

## Applying the CyberSKA to the ALMA Data Problem

---

**Erik Rosolowsky**, Russ Taylor, Cameron Kiddle, Jim Cordes,  
Olivier Eymere, Victoria Kaspi, Ingrid Stairs, Tony Willis

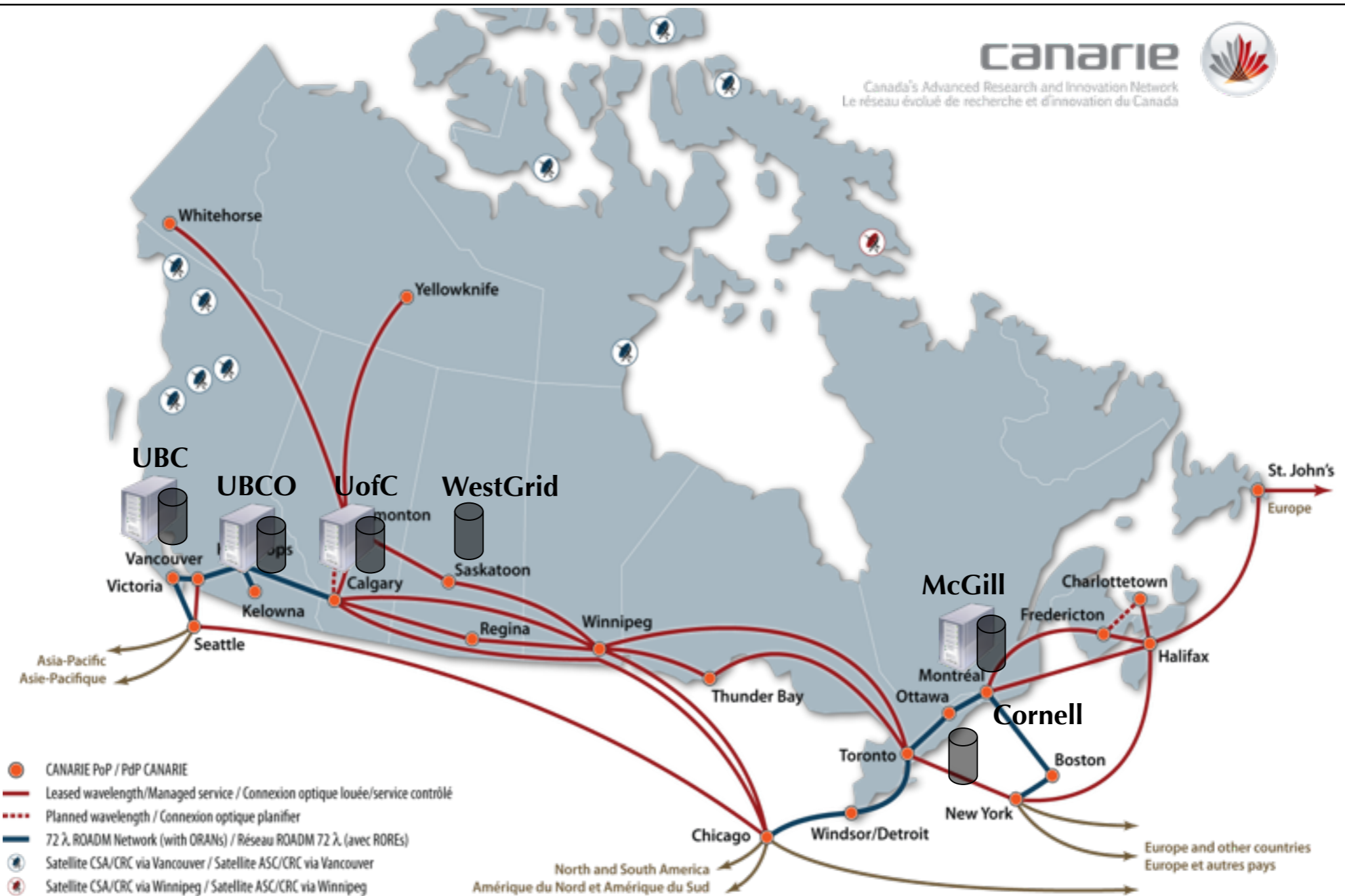


# What is CyberSKA?

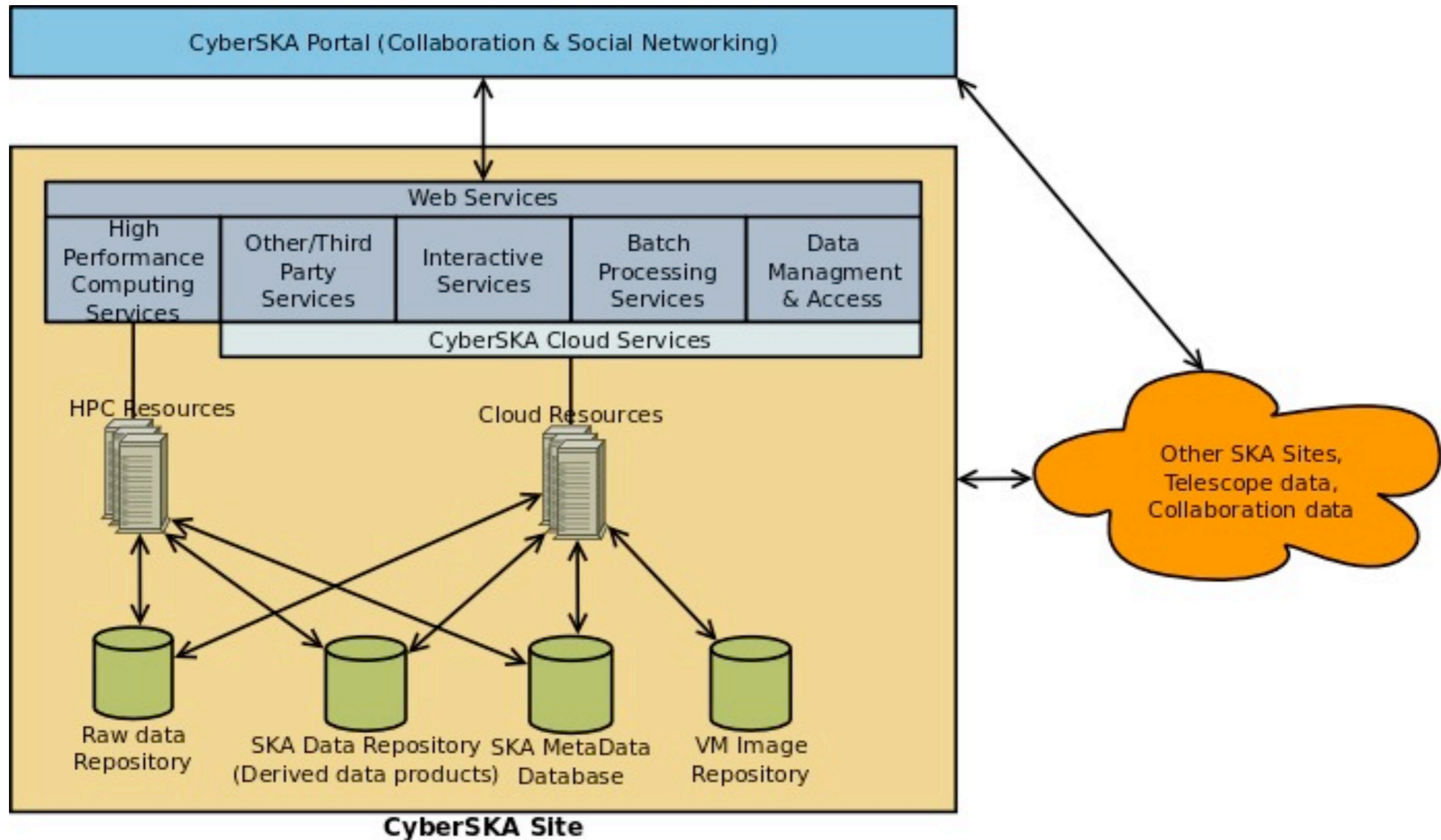
---

- Cyberinfrastructure platform for data-intensive radio astronomy
- Platform support for:
  - Collaboration
  - Data Storage
  - Data Management
  - Data Distribution
  - Processing / Analysis
  - Visualization
- Canadian funding for CyberSKA provided by CANARIE, as part of their Network Enabled Platforms (NEP) program, and Cybera

## CyberSKA Sites



# High Level Architecture



# Collaborative Portal

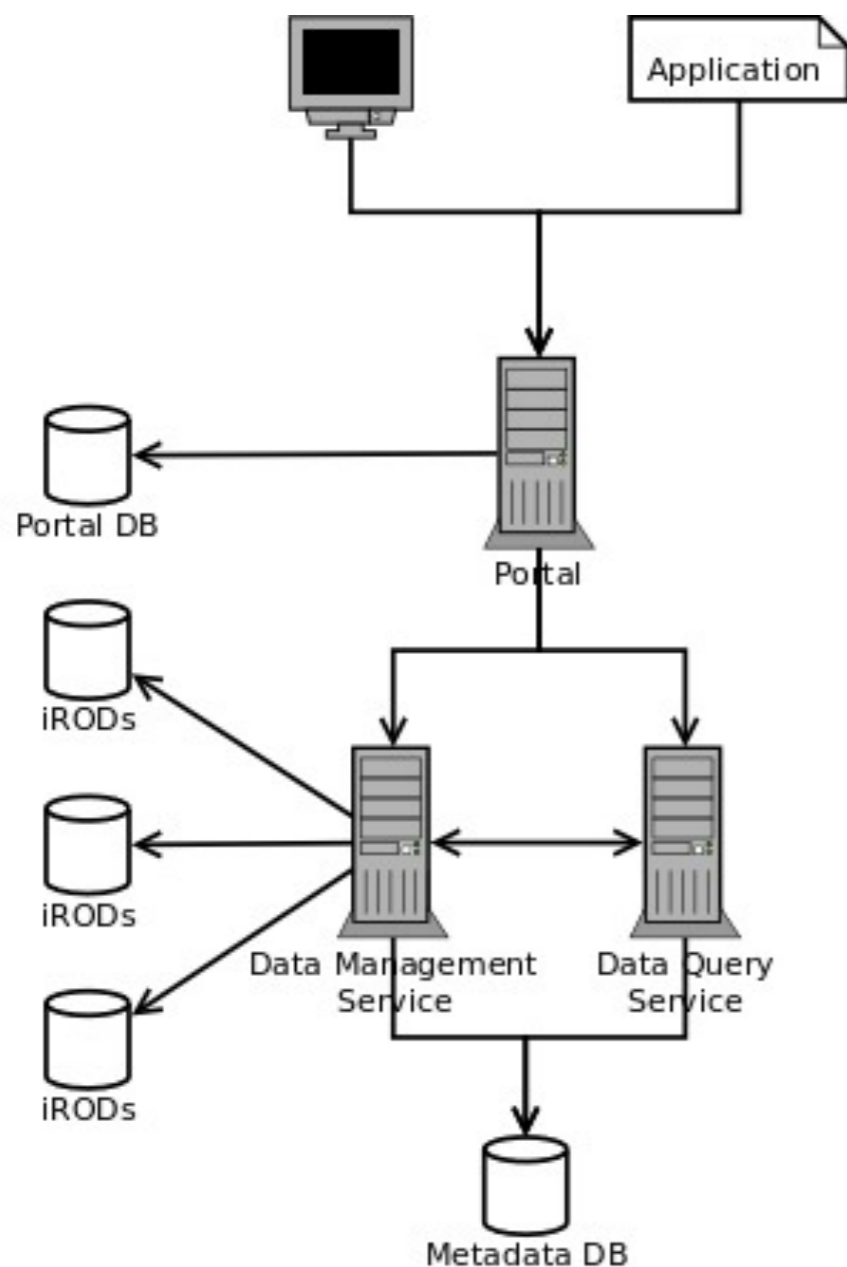
- Portal built on top of the Elgg open source social networking platform

The screenshot displays the CyberSKA Collaborative Portal interface, which is built on the Elgg open source social networking platform. The interface features a dark navigation bar at the top with links for Home, Profile, Settings, myDashboard, myGroups, Tools, About, and Help. A search bar and a Log out button are also present in the top right corner.

The main content area is organized into several widgets:

- Left Sidebar (User Profile):** Displays the user's name (Demo User) and options to subscribe to the feed or bookmark the page. It also includes a link to edit the dashboard layout and a list of user-specific links: My applications, My blog, My bookmarked items, My event calendar, My files, My pages, My publications, and My tasked items. At the bottom, it shows My activity, Contact's activity, and Site activity.
- Group membership:** A list of groups the user is a member of, including GALFACTS Pipeline, CASA Pipeline, Application Developers, CASA Users, PALFA Survey Group, GALFACTS Consortium, and GALFACTS Technical Team.
- Activity:** A section for recent updates, currently showing a contact update by Cameron Kiddle regarding the CyberSKA CANARIE NEP Progress.
- Event calendar:** A section for upcoming events, currently empty.
- Applications:** A list of applications registered by users, including Source Counts Application, GALFACTS Pipeline, CASA Pipeline - Test Application, and CyberSKA Visibility Processing Pipeline.
- Recent Astro-ph Eprints:** A list of recent astronomical eprints, including articles on gravitational wave detection, TeV neutrinos, and radio pulsars.
- Active Users:** A section showing the current active user, Demo User, and their recent activity, including a link to IBM's BigSheets for Big Data.
- Bookmarks:** A section for user bookmarks, currently showing two movie player demo files.

# Distributed Data Management System



- Based on iRODS (Integrated Rule-Oriented Data System)
  - Data locate abstracted from users
  - Automatic replication
  - Use-based distribution
- Advanced upload/download tools
- Portal recognizes FITS and CASA uvw and image datasets as astronomical formats

# Distributed Data Management System

CYBERSKA

A Cyberinfrastructure platform to meet the needs of data intensive radio astronomy on route to the SKA

[Home](#) [Profile](#) [Settings](#) [myDashboard](#) [myGroups](#) [Tools](#) [About](#) [Help](#)

[Log out](#)

Demo User

[Subscribe to feed](#)  
[Bookmark this](#)

**My folders**  
**My files**  
Your contacts' files  
All site files  
Upload a file  
Advanced Upload

All  
General  
FITS images

Search files

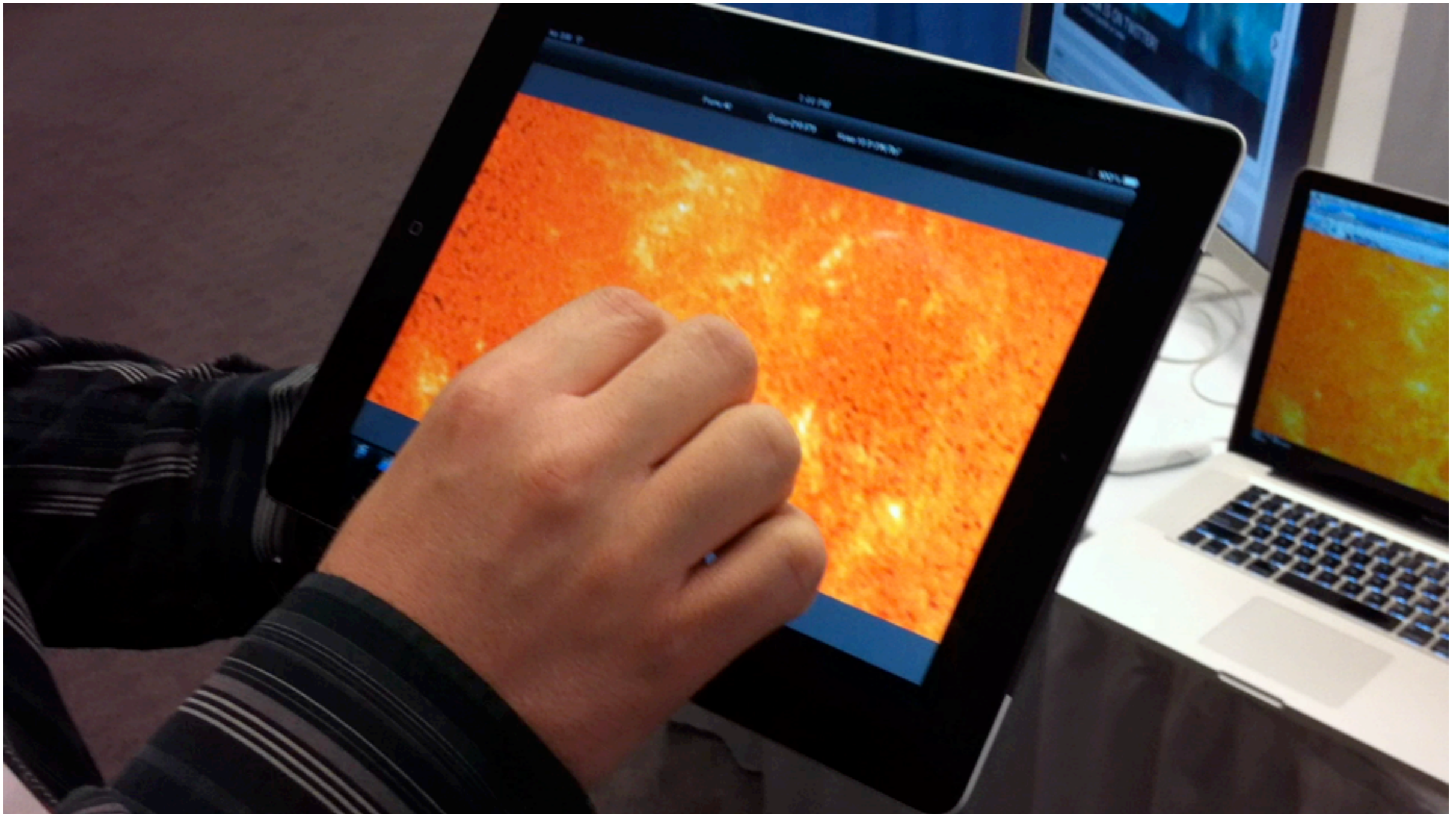
**My files**

12Next »

	Antennae_South.cal.ms.tar.gz Demo User 58 minutes ago	<a href="#">Manage</a> <a href="#">Delete</a>
	2229.cont.im.fits Demo User 43 days ago	<a href="#">Manage</a> <a href="#">Delete</a>
	CANARIE Site Visit Test Demo User 43 days ago	<a href="#">Manage</a> <a href="#">Delete</a>
	g192_src.split.im.fits Demo User 45 days ago	<a href="#">Manage</a> <a href="#">Delete</a>
	G24_92A_11.fits Demo User 46 days ago	<a href="#">Manage</a> <a href="#">Delete</a>
	g19_12coall.fits Demo User 46 days ago	<a href="#">Manage</a> <a href="#">Delete</a>
	3C84.fits Demo User 46 days ago	<a href="#">Manage</a> <a href="#">Delete</a>
	3c273.im.fits5 Demo User 96 days ago	<a href="#">Manage</a> <a href="#">Delete</a>
	b2.spw_3.corrected.im.fits Demo User 97 days ago	<a href="#">Manage</a> <a href="#">Delete</a>
	SKA2011DemoSlides.ppt Demo User 97 days ago	<a href="#">Manage</a> <a href="#">Delete</a>

## Visualization

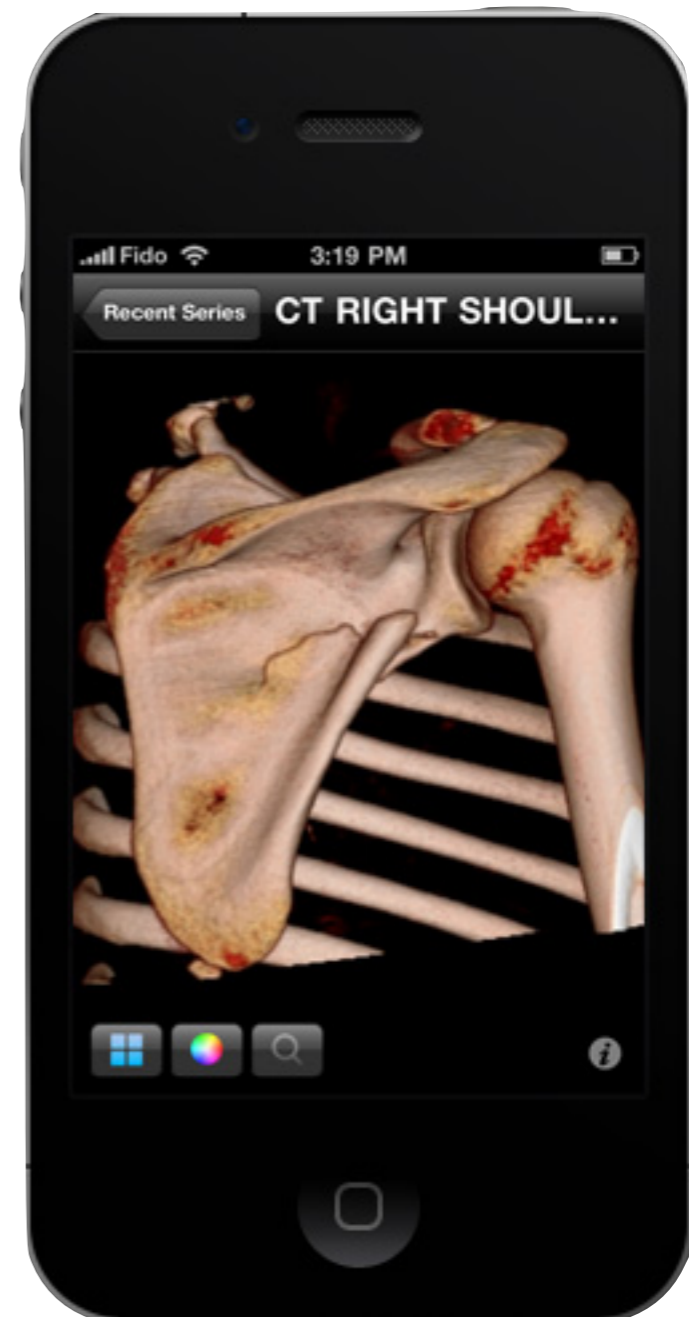
---



# Visualization

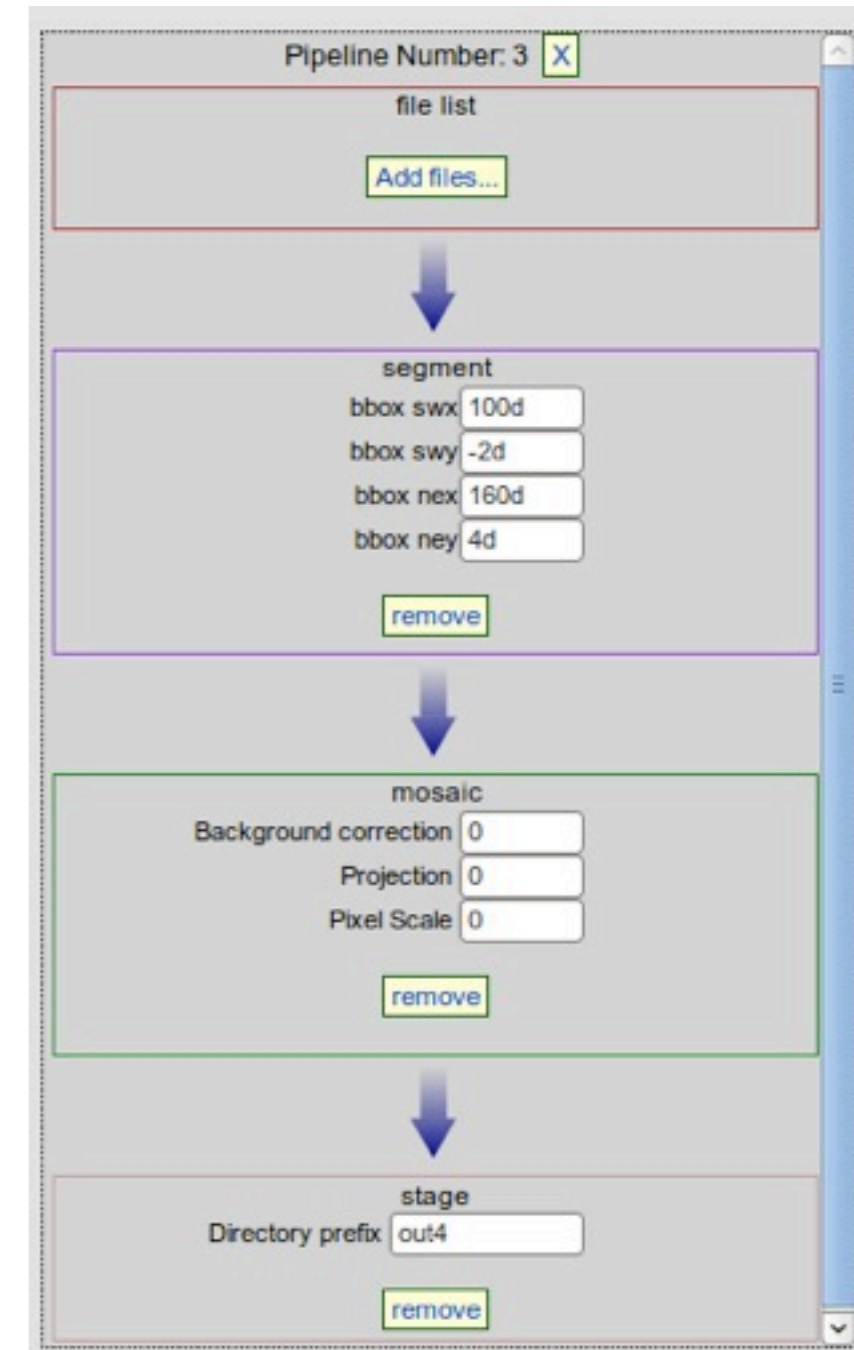
---

- Visualization tools have emphasize server-side back end and collaborative aspects
- Partnership with Calgary Scientific
- Astronomically-savvy medical imaging platforms being implemented



# Building an ALMA Ready System

- ❑ Extraction of metadata into VOTable-compliant format
- ❑ Builds a database for queries astronomical files
  - Rapid metadata-based search
  - Location, Time, Frequency queries
  - Object, Telescope, Quality queries
  - MS / FITS uvw data native



# Building an ALMA Ready System

- ❑ Access/download subcubes
- ❑ Requested data generated in VM pool managed with Condor

The screenshot displays the CYBERSKA web interface. At the top, the CYBERSKA logo is followed by the tagline "A Cyberinfrastructure platform to meet the needs of data intensive radio astronomy on route to the SKA". Below this is a navigation bar with links: Home, Profile, Settings, myDashboard, myGroups, Tools, About, and Help. A search bar and a "Log out" button are also present.

The main content area is divided into two panels. The left panel, titled "GALFACTS Consortium", contains a sidebar with links for "GALFACTS Consortium Group", "Bookmark this", "Create new download", "Download requests", "Group pages", "Group bookmarks", "Group files", "Group blog", "Group calendar", "Group tasks", "Group applications", "Group data", and "Data Discussion".

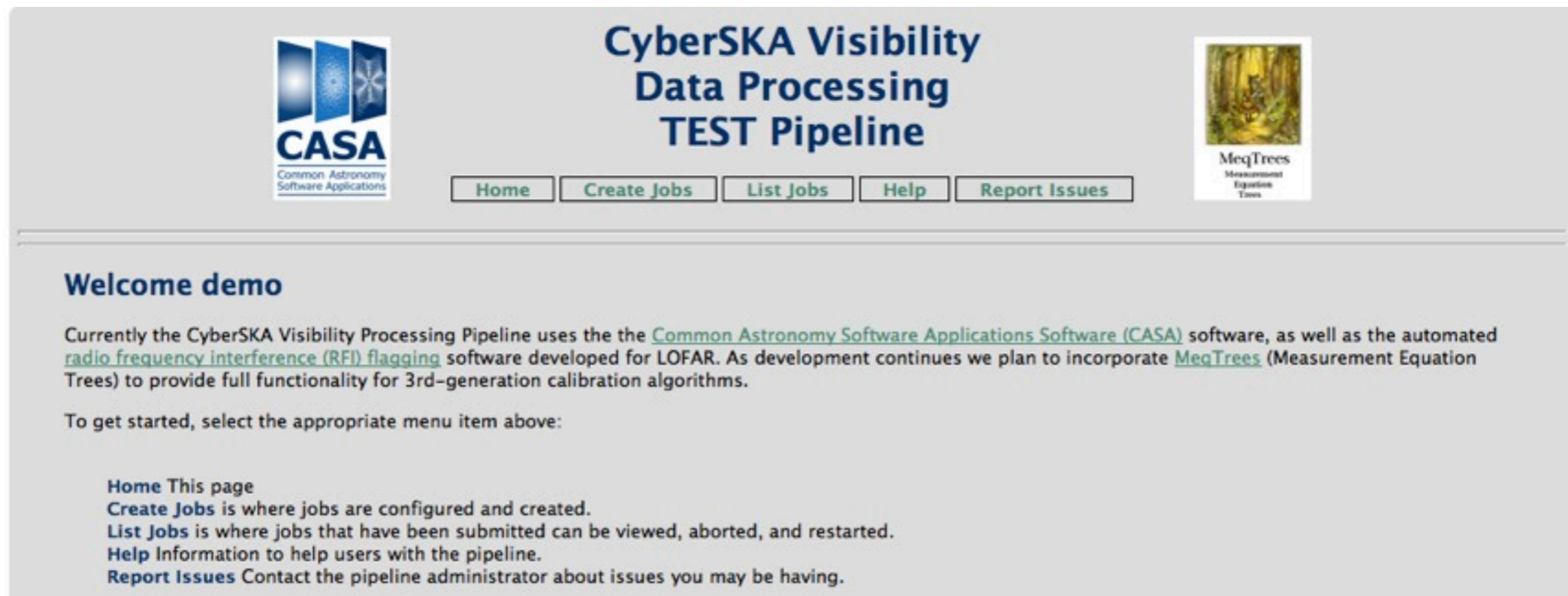
The right panel, titled "10ALFA D Field", shows a large image of a radio astronomy field with a green rectangular region of interest. Below the image are various controls for data selection and download. These include a "Display cube" dropdown set to "I", a "Cubes" section with checkboxes for "I", "Q", "U", "V", and "W", and a "Download information" box showing an estimated download size of 145.54 MB for 2 cubes (943 x 289 x 70 (1750/25)).

At the bottom, there are controls for "Frequency range" (Start channel #: 0, End channel #: 1749) and "Spectral averaging" (Averaging width  $\Delta n$ : 25,  $\Delta f$ : 1.22 MHz). "Submit" and "Cancel" buttons are at the very bottom.

# Building an ALMA Ready System

---

- A workflow-based interface to CASA
- Operates on measurement sets in the DMS
- Uses HPC resources linked with storage volumes
- Overlap with CANFAR Project @ CADC



The screenshot shows the web interface for the CyberSKA Visibility Data Processing TEST Pipeline. At the top, there is a header with the CASA logo on the left, the title "CyberSKA Visibility Data Processing TEST Pipeline" in the center, and the MeqTrees logo on the right. Below the title is a navigation bar with five buttons: "Home", "Create Jobs", "List Jobs", "Help", and "Report Issues". The main content area starts with a "Welcome demo" section. It contains a paragraph explaining that the pipeline uses CASA software and automated radio frequency interference (RFI) flagging software. It also mentions plans to incorporate MeqTrees for 3rd-generation calibration algorithms. Below this, a line of text says "To get started, select the appropriate menu item above:". At the bottom, there is a list of links with descriptions: "Home This page", "Create Jobs is where jobs are configured and created.", "List Jobs is where jobs that have been submitted can be viewed, aborted, and restarted.", "Help Information to help users with the pipeline.", and "Report Issues Contact the pipeline administrator about issues you may be having."

**CyberSKA Visibility  
Data Processing  
TEST Pipeline**

[Home](#) [Create Jobs](#) [List Jobs](#) [Help](#) [Report Issues](#)

**Welcome demo**

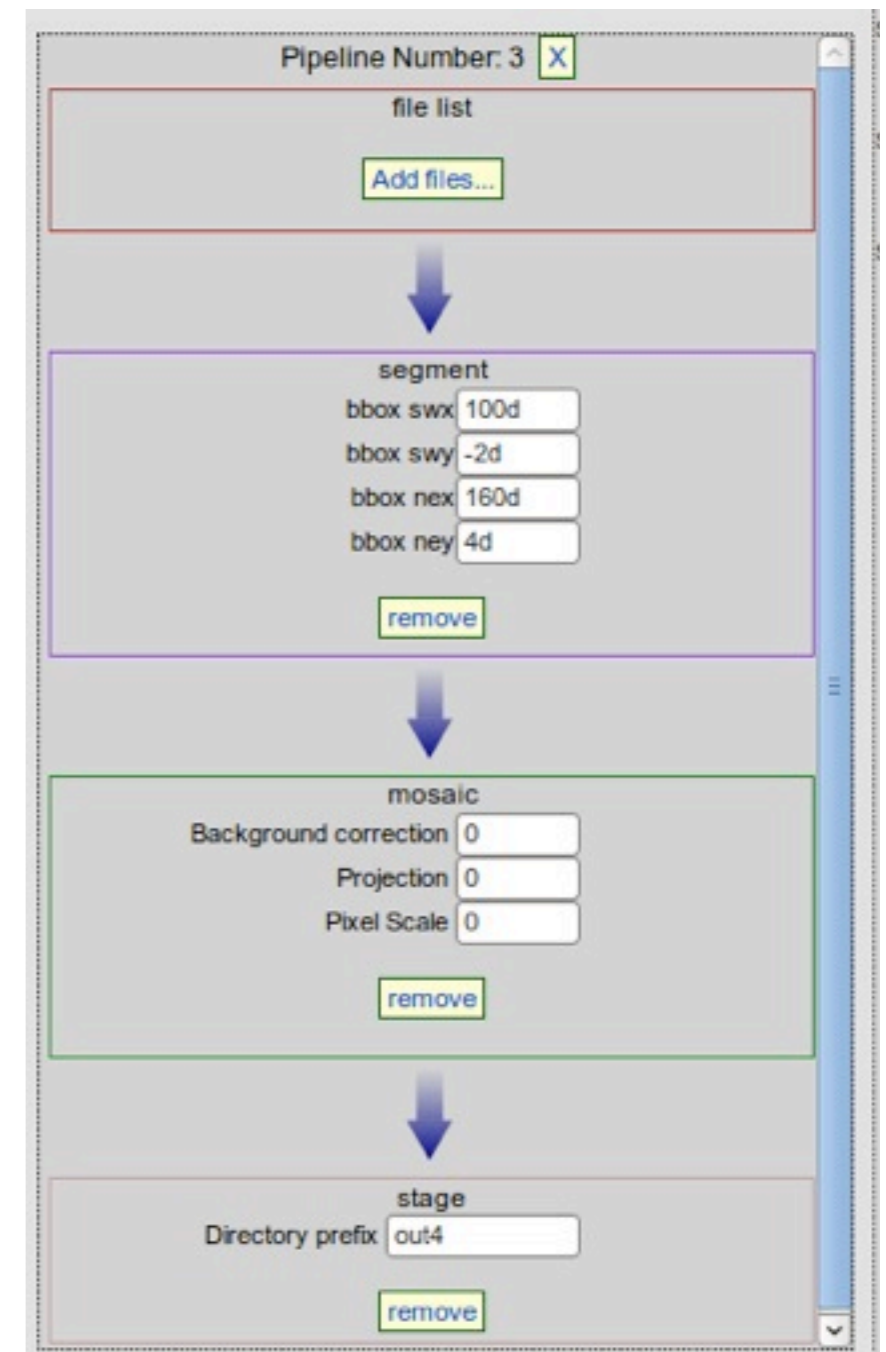
Currently the CyberSKA Visibility Processing Pipeline uses the the [Common Astronomy Software Applications Software \(CASA\)](#) software, as well as the automated [radio frequency interference \(RFI\) flagging](#) software developed for LOFAR. As development continues we plan to incorporate [MeqTrees](#) (Measurement Equation Trees) to provide full functionality for 3rd-generation calibration algorithms.

To get started, select the appropriate menu item above:

**Home** This page  
**Create Jobs** is where jobs are configured and created.  
**List Jobs** is where jobs that have been submitted can be viewed, aborted, and restarted.  
**Help** Information to help users with the pipeline.  
**Report Issues** Contact the pipeline administrator about issues you may be having.

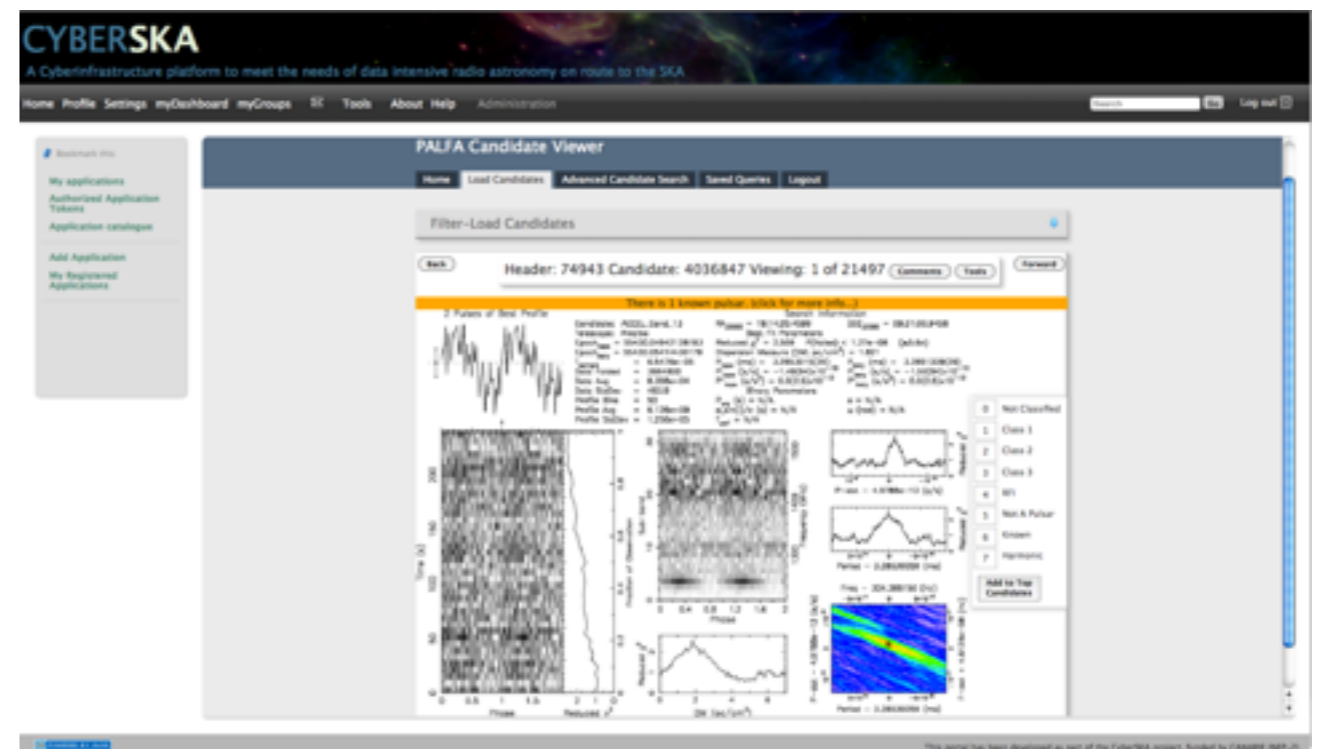
# Building an ALMA Ready System

- On-site image manipulation workflows (e.g., segmentation, mosaicking, reprojection)
- Future workflow development will incorporate the inclusion of user submitted Python- and C-based modules
- Dynamically spawned VMs for processing



# Extensible via Third-Party Apps

- API for integrating third party / remotely hosted applications
- Single sign-on to applications enabled using OAuth
- Current applications include PALFA Candidate Viewer, PALFA Top Candidates, PALFA Observation Scheduler, PALFA Diagnostics Tool, GALFACTS Pipeline, Source Counts



# ALMA Opportunities

---

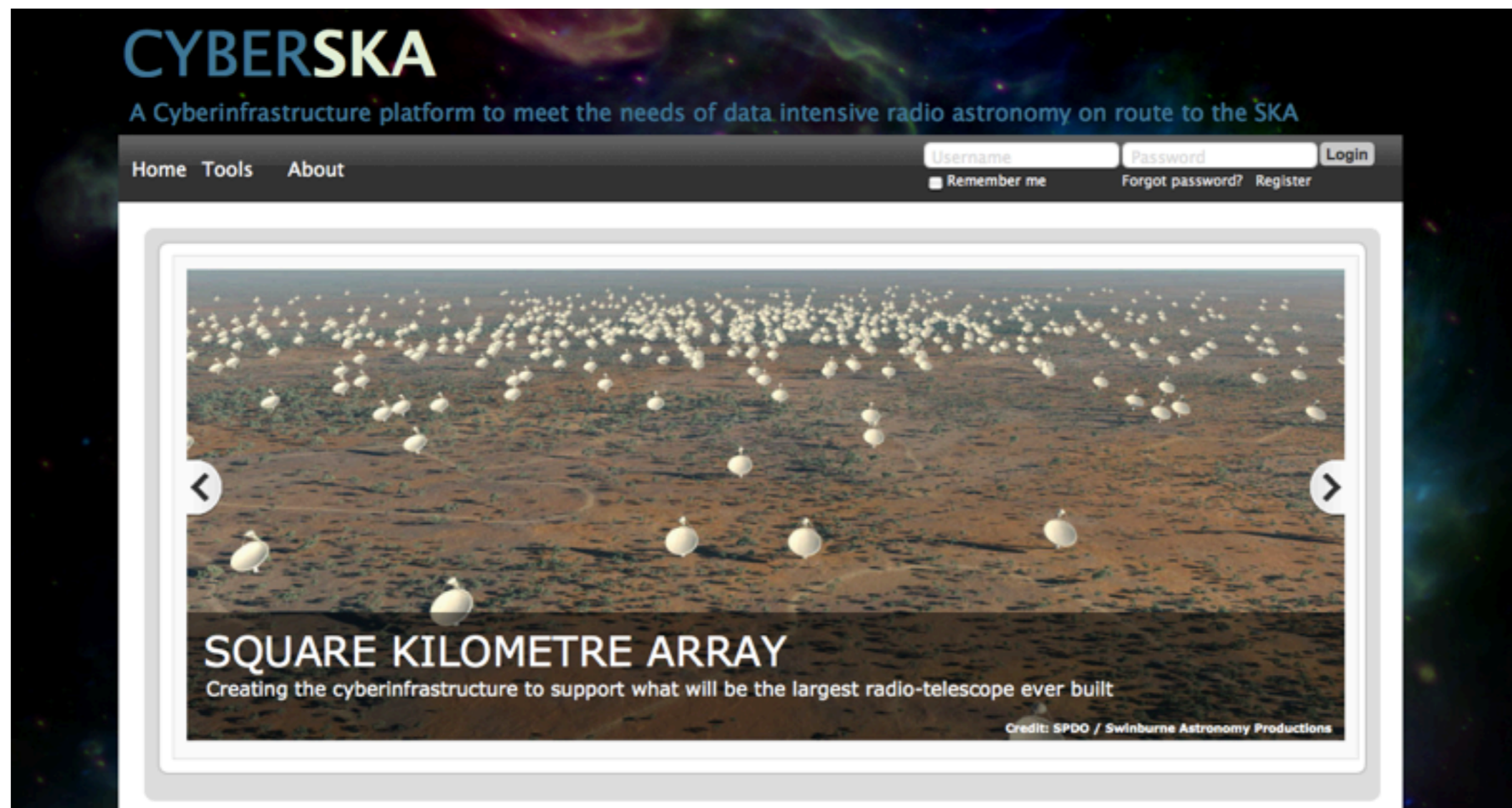
- Cyberinfrastructure platform directly addresses early ALMA needs
  - Archives can be mirrored into distributed data system
  - Development of CASA pipeline and re-reduction capacities
  - Visualization platform
  - Collaboration platform

## Contact Information

---

Portal: <http://www.cyberska.org/>

E-mail: [info@cyberska.org](mailto:info@cyberska.org)



You are invited to the ALMA Integration Group