The Australia Telescope Online Archive and using rich metadata to drive scientific processing

Peter Lamb
CSIRO ICT Centre
Australia Telescope Online Archive

- Introduction
- The Australia Telescope Compact Array
- Building the AT Online Archive
- Radio Interferometry data
- Virtual Observatories: Service Description, Composition & Execution
- It’s not just Virtual Observatories...
The Grid & eScience

• Convergence of high-performance computing, huge data stores and high-bandwidth computing to:
  • share resources (data/computation/instruments)
  • enhance collaboration
  • allow for new ways of conducting research
Virtual Observatories & the IVOA

- Global electronic access to astronomical data archives of space and ground-based observatories & sky surveys
- Coordination data analysis using common standards, high-bandwidth networking & state-of-the-art analysis tools
International Virtual Observatories Alliance

- UK – AstroGrid
- Australia – Aus-VO
- EU – AVO
- China – ChinaVO
- Canada – CVO
- France – VOFrance
- Germany – GAVO
- Italy – DRACO
- Japan – JVO
- USA – NVO
- Russia – RVO
- India – VO-I
- Korea – KVO

http://www.ivoa.net/
Aus-VO Projects

- HI Parkes All-Sky Survey
- 2QZ – spectroscopic quasar survey
- MACHO archive
- ATCA archive
- Machine learning

- SUMSS 843MHz southern sky survey
- RAVE radial velocity survey
- Remote Visualisation

http://www.aus-vo.org/
The Australia Telescope Compact Array

- 6×22m dish interferometer telescope
- Maximum baseline 6km; minimum 30m
- 5 antennas can be positioned at 15m intervals on 3km of E-W track
- 214m N-S spur track
- Receivers for 20cm, 13cm, 6cm, 3cm, 12mm, 3mm (3 receivers only)
The Australia Telescope Compact Array

- Started operations June 1990
- Generates ~0.5GB of data/day
- Base archive ~2700 CDROMs + a few Exabyte tapes at Narrabri
- ~1.7TB
Building the archive

- ATCA has 128kb ISDN link as its only external network
- ~1200 days to transfer data
- Transfer on disk? tape?
- Tape: creates additional copy of archive
- Data transferred on 54 DLT tapes
Problems transferring the data

- **Tape read problems:**
  - one DLT tape was unreadable
  - several were readable only on a DLT drive and not the SDLT drive
- **Two RAID failures with complete loss of data on the failed RAID**
- **Data corruption on Reiser filesystem**
Problems building the archive

- Unfamiliarity with radio interferometry
- Unfamiliarity with data formats
- Unfamiliarity with data curation
- Checksums on data files collected too late in process
Problems building the archive

- Data was not collected with a view to use in an archive.
- Important metadata was:
  - incomplete and/or inconsistent
  - some metadata sets in varying formats
  - not in computer-readable form
  - departed with observer
State of the Australia Telescope Online Archive

- All ATCA data up to Feb 2003 loaded
- Clean backup on 18 SDLT tapes
- Metadata extracted from RPFITS data files into Oracle database
- Currently preparing proposals data for loading into database
• Scripts for inferring some observation metadata (calibrators, logical groupings of scans, mosaics) in prototype

• Positions database load script in prototype
State of the Australia Telescope Online Archive

- Web front end to search metadata and download files
- Experimental Grid front end using OGSA-DAI
- Need a computable definition of proprietary data period rules
- When data availability rules are implemented, can go public
ATOA Data Model
IT Challenges of Virtual Observatories

- Provide a uniform model of observational, derived and modelled data for an entire scientific discipline.
- Provide a model of computation for tools used to reduce the data.
- Allow for automatic, assisted and convenient manual assembly of data sources and tools.
Radio Interferometry Imaging
Raw data to source extraction

Amplify, filter, correlate

Calibrate, Inverse Fourier Transform

Many details omitted!

Deconvolve

Source extraction
Raw data - visibilities

- Visibilities are samples (per frequency channel, polarisation) of the Fourier Transform of the image
- Visibilities are the data stored in the ATCA archive
- Can derive images, spectral data cubes, time series, large mosaiced images, etc
"Dirty" Image

Calibrate, Inverse Fourier Transform

• Formed from the Inverse Fourier Transform on the visibilities

• Has high level of artefacts because of incomplete sampling of the Fourier plane
Deconvolve

- Deconvolve model of interferometer beam (derived from visibility sample locations) against dirty image, replacing the “dirty” beam with a “clean” Gaussian beam

- Highly underdetermined!
Source Extraction

- Fit Gaussian models to bright areas of the image
- Requires classification of the sources – true source, artefact, noise...
The Virtual Observatory is across the whole spectrum

Search optical VO data

Image above: Copyright © 1995 A. M. Koekemoer et al. Australia Telescope National Facility
IT Demands

- Large distributed database (ATCA data alone is ~1.7TB)
- Need common data model
- Need to
  - describe
  - compose
  - execute
data queries and processing
Service description

Source detection

Radio

Optical

Data query

X-Ray

Radio imaging

© CSIRO ICT Centre 2003
Service composition

User task description  User constraints

Service metadata  Service composition  Composition rules

Radio  Data query  Radio imaging  Source detection

Data query  Source detection  Data query
Where we are

- Select an area of interest in the sky
- Assemble or acquire the data sets (does the data I need exist already)?
- Process/reduce
- Combine
- Analyse & check hypothesis
- Double-check data and processing

✓ Scientific results
Where we want to be

- State the data and processing requirements
- Receive the data processed as needed
- Check the hypothesis
- Double-check the data and processing

✓ Scientific results; sooner & with less effort
What’s in the way

- We have archives, not data access nodes
- Many kinds of data: radio, optical, X-ray; images, spectra, polarisation maps
- Much data processing/reduction needed to make inspection, search & fusion possible
- Data treatment is a craft
- Still a lot of work to do on service description
Directions

- Complete metadata construction for Australia Telescope archive.
- Paper trial of representative sample of existing service composition tools.
- Execution trial of reduced sample (no more than 2-3) composition tools.
Directions

• Deliver:
  • composition requirements and constraints
  • service discovery and testing
  • composition debugging and testing
  • service availability
  • resource optimisation
  • exception handling
  • recomposition
  • service compositions available as services.
Not just Virtual Observatories

- Smart, dynamic service composition for both Web Services and on the Grid is a challenging problem.
- Its application goes well beyond Astronomy and e-Science, and into the composition of complex services of any kind.
Australia Telescope Compact Array

Images © Copyright CSIRO Australia, 1995-2003