CYBERSKA

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

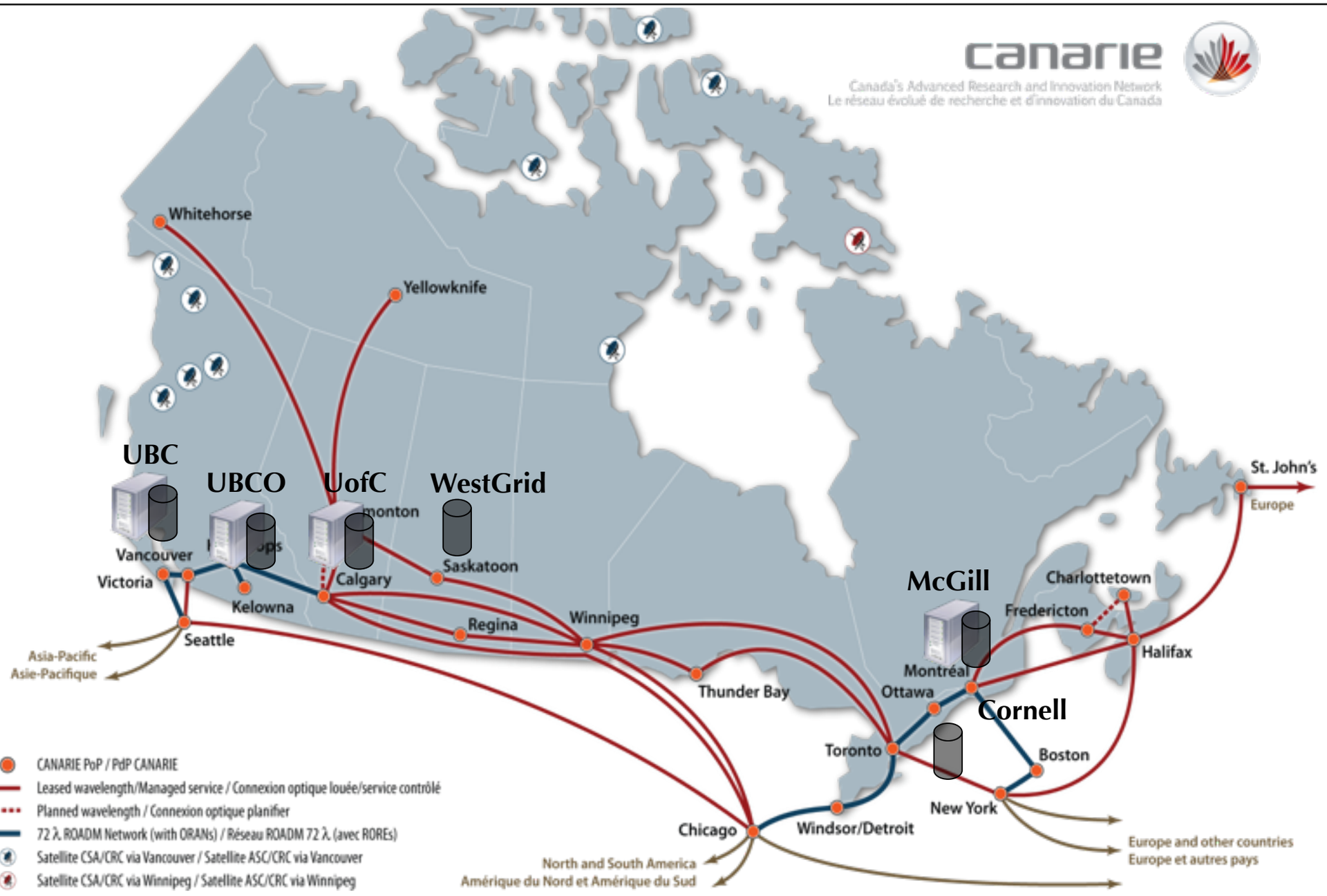
Applying the CyberSKA to the ALMA Data Problem

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

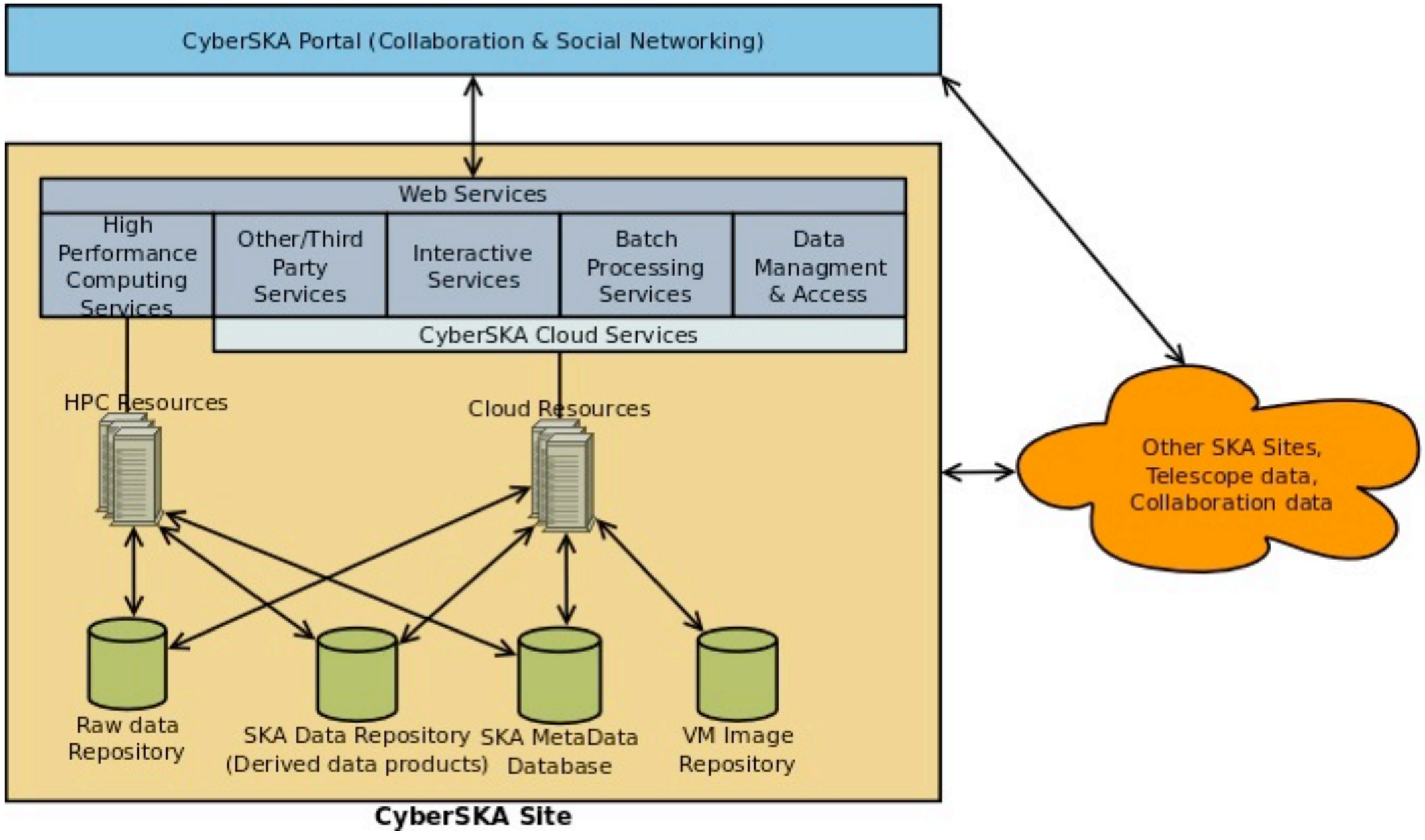
What is CyberSKA?

- 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

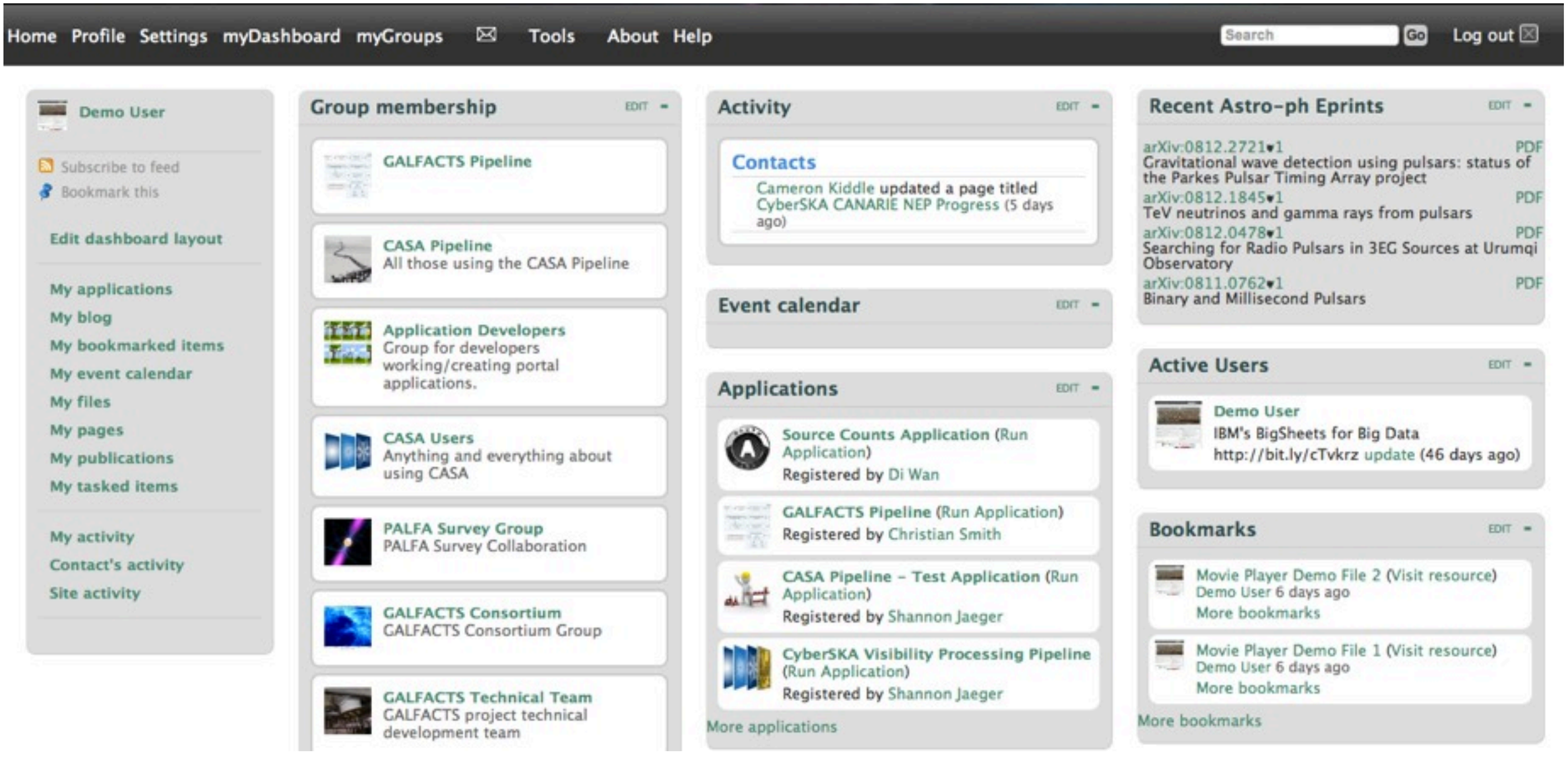


High Level Architecture

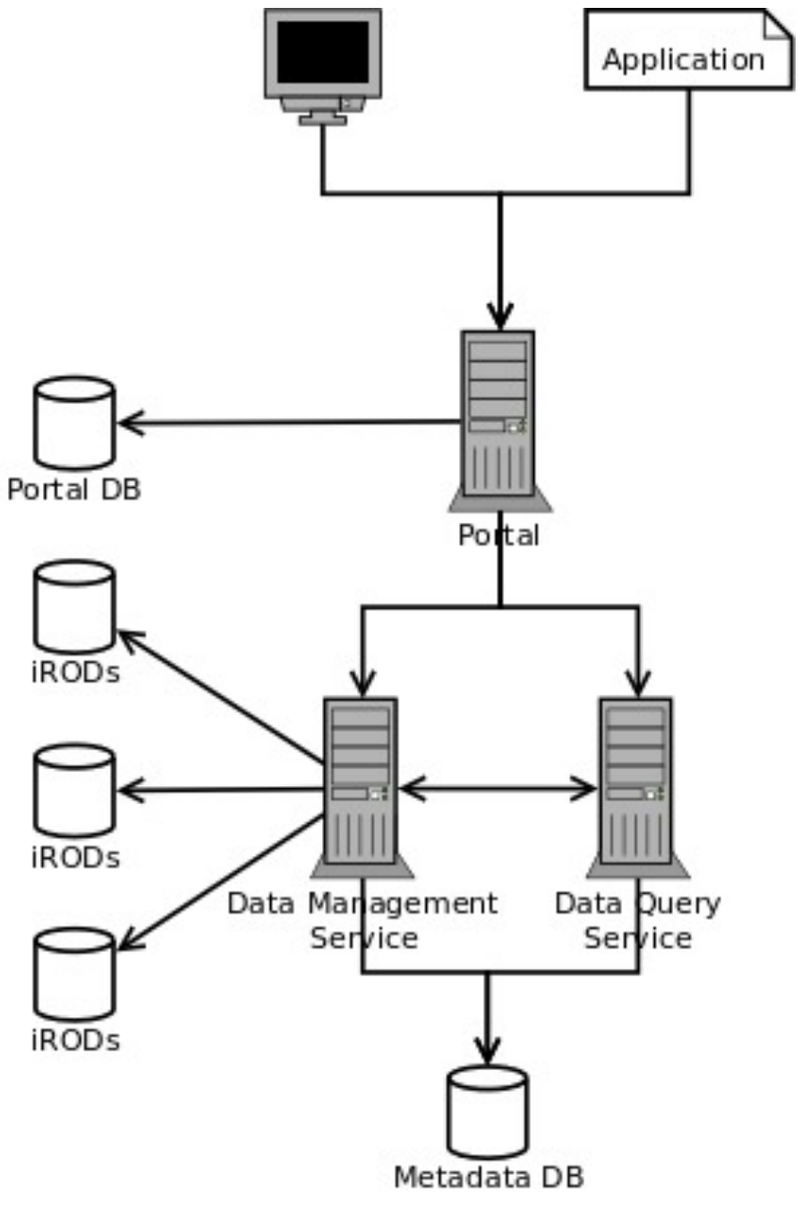


Collaborative Portal

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

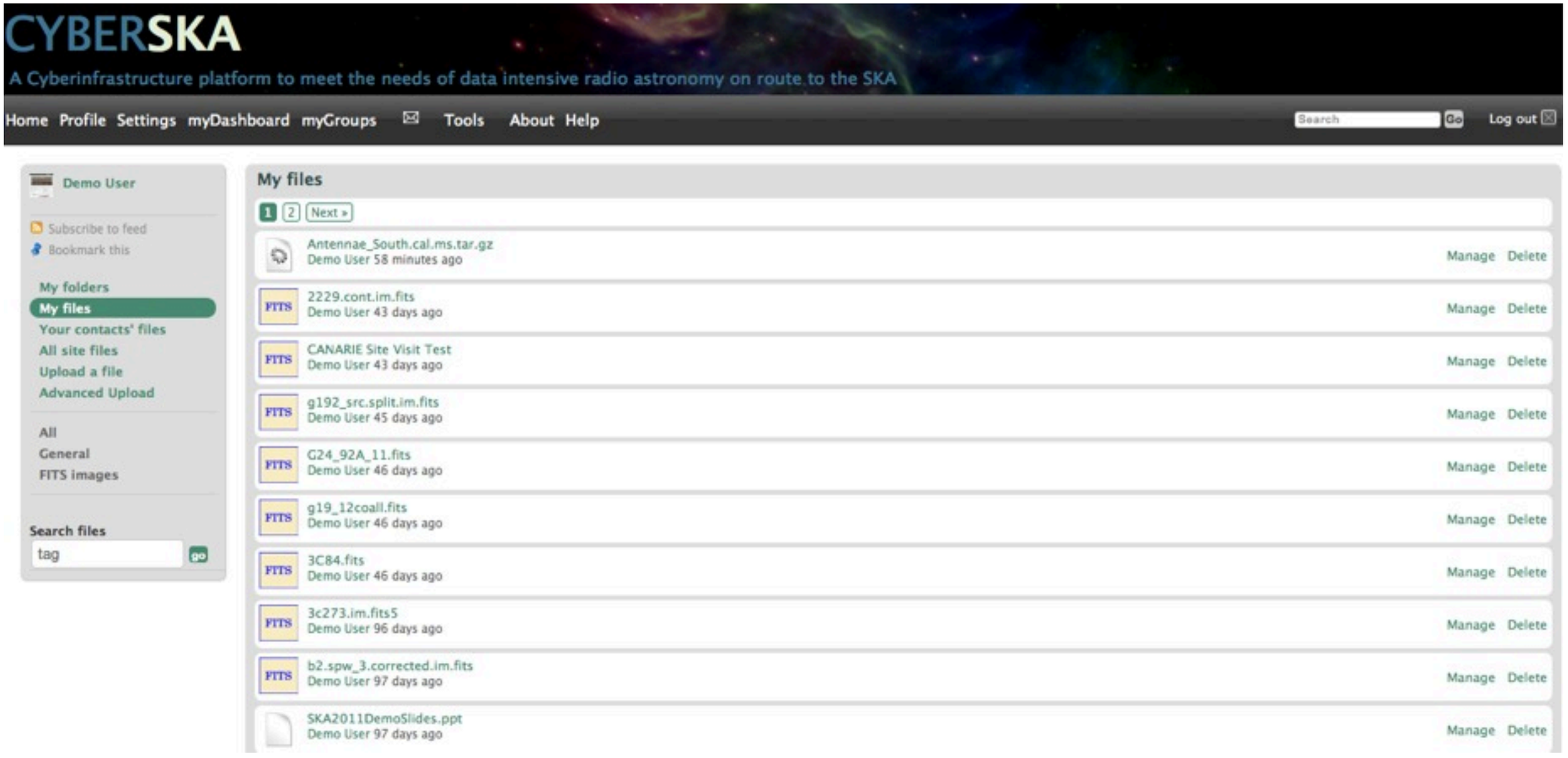


Distributed Data Management System

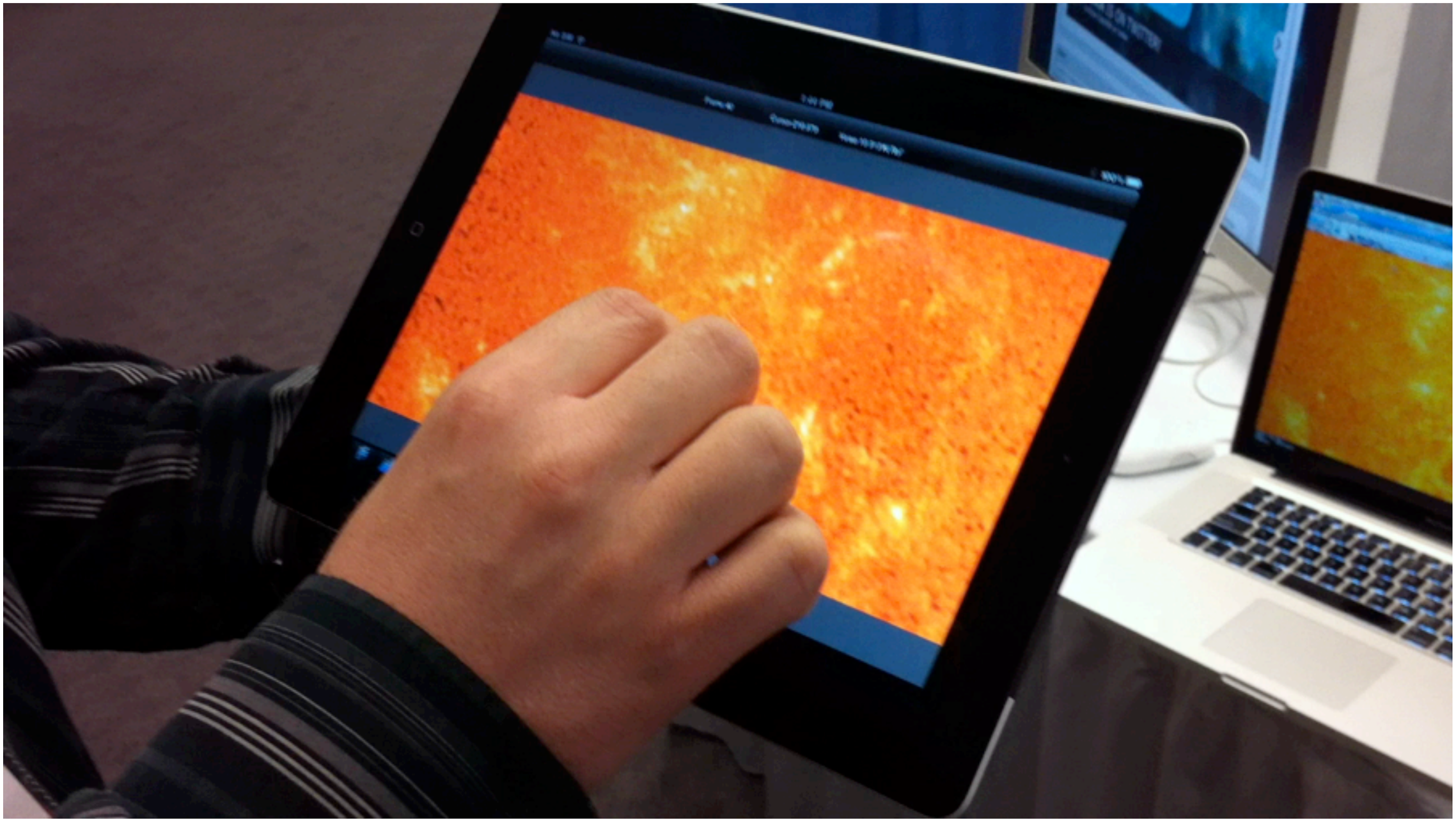


- Based on iRODS (Integrated Rule-Oriented Data System)
  - Data locate abstracted from users
  - Automatic replication
  - Use-based distribution
- Advanced upload/download tools
- Portal recognizes FITS and CASA uvw and image datasets as astronomical formats

Distributed Data Management System

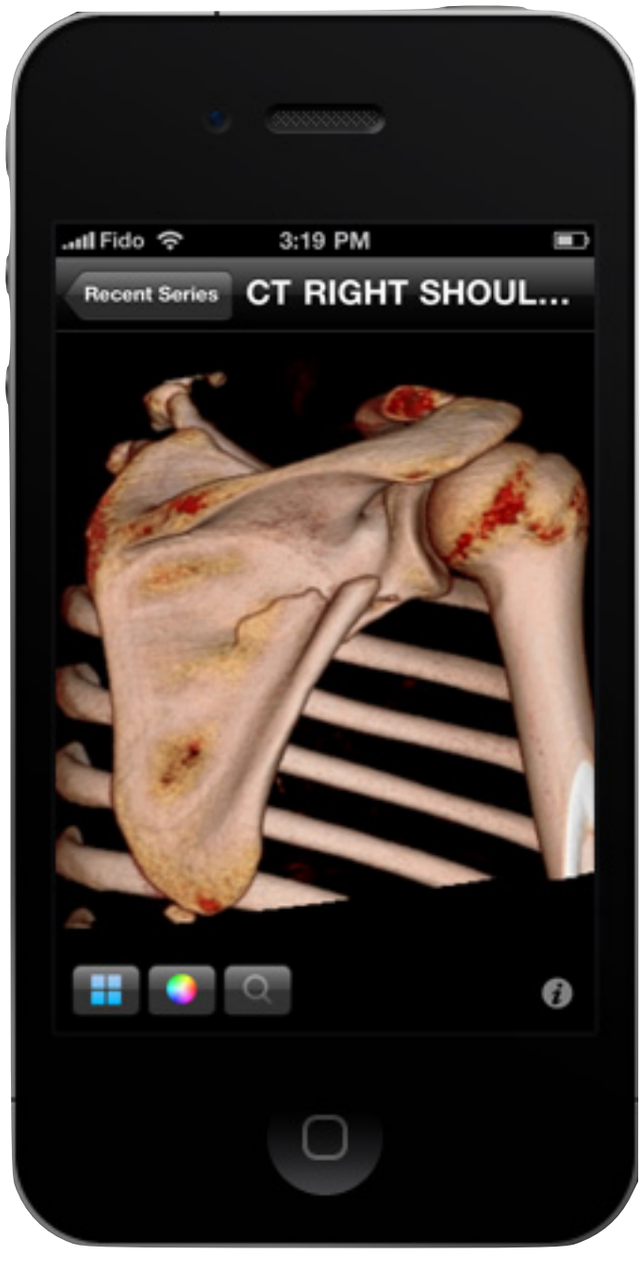


Visualization



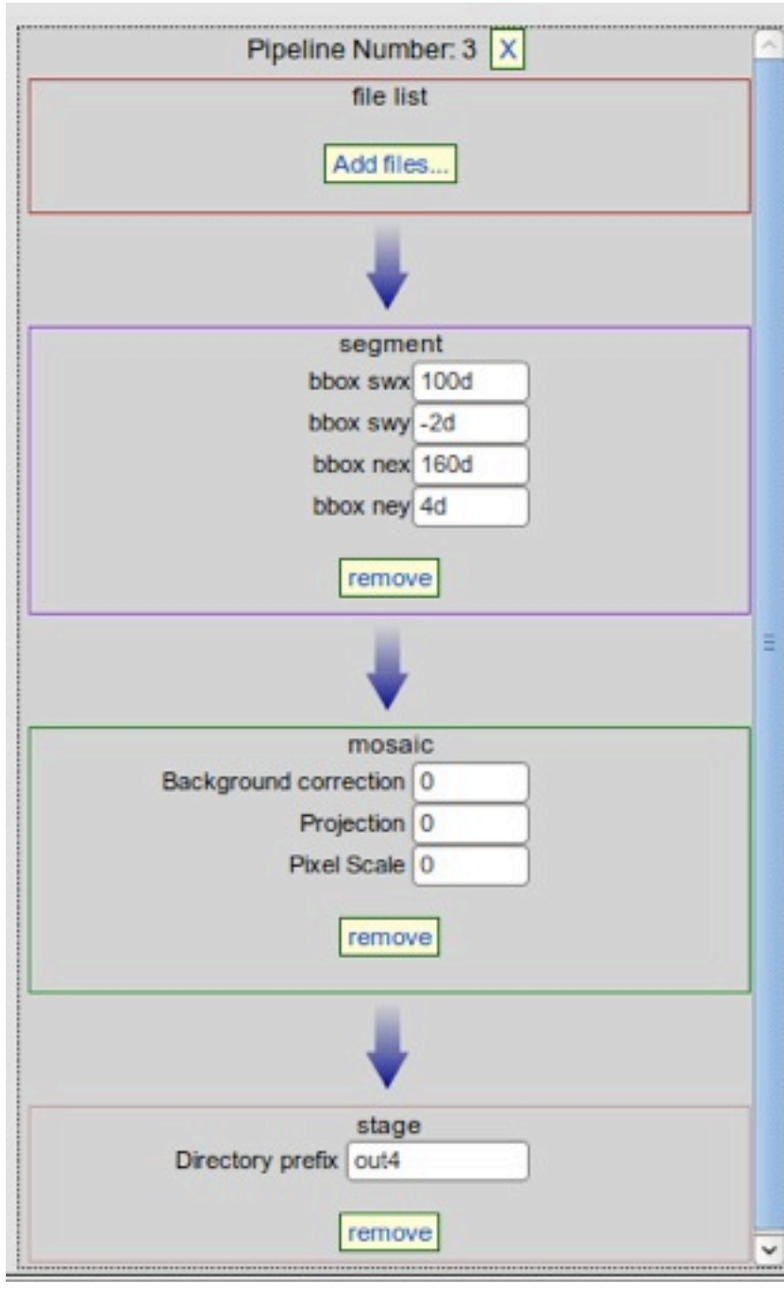
Visualization

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



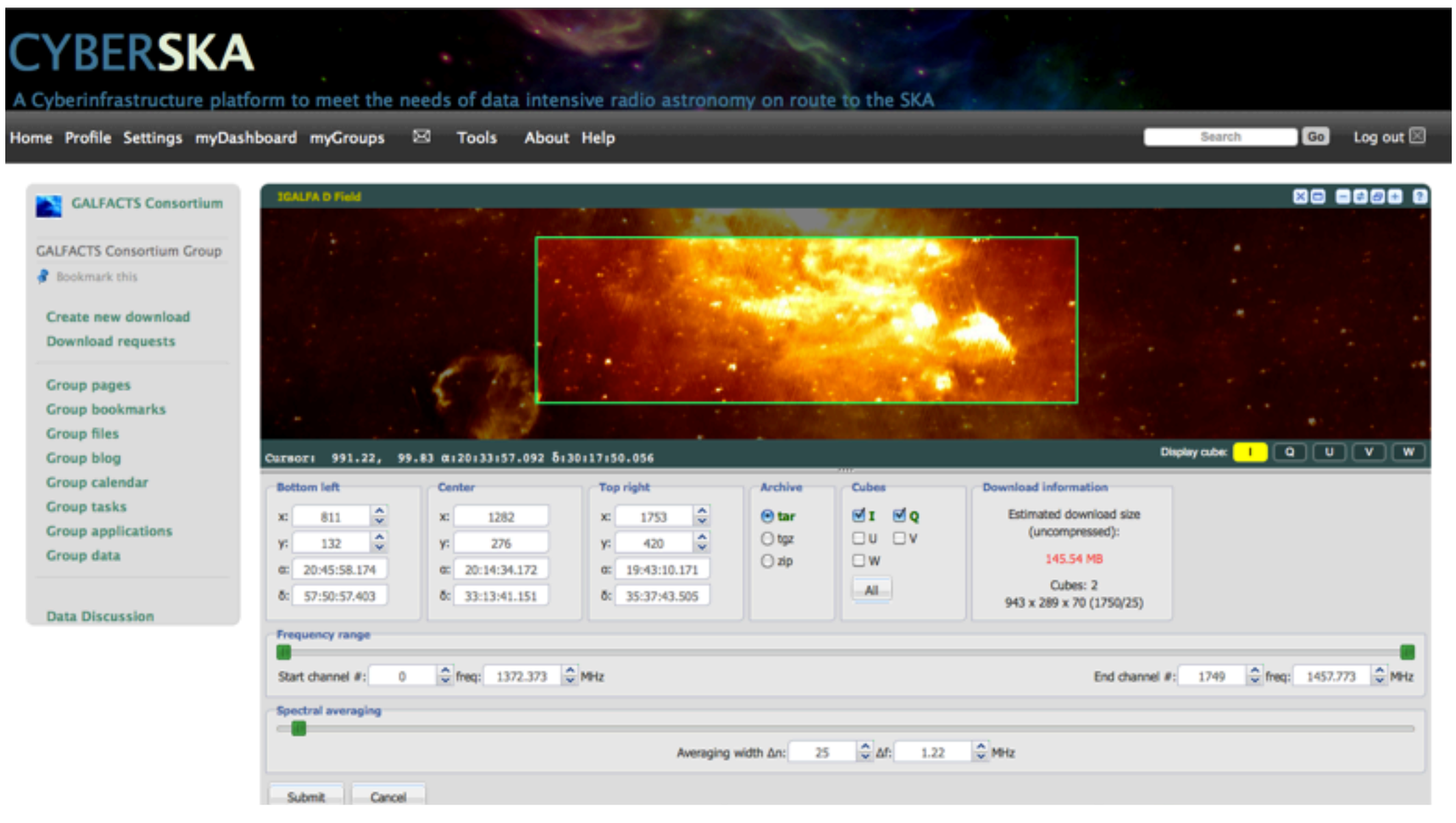
Building an ALMA Ready System

- 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



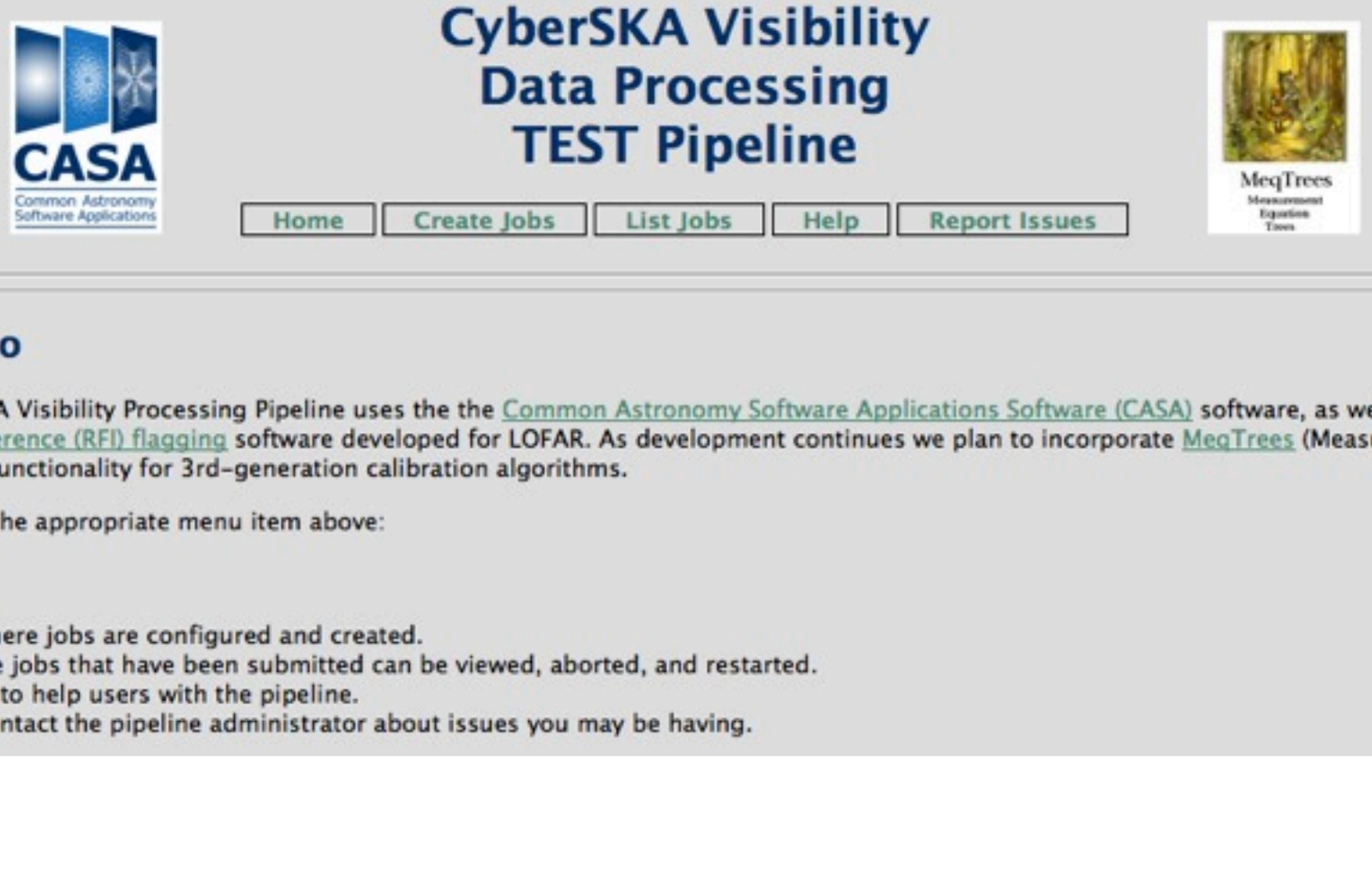
Building an ALMA Ready System

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



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

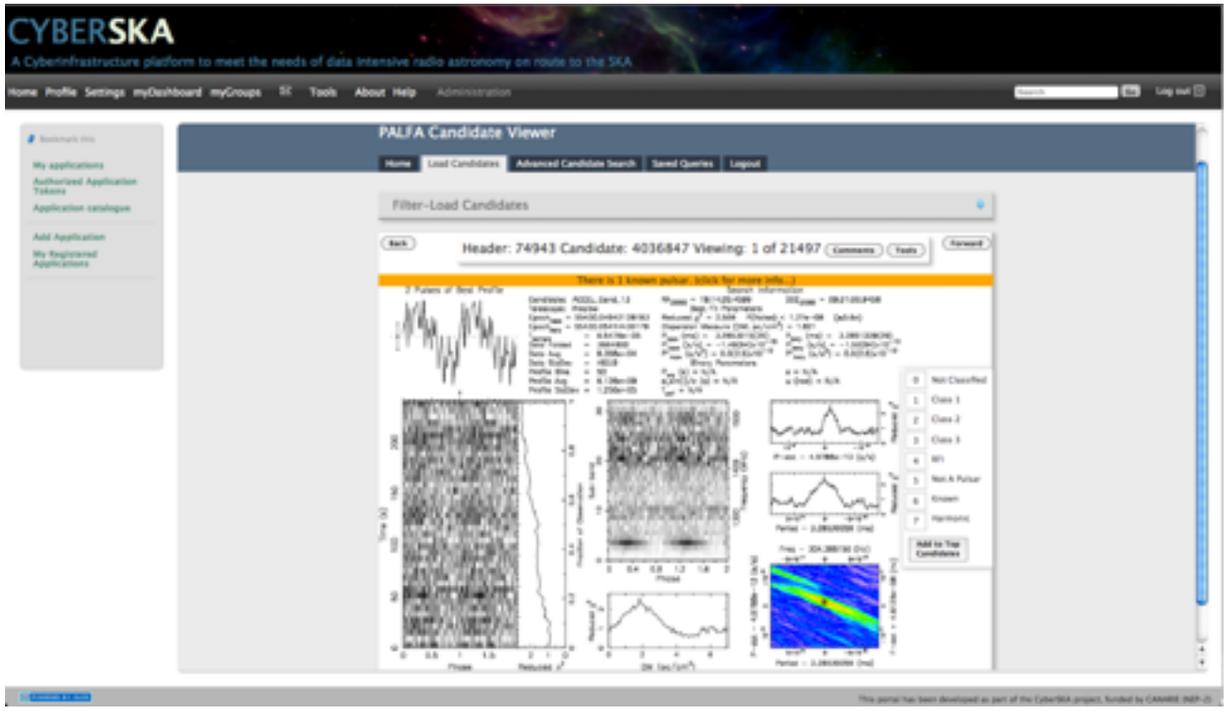


Building an ALMA Ready System

- 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

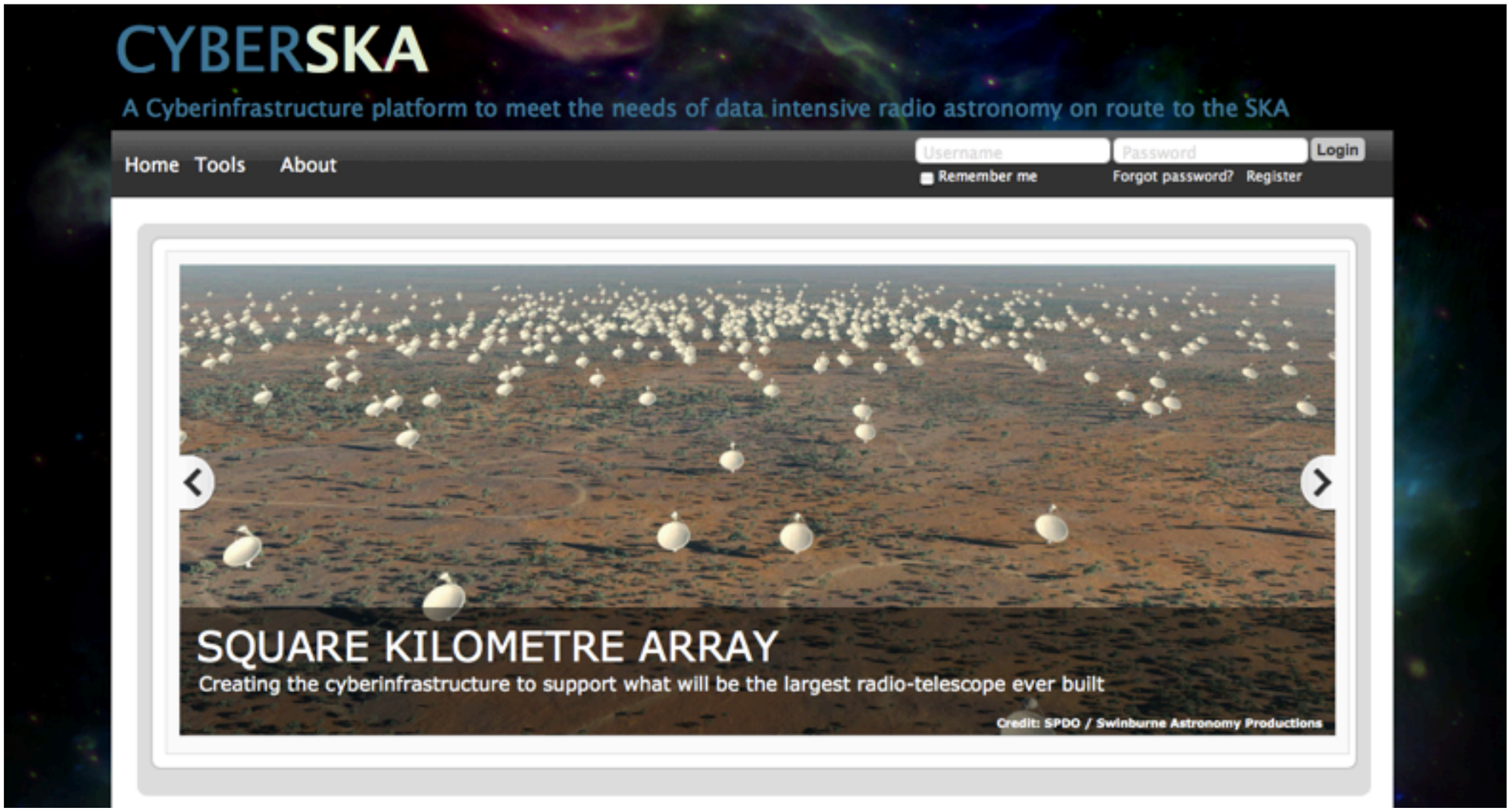


ALMA Opportunities

- 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



You are invited to the ALMA Integration Group