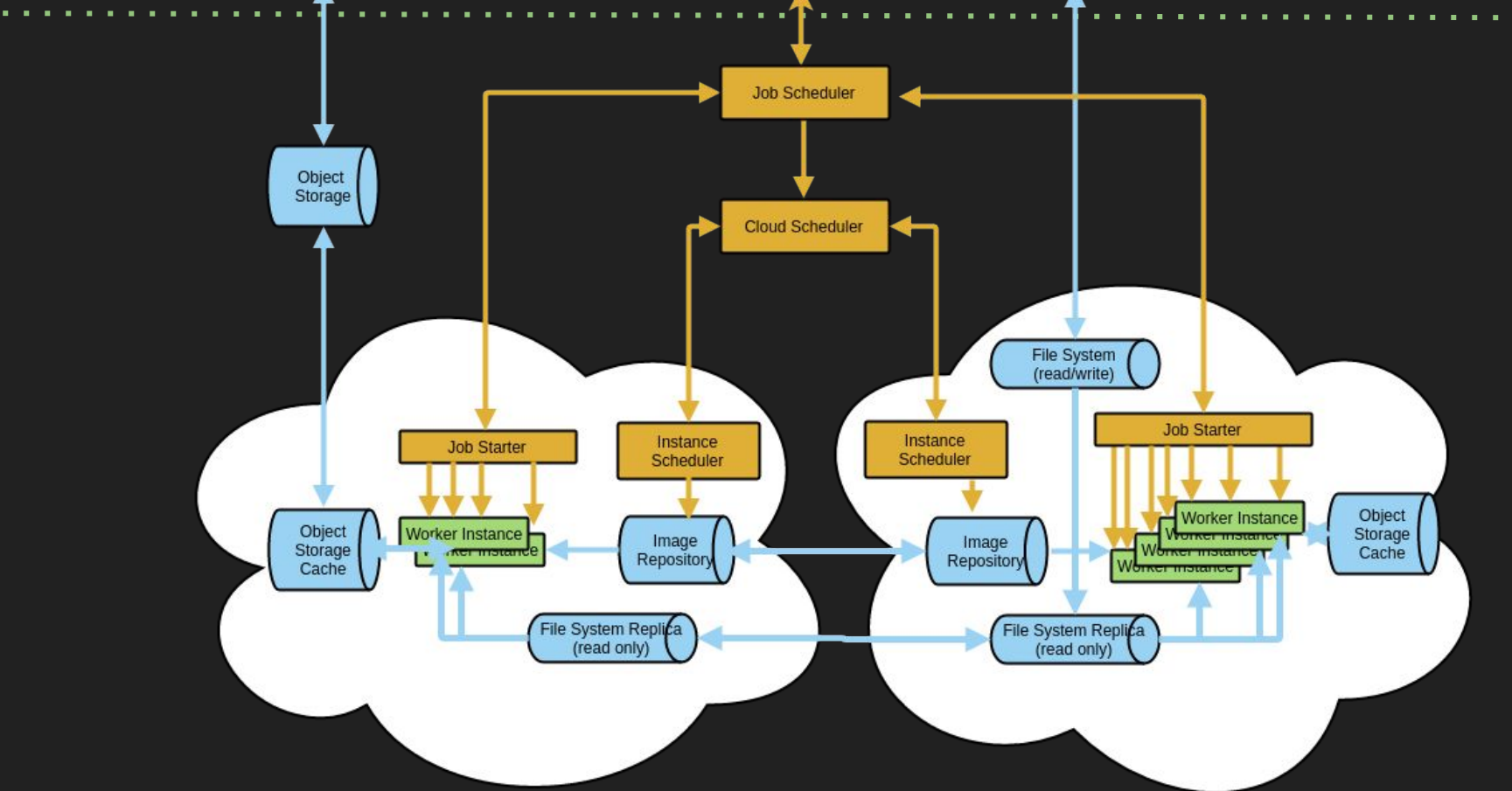


Users

Object Storage Client

CADC Batch Client

POSIX access



cloud processing challenges

- overheads
 - virtualization (sold: 5%, measured: 18%)
 - VM size, boot time (hard on short time jobs)
 - mindsets, sysadmin astronomers
- shared storage between VMs
- network tuning
- embarrassingly parallel only so far
- technology evolution: containers, kubernetes, serverless...

services integration



access control and group management at core

- user based management
- IVOA SSO + credential deleg.
- group management web service

CANFAR
Canadian Advanced Network for Astronomical Research

Portal Documentation Login

Create Group

Showing 9 rows (9 before filtering).

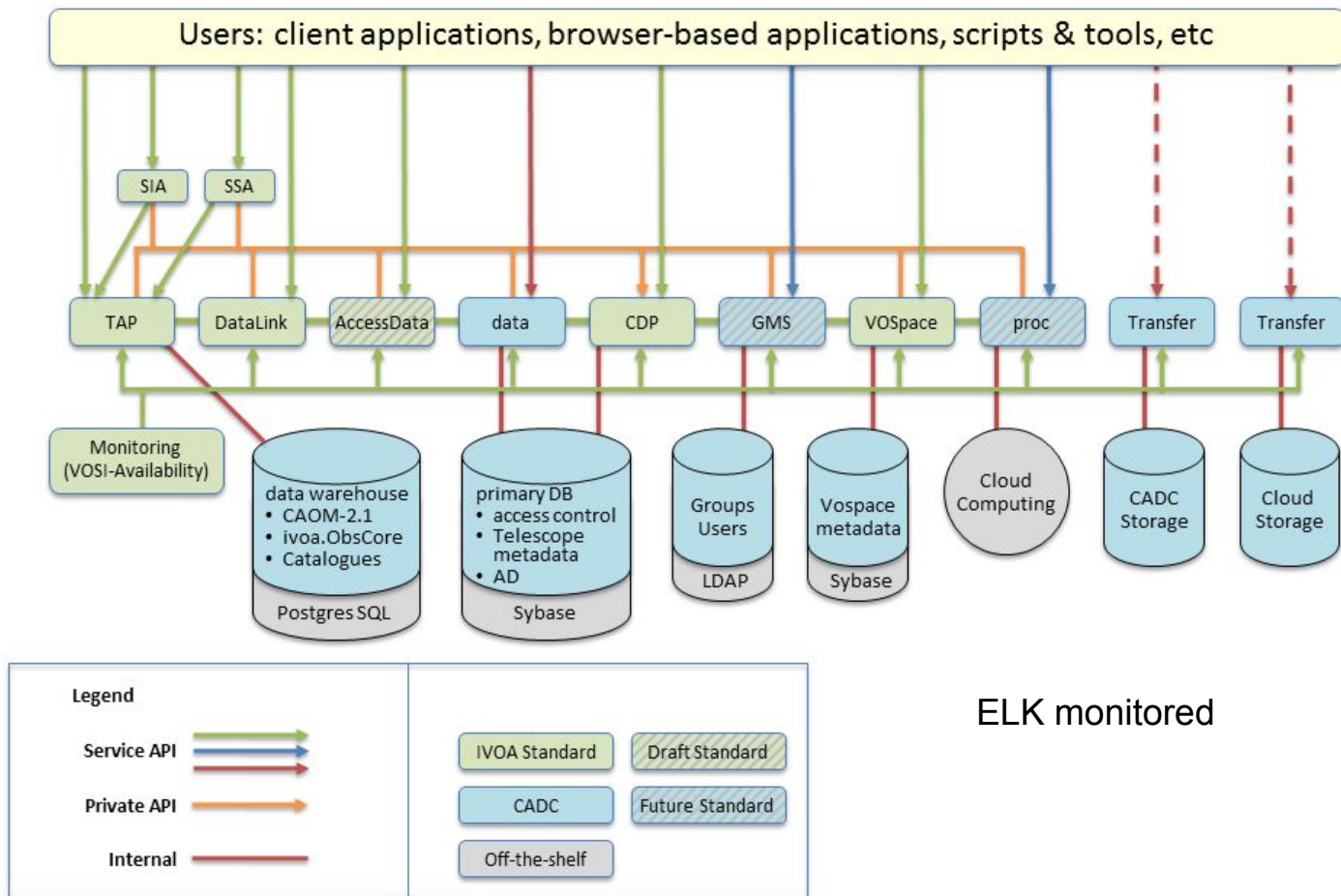
Name +	Owner Name	Administrators	Members	Description
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
AdminTest	Adrian Damian	View	View	This is an admin test group
AdrianTestGroup	CADC Regtest1	View	View	Adrian's Group used for testing
CANFAR-Staff	John A. Ouellette	View	View	CANFAR User Group
GPID	Severin Gaudet	View	View	Gemini PI Distribution
Gemini-bad-files	Severin Gaudet	View	View	Files that can't be retrieved by Paul Hirst
IVOA_photo	Severin Gaudet	View	View	Upload of IVOA photos
MAP	Severin Gaudet	View	View	Multi-Archive Project
gaudet-VOS-write	Severin Gaudet	View	View	Write group for gaudet's VOSpace group
gsadist	John A. Ouellette	View	View	

Date modified: 2015-03-02

About
CANFAR Project
Help

Contact
CANFAR on GitHub
✉ E-mail

services architecture



ELK monitored

integration challenges

- data transfer dominated
- astronomy data centres federation
- legacy software for modern architectures
- batch processing competing with web services
- web services layer can be heavy

are we ready to scale?

- small surveys storage: 2016: +1PB, 2017:+4PB
- large surveys (SKA, Euclid, LSST) Canadian involvement



thanks!

www.cadc-ccda.hia-ihc.nrc-cnrc.gc.ca

www.canfar.net



github.com/canfar

github.com/opencadc

