The Virtual Observatory – enabling data access – key science drivers

Nicholas Walton
AstroGrid and Euro-VO Technology Centre
Project Scientist
(Institute of Astronomy, University of Cambridge)
The Virtual Observatory

• The Challenge: Facilitate Better Science
  – Science: range of astronomy, solar to cosmology
  – Technical: global system
  – Political: global community

• Building a solution: The Virtual Observatory
  – standards
    • IVOA
    • GGF
  – implementations

• Context for this workshop
  – AstroGrid: an example VO implementation
The Challenge of Data

- Astronomy is an observational science
- Progress is made via understanding gained from the study of the cosmos
- Powerful observatories exist producing observational data across the wavelength domain
  - data comes in many formats
  - levels of complexity
- Data & applications held globally: USA, Europe, Asia
  - heterogeneous data archives
- Research partnerships are also global
- Connecting researchers with data and applications is the challenge for the Virtual Observatory
The Need for Virtual Observatories: Managing Technological Change

- The massive **Growth of Data**
  - Number + size of telescopes
    - Optical: ESO's 4x8m VLT, 2x8m Gemini
    - X-ray: XMM-Newton
    - sub-mm: ALMA
  - Increase in **size and multiplex** capabilities of instrumentation:
    - Infra-Red: VISTA > 100 GB/nights
    - Radio: e-Merlin > data rates ~320 Gbps
    - All sky at 0.1 arcsec – 100 TB
Supernova Remnant
Cassiopeia-A – a 300 year old Supernova

The Challenge and Opportunity of multi-Wavelength data:

Shocks seen in the X-ray Chandra image

Dust shows in the IR

Heavy elements seen in the optical

Mapping $\bar{e}_s$ in the magnetic field as revealed by Radio data

Images from Chandra Science Centre
What happens to the Earth's magnetosphere during a coronal mass ejection?

Event imaged by space based solar observatory

Effect detected later by satellites and ground radar

New & Improved Science from VO's:
Space Weather

SOHO/EIT – EUV
Yohkoh – Xray
New & Improved Science: Cosmology

Multiple large image sources: registration & association

Automatic cluster finding techniques

Multi-TB $\lambda$CDM models, e.g. Millennium Sim

Generate Shear Maps c.f. CDM models > DM distribution with redshift

Remove stars correlate gals with $z$

X-ray cluster: Chandra X-ray (Mullis) overlaid on a deep BRI image (Clowe & Luppino).

Source ID from multiplexed spectral data

Colour-Colour relationships classification in multi-phase space

Figure 7. Example of astrometric & Wide Field Camera on WF.

ESO
Science Challenge

• Multi-Wavelength data required
  – key science: e.g. planets, large scale structure, galaxy formation: all needs a combination of data

• More data required
  – survey telescopes create ~TB/night

• More use of database organised resources used

System needs to be fast and easy for the astronomer to use and affordable for the data providers to operate

System use drives increased (sustainable) funding for operations and development of data access system
Solutions: a Flexible Framework

• Create a system that recognises:
  – no one data provider or repository: thus data interoperability
  – application provision

• Requires
  – a system built upon agreed interoperability standards

• Exploits
  – wider IT developments: Grid and WS technologies
    • power of XML/ SOAP etc
  – access to high speed networks
    • but note: backbones ~10Tb/s, desktops ~100Mb/s
  – reduced costs of h/w: all data now on spinning disks
Building the Virtual Observatory:

Global scope
International partnerships
Agreeing interoperability standards
Building regional implementations
Based on new computational technologies
Deployed on the fastest networks
IVOA: Stds Enabling Interoperability

- The International Virtual Observatory Alliance
  http://www.ivoa.net
- A global partnership
- Projects represent global astronomy data providers
- IVOA a forum for interoperability standards
- VO projects build on these standards
- Global reach
IVOA Architecture Analysis

Analysis of a VO:
- Multi-layer
- Complex
- User interfaces thru a portal
- Astro-apps interface to a VO abstraction layer
- Lower level middleware provided by the 'grid' world
  - e.g. SRB
- Hardware at bottom layer

IVOA has working groups to address 'astro' specific 'boxes'

Ref: IVOA Architecture Overview: Williams et al, 2004
IVOA Note 2004-06-14:
http://www.ivoa.net/Documents/Notes/IVOArch/IVOArch-20040615.html
VO Standard Areas

- Standard vocabulary (semantics)
- Standard ontology describing how terms are related
- Standard data model (encoding format) for each type of measurement
- Standard query language for issuing spatial, temporal, and semantic queries across the catalogs.
- Standard access services for retrieving catalog records or image cutouts.
- Standard mechanisms for interacting with storage systems (VOSpace)
- Standard authentication/authorisation mechanisms
- Standard event notification services.
IVOA Working Groups: [http://www.ivoa.net/forum](http://www.ivoa.net/forum)

- **Registry:**
  - how to 'register' resources: concept of VOResources
- **Data Access Layer:**
  - Standards for remote data access: e.g. SIAP, SSA
- **Data Model:**
  - Standards for the actual data: e.g. XML'ing of FITS
- **VO Query Language:**
  - Standards for 'astro' database access: e.g. Openskyquery, 'circle'
- **Unified Content Descriptors: std Vocabulary**
  - Standards for common ways of describing data: metadata
- **VOTable:**
  - XML representation of tabular data
- **Grid & Web Services:**
  - Interfaces to Grid and Web Service stds: e.g. 'Heartbeat'
GGF and Astronomy

• Current IVOA implementations largely utilise web services, however GGF standards are becoming relevant in a number of areas:
  – Authorisation
  – Transport, e.g. GridFTP
  – Persistent Storage
  – Job Submission

• Use of grid service standards actively under investigation by a number of VO projects
  – AstroGrid
  – USA's NVO
Science Driven Virtual Observatories
US-VO: NVO http://www.us-vo.org

- Partnership of major data/compute centres in the USA:
  - IPAC, NASA-HESARC, NASA-JPL, NRAO, NOAO, SDSS, SAO, STScI
  - SDSC, NCSA, Pittsburgh
  - Globus, MS

- Webservices based
- Initially tools
  - Now moving to an architecture
Japan-VO: JVO  [http://jvo.nao.ac.jp](http://jvo.nao.ac.jp)

- **Partners**
  - NAOJ, JAXA, ICRR, Ochanomizu U, Osaka U, Titech, Fujitsu
  - Data from Subaru, Astro-F, Nobeyama, Alma

- **1st JVO prototype based on Globus Toolkit 2 – 'grid'**
- **2nd JVO prototype based on GT3 – grid services, improved performance**

---

**Demo 1: Cross match & Image request**

Cross match of the optical and X-ray catalogs of SXDS and image retrievals.

- (1) Search/Request (JVOQL)
- (2) cross match result
- (3) catalog search
- (4) xmatch request
- (5) image
- (6) image
- (7) xmatch request
- (8) result

**Demo 2: Cosmic String Search**

Data request to the SXDS optical catalog, GL candidate selection, String search by pattern recognition.

- (1) GL candidate selection
- (2) catalog search
- (3) result
- (4) GL candidate selection
- (5) GL cand. selection
- (6) String search
- (7) String search
- (8) result
AstroGrid: UK's Virtual Observatory

Empowerment of scientists

- Improve the quality, ease, speed and cost effectiveness of on-line astronomy
- Make comparison and integration of data seamless
- Removing barriers to multi-wavelength astronomy
- Enable access to very large data sets

Project: 2001-2007: key data and resource providers in consortium
Infrastructure, new tools, resource discovery, data mining+visualisation
AstroGrid: An example implementation
AstroGrid 2006.2 Release: Apr 2006

http://www.astrogrid.org/launch
Workbench

- User Interface to VO services
- Delivery via Java Webstart technology
- Components
  - Registry
  - Find Data
  - Work with Apps
  - Workflows
  - Client Visualisation
- Enables Science
Wider Relevance

• Use of IVOA stds vs use of more generic GGF stds
  – implementations in use – supporting large data flows, complex computational environments
• Comparison with other domains, e.g. Medical
  – workflows and application environment
  – server based system with plugable clients for data visualisation
  • potential use from medicine to literature
• Supporting People Systems and Technical Systems
• Following talks and presentations look at specific areas where GGF standards may have relevance in supporting the construction of the Virtual Observatory
Closing

• The Virtual Observatory is developing a science driven infrastructure to allow access to distributed data and applications required to perform analysis and interpretation
  – 2006 is the year of rollout of significant systems in the USA, Europe, Asia and the UK

• Key Links
  – Astro-RG @ GGF: https://forge.gridforum.org/projects/astro-rg/
  – IVOA: http://www.ivoa.net
  – AstroGrid: http://www.astrogrid.org
  – NVO: http://www.us-vo.org
  – Euro-VO: http://www.euro-vo.org