a place of mind

THE UNIVERSITY OF BRITISH COLUMBIA

A Distributed Datacube Analysis Service for Radio Telescopes

University of British Columbia
Okanagan Campus

Venkat Mahadevan

Dr. Erik Rosolowsky

ADASS 2010

a place of mind

THE UNIVERSITY OF BRITISH COLUMBIA

Overview

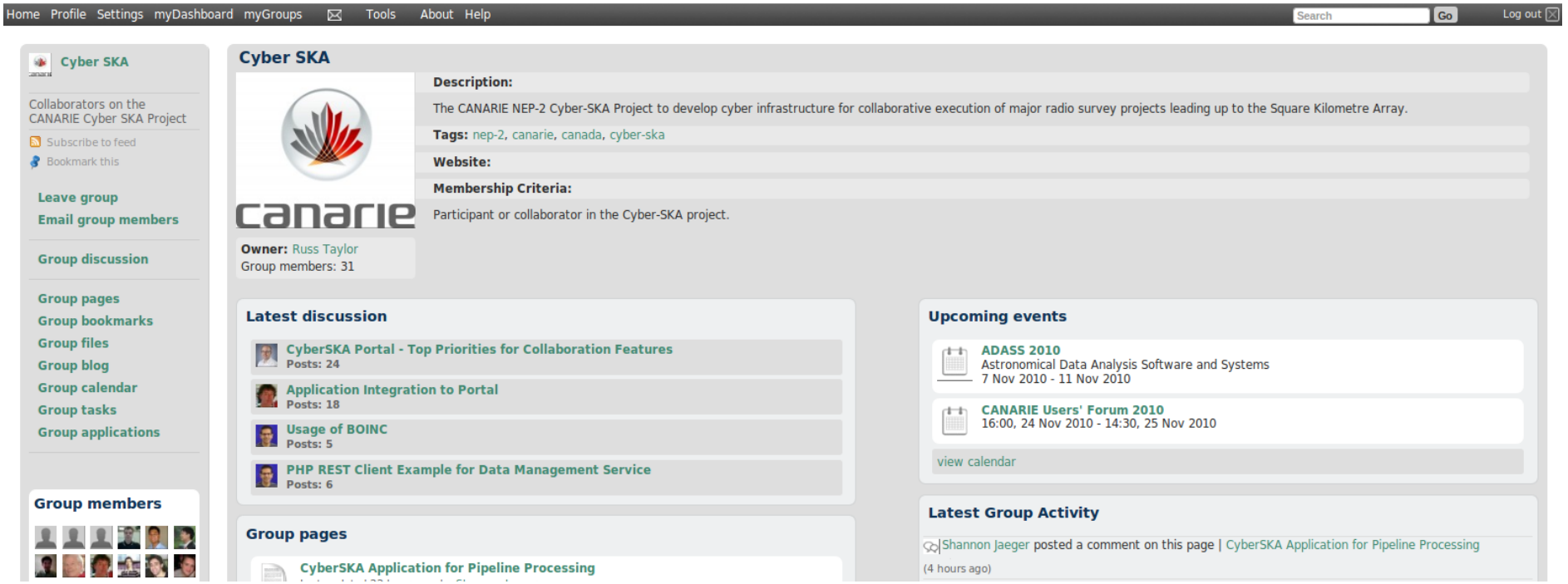
- Within the context of the overall CyberSKA Project, we are currently developing the cyberinfrastructure to enable distributed storage and processing of data in the form of FITS data cubes.
- Our primary motivations are:
  - to provide transparent access to cloud computing resources.
  - to provide users with access to their data via a web based science portal so they can manage and analyze large data sets.

Overview (cont.)

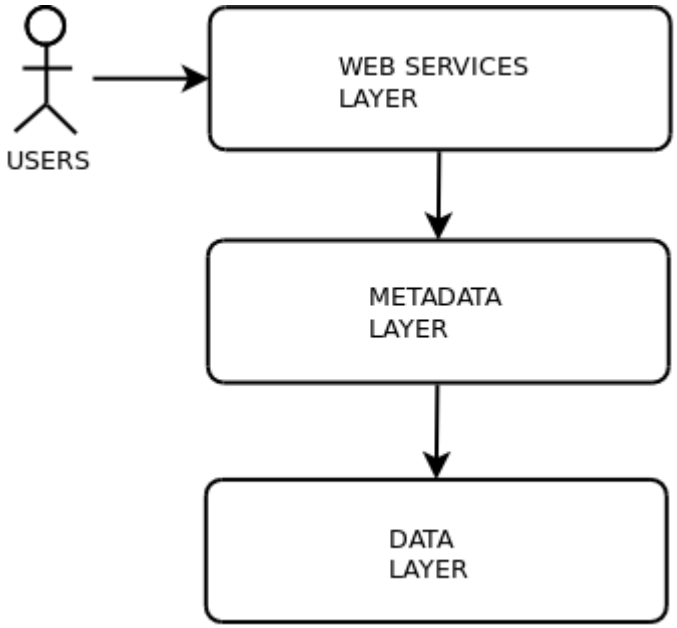
- A web portal has been established at www.cyberska.org.
- Users can access web-based data analysis applications as well as various data sets.
- Users can also participate in group discussions and project development.

Overview (cont.)

CyberSKA



High-level System Architecture



High-level System Architecture (cont.)

- We want to provide a software stack with the 3 layers that can be setup at remote sites to allow them to join the CyberSKA “federation”.
- Data storage and data processing services will be aggregated and provided to end users on demand.

Data Layer

- FITS files are stored in a data grid based on iRODS (Integrated Rule-Oriented Data System).
- iRODS is a “hands-off” distributed data system i.e. a data grid management system that supports:
  - Data replication and cross-site backups.
  - Abstraction of data location from the user.
  - High speed data transfers using multiple TCP streams.

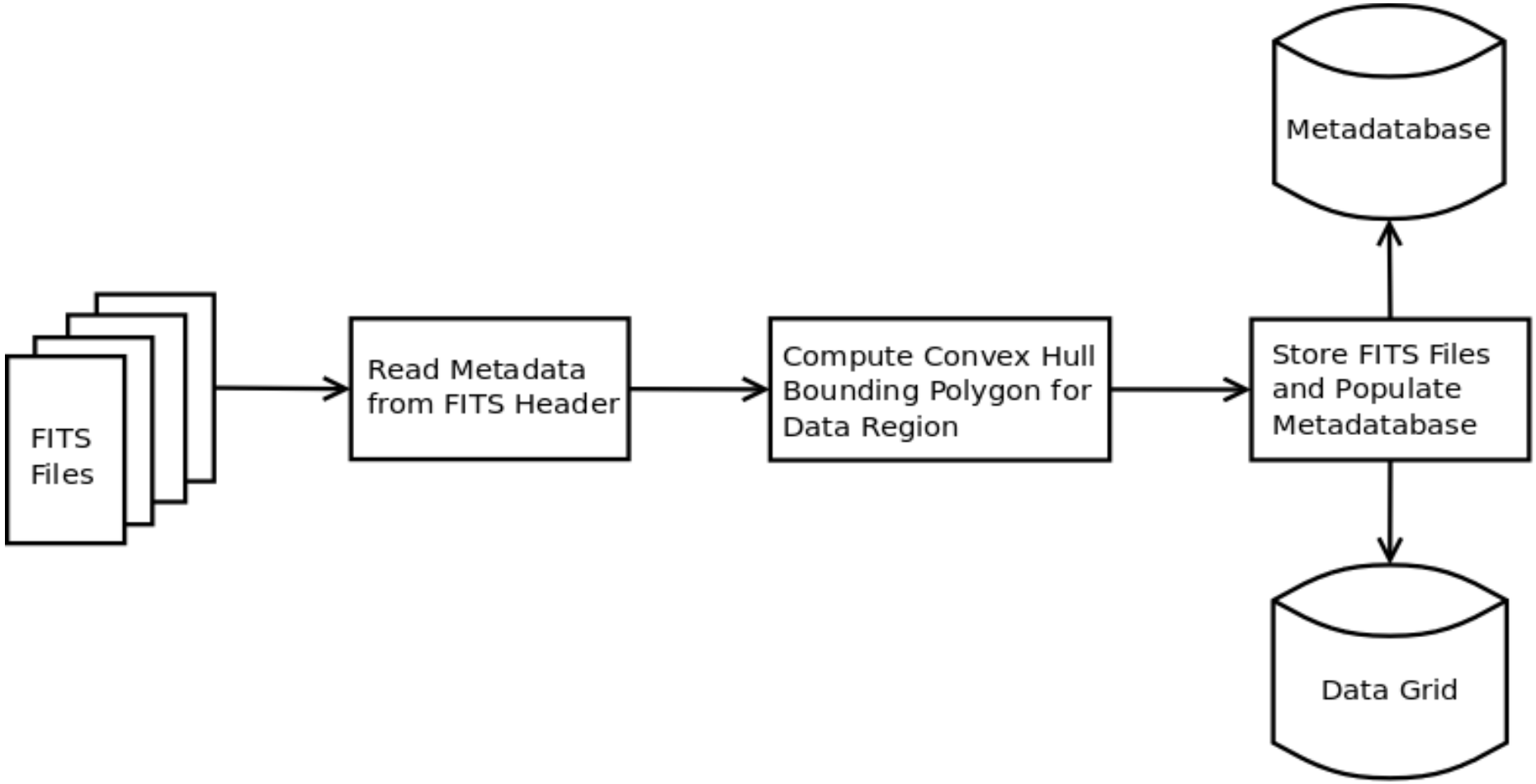
Data Layer (cont.)

- iRODS also has an advanced "rules engine" to automate administrative tasks. For example, a rule can be used to perform certain processes on a file at check-in time.
- We are primarily using it to store FITS data files at different locations (one at UBC, the other at the University of Calgary).
- Eventually, there will be multiple sites each housing various collections of data.

Metadata Layer

- Data from FITS files is “ingested” into the metadatabase and can be queried using a combination of:
  - Spatial coordinate parameters.
  - Spectral frequency and stokes parameters.
  - Temporal date parameters.

Metadata Layer (cont.)



Metadata Layer (cont.)

- A spatially enabled PostgreSQL/PgSphere database is used to maintain resource metadata.
- The schema is based on IVOA Resource Metadata recommendations.
- Large volumes of data can be stored in PostgreSQL:
  - Unlimited maximum database size.
  - 32 TB maximum table size.

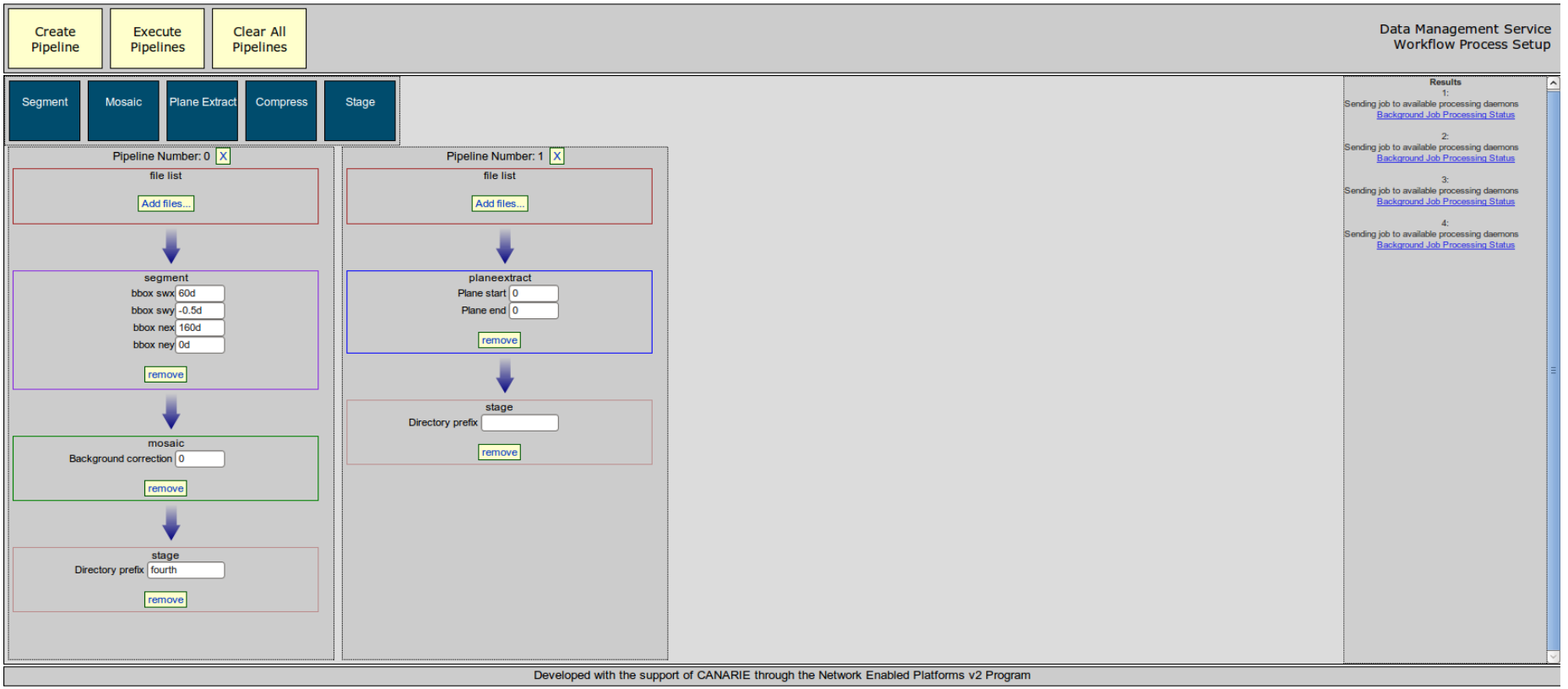
Metadata Layer (cont.)

- Using a spatially enabled database has key advantages when working with astronomical data:
  - Spatial data types and queries: e.g. polygon contains/overlaps, circle contains/overlaps, etc.
  - Ability to generate complex "astrospatial" queries for data using a more natural SQL syntax.
- GiST (Generalized Search Tree) indexes can be used to speedup spatial queries on large databases.

Web Services Layer

- We have developed a web based workflow builder that currently supports image segmentation, image mosaicking (based on the excellent Montage package), spatial reprojection, and plane extraction from data cubes.
- While leveraging distributed data storage and data processing facilities in the background, the user's experience is abstracted away from these details.
- Data is shipped to where the processing facilities exist automatically.

Web Services Layer (cont.)



Future Developments

- The workflow builder will be expanded to include additional computational modules such as:
  - Image convolution.
  - Object identification.
  - Image statistics.
  - Fourier transforms and spectral analysis.
  - Basic pixel array manipulation.
- An API for community developed modules will be developed.

Acknowledgements

- We would like to acknowledge usage of the following open-source software packages (in no particular order):
  - PostgreSQL (http://www.postgresql.org)
  - PgSphere (http://pgsphere.projects.postgresql.org)
  - Bitnami Lamp Stack (http://bitnami.org)
  - CodeIgniter (http://codeigniter.com)
  - jQuery (http://jquery.org)
  - Lightbox+ (http://serennz.sakura.ne.jp/toybox/lightbox)

Acknowledgements (cont.)

- ImageMagick (http://www.imagemagick.org)
- YUI 3 (http://developer.yahoo.com/yui/3)
- Prototype (http://www.prototypejs.org)
- TableKit (http://www.millstream.com.au/upload/code/tablekit)
- iRODS (http://www.irods.org)
- Montage (http://montage.ipac.caltech.edu/)
- WCS Tools (http://tdc-www.harvard.edu/wcstools/)

a place of mind

THE UNIVERSITY OF BRITISH COLUMBIA