

An aerial photograph of a vast, flat, reddish-brown desert landscape. The ground is covered with numerous small, white, spherical objects, likely balloons or buoys, scattered across the terrain. The objects are densely packed in some areas and more sparse in others, extending towards the horizon. The sky is a clear, pale blue. The overall scene suggests a large-scale field experiment or data collection project in a remote, arid environment.

CyberSKA

Russ Taylor

Institute for Space Imaging Science
University of Calgary

Technology Challenges en route to SKA



- **Large collecting Area**
- **Large field-of-view**
- **Large bandwidth**
- **Power consumption**
- **Massive Data**



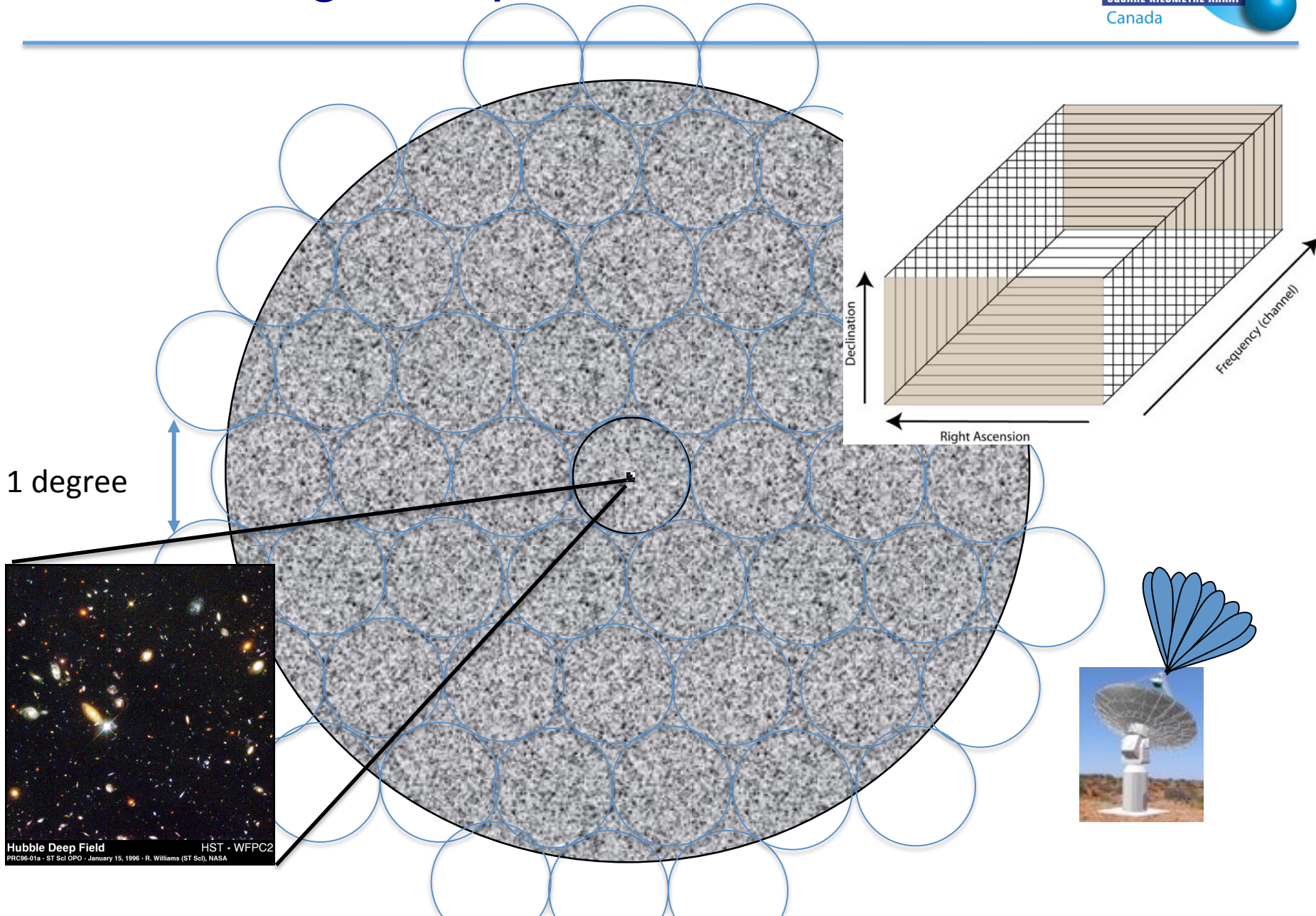
What we need



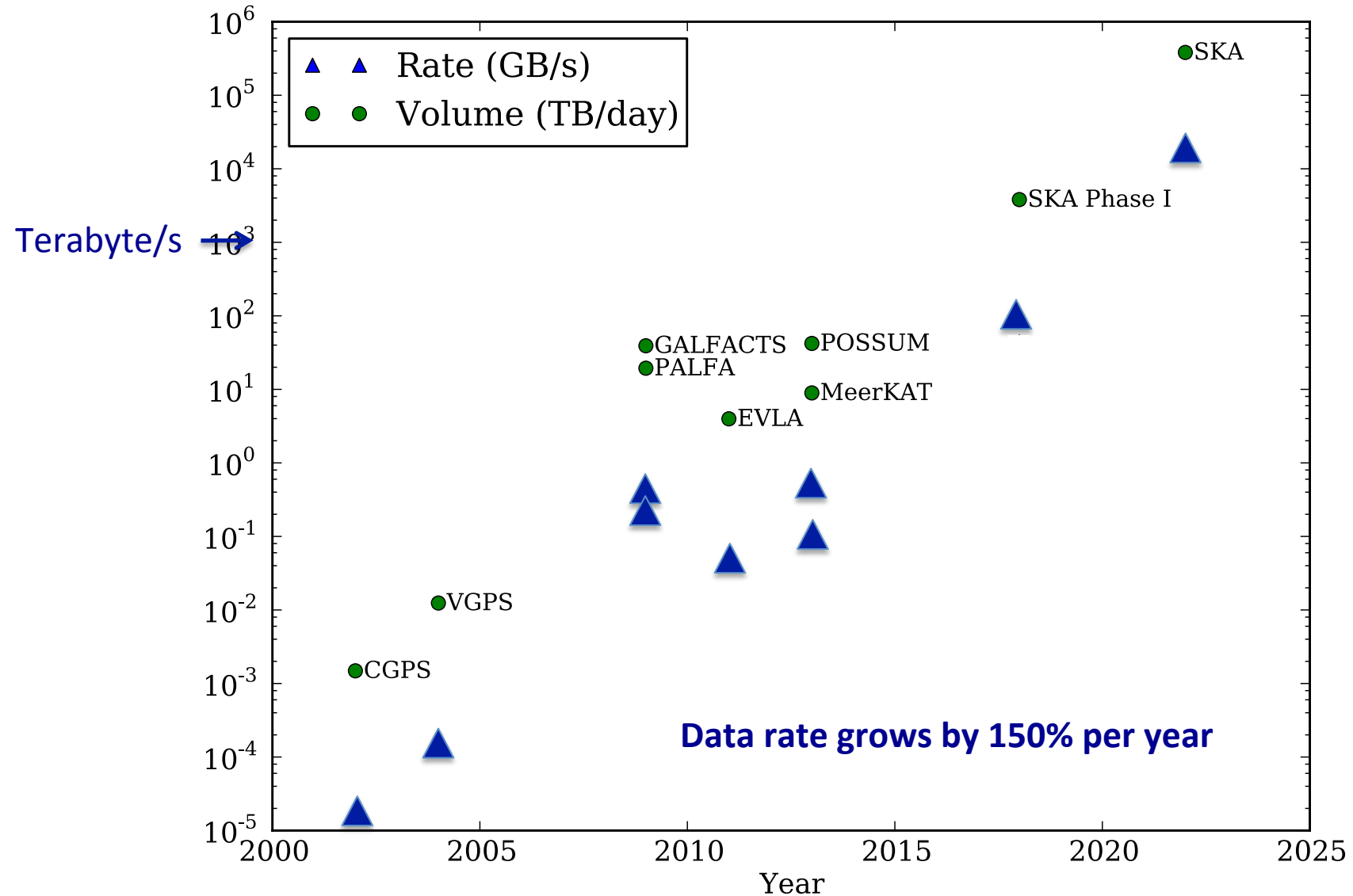
...to develop a new generation of computer technology to store and process the data soon to be captured by the Square Kilometre Array, a new radio telescope that will collect each day twice the amount of information presently generated on the entire World Wide Web.

Al Gore,
The Future: Six Drivers of Global Change
Random House, NY 2013

FPA Image Outputs

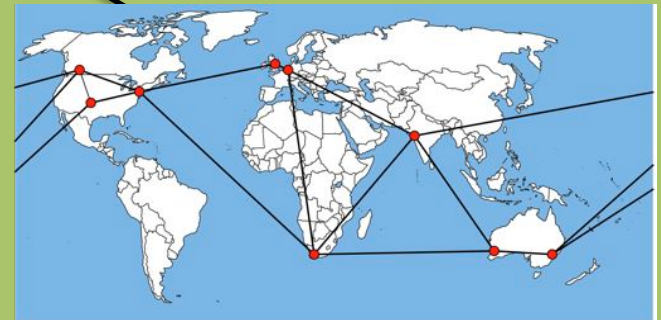


Survey Raw Data Rates out of Correlator



Sociology of Radio Astronomy

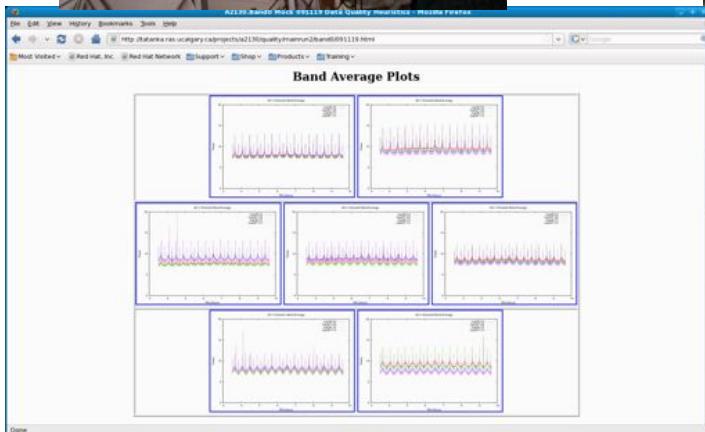
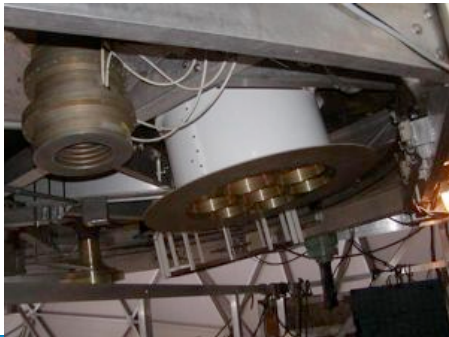
- Much of the key science en route to the SKA will be achieved via large-scale survey mode observing programs executed by globally distributed teams of researchers



Arecibo ALFA Surveys

- ALFALFA (HI)
- PALFA (Pulsars)
- GALFACTS (imaging Spectro-polarimetry)

GALFACTS and PALFA
Aggregate rate 500 MB/s



LOFAR Survey Science

- Sky surveys at 15, 30, 60, 120, 200 MHz
 - Galaxy formation
 - Intergalactic magnetic fields
 - Star formation in early universe
 - Expansion of discovery parameter space



ASKAP Survey Science

- WALLABY (HI emission)
- EMU (continuum)
- POSSUM (polarization)
- FLASH (HI absorption)
- VAST (slow transients and variables)
- GASKAP (Galactic HI)
- CRAFT (fast transients)
- DINGO (Deep HI)
- COAST (pulsar and timing survey)
- VLBI (high resolution science)



Some project will be commensal

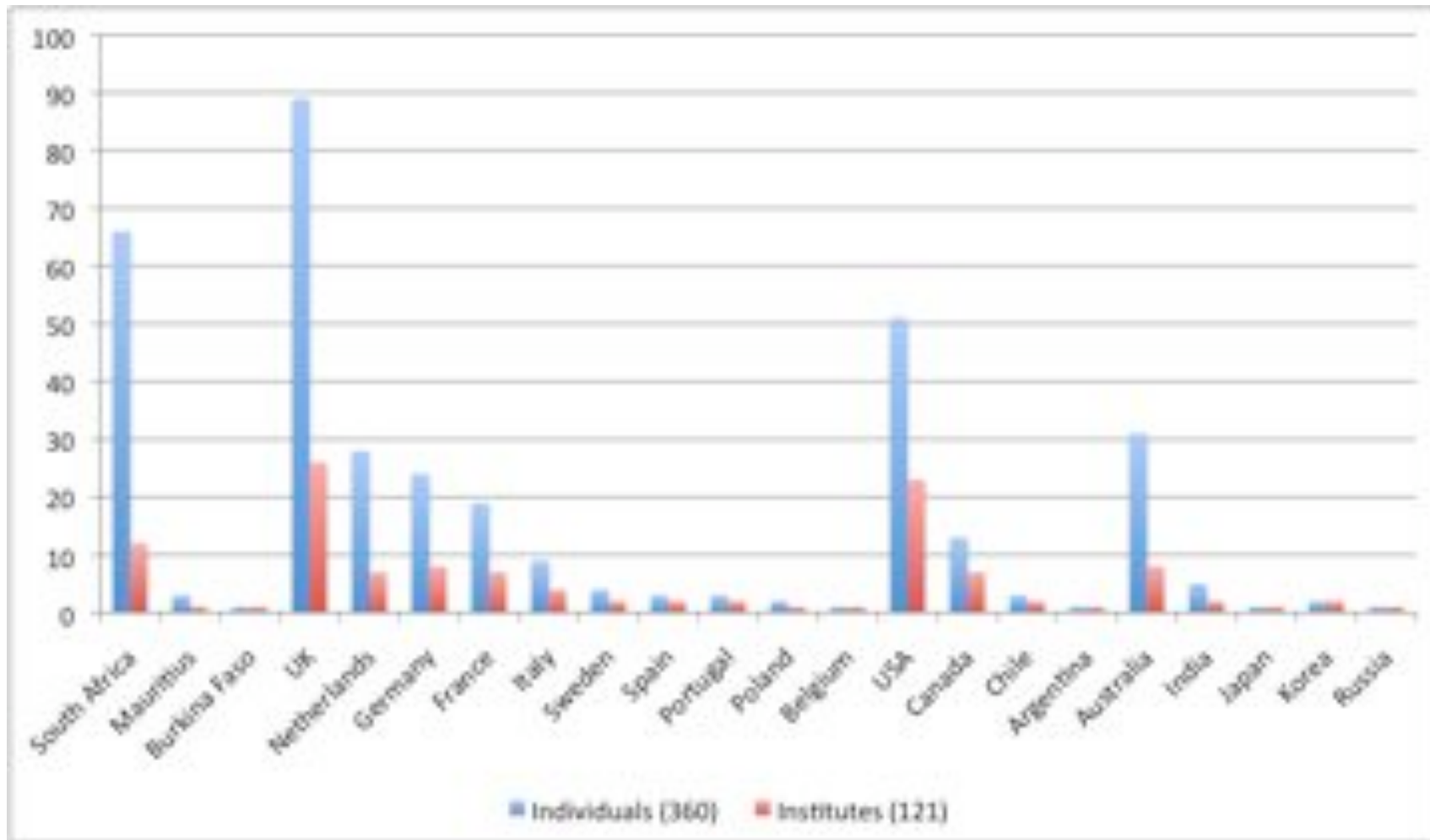
MeerKAT Survey Science

- Pulsar Timing
- LADUMA (Deep HI)
- MESMER (High-z CO)
- MeerKAT Absorption Line Survey
- MHONGOOSE (Nearby HI)
- TRAPUM (pulsar search)
- MeerKAT HI Survey of Fornax
- MeerGAL (Galactic Plane Survey)
- MIGHTEE (Deep continuum and polarization)
- ThunderKAT (variables and transients)



Some project will be commensal

MeerKAT Large Surveys (43,000 hours allocated)



22 countries

The Challenge

“Survey” mode observations drive:

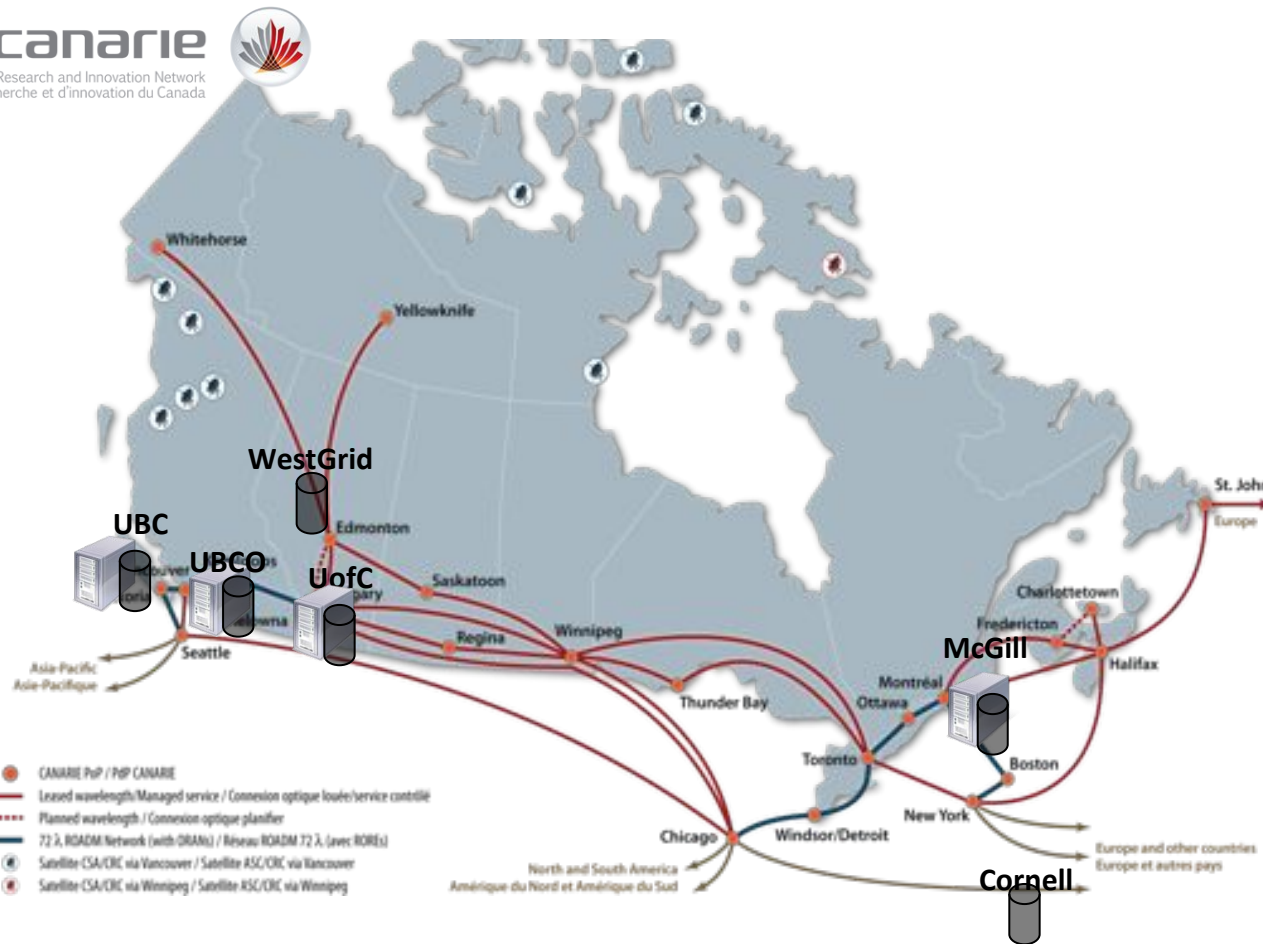
- Very high data rates and volumes
 - Storage, transfer, access
 - Delivery of data to end users not practical
- Complex, multi-purpose, processing and analysis
 - Processing, analysis, visualization, data mining
 - Multiple processing and analysis chains
- Collaborative execution by globally distributed teams of researchers
 - Distributed and remote science community
 - Distributed collaboration in data processing, analysis and science

The CyberSKA Project

Initiative to develop a scalable and distributed cyberinfrastructure platform to meet evolving needs of data-intensive radio astronomy en route to the SKA



Canada's Advanced Research and Innovation Network
Le réseau évolué de recherche et d'innovation du Canada

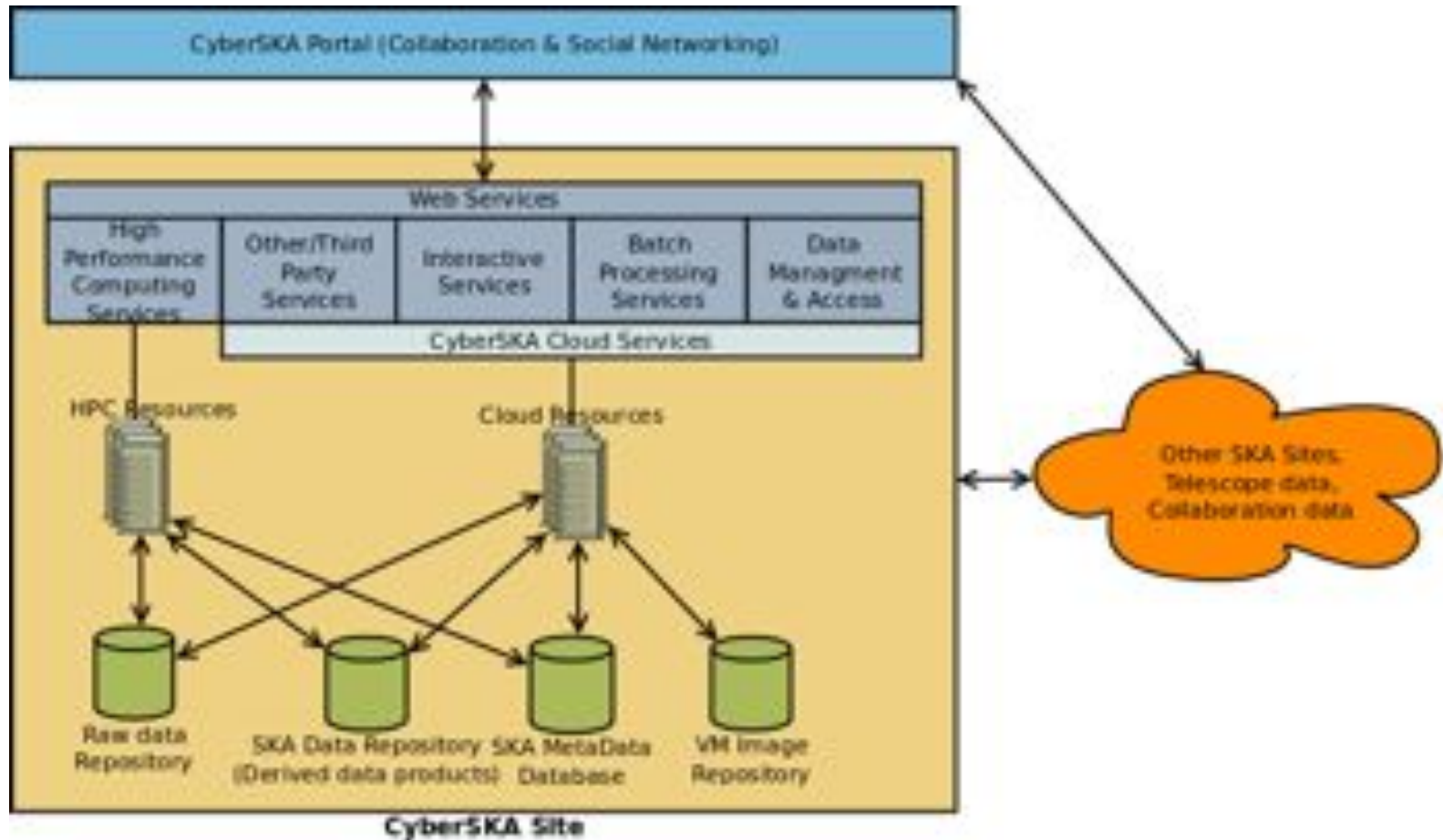


Focus Areas



- Collaboration
 - Portal built on social networking technologies
- Data Management
 - Scalable collaborative access, sharing and searching of distributed (BIG) data sets
- Data Processing
 - Framework for executing algorithms and workflows for time-domain and image-domain astronomy
- Data Visualization and Visual analytics
 - On-line interactive visualization of remote Big Data
- Third Party Applications
 - Community driven site with common API

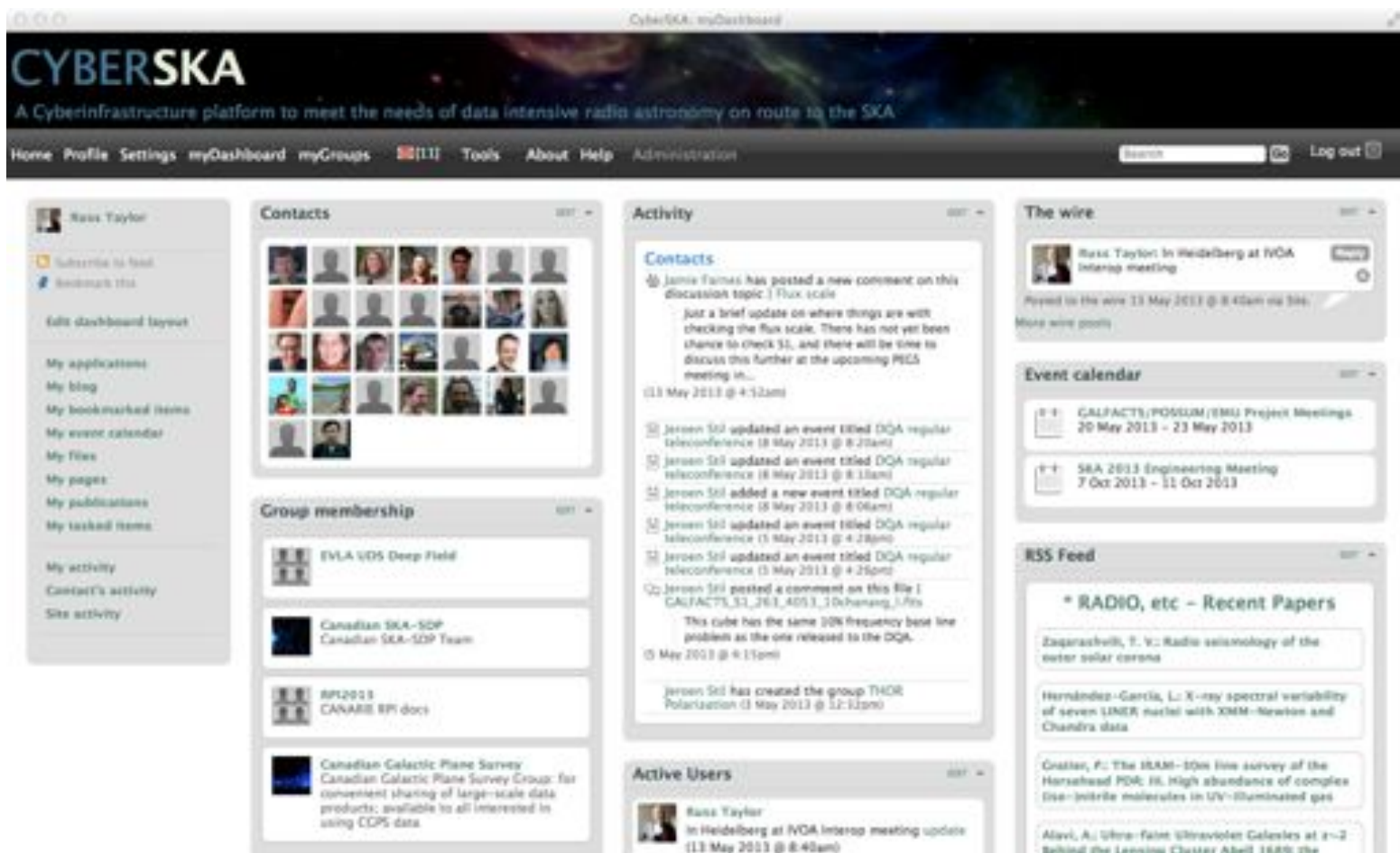
CyberSKA High Level Architecture



Collaborative Portal

Built on top of the Elgg open source social networking platform

- Provides many features including: tags, bookmarks, profiles, blogs, wikis, contacts, groups, document sharing, discussions, messaging, calendars, status, activity feeds, etc.



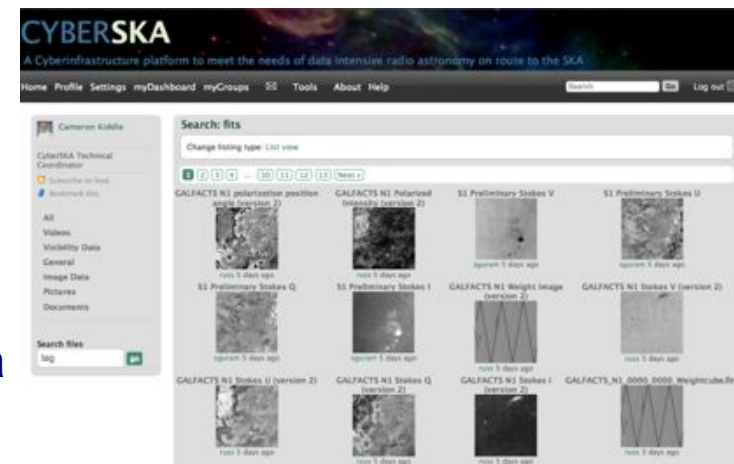
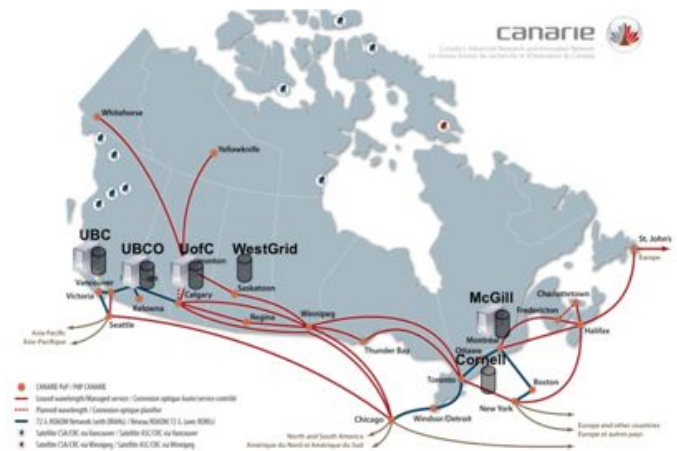
The screenshot displays the CYBERSKA web interface. At the top, the header includes the site name 'CYBERSKA' and a tagline: 'A Cyberinfrastructure platform to meet the needs of data intensive radio astronomy on route to the SKA'. Below the header is a navigation menu with links for Home, Profile, Settings, myDashboard, myGroups, Tools, About, Help, and Administration. A search bar and a 'Log out' button are also present.

The main content area is divided into several widgets:

- Profile:** Shows the user 'Russ Taylor' with options to subscribe to their feed, bookmark their site, and edit the dashboard layout. A list of 'My applications' and 'My pages' is visible.
- Contacts:** A grid of profile pictures of other users.
- Activity:** A feed of recent updates, including comments on 'Flux scale' and event updates for 'DQA regular teleconference'.
- The wire:** A news item titled 'Russ Taylor in Heidelberg at IVOA Interop meeting' with a 'More wire posts' link.
- Event calendar:** A calendar view showing events like 'CALFACTS/POSUM/OSU Project Meetings' (20 May 2013 - 23 May 2013) and 'SKA 2013 Engineering Meeting' (7 Oct 2013 - 11 Oct 2013).
- RSS Feed:** A section titled '* RADIO, etc - Recent Papers' listing scientific articles such as 'Radio astrometry of the outer solar corona' and 'High abundances of complex diisobutylene molecules in UV-illuminated gas'.
- Group membership:** Lists groups the user belongs to, including 'EVLA VDS Deep Field', 'Canadian SKA-SDP Canadian SKA-SDP Team', 'BPI2013 CANARIS BPI docs', and 'Canadian Galactic Plane Survey'.
- Active Users:** Shows a list of currently active users, including 'Russ Taylor' with a recent update about the IVOA meeting.

Distributed Data System

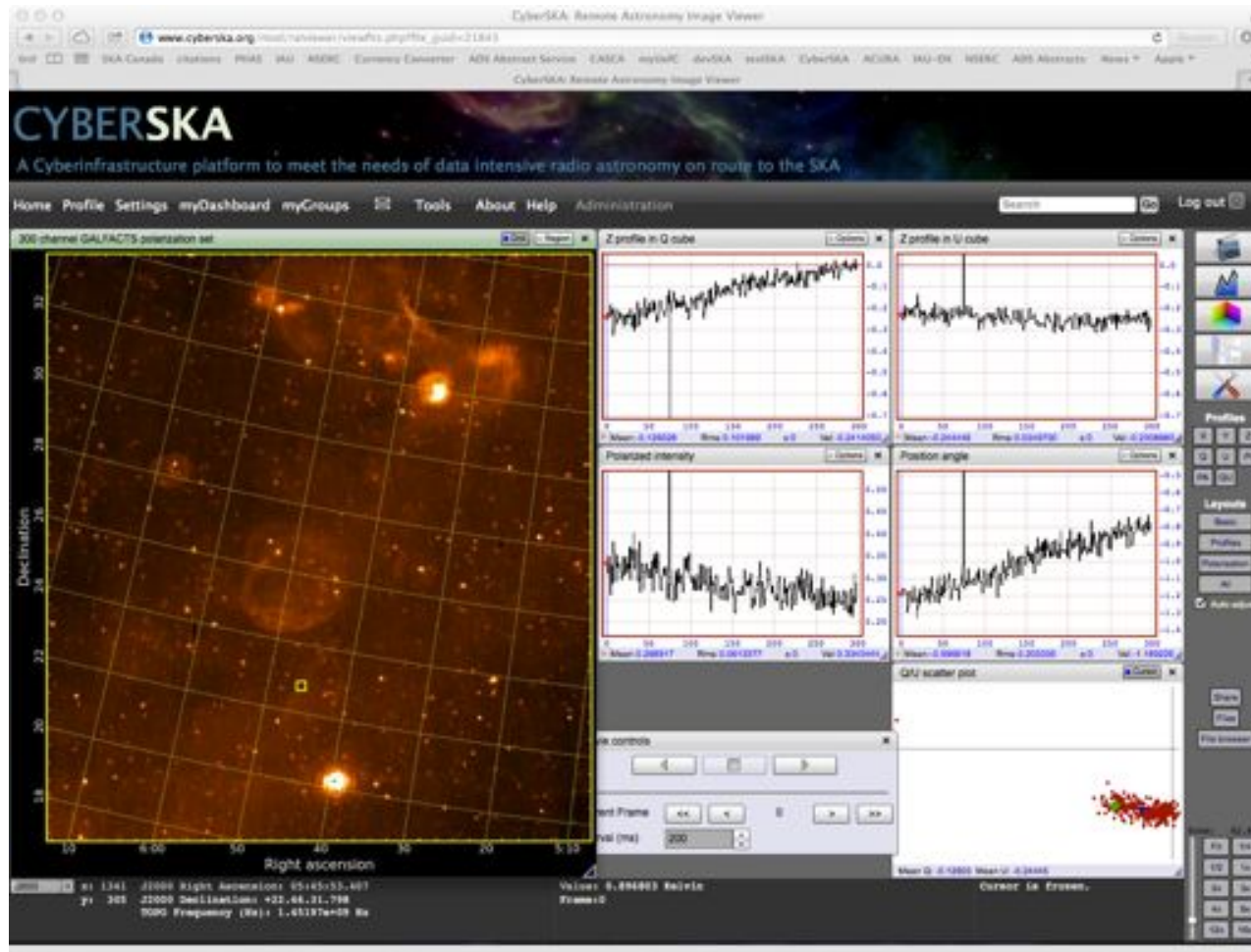
- Based on iRODS (Integrated Rule-Oriented Data System)
 - Abstracts data location
 - Supports data replication / cross-site backup
 - Efficient WAN data transfer
 - Rule engine to automate various tasks
- Upload/download tools
 - Java Applet / Java Web Start based
 - Supports “large” data uploads/downloads
- Automated mime type recognition
 - For many common file types
 - FITS and Measurement Set (CASA) image data or visibility data
 - Automated header extraction and thumbnail generation



On-line visualization of Big Data

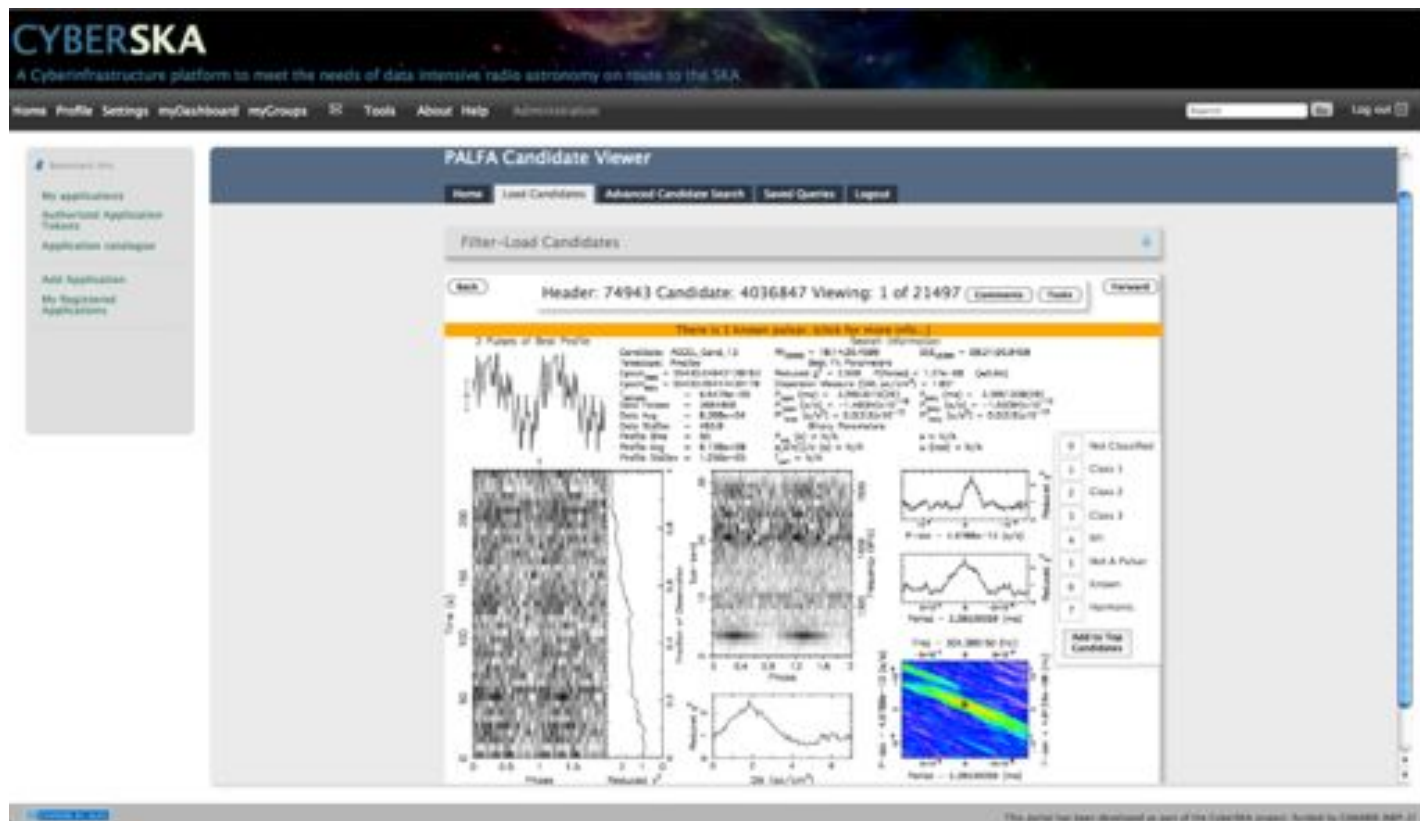
VM based on-line interactive visual analytics of large, multi-dimensional image cubes

- 0.5 TB full Stokes I, Q, U (4D) image cube sets
- Collaboration, screen sharing, platform independent

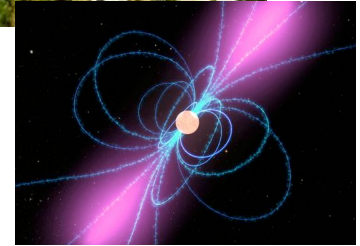
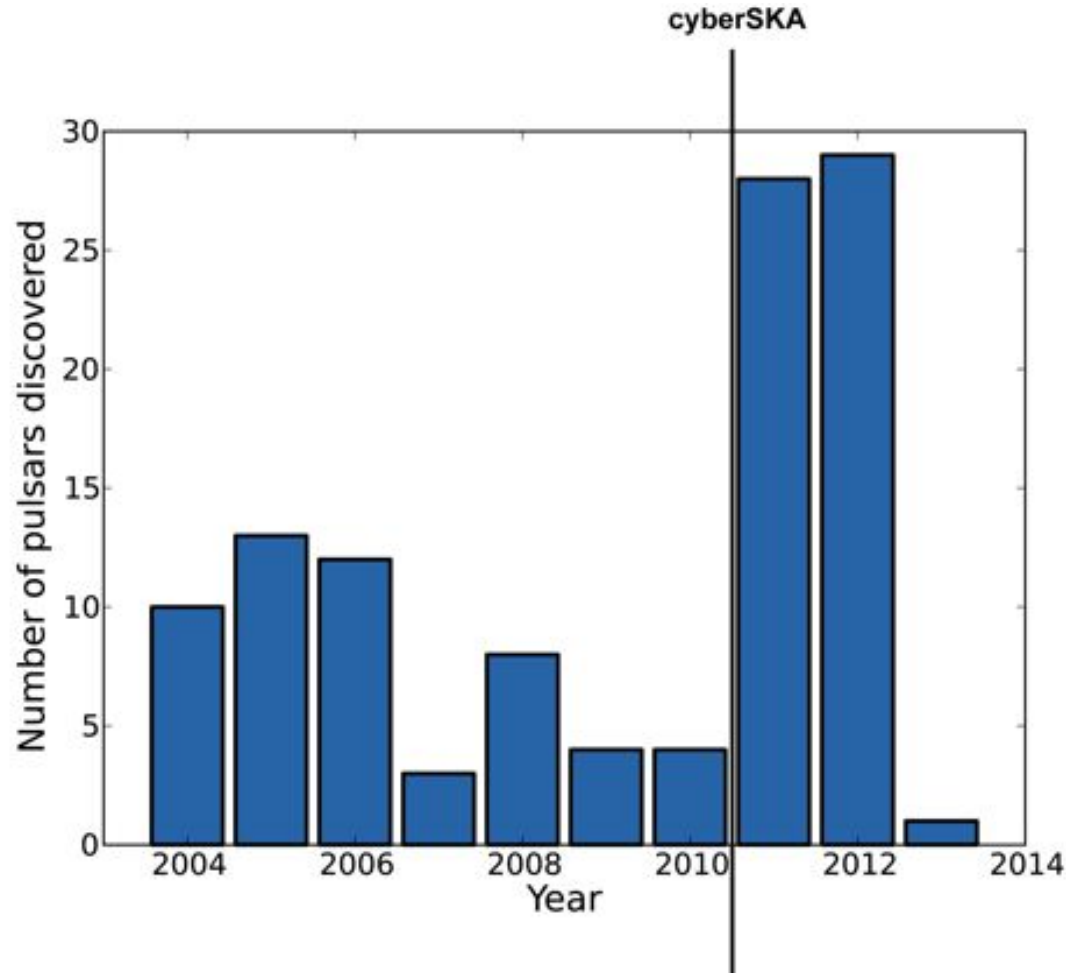


Third Party Application Interface

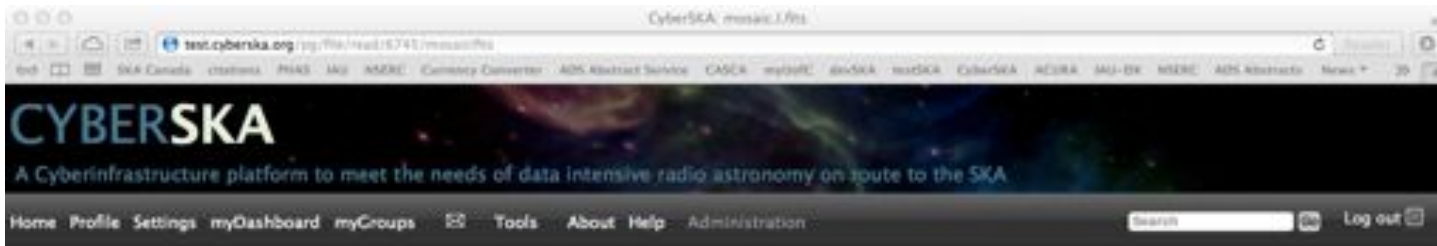
- API for integrating third party / “remote” server applications
- Single sign-on to applications enabled using Oauth
- Push/pull information and data to/from portal



Time-Domain Astronomy Co-development



cyberSKA Data Access Control



Russ Taylor

- Subscribe to feed
- Bookmark this
- My folders
- My files
- My contacts' files
- All site files
- Upload a file
- Advanced Upload

mosaic.l.fits

Russ Taylor
21 January

Deep mosaic image of ELAIS N1, grmt

Download Advanced Download

Image Header

FILETYPE =
SINGLE =
BITPIX =
NAXIS =
NAXIS1 =
NAXIS2 =
NAXIS3 =
NAXIS4 =
BSCALE =
BZERO =
SHAPE = 0.0
BUNIT = 0.0
BPA =
BTYPE =
OBJECT =

Share Manage Replace

Comment

CyberSKA Manage File

test.cyberska.org /sg/Files/Head/15743

Manage File

Title: mosaic.l.fits

Description: Deep mosaic image of ELAIS N1 at 610 MHz from the CMRT

Path:

Tags: elais n1, grmt

Select a parent folder: Main folder

Access: Public

- Enable Virtual Observatory access
- Enable download of this file

Save

Access

Public

- Enable Virtual Observatory access
- Enable download of this file

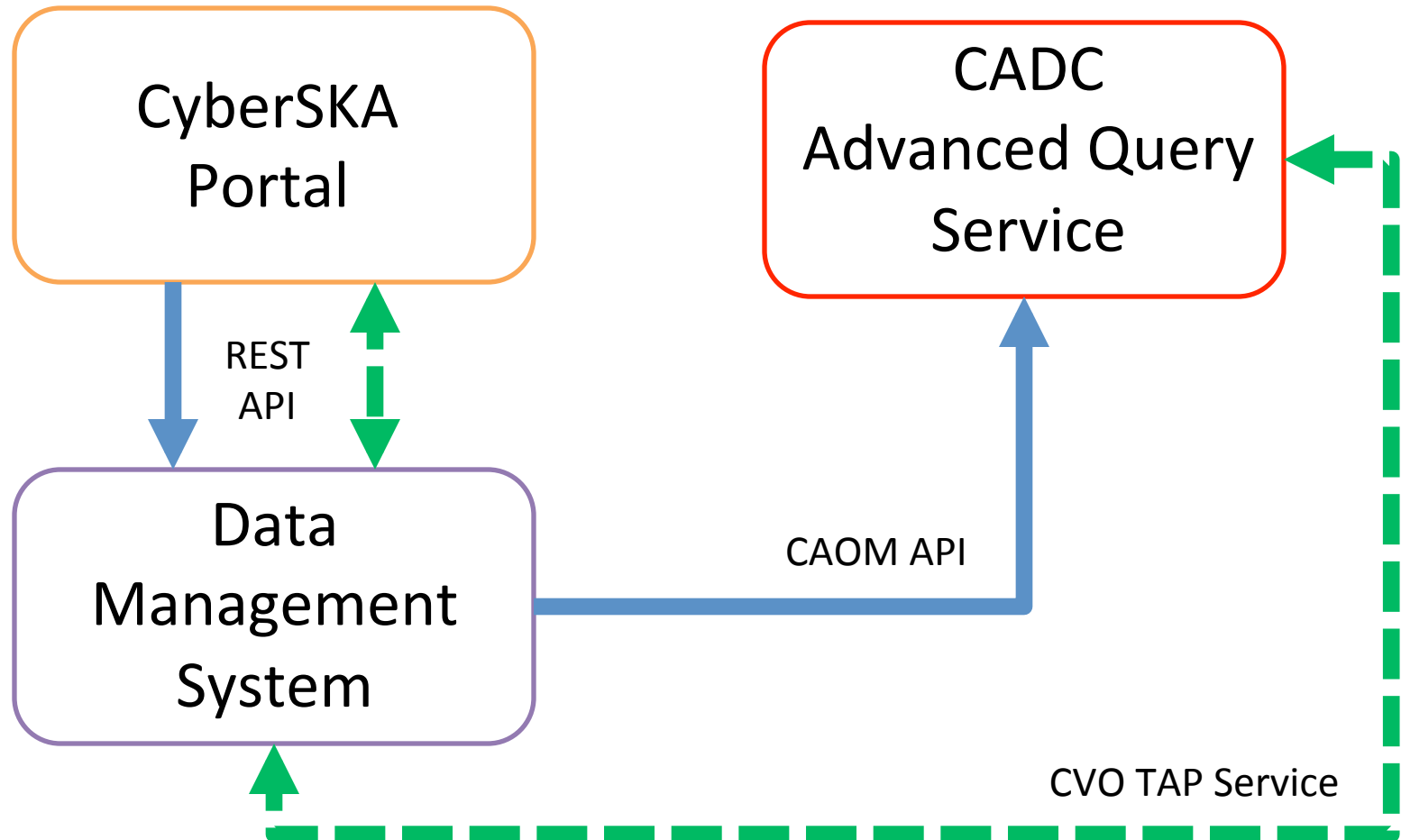
Access

Public

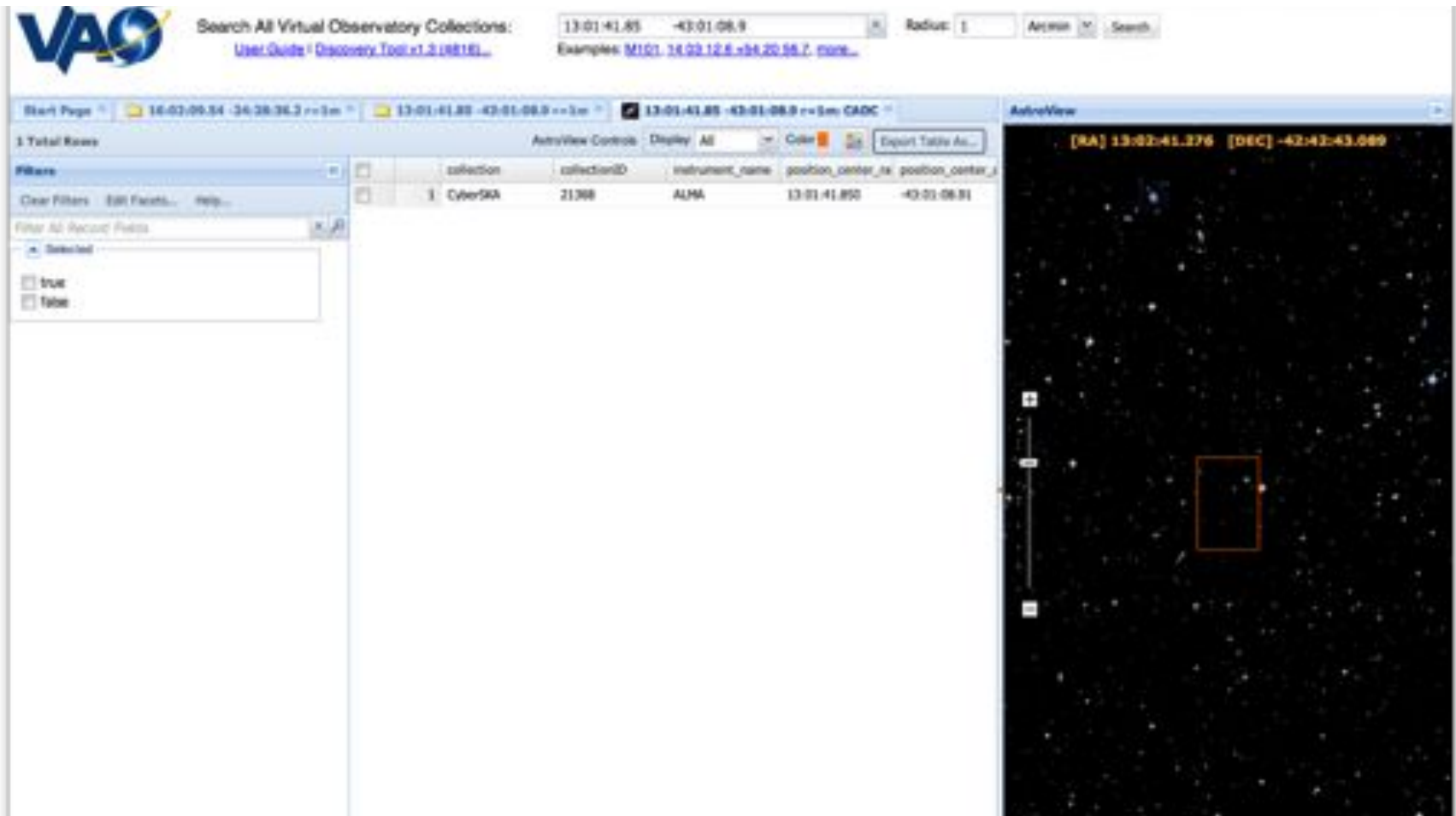
- Enable Virtual Observatory access
- Enable download of this file

Save

VO Interactions



CyberSKA IVOA Collection



VAO Search All Virtual Observatory Collections: 13:01:41.85 -43:01:08.9 Radius: 1 Arcmin Search

User Guide | Discovery Tool v1.3.1 (8/18/13) Examples: M101 14:02:12.4+34:20:58.7 core...

1 Total Rows

collection	collectionID	instrument_name	position_center_ra	position_center_d
1 CyberSKA	21368	ALMA	13:01:41.850	-43:01:08.91

AstroView Controls: Display: All Color: Export Table As...

[RA] 13:02:41.376 [DEC] -43:43:43.089

CyberSKA Usage

- 362 members globally distributed
- 40+ “groups” (GALFACTS, PALFA, RM Synthesis, EVLA Deep Polarization Field, GMRT Deep Polarization Field, CASA Users, ...)



On-going and future work



- Study funded by North American ARC in collaboration with Harvard to adapt on-line visual analytics for ALMA on-line data system
- Further development of on-line collaborative visual analytics of remote Big Data sets
- Completion of user interface for data pipeline tool to allow user developed pipeline processing
- Collaboration with CADC on next generation IVOA interface to include visibility data sets (CAOM2)
- Incorporation of distributed HPC-based cloud architecture for scalability to multi-site petascale (CADC/CANFAR, Compute Canada, IBM Watson).



The Square Kilometre Array

A Global Observatory

A Global Solution to BIG Data