

# CyberSKA

## Cyberinfrastructure for Radio Astronomy

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

Cameron Kiddle

Research Fellow, Grid Research Centre, University of Calgary  
Technical Coordinator, CyberSKA



McGill



- 
- 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



## University of Calgary

- Russ Taylor (Professor, Lead PI)
- Eric Donovan (Associate Professor)
- Robert A. Este (Project Manager)
- Cameron Kiddle (Technical Coordinator)
- Mircea Andreucut (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)



## University of British Columbia, Okanagan

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



## Cornell University

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



## IBM Canada

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



## National Research Council Canada

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

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



## ■ 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

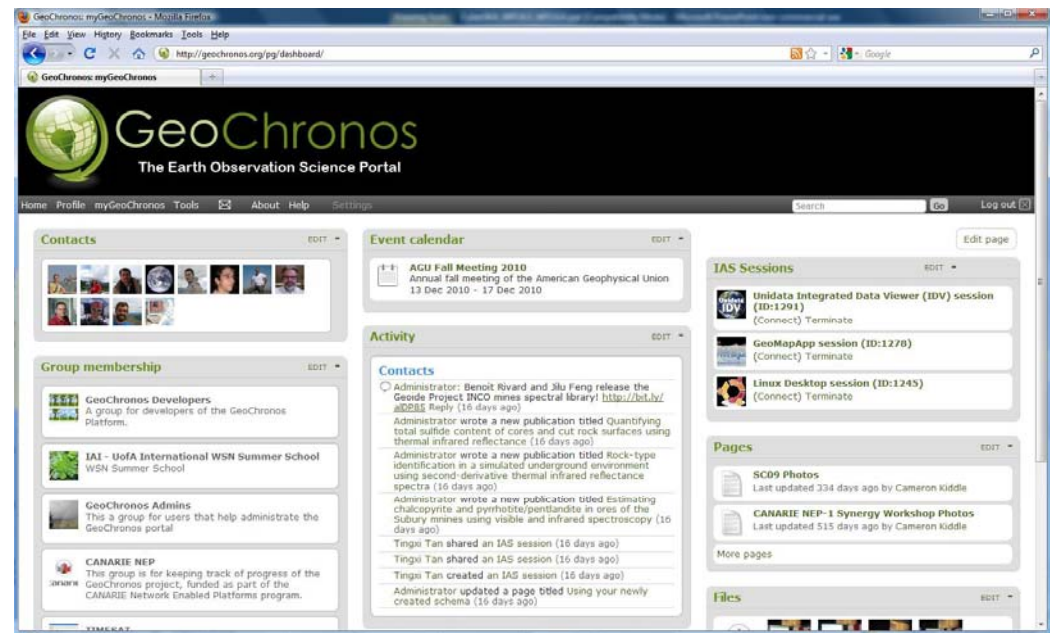
## ■ Automated

- Automation and dynamic reconfiguration of services and data workflows in response to user demand, changing user objectives, available data and resource availability

- 
- **Transparent**
    - provide users with transparent access to data, computing resources and services
  - **Web-enabled**
    - a 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)



- 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



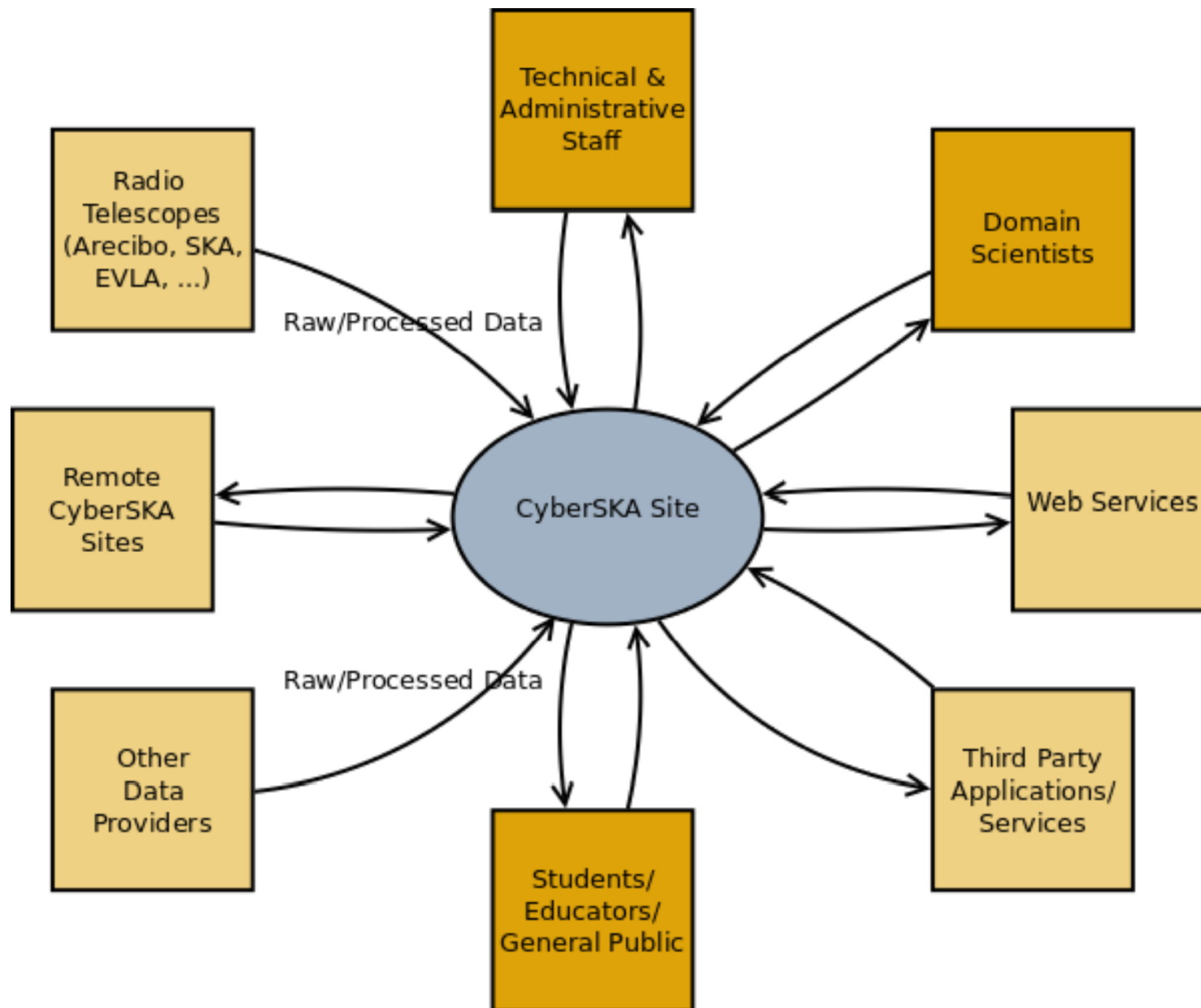
### ■ Why social networking?

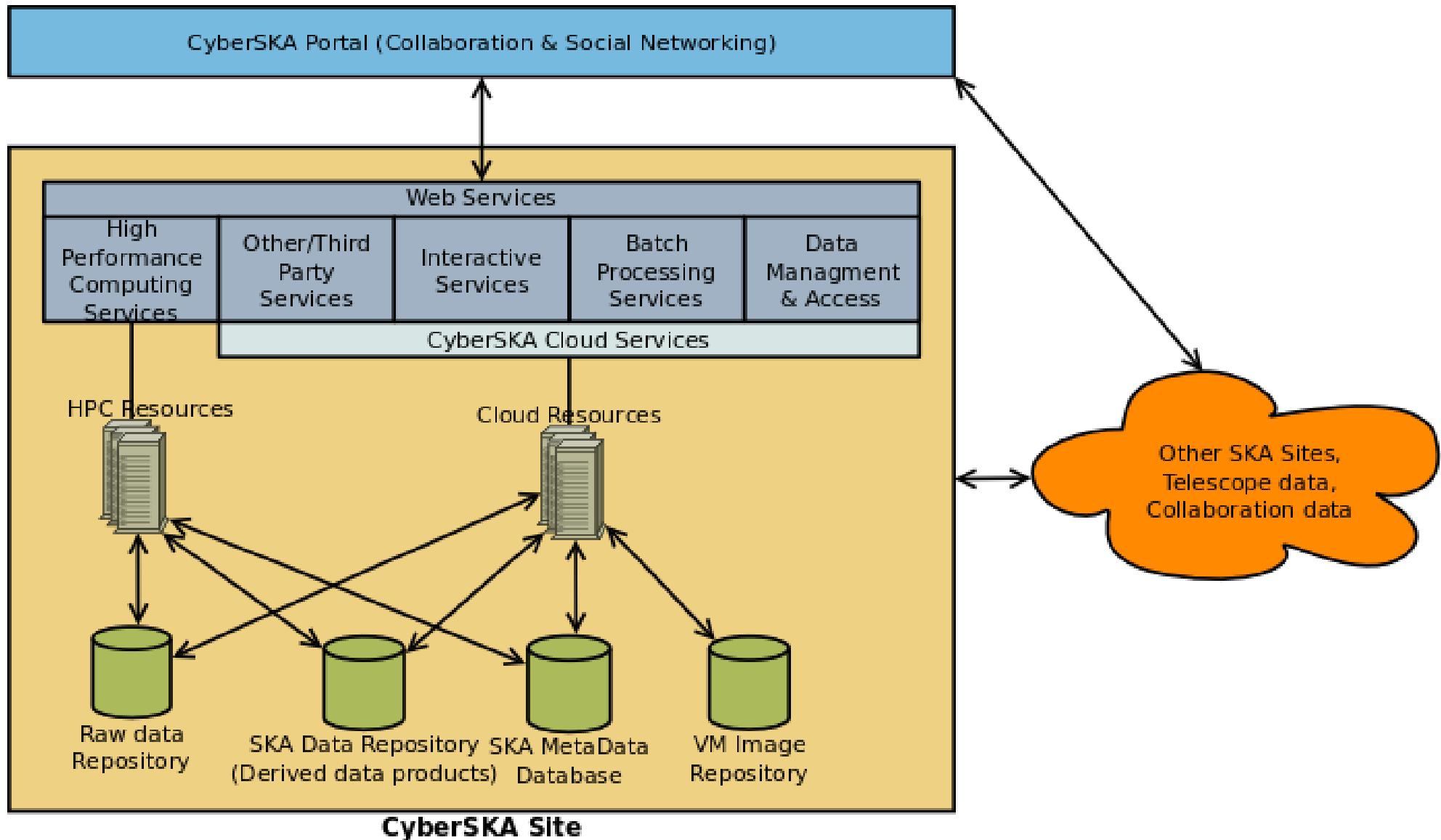
- can enhance collaboration capabilities around data and applications - “*Facebook for Scientists*”

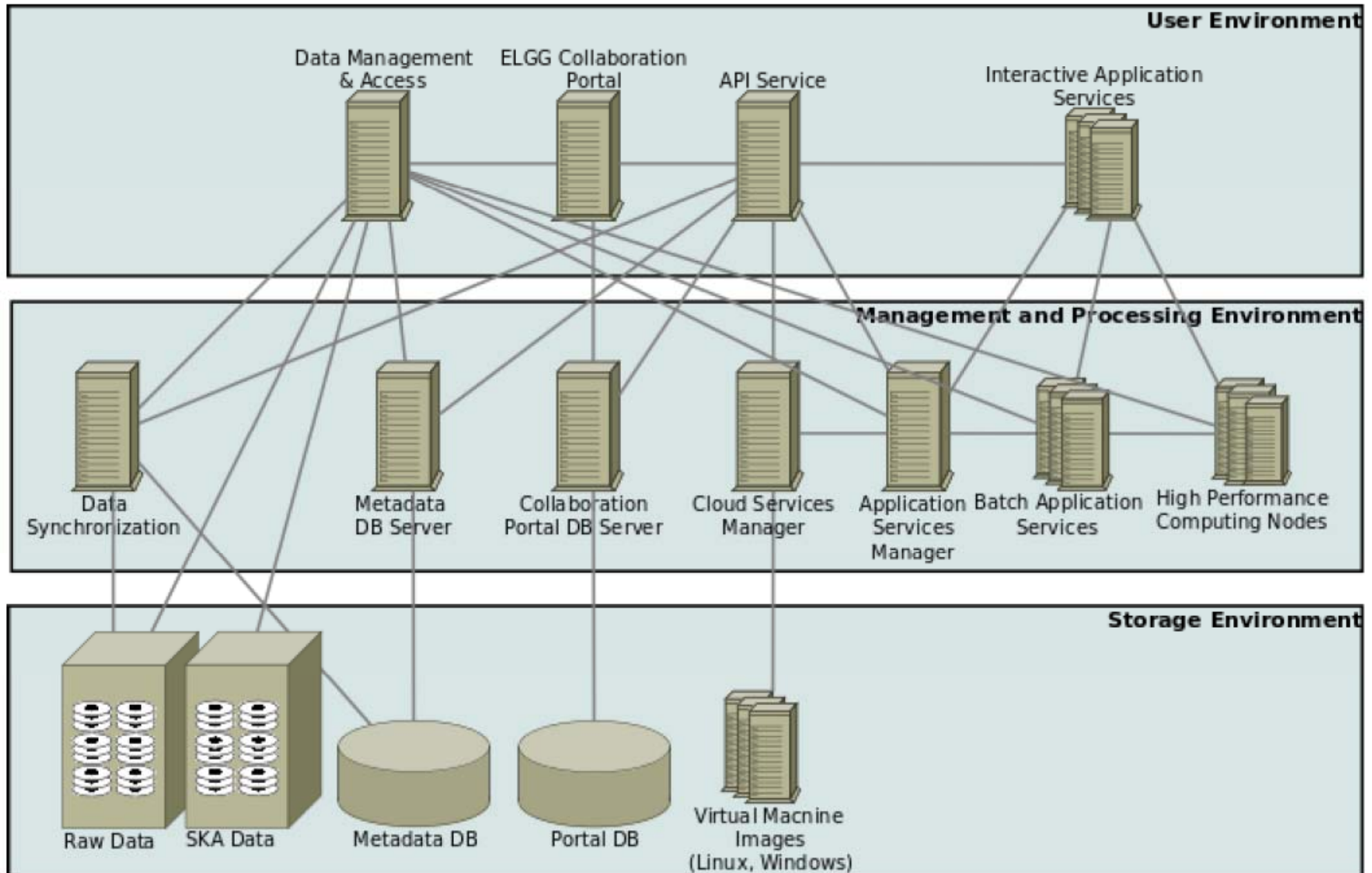
### ■ Facebook analogy

- a 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

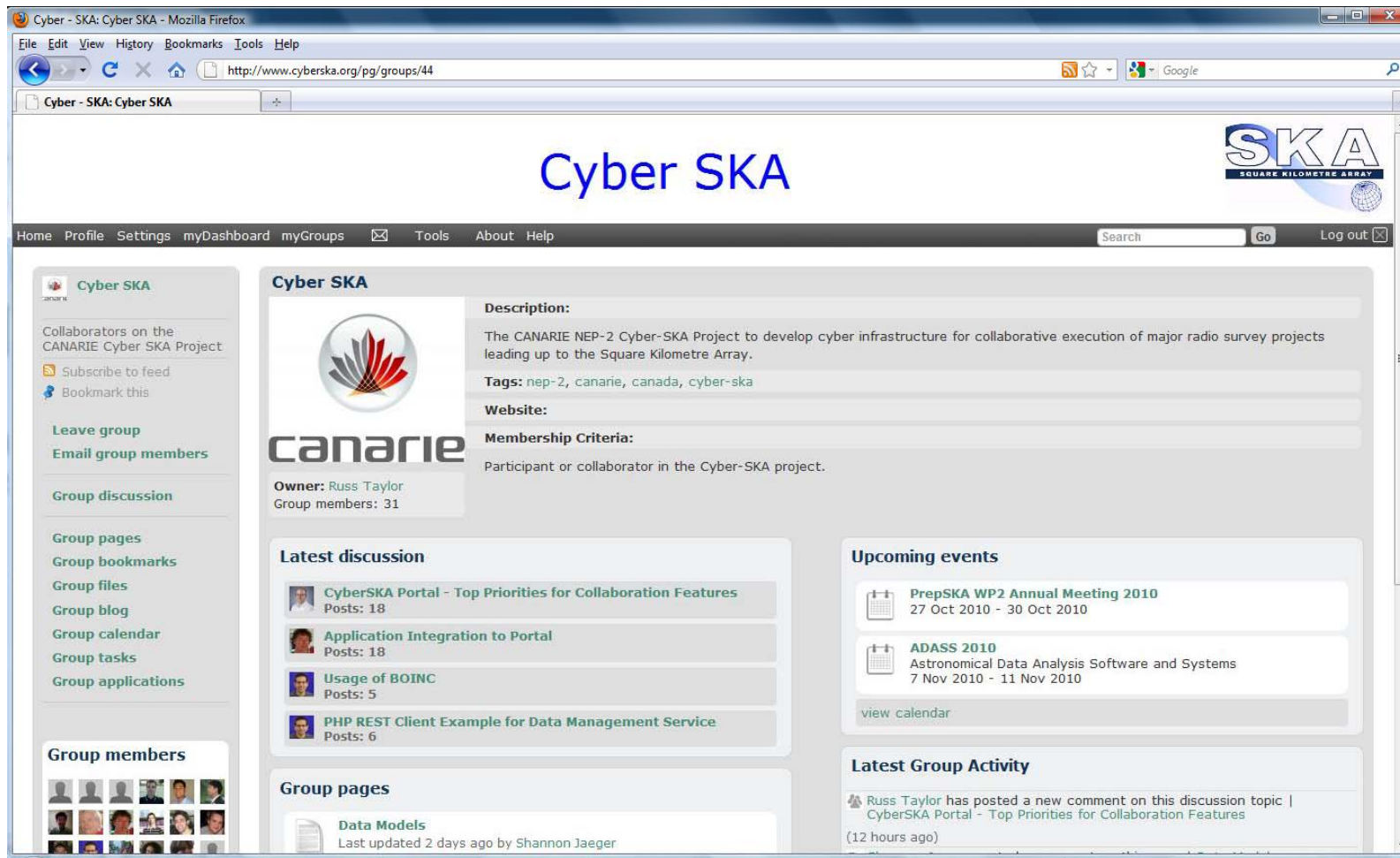




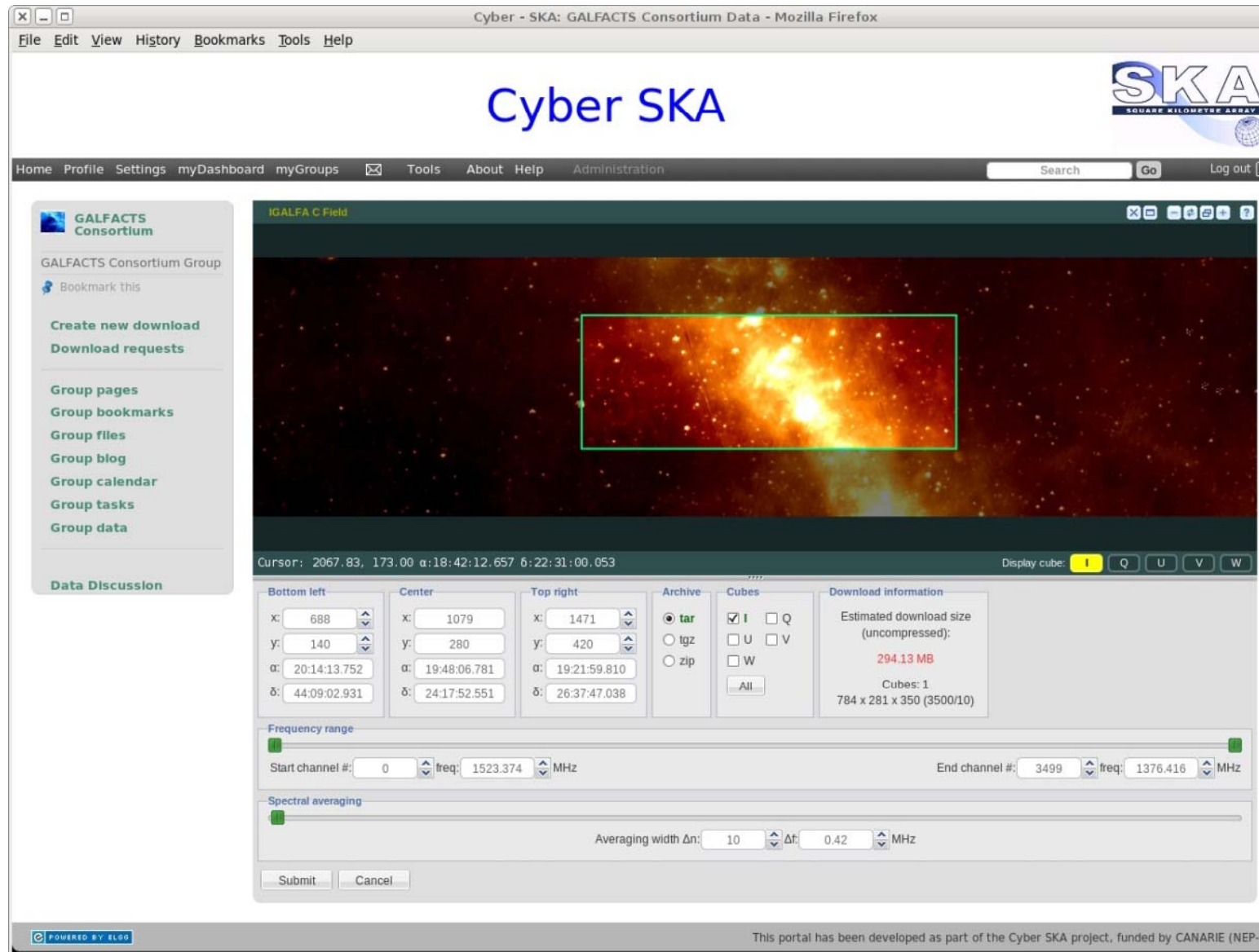




- Portal (<http://www.cyberska.org>) built on top of the Elgg open source social networking platform
  - provides many Facebook-like features including: tags, bookmarks, profiles, blogs, wikis, friends/contacts, groups, media/document sharing, discussions, message boards, calendars, status, activity feeds



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

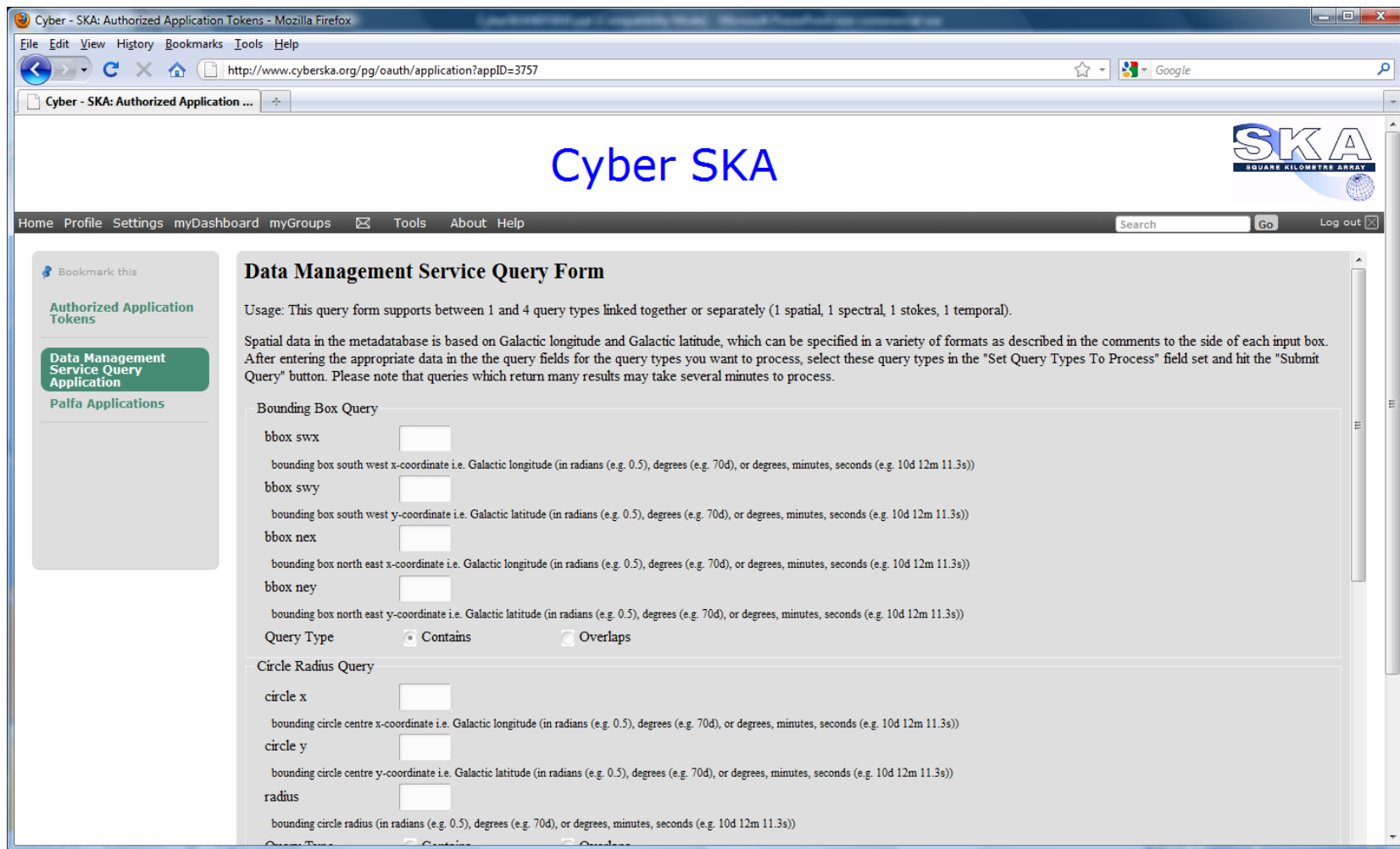


The screenshot shows the Cyber SKA web interface. The browser window is titled "Cyber - SKA: GALFACTS Consortium Data - Mozilla Firefox". The page has a header with the SKA logo and "Cyber SKA" text. Below the header is a navigation bar with links: Home, Profile, Settings, myDashboard, myGroups, Tools, About, Help, Administration. A search bar and "Log out" link are also present. On the left, there is a sidebar for the "GALFACTS Consortium" with options like "Create new download", "Download requests", and "Data Discussion". The main content area displays a radio image of a galaxy with a green rectangular region of interest. Below the image, there are input fields for coordinates (X, Y, RA, Dec) and a "Display cube" button. A "Download information" box shows an estimated download size of 294.13 MB. At the bottom, there are fields for "Frequency range" (Start channel #, End channel #, frequency) and "Spectral averaging" (Averaging width  $\Delta n$ ,  $\Delta t$ ). The footer mentions "POWERED BY ELGO" and "This portal has been developed as part of the Cyber SKA project, funded by CANARIE (NEP-2)".



## ■ 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 titled "Cyber - SKA: Authorized Application Tokens - Mozilla Firefox". The address bar shows the URL "http://www.cyberska.org/pg/oauth/application?appId=3757". The page features the SKA logo in the top right corner and a navigation menu 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 titled "Data Management Service Query Form". It 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)." and a detailed explanation of the spatial data format: "Spatial data in the metadatabase 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 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 main 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. 70d), or degrees, minutes, seconds (e.g. 10d 12m 11.3s))
- bbox swy**: bounding box south west y-coordinate i.e. Galactic latitude (in radians (e.g. 0.5), degrees (e.g. 70d), or degrees, minutes, seconds (e.g. 10d 12m 11.3s))
- bbox nex**: bounding box north east x-coordinate i.e. Galactic longitude (in radians (e.g. 0.5), degrees (e.g. 70d), or degrees, minutes, seconds (e.g. 10d 12m 11.3s))
- bbox ney**: bounding box north east y-coordinate i.e. Galactic latitude (in radians (e.g. 0.5), degrees (e.g. 70d), or degrees, minutes, seconds (e.g. 10d 12m 11.3s))

**Query Type**: ☒ Contains ☐ Overlaps

**Circle Radius Query**

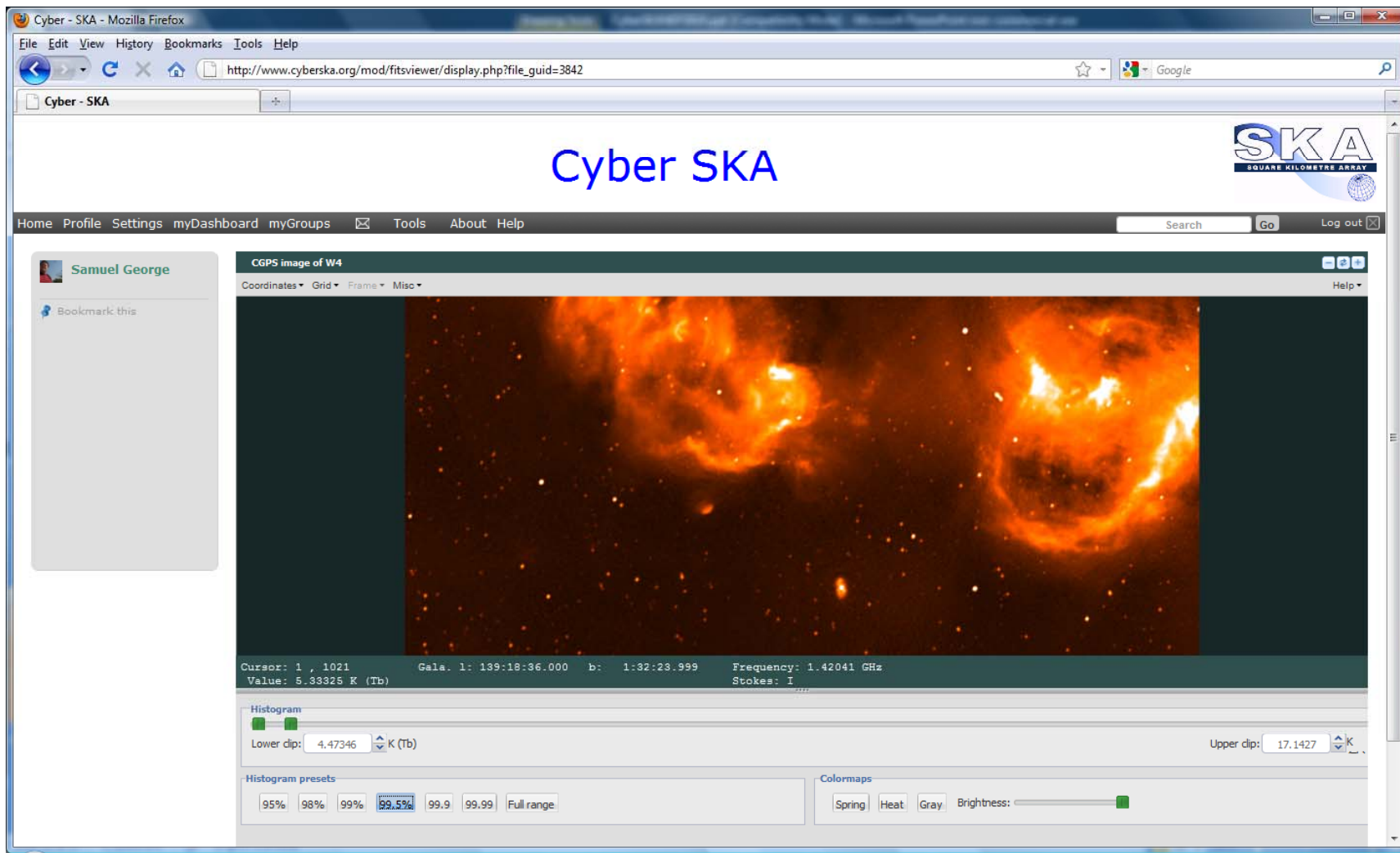
- circle x**: bounding circle centre x-coordinate i.e. Galactic longitude (in radians (e.g. 0.5), degrees (e.g. 70d), or degrees, minutes, seconds (e.g. 10d 12m 11.3s))
- circle y**: bounding circle centre y-coordinate i.e. Galactic latitude (in radians (e.g. 0.5), degrees (e.g. 70d), or degrees, minutes, seconds (e.g. 10d 12m 11.3s))
- radius**: bounding circle radius (in radians (e.g. 0.5), degrees (e.g. 70d), or degrees, minutes, seconds (e.g. 10d 12m 11.3s))

At the bottom of the form, there are buttons for "Query Types", "Contains", and "Overlaps".

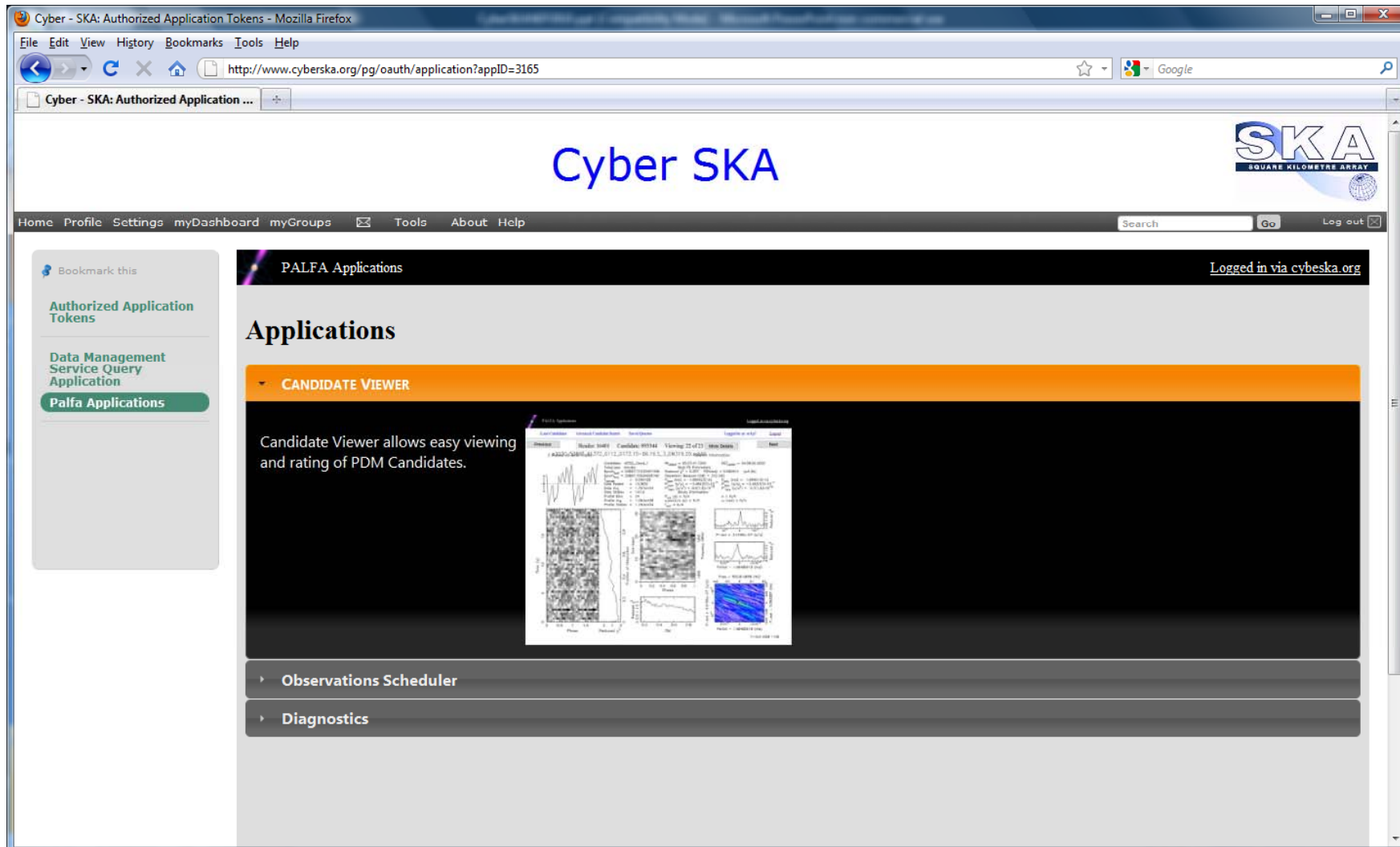


## ■ On-line visualization of multidimensional FITS files:

- Supporting: interactive panning & zooming, histogram correction, color map adjustments, display pixel data value, multiple coordinated systems, grids, selection of frame for multi-dimensional images



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



The screenshot shows a web browser window titled "Cyber - SKA: Authorized Application Tokens - Mozilla Firefox". The address bar displays the URL: <http://www.cyberska.org/pg/oauth/application?appID=3165>. The page header features the SKA logo and the text "Cyber SKA". Below the header is a navigation bar with links: Home, Profile, Settings, myDashboard, myGroups, Tools, About, Help. A search bar and a "Log out" link are also present.

The main content area is titled "PALFA Applications" and includes a "Logged in via cyberska.org" status. The "Applications" section is highlighted with an orange bar and contains a "CANDIDATE VIEWER" sub-section. The text states: "Candidate Viewer allows easy viewing and rating of PDM Candidates." Below this text is a preview of the candidate viewer interface, which displays various plots and data for a specific candidate.

On the left side of the page, there is a sidebar with the following links: "Bookmark this", "Authorized Application Tokens", "Data Management Service Query Application", and "Palfa Applications" (which is highlighted with a green bar).

At the bottom of the main content area, there are two more sections: "Observations Scheduler" and "Diagnostics".

## ■ Infrastructure

- Acquisition of hardware at participating sites to establish prototype testbed
- Set up cloud computing environments and key services at each site
  - Cloud platforms under consideration include ASPEN, Eucalyptus, Nimbus, OpenStack

## ■ 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)

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

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

