Discovering the Unexpected in Large Surveys

Brian Schmidt (ANU)

CAASTRO
ARC CENTRE OF EXCELLENCE
FOR ALL-SKY ASTROPHYSICS

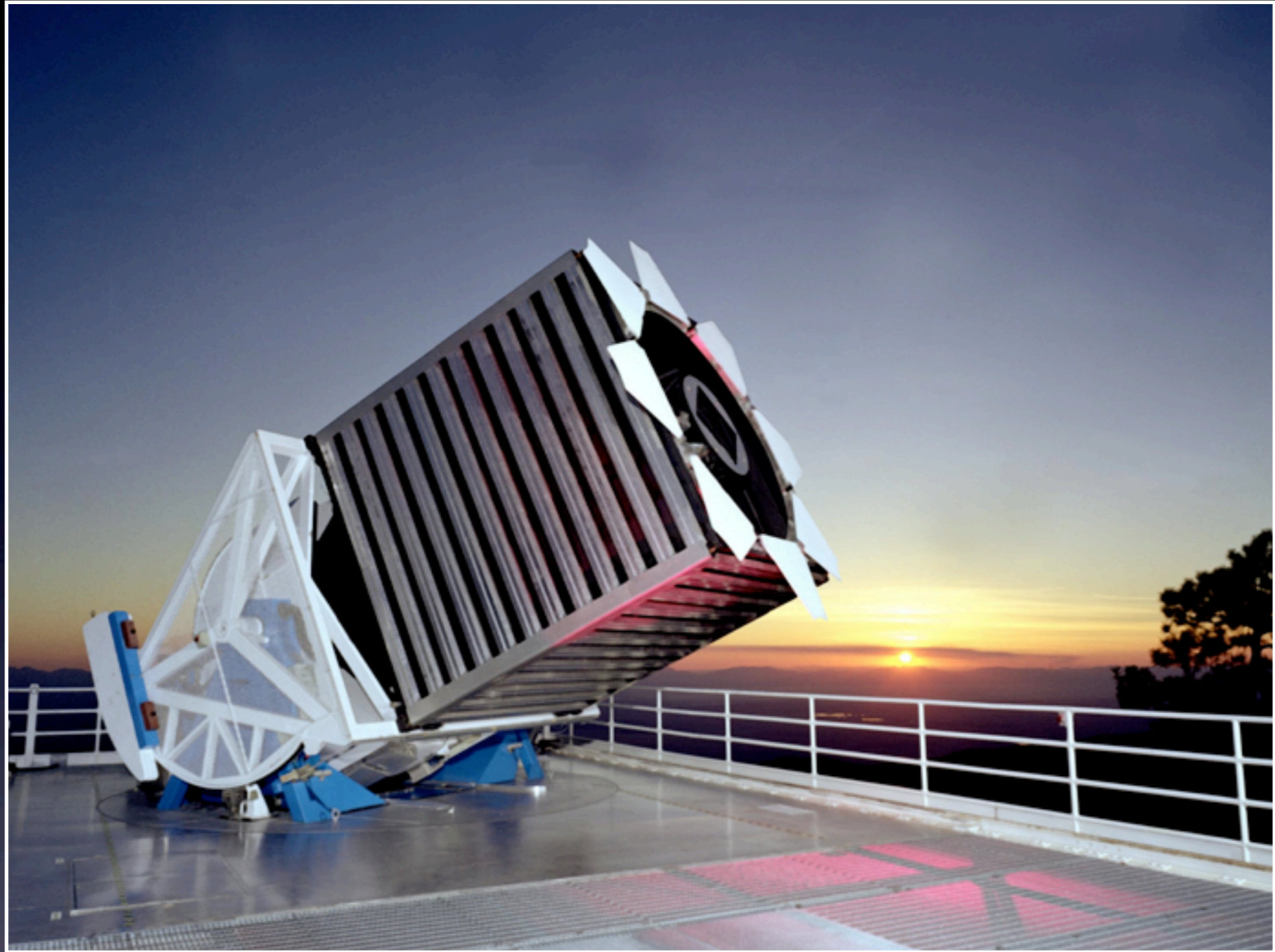Australian
National
University
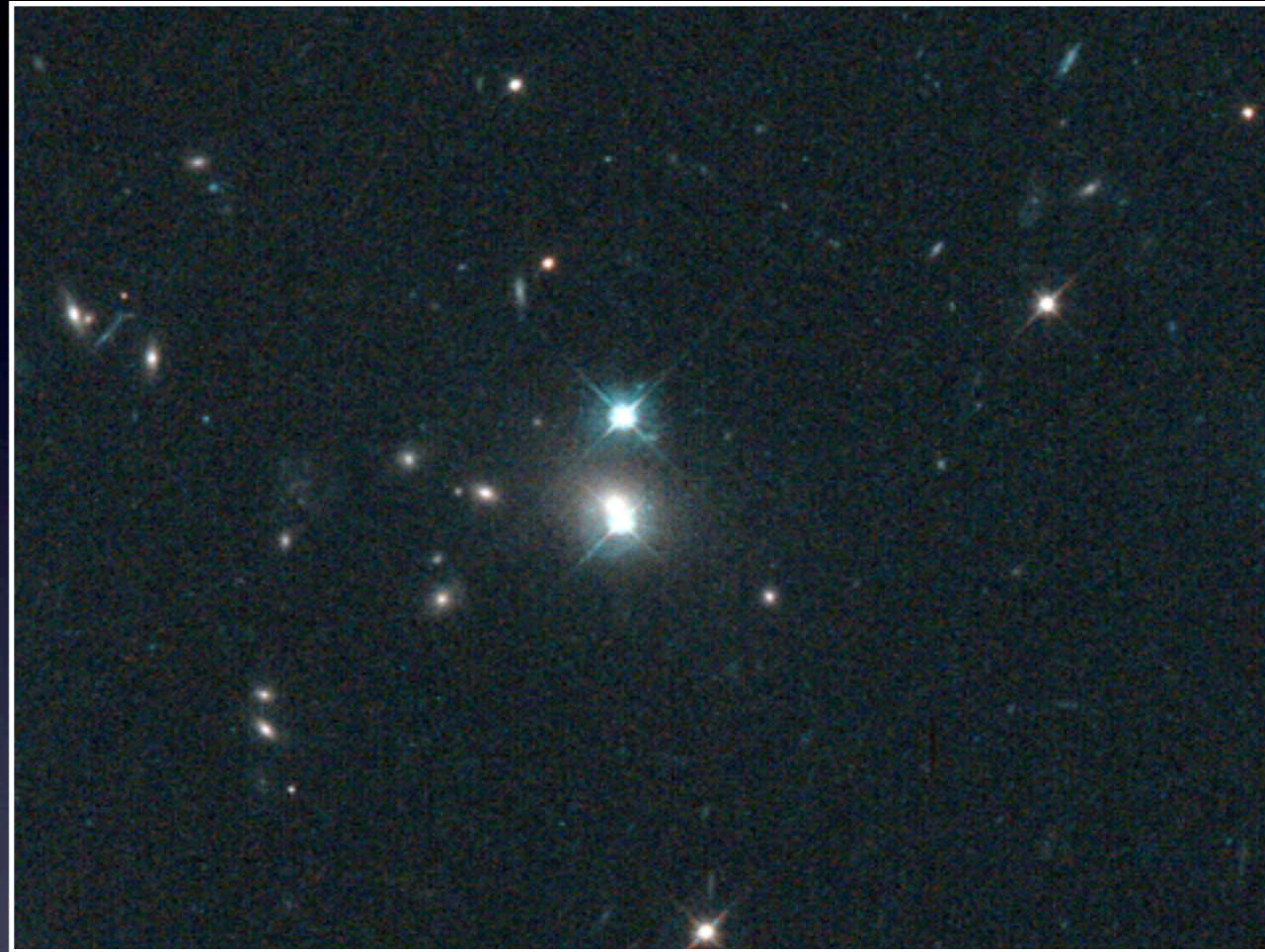
Big Discoveries in Astronomy

Hero Mode

- Hubble - A little help from your friends

- Baade - Dark Nights on the 200inch

- Zwicky - Always willing to thinking outside the Box

- Penzias & Wilson - When Bat Guano is eliminated, look to Princeton

- Bell & Hewish - Not Little Green Men

Most Cost Effective Facility of All Time? 5600 papers

Major SDSS Papers

- z>6 QSOs and the end of the epoch of reionisation

- Brown Dwarfs (with 2MASS)

- Many New Gravitational Lenses

- Sub-Structure of Milky Way

- Smallest Low Surface Brightness Galaxies (non-SDSS)

- Dynamical Asteroid Families

- Hyper-velocity Stars (non-SDSS)

- Baryon Acoustic Oscillations

Common Theme

- Not the Major Selling Point of SDSS

- Made Possible by giving Access to the Data to Many Astronomers

- Most events could not be anticipated by the survey Developers...

- People were heavily invested in Program

Key is Allowing Wide Spread Exploration - NOT Blindly Exploring
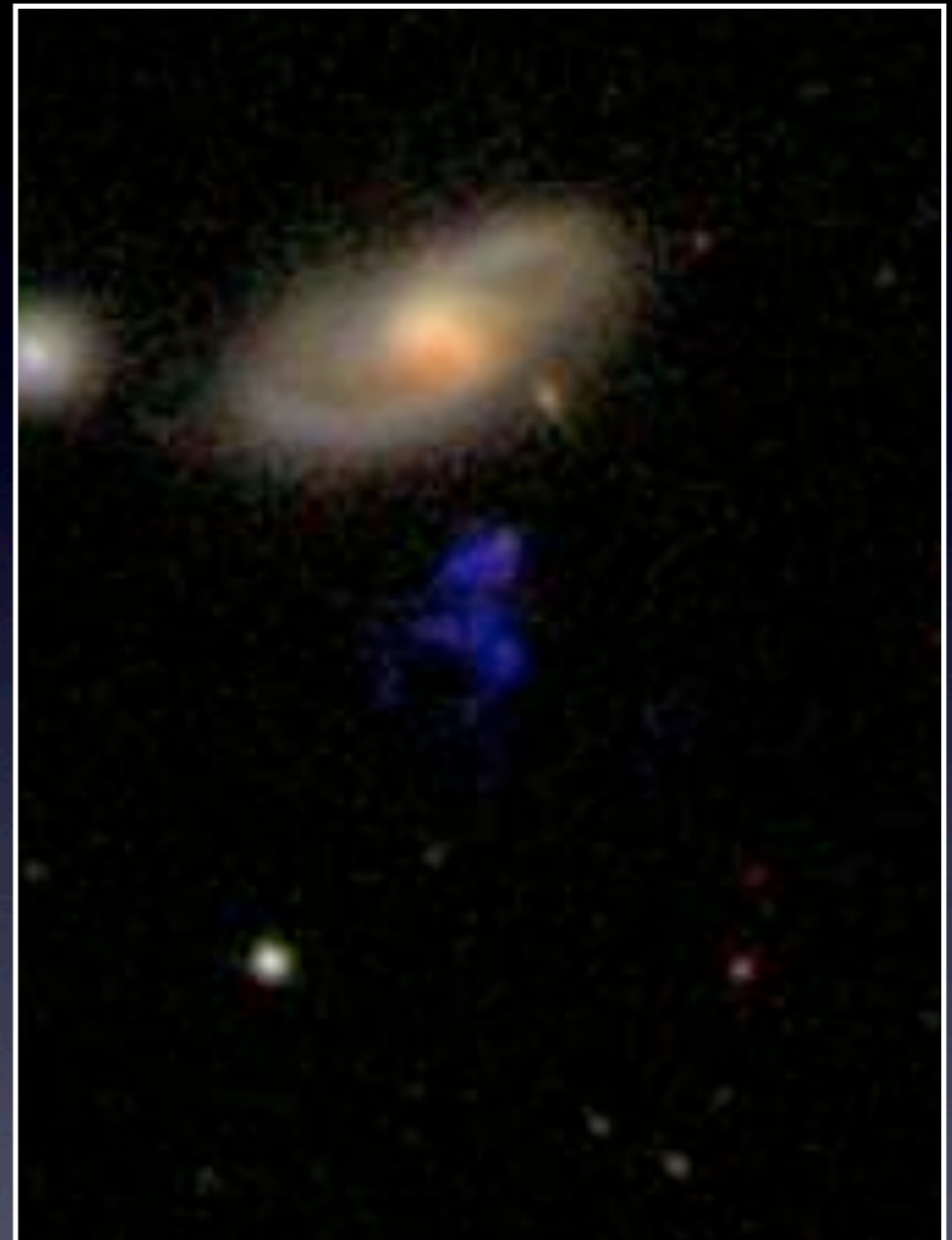
What SDSS did Right

- Large Invested Research Community
- Data Made Available As Easily As Possible
- Unique but interesting Data

Unique Datasets

- SDSS increased by 1000-fold the precise multi-colour photometric data

- Data identified the nature of 10^9 objects allowing statistic studies to be done in almost any area of astrophysics

- Essentially Impossible to Repeat

But Every Dataset worth Taking is Unique

- Science Programs operate to do unique science

- Linking Each significant Dataset to the Greater Astronomical World provides opportunities

Connecting Data Sets

- We are currently at a time where we are collecting information at different wavelengths of every object in the sky to a given flux limit (surveys)

- New opportunties come from not just the uniquess of each data set, but also putting together these disparate sets of information to answer new questions

Accessing Data

- Being able to nimbly manipulate and explore data is a key-factor to scientific creativity

- Databases for easy access

- Tools for easy access

Tool-kits instead of Tools

- Tools of my PhD Studies

  - IRAF

  - DOPHOT/DAOPHOT/SEXTRACTOR

  - MONGO/SM

  - SAOIMAGE/DS9

- My own C/csh programs and libraries

  How I wished all of these tools and parts there-of were easily linkable to my own code

The New World

- Astronomers are currently using PYTHON

- Provide a tool-kit written in whatever (C/C++/Python/F90/R ...) linked to PYTHON

- Build Tools from this tool-kit for use - not stand-alone tools.

- Allows others to build and share new tools built up from this common tool-kit

Science works Best When it builds on itself

- Value-Added Datasets

  - Users create new Datasets from subsets of other Datasets or combinations of datasets

- These new Databases need to be published and easily inter-linked to everyone

My Vision

Primary Data

Tool Kits

Tools

Secondary Data
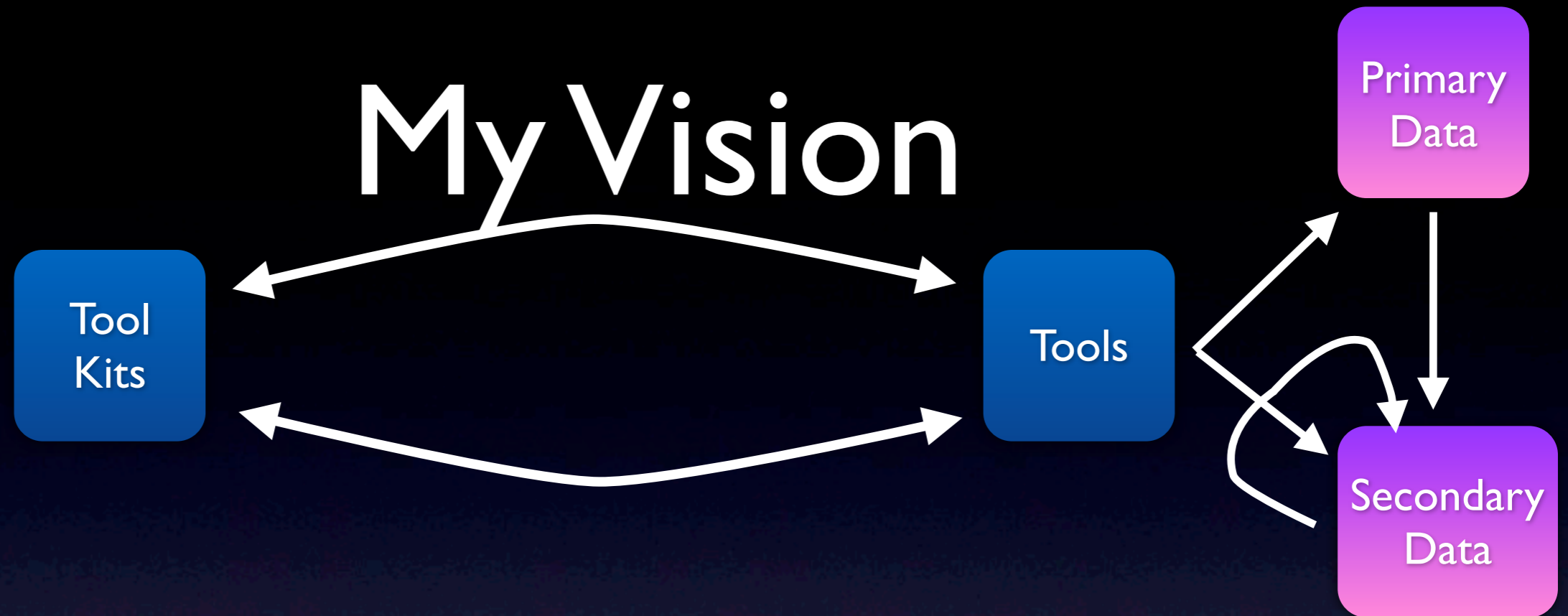
- Inter-operable tool-kits built into sharable tools by the community, operating on

- Inter-operable datasets which can be built into new datasets, which are once again inter-operable and shared worldwide

SkyMapper Dataset

- All objects of the Southern Sky to SDSS Depth

- Provide tools within the VO framework for accessing Datasets

  - But we do not have the resources to build our own toolkits

- Allow Users to build their own value-added data sets - (identification, phot-z, etc) - but how do we easily link up for everyone?

Your Vision

- How do we get the most out of data?

- How do we build on each others work?

- How do we build datasets and tools which gracefully age?

- These I see as the key questions of Astro-infomatics