

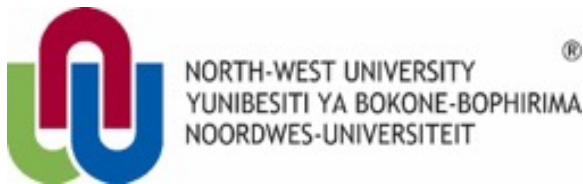


Inter-University Institute
for Data Intensive Astronomy

from big data to big ideas

Implications of supporting reproducibility for IDIA

Rob Simmonds
IDIA Associate Director
Professor, Dept. Computer Science, UCT.

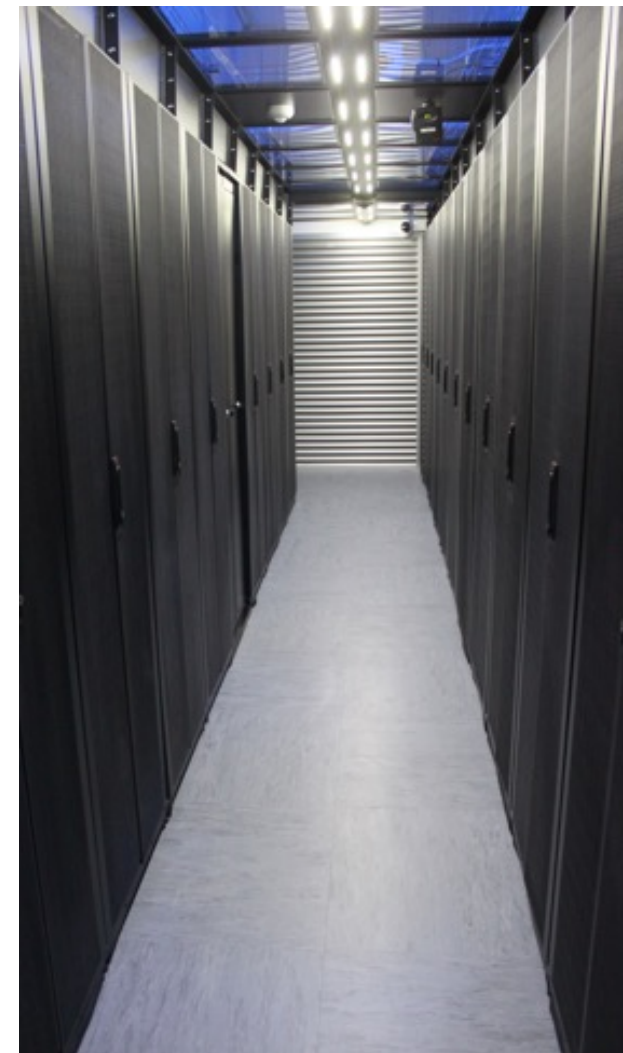


UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



Introduction

- Use of community Science Gateway
- Supporting common environments
- Research Data Management exploration





CyberSKA portal

- Use of portal / science gateway to access resources
 - Enables services to be maintained for all users
- Have user base that use this for part of their work
 - Data management and interactive visualization well supported
 - Workflow management needs replacing
- More effort would have been required to support some work
 - Development effort focused on primary user's use cases

The screenshot shows the CyberSKA portal interface. At the top, the header reads "CYBERSKA" and "A Cyberinfrastructure platform to meet the needs of data intensive radio astronomy on route to the SKA". Below the header is a navigation bar with links for Home, Profile, Settings, myDashboard, myGroups, Tools, and About Help. A search bar and a "Log out" button are also present. The main content area is divided into several sections:

- Profile:** Cameron Kiddle, with options to subscribe to the feed, bookmark the page, and edit the page layout. A sidebar lists "My applications", "My blog", "My bookmarked items", "My event calendar", "My files", "My pages", "My publications", and "My tasked items".
- Contacts:** A grid of user avatars.
- Event calendar:** A list of events, including "Imaging Science Technical Meeting" (10:00 - 11:00, 28 Jun 2011), "CANARIE Site Visit (Tentative)" (1:00 - 3:00, 29 Aug 2011), and "ADASS XXI: Astronomical Data Analysis Software and Systems Conference" (6 Nov 2011 - 10 Nov 2011).
- Recent Astro-ph Eprints:** A list of recent publications, such as "Testing the cosmological evolution of magnetic fields in galaxies with the SKA" and "A generalised Measurement Equation and van Cittert-Zernike theorem for wide-field radio astronomical interferometry".
- Active Users:** A list of active users, including Cameron Kiddle (busy preparing for SKA 2011) and Samuel George (waiting for processing).
- Group membership:** A list of groups, including "DMS support of Astronomical Data", "CyberSKA Sys Admins", "Application Developers", and "Portal Support".
- Activity:** A list of recent activity, including "Samuel George bookmarked Detection Thresholds and Bias Correction in Polarized Intensity" and "Russ Taylor updated a page titled SubGroups".
- Pages:** A list of pages, including "CyberSKA Updated Collaboration Requirements - Phase II".



Common code environments

- Use of cloud resources more attractive to community than using HPC
- Repository of maintained VM images results in same images being used on multiple servers / clusters
- Repository of Docker images to deploy software in a reproducible way
- Repository of iPython notebooks being created to enable reuse
- GitHub used to track changes in notebooks, recipes and scripts
- Aim to produce standardized data models and format

Going forward: Research Data Management

- Greater automation in linking code, images and workflows would be useful long term
- Possible through Gateway; harder to do with user code that is run independently
- Started working with UCT library on Research Data Management (RDM)
 - Part DIRISA Tier 2 facility
 - Funding for Data Scientist working on RDM issues
 - FigShare used to link publications, data and code
 - Interested in collaboration on open source solutions to creating Research Objects
 - already hosting ADS replica for South Africa

