

# Looking Towards the Future of Radio Astronomy with the CyberSKA Collaborative Portal

---

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



McGill



canarie



CYBERA

# CyberSKA:

*Creating the cyberinfrastructure to support what will be the largest radio-telescope ever built: the Square Kilometre Array (SKA)*

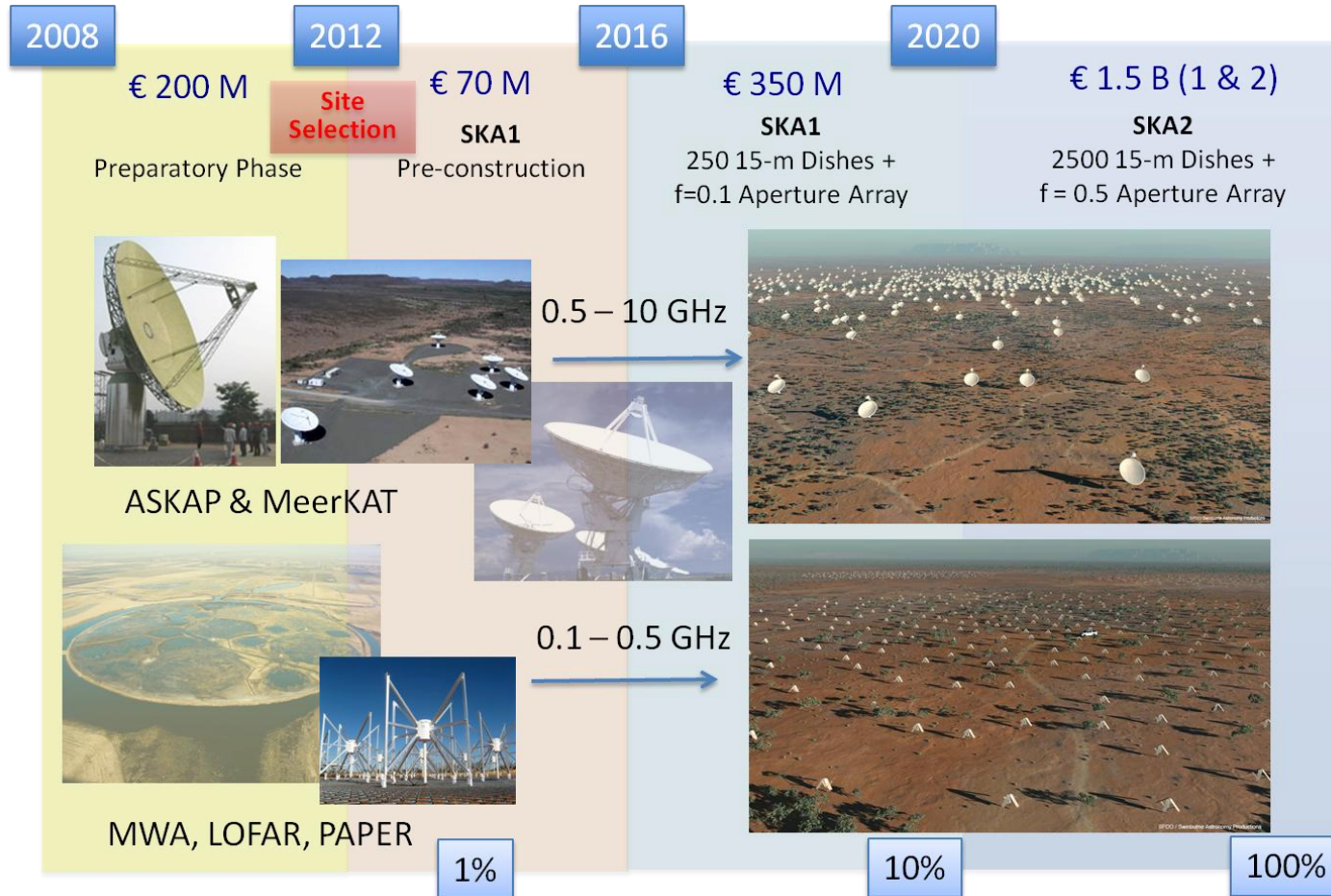
---



**Artists impression of the core of the SKA**

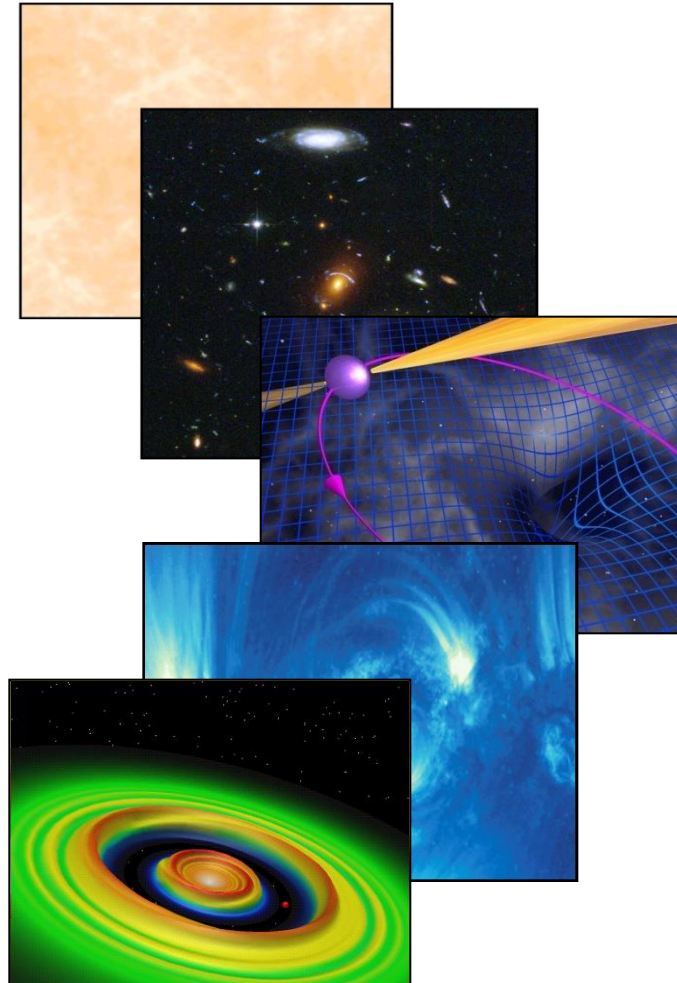
Image credit: SPDO / Swinburne Astronomy Productions

# SKA Timeline

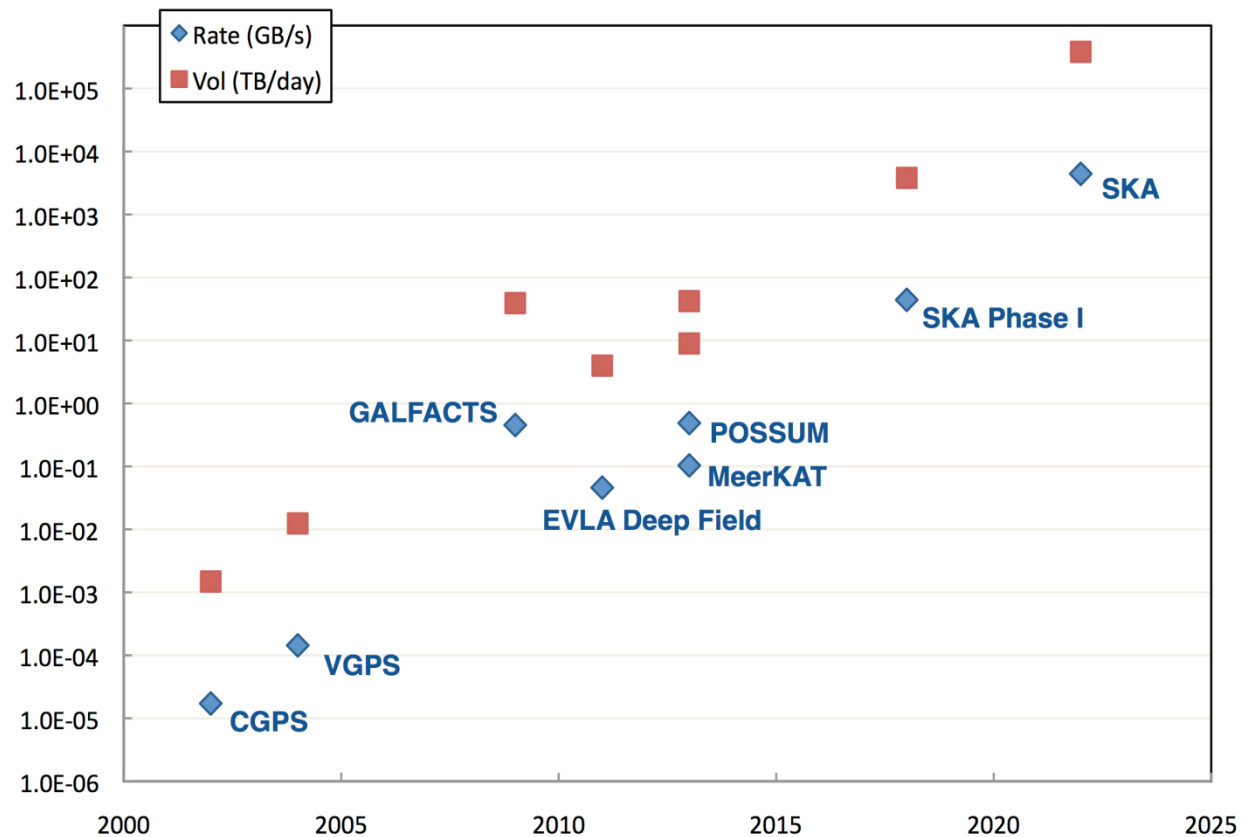


# SKA Key Science Questions

- **Probing the Dark Ages**
  - When & how were the first objects formed?
- **Cosmology and Galaxy Evolution**
  - Nature of Dark Energy and Dark Matter
  - Formation and evolution of galaxies and large scale structure
- **Strong-field tests of General Relativity**
  - What happens to space-time under extreme conditions?
- **Origin & Evolution of Cosmic Magnetism**
  - Where does magnetism come from?
  - What is its role in cosmic evolution?
- **Cradle of Life**
  - What and where are the conditions for life?
  - Does intelligent life exist elsewhere?



# Radio Imaging Survey Data Rates



# CyberSKA Motivation

---

- Most SKA Key science goals will be achieved via large-scale survey type observing programs
  - Very high data rates and volumes
  - Complex, multi-purpose, processing and analysis
  - Executed by globally distributed teams of researchers
- Drives the need for cyberinfrastructure solutions for
  - Collaboration tools
  - Data storage, management and distribution
  - Data processing, analysis and visualization

# CyberSKA Overview

---

- Initiative to develop a scalable and distributed cyberinfrastructure platform to meet evolving science needs of the SKA
- Led by the University of Calgary with several partner institutions from North America currently
- Canadian funding for CyberSKA provide by CANARIE, as part of their Network Enabled Platforms (NEP) program, and Cybera
- Starting by establishing cyberinfrastructure to support current large-scale astrophysical data needs generated by GALFACTS, PALFA and other high data volume SKA Pathfinder projects

# CyberSKA Project Team



UNIVERSITY OF  
CALGARY

## University of Calgary

- Russ Taylor (Professor, Lead PI)
- Eric Donovan (Associate Professor)
- Robert A. Este (Project Manager)
- Cameron Kiddle (Technical Coordinator)
- Mircea Andrecut (Developer)
- Roger Curry (Developer - Grid Research Centre)
- Pavol Federl (Developer)
- Arne Grimstrup (Developer)
- Sukhpreet Guram (PhD Student)
- Paolo Pragides (Developer)
- Dina Said (PhD Student)
- Christian Smith (System Administrator)
- Tingxi Tan (Developer – Grid Research Centre)



McGill

## McGill University

- Victoria Kaspi (Professor)
- Rafal Klodzinski (Developer – Sequence Factory)
- Patrick Lazarus (MSc Student)
- Shibl Mourad (President – Sequence Factory)
- Alex Samoilov (Developer – Sequence Factory)



## University of British Columbia

- Ingrid Stairs (Associate Professor)
- Bryan Fong (Developer)
- Mark Tan (Developer)



OKANAGAN

## University of British Columbia, Okanagan

- Erik Rosolowsky (Assistant Professor)
- Venkat Mahadevan (Developer)



Cornell University

## Cornell University

- Jim Cordes (Professor)
- Adam Brazier (Research Associate)
- Shami Chatterjee (Research Associate)
- Eric Chen (Analyst Consultant)



## National Research Council Canada

- Tom Landecker (Principal Research Officer)
- Tony Willis (Senior Research Council Officer)

## IBM Canada

- Don Aldridge (General Manager, Research & Life Sciences)
- Olivier Eymere (IT Architect)

# Experience/Background

- Leveraging knowledge and experience of the Grid Research Centre at the University of Calgary, IBM, and a large technical team
- Adapting, customizing and extending technologies used by GeoChronos (<http://geochronos.org>)
  - A platform developed by the Grid Research Centre
  - Enables Earth observation scientists to access and share data and applications and collaborate more effectively
  - Employs social networking, cloud computing and data management technologies
- Making use of other existing tools and technologies where possible



# Requirements

---

- Distributed
  - Provide access to distributed data, computing resources and services
- Scalable
  - Must be able to scale to support increasing data and processing needs
- Deployable
  - Different sites should be able to deploy developed tools and participate in CyberSKA relatively easily
- Heterogenous
  - Provide a framework to enable interaction with different types of data, computing resources and services and to add/execute different processing algorithms and workflows
- Automated
  - Automation and dynamic reconfiguration of services and data workflows in response to user demand, changing user objectives, available data and resource availability

# Requirements - II

---

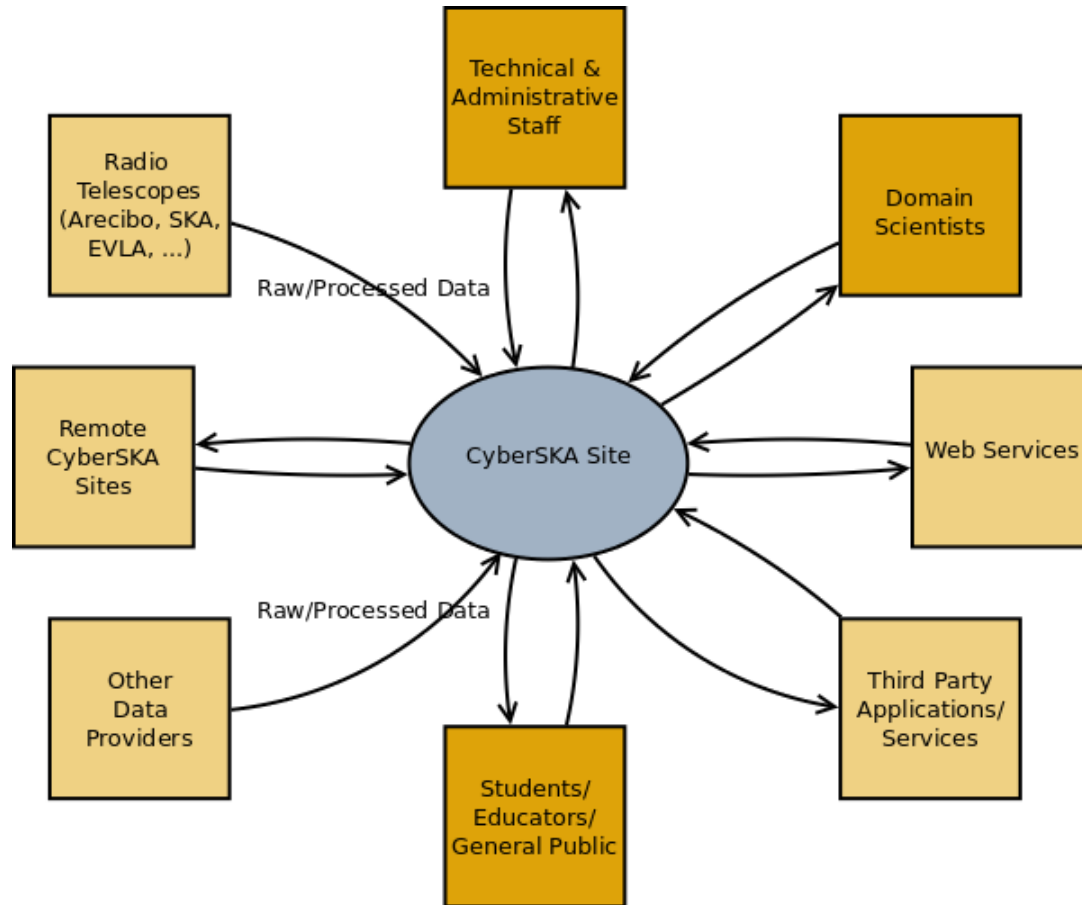
- Transparent
  - Provide users with transparent access to data, computing resources and services
- Web-enabled
  - Web-based platform that users can access from anywhere with Internet access
- Collaborative
  - Enable international/distributed teams to collaborate and communicate effectively
- Interactive
  - Enable on-line interactive visualization of data
- Auditable
  - Be able to track where data has come from and processes applied to it (data provenance)
- Interoperable
  - Compliant with existing standards such as Virtual Observatory (VO)

# Why social networking?

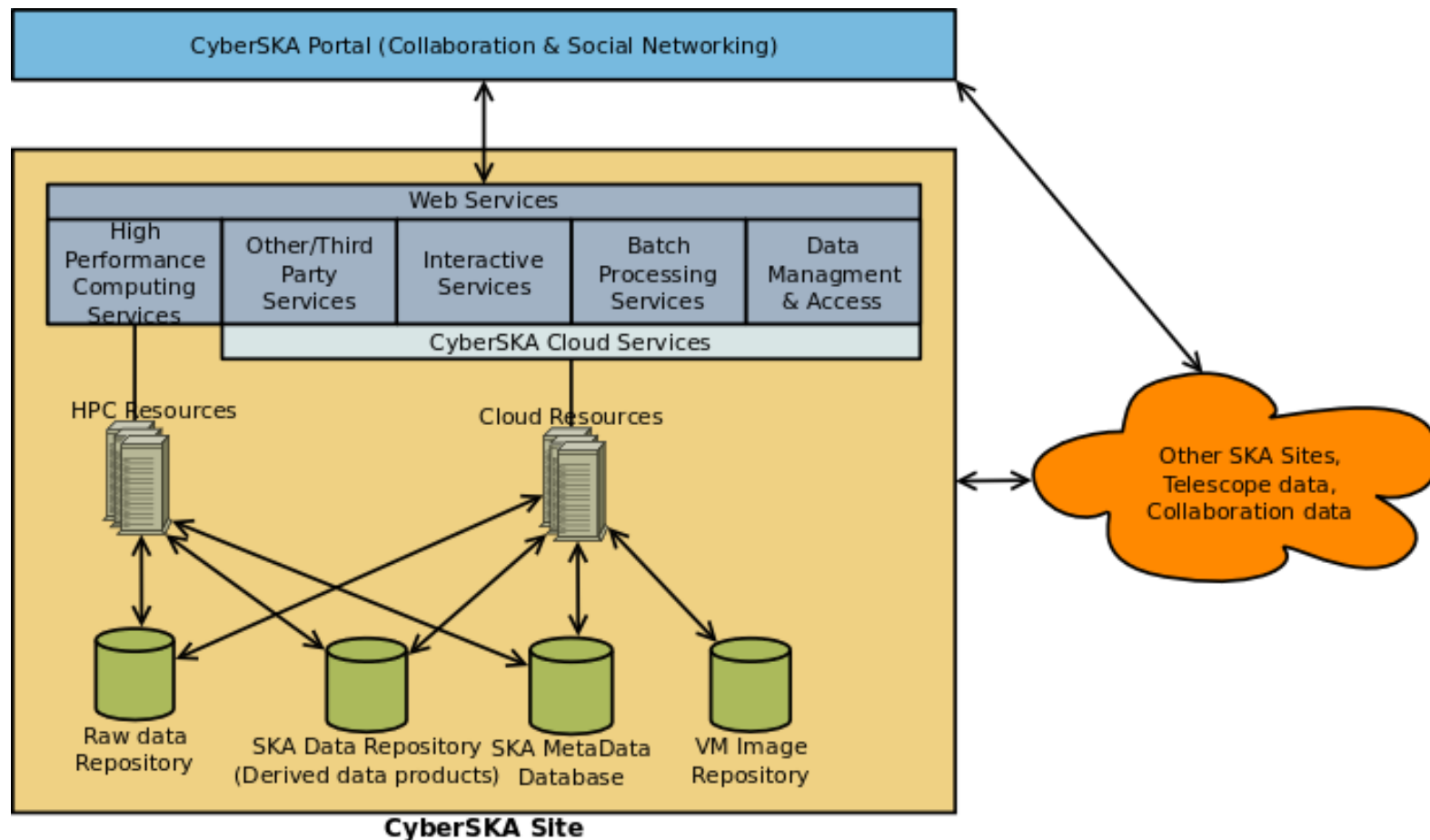
---

- Can enhance collaboration capabilities around data and applications
  - “*Facebook for Scientists*”
  
- Facebook analogy
  - Platform dealing with large scale in terms of users, data and applications
    - > 500 million users (50% log on to Facebook on any given day)
    - > 30 billion pieces of content shared each month
    - > 550 thousand applications on Facebook Platform

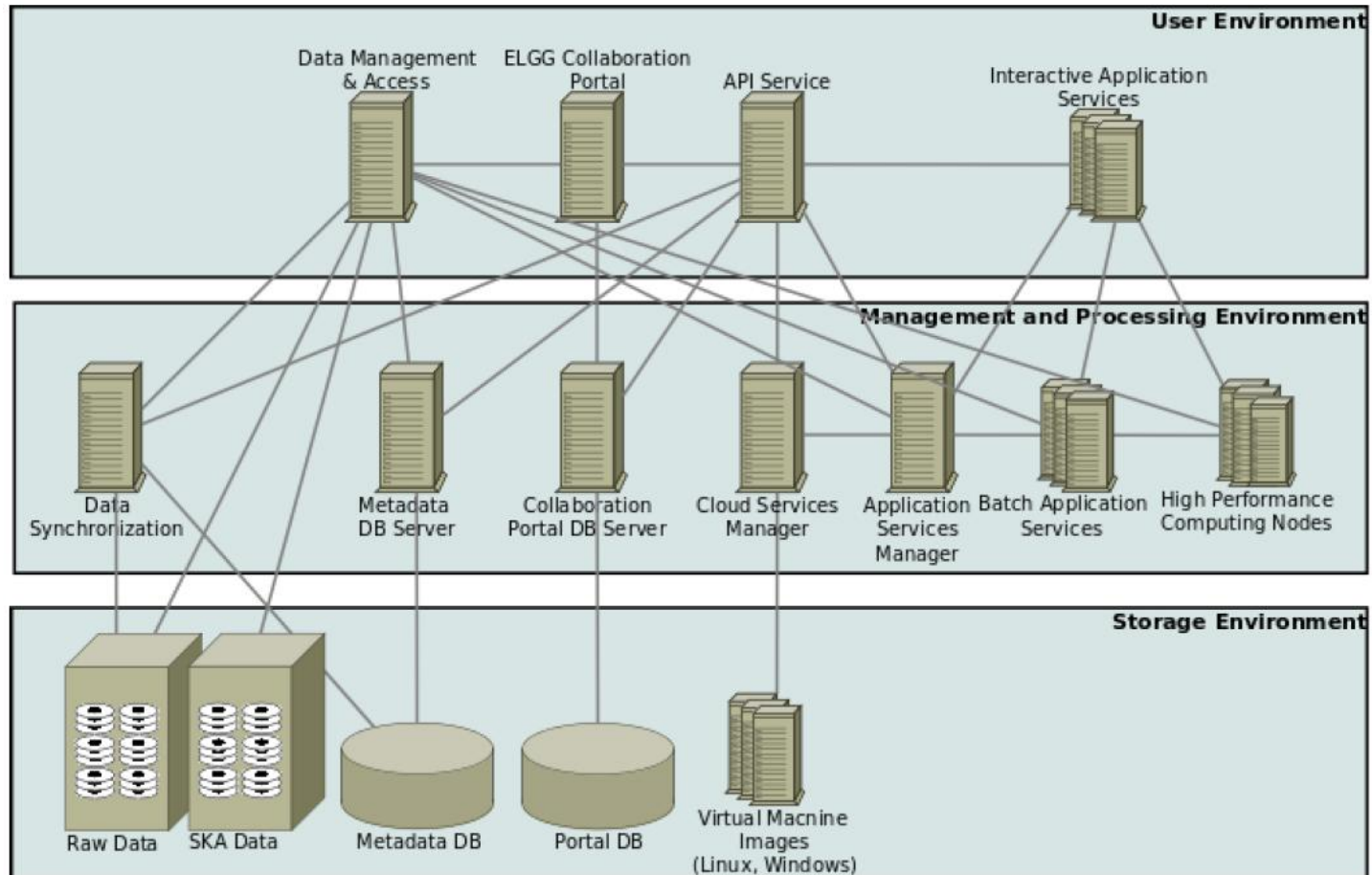
# System Context Model



# High Level Architecture

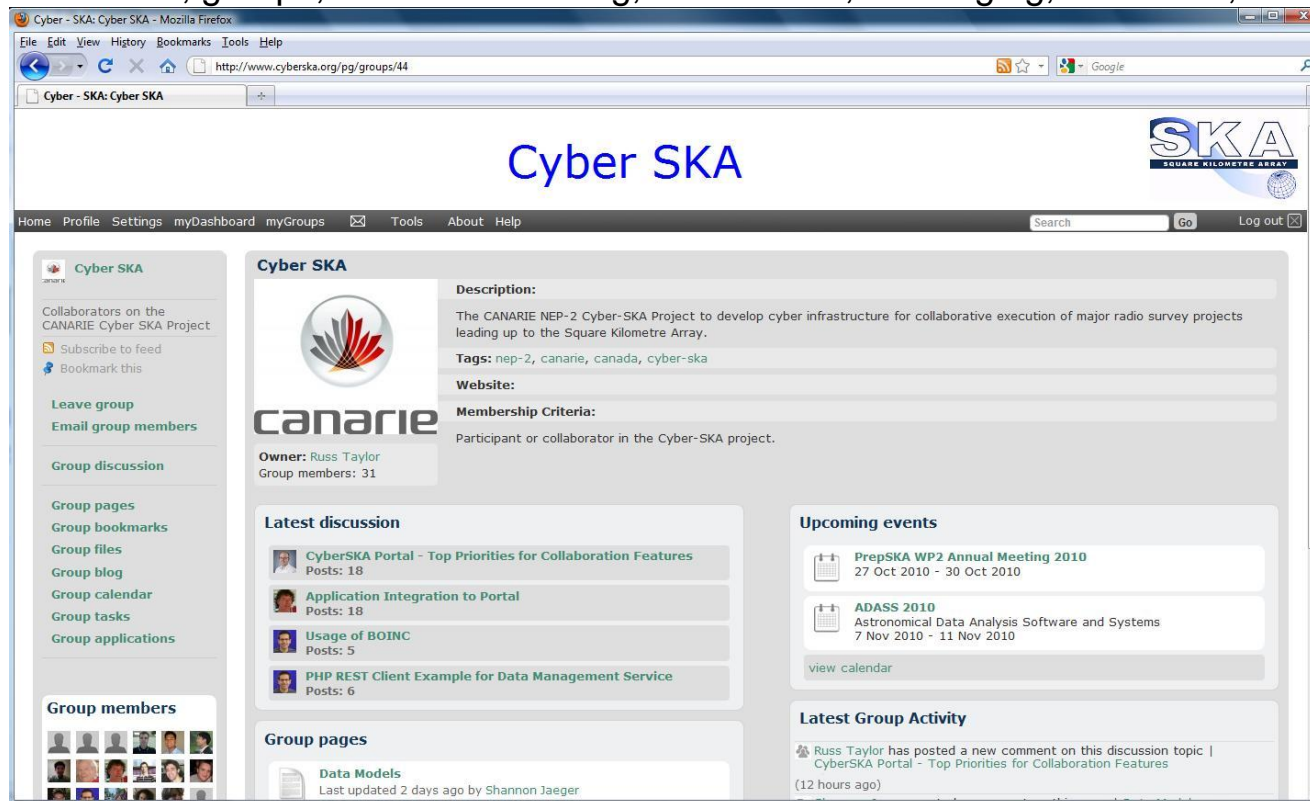


# Logical Operational Model



# Solutions - Collaboration

- Portal built on top of the Elgg open source social networking platform
  - Provides many Facebook-like features including :tags, bookmarks, profiles, blogs, wikis, contacts, groups, document sharing, discussions, messaging, calendars, status, activity feeds



# Solutions - Data

- Access/download data for selected parameters and region of interest
- Requested data generated in virtualized Condor pool on server side

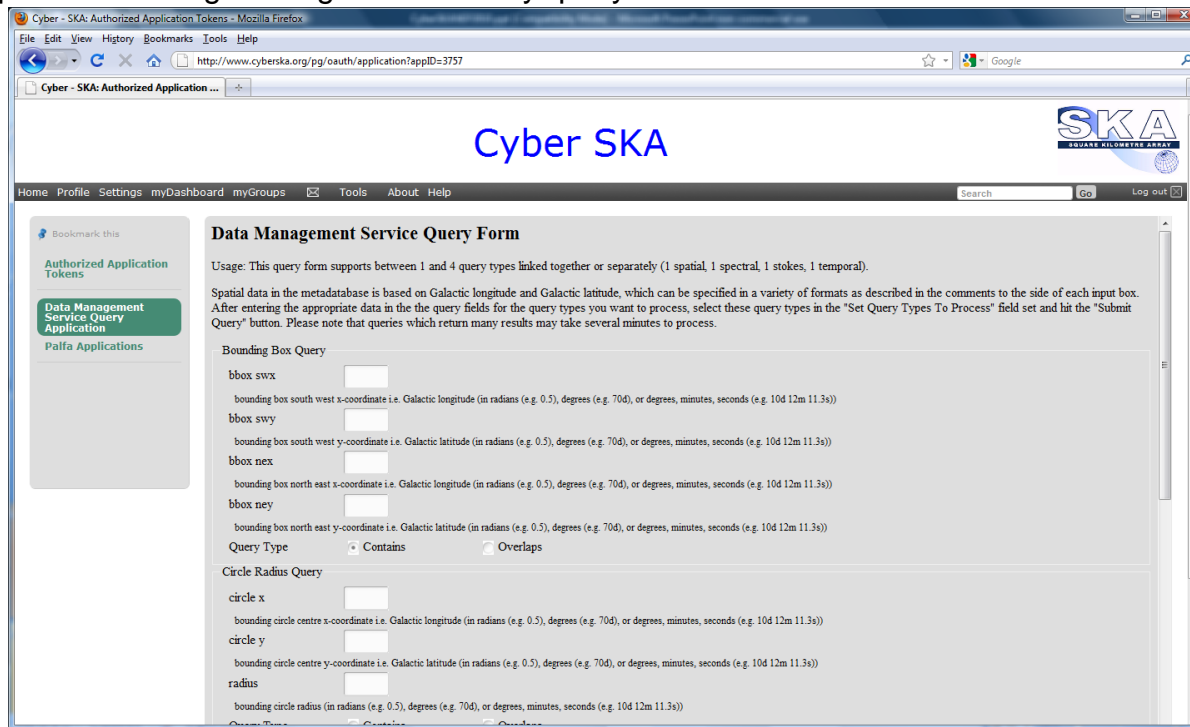
The screenshot displays the Cyber SKA web interface in a Mozilla Firefox browser window. The page title is "Cyber - SKA: GALFACTS Consortium Data". The main content area features a large astronomical image of a galaxy field with a green rectangular region of interest highlighted. Below the image, there are several control panels for data selection:

- Cursor:** 2067.83, 173.00  $\alpha$ :18:42:12.657  $\delta$ :22:31:00.653
- Display cube:** I Q U V W
- Bottom left:** x: 688, y: 140,  $\alpha$ : 20:14:13.752,  $\delta$ : 44:09:02.931
- Center:** x: 1079, y: 280,  $\alpha$ : 19:48:06.781,  $\delta$ : 24:17:52.551
- Top right:** x: 1471, y: 420,  $\alpha$ : 19:21:59.810,  $\delta$ : 26:37:47.038
- Archive:**  tar,  tgz,  zip
- Cubes:**  I,  Q,  U,  V,  W,
- Download information:** Estimated download size (uncompressed): 294.13 MB, Cubes: 1, 784 x 281 x 350 (3500/10)
- Frequency range:** Start channel #: 0, freq: 1523.374 MHz, End channel #: 3499, freq: 1376.416 MHz
- Spectral averaging:** Averaging width  $\Delta\nu$ : 10,  $\Delta t$ : 0.42 MHz

At the bottom of the interface, there are "Submit" and "Cancel" buttons. A footer note states: "This portal has been developed as part of the Cyber SKA project, funded by CANARIE (NEP-2)".

# Solutions – Data II

- Distributed data management service
  - Built on iRODS (Integrated Rule-Oriented Data System)
  - Running at two sites currently (University of British Columbia Okanagan & University of Calgary)
  - PostgreSQL database for image metadata (Adherent to VO metadata standards)
  - Query service with RESTful API (spatial, temporal and spectral queries supported)
  - Supports mosaicing of images returned by query



The screenshot shows a web browser window with the URL <http://www.cyberska.org/pg/oauth/application?appId=3757>. The page title is "Cyber SKA" and features the SKA logo. The main content area is titled "Data Management Service Query Form" and includes a usage instruction: "Usage: This query form supports between 1 and 4 query types linked together or separately (1 spatial, 1 spectral, 1 stokes, 1 temporal). Spatial data in the metadata is based on Galactic longitude and Galactic latitude, which can be specified in a variety of formats as described in the comments to the side of each input box. After entering the appropriate data in the the query fields for the query types you want to process, select these query types in the "Set Query Types To Process" field set and hit the "Submit Query" button. Please note that queries which return many results may take several minutes to process."

The form is divided into two sections: "Bounding Box Query" and "Circle Radius Query".

**Bounding Box Query**

- bbox swx:  bounding box south west x-coordinate i.e. Galactic longitude (in radians (e.g. 0.5), degrees (e.g. 704), or degrees, minutes, seconds (e.g. 104 12m 11.3s))
- bbox swy:  bounding box south west y-coordinate i.e. Galactic latitude (in radians (e.g. 0.5), degrees (e.g. 704), or degrees, minutes, seconds (e.g. 104 12m 11.3s))
- bbox nex:  bounding box north east x-coordinate i.e. Galactic longitude (in radians (e.g. 0.5), degrees (e.g. 704), or degrees, minutes, seconds (e.g. 104 12m 11.3s))
- bbox ney:  bounding box north east y-coordinate i.e. Galactic latitude (in radians (e.g. 0.5), degrees (e.g. 704), or degrees, minutes, seconds (e.g. 104 12m 11.3s))

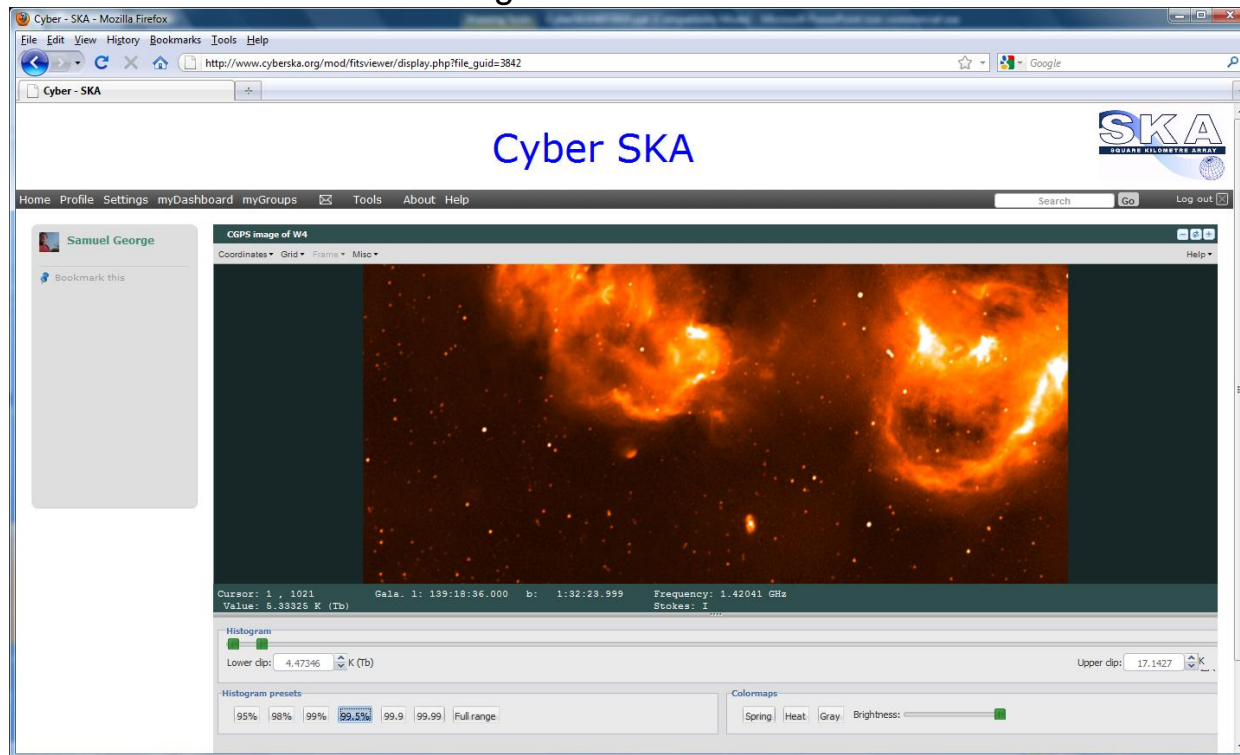
Query Type:  Contains  Overlaps

**Circle Radius Query**

- circle x:  bounding circle centre x-coordinate i.e. Galactic longitude (in radians (e.g. 0.5), degrees (e.g. 704), or degrees, minutes, seconds (e.g. 104 12m 11.3s))
- circle y:  bounding circle centre y-coordinate i.e. Galactic latitude (in radians (e.g. 0.5), degrees (e.g. 704), or degrees, minutes, seconds (e.g. 104 12m 11.3s))
- radius:  bounding circle radius (in radians (e.g. 0.5), degrees (e.g. 704), or degrees, minutes, seconds (e.g. 104 12m 11.3s))

# Solutions - Visualization

- On-line visualization of multi-dimensional FITS files
  - Supports interactive panning & zooming, histogram correction, color map adjustments, display pixel data value, region statistics, multiple coordinate systems, grids, selection of frame for multi-dimensional images



# Solutions - Applications

- API for integrating third party / remotely hosted applications
- Single sign-on to applications enabled using OAuth

The screenshot shows a web browser window with the URL <http://www.cyberska.org/pg/oauth/application?appId=4053>. The page title is "Cyber SKA" and the logo is visible in the top right. The navigation menu includes "Home", "Profile", "Settings", "myDashboard", "myGroups", "Tools", "About", and "Help". The main content area is titled "PALFA Applications" and is logged in as "kiddiec". The page contains several sections:

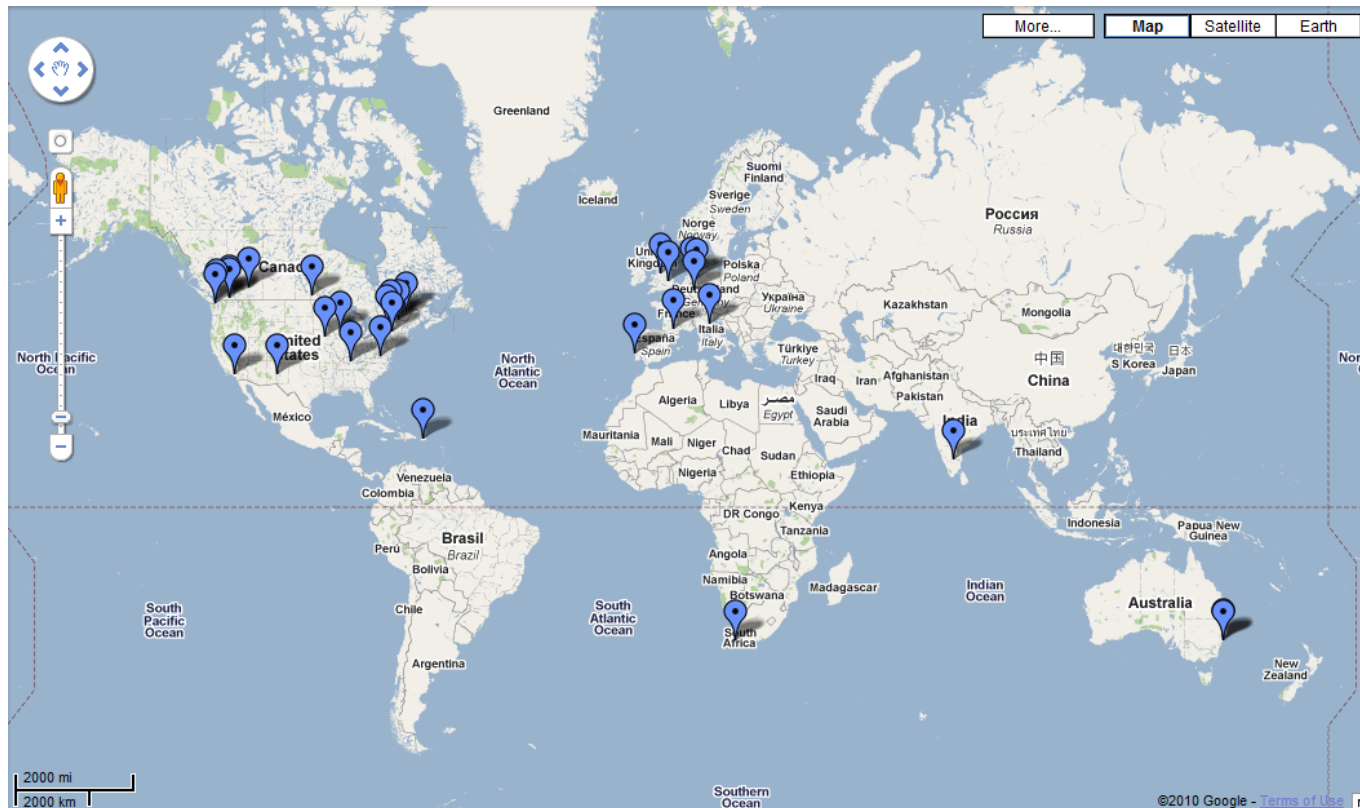
- Authorized Application Tokens**: A sidebar menu with options like "Test Remote Application", "Data Management Service Query Application", and "Add Application My Registered Applications".
- Application Information.**: A section with a "Control Panel" and instructions on how to use the application's features.
- Form Query.**: A section with instructions on how to load Pulsar Candidates into the application using a filter.
- Query.**: A section with instructions on how to load Pulsar Candidates manually using a MySQL query.
- Shortcut Keys.**: A section with a list of shortcut keys for controlling the rating process.

The shortcut keys list is as follows:

- 0 - Rates the Pulsar Candidate as: **Not Classified**
- 1 - Rates the Pulsar Candidate as: **Class 1**
- 2 - Rates the Pulsar Candidate as: **Class 2**
- 3 - Rates the Pulsar Candidate as: **Class 3**
- 4 - Rates the Pulsar Candidate as: **RFI**
- 5 - Rates the Pulsar Candidate as: **Not A Pulsar**
- 6 - Rates the Pulsar Candidate as: **Known**
- 7 - Rates the Pulsar Candidate as: **Harmonic**
- n - Move to next Candidate
- b - Move to previous Candidate

# CyberSKA Portal Usage

- 100+ members from around the world
- 15+ groups (GALFACTS, PALFA, EVLA, GMRT, CASA Users...)



# Next Steps

---

- Infrastructure
  - Acquisition of hardware at participating sites to establish prototype testbed
  - Set up cloud computing environments and key services at each site
- Collaboration
  - Refinement and development of collaboration features based on user feedback
- Data Management
  - Expansion of distributed data management system to other sites
  - Better integration of data management system with other CyberSKA tools and services
- Data Visualization
  - Provide server side support and improve scalability
- Data Processing
  - Establish dynamic batch based processing and interactive service environments on cloud platform
  - Establish framework for adding and integrating different processing algorithms and workflows
- Applications
  - Extension of third party application API to enable two way interaction between portal and applications (i.e. pull data/information from portal, push news feeds to portal based on application activities)

# Contact Information

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

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

