Exploiting Virtual Observatory and Information Technology: Techniques for Astronomy

Nicholas Walton
AstroGrid Project Scientist
Institute of Astronomy,
The University of Cambridge

Lecture #1 Goal:

Explore the emerging new distributed data resources, and mechanisms to (remotely) analyse that data

Course Overview

- Lecture One: VO: Virtual Observatory overview, the technology (grids), science drivers, VO standards, VO Projects
- Lecture Two: Data: databases, science archives, federating data, registries, xml. Science: hunting for brown dwarfs
- Lecture Three: Applications: workflows, data analysis systems, theory services. Science: photometric redshifts
- Lecture Four: SQL: data mining with SDSS DR3, open sky query, Sloan Web Services [Richard McMahon]
- Lecture Five: Demonstrations of VO Tools: I: apps
- Lecture Six: Demonstrations of VO Tools: II: workflows
- Lecture Seven: Radio/sub-mm I: ALMA, SKA, eVLA, eMerlin; challenges and solutions [John Richer]
- Lecture Eight: Radio/sub-mm II: applications and pipelines, analysis algorithms [John Richer]

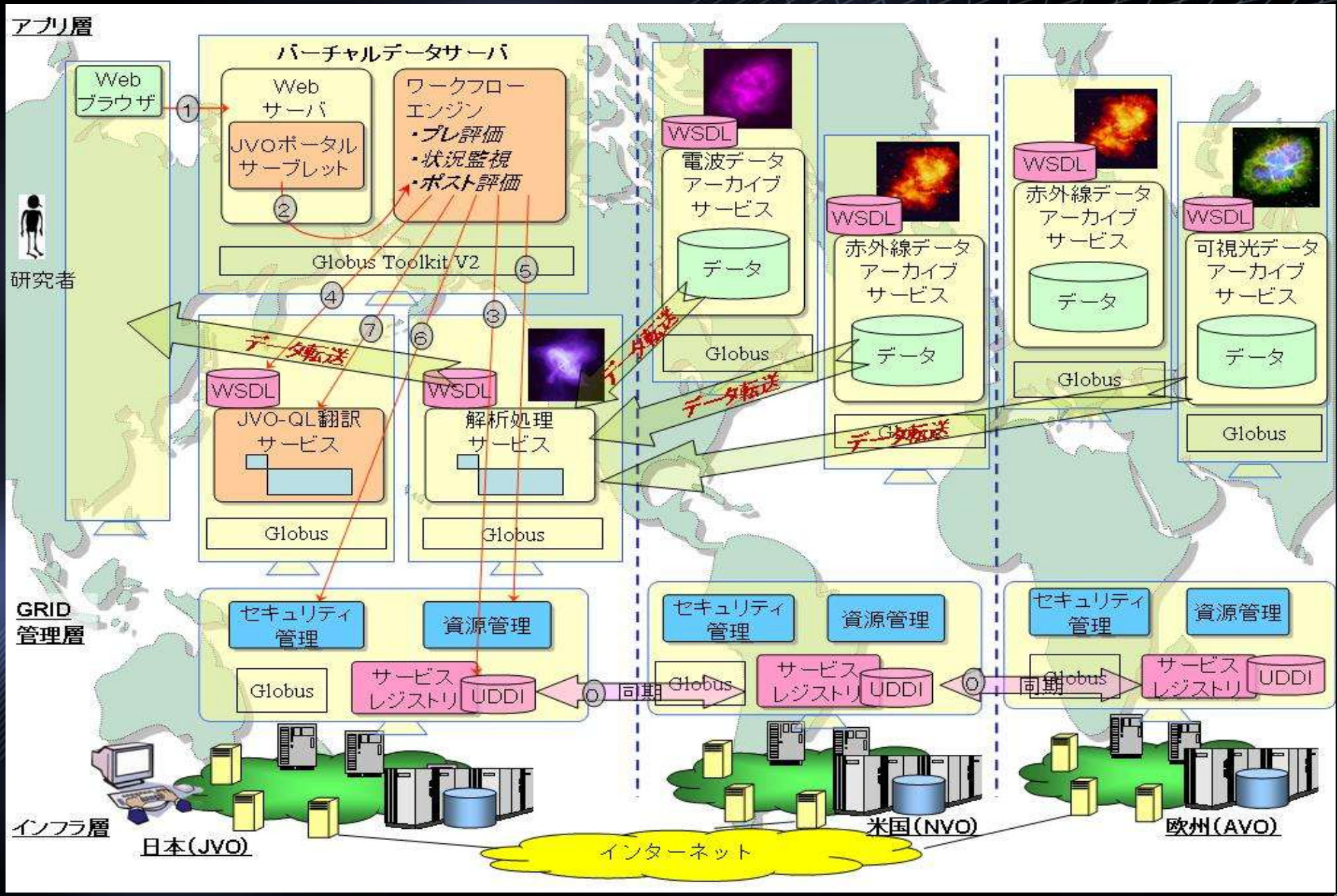
Summary: Lecture #1

- Introduction and Background
- Virtual Observatories
  - Science Drivers
  - Technological Drivers
- Distributed Computing
  - Networks
  - The Grid
- Supporting Computational Technologies
  - Web Services
  - Grid Services
- VO Technologies
  - VO Standards

Introduction and Background

- This course is *not* about programming per se,
- It is about:
  - Locating data and information resources
  - Discovering newly emerging Virtual Observatory Systems
  - Learning use of these systems
  - Moving on to use these systems in your science
  - ... and then possibly doing some of your own development
- This lecture introduces the Virtual Observatories
  - Following lectures explore the use of these facilities

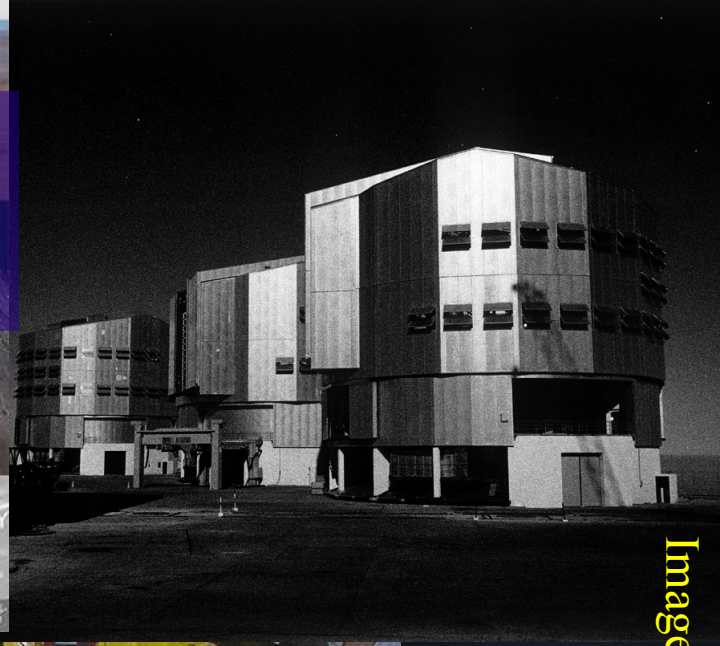
But first ... VO in a nutshell ...



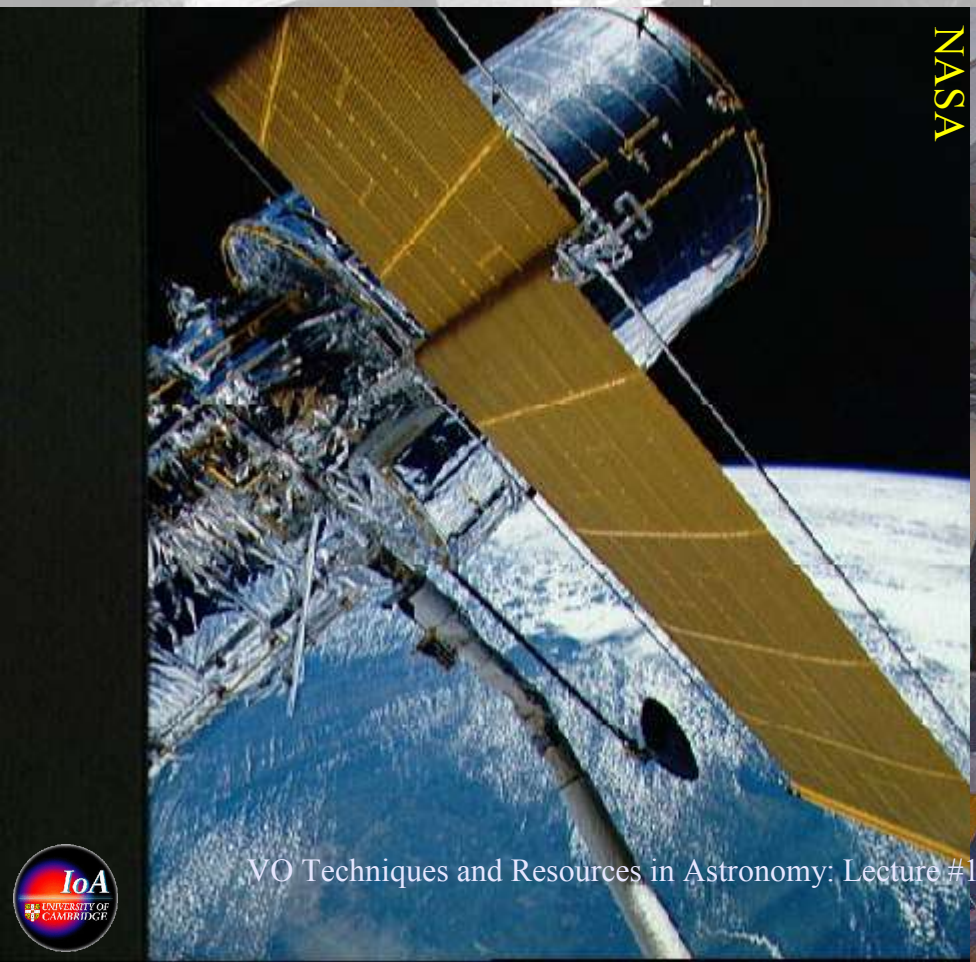
Virtual Observatories: Overview

From WHY to HOW and WITH WHAT

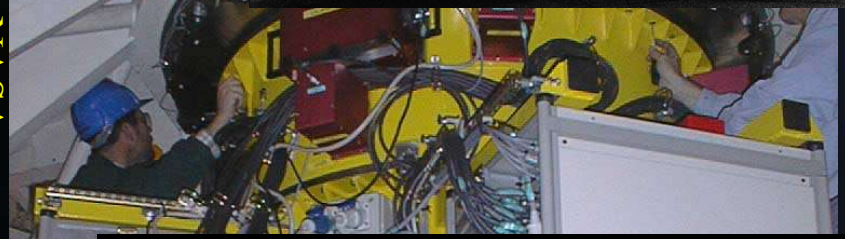
Traditional Observatories

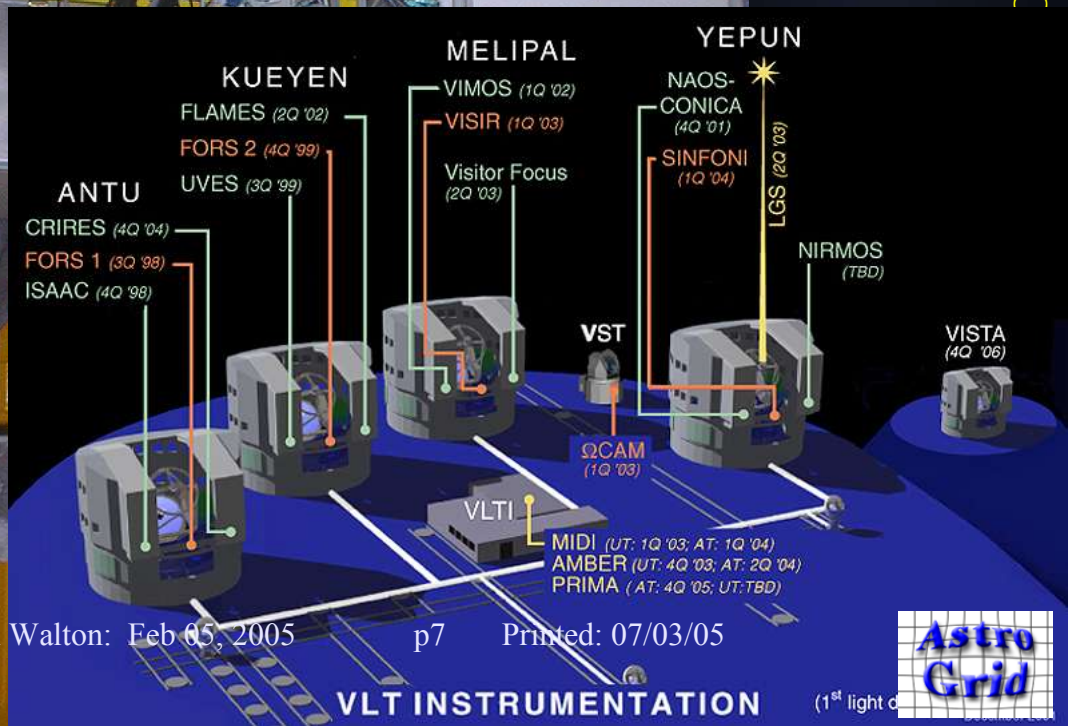


Images from ESO

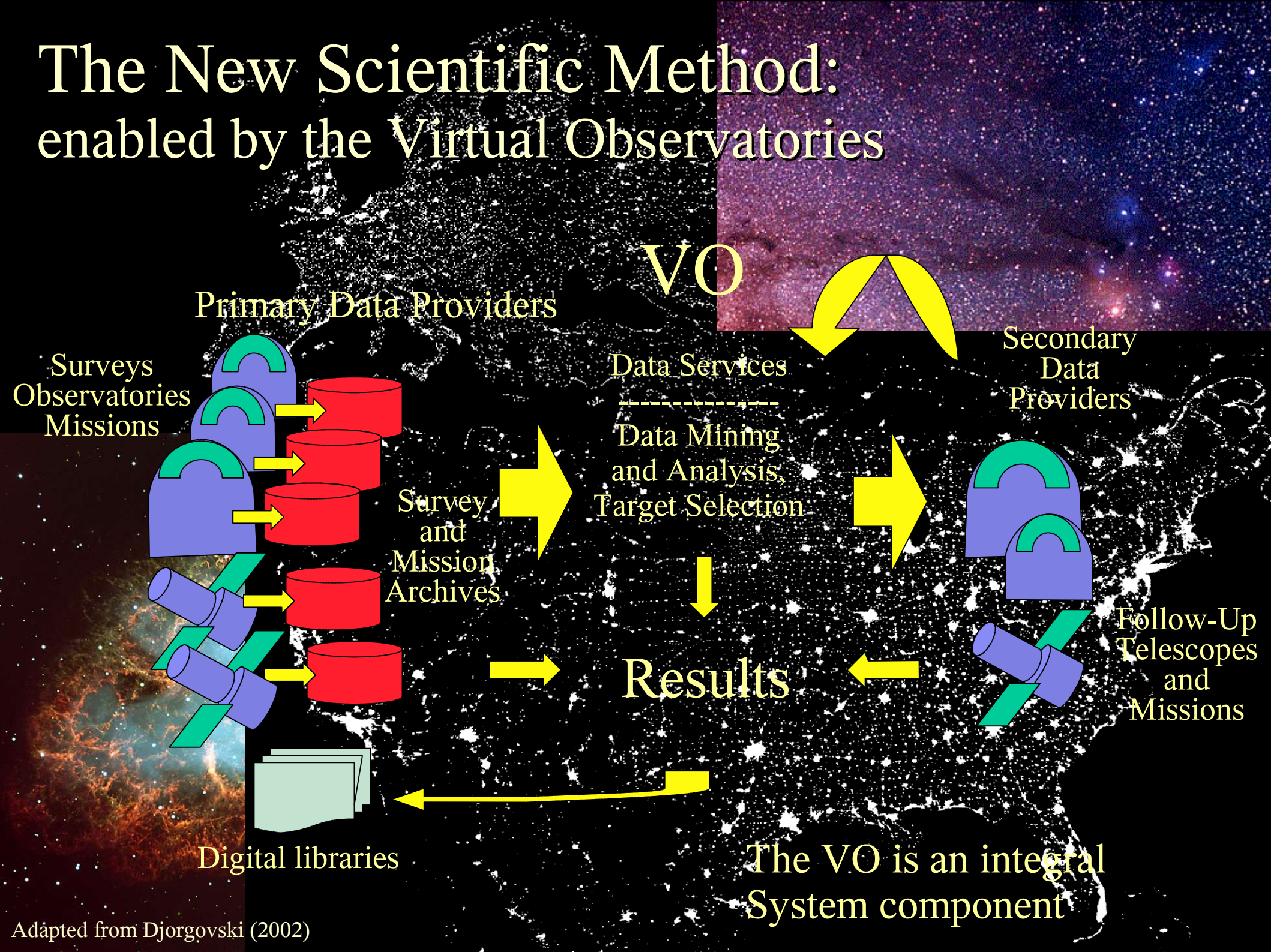


NASA





The New Scientific Method:
enabled by the Virtual Observatories

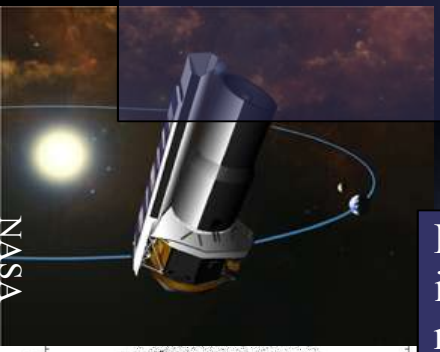


Adapted from Djorgovski (2002)

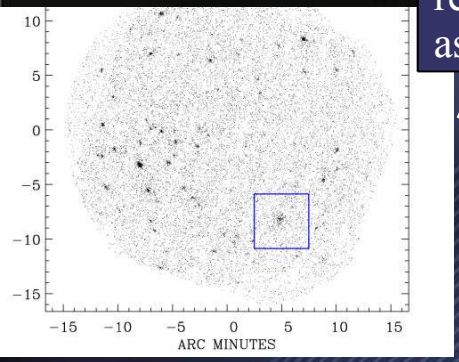
Science Drivers for a VO

- Enable new science
  - Perhaps through doing things that weren't thought of before
  - By making possible science that was impossible to do before
    - e.g. High precision cosmology, where mass data handling is required
  - By making it easier to what is currently hard to do
    - Thus giving the researcher – you – more time for interpretation of results
- All VO projects list a number of key use cases

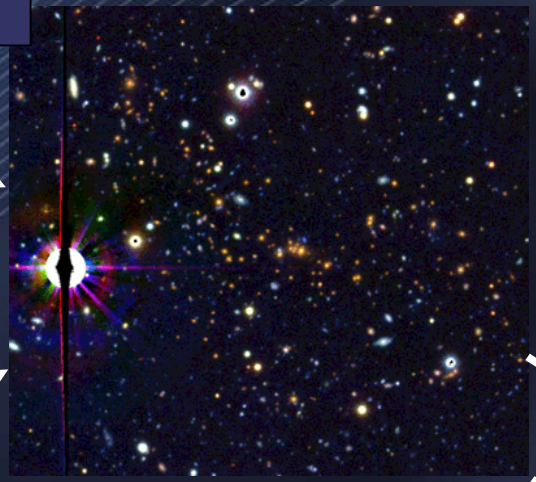
New Science from VO's: Cosmology



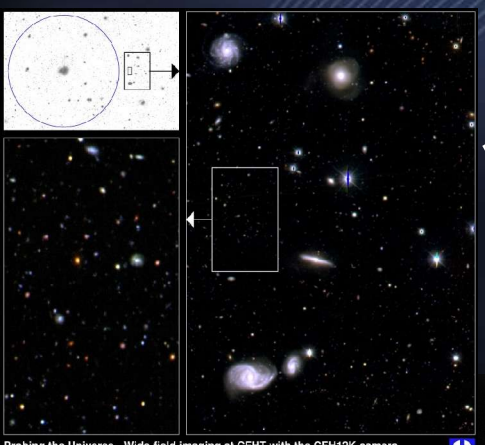
Multiple large image sources: registration & association

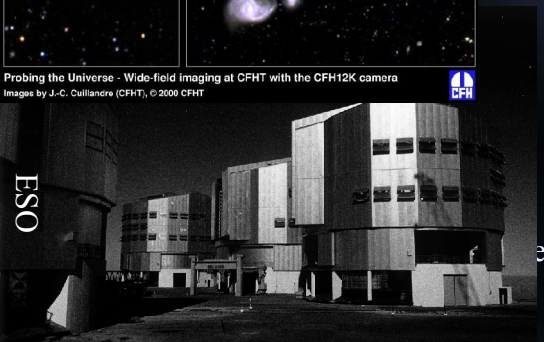


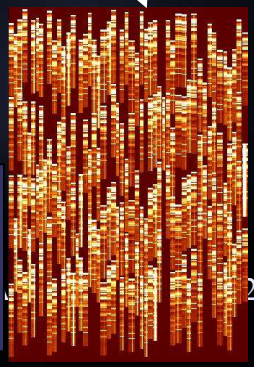
Automatic cluster finding techniques



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

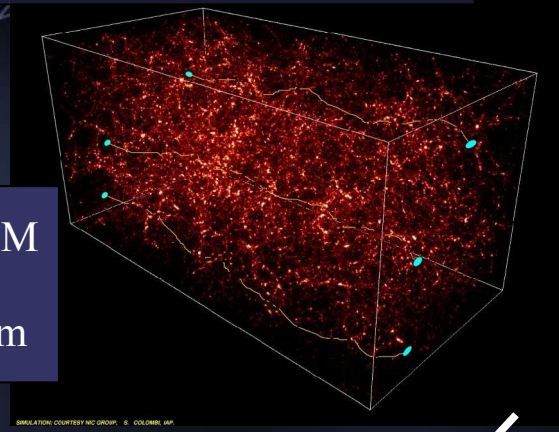




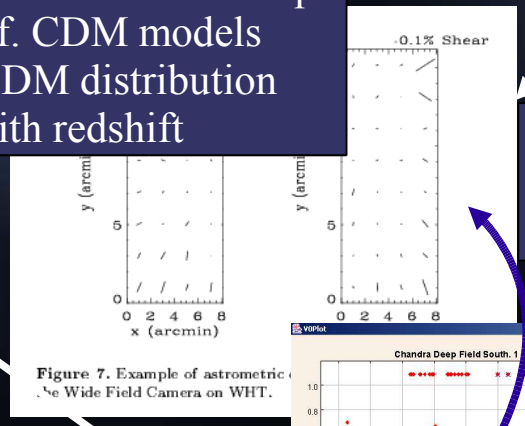


Source ID from multiplexed spectral data

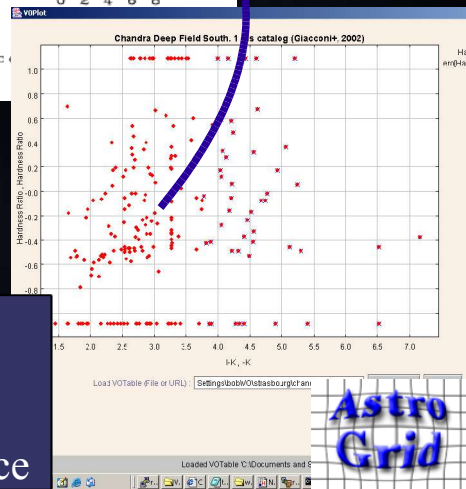
Multi-TB λCDM models, e.g. Millennium Sim



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



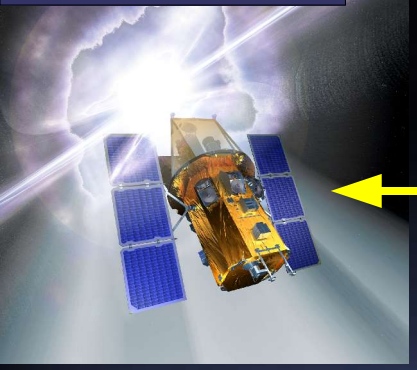
Remove star correlate gal with z



Colour-Colour relationships classification in multi-phase space

Gamma Ray Bursts

SWIFT satellite observes gamma ray burst



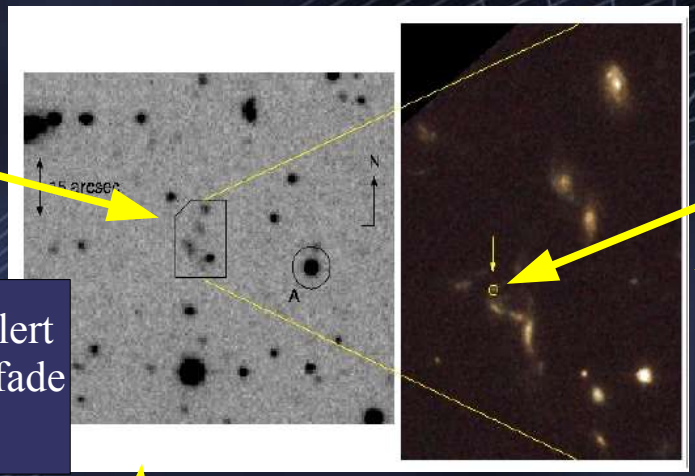
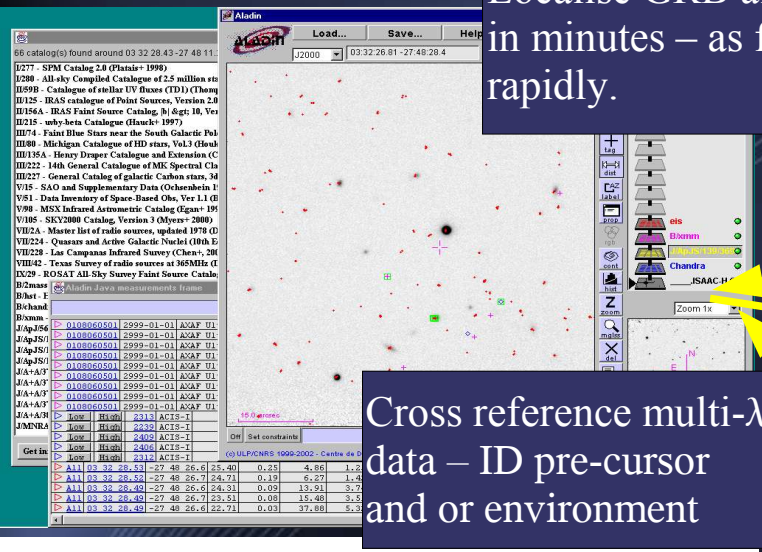


Image from ESO

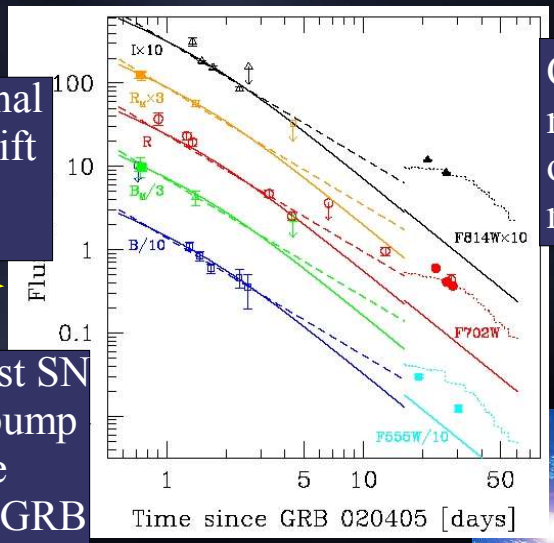
Interaction with observatory pipe-lines

Localise GRB alert in minutes – as fade rapidly.



Large computational photometric redshift calcs on multi-λ > gives distance

Cross reference multi-λ data – ID pre-cursor and or environment



Collate data from multiple telescopes over months - meta data issues

Compare against SN light curves – bump shows evidence for a SN in the GRB (Price et al, 2002)

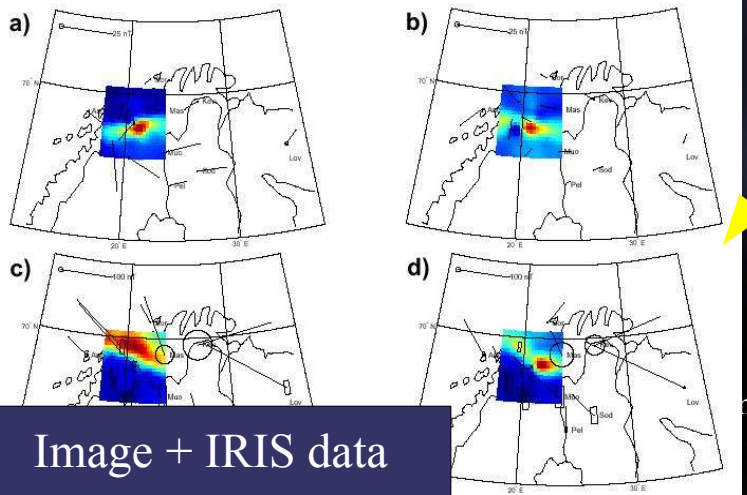


Image + IRIS data

Reprocessing of ionospheric STP data change coords from earth to celestial

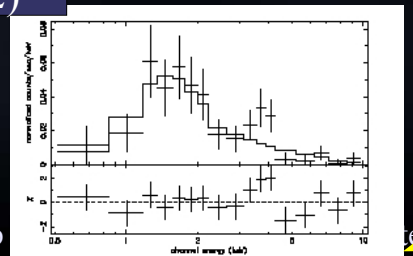


Fig. 2. EPIC-pn spectrum of the afterglow of GRB 010220, 20 ks exposure, fit with a power-law with Galactic absorption.

D. Ducros, ESA

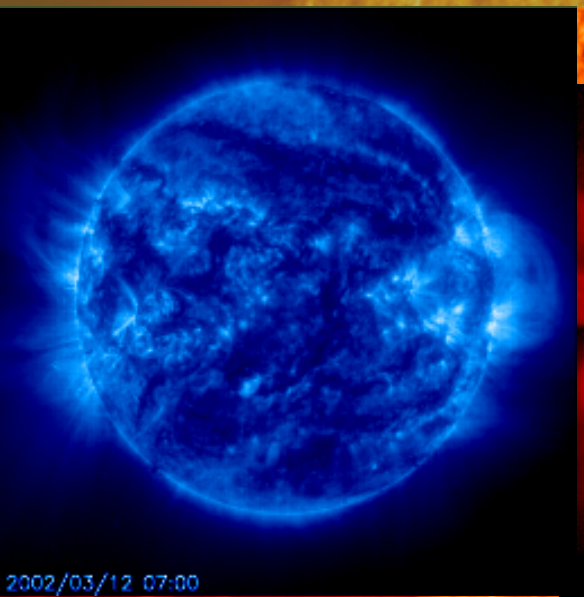
IoA UNIVERSITY OF CAMBRIDGE

New & Improved Science from VO's: Space Weather

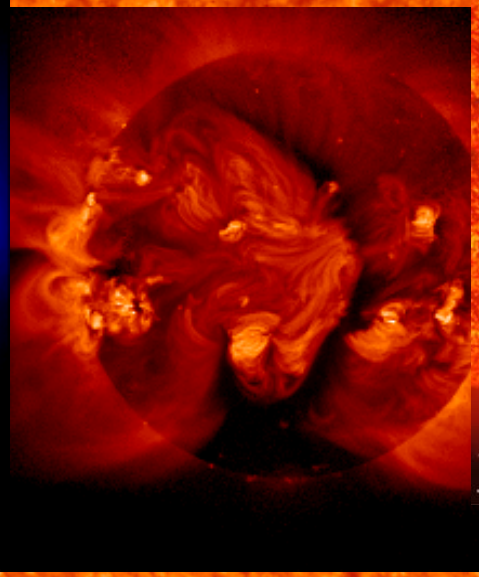
What happens to the Earth's magnetosphere during a coronal mass ejection ?

Event imaged by space based solar observatory

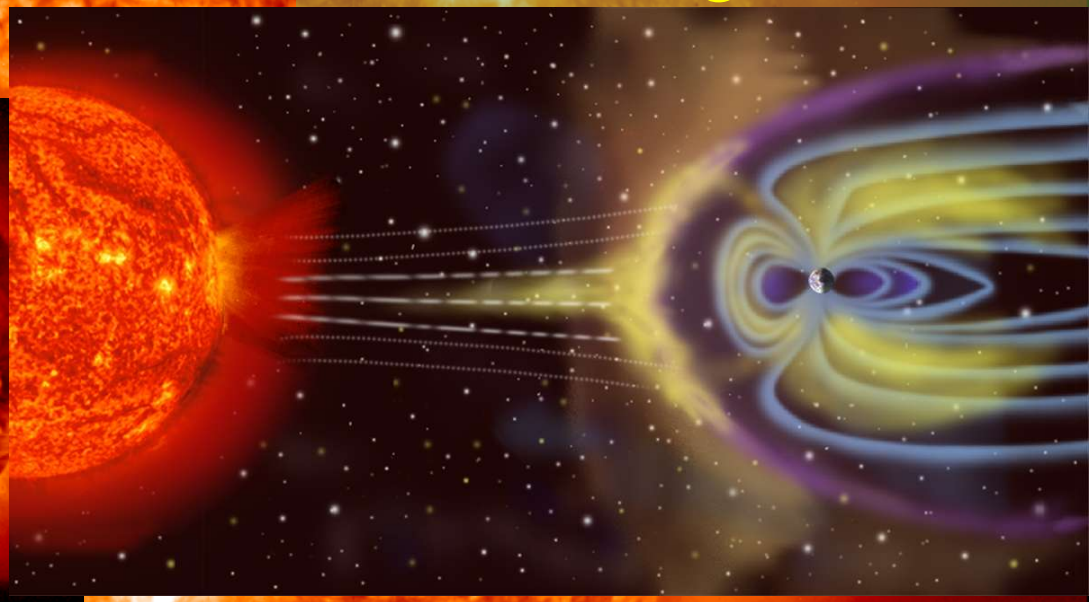
Effect detected later by satellites and ground radar



SOHO/EIT – EUV



Yohkoh – Xray



NASA: Living With a Star – http://lws.gsfc.nasa.gov

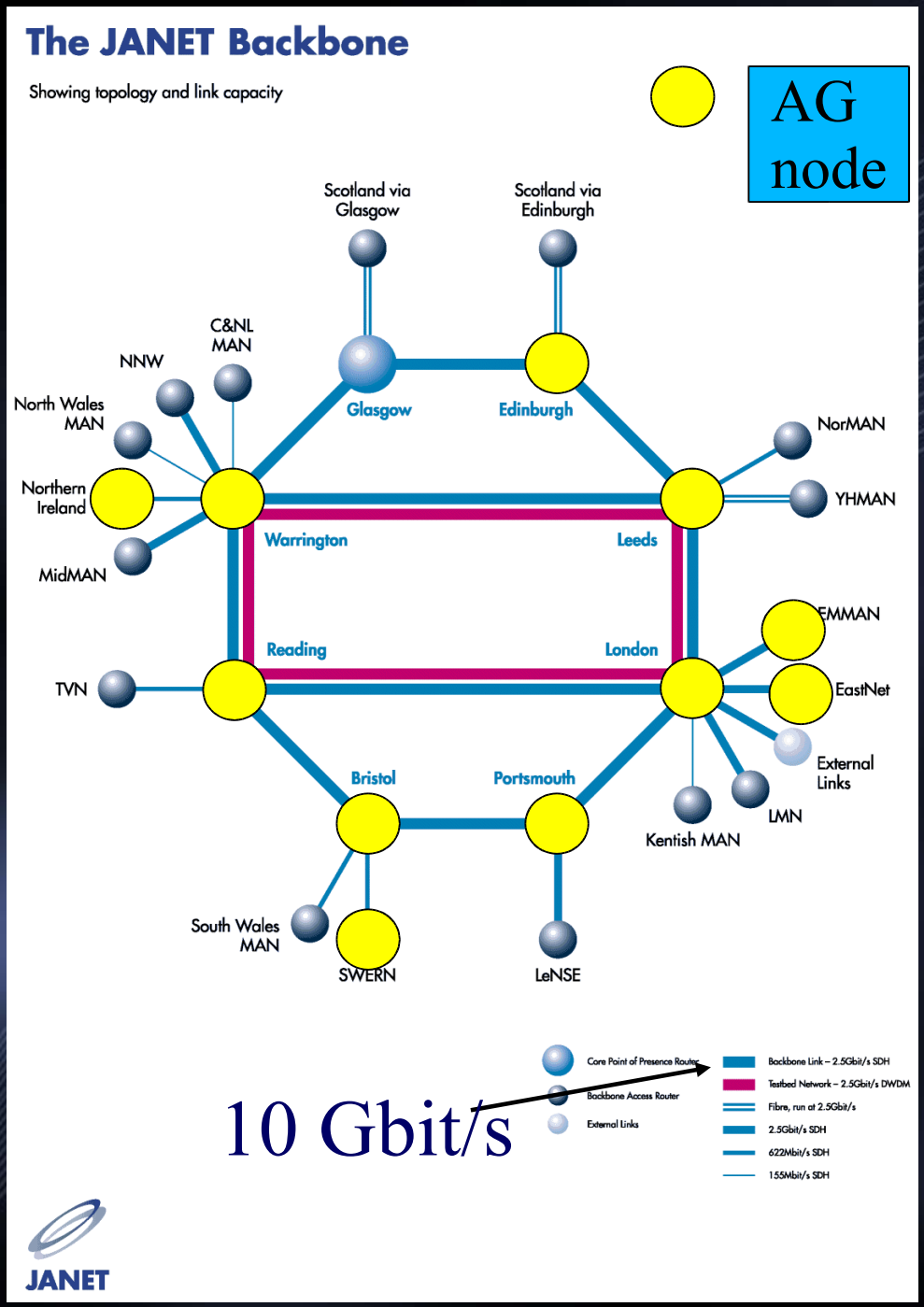
The Need for Virtual Observatories: Managing Technological Change

- 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

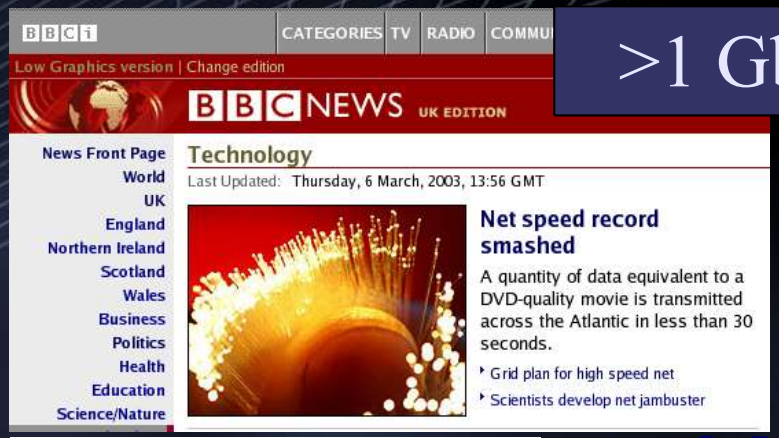
New Computational Opportunities

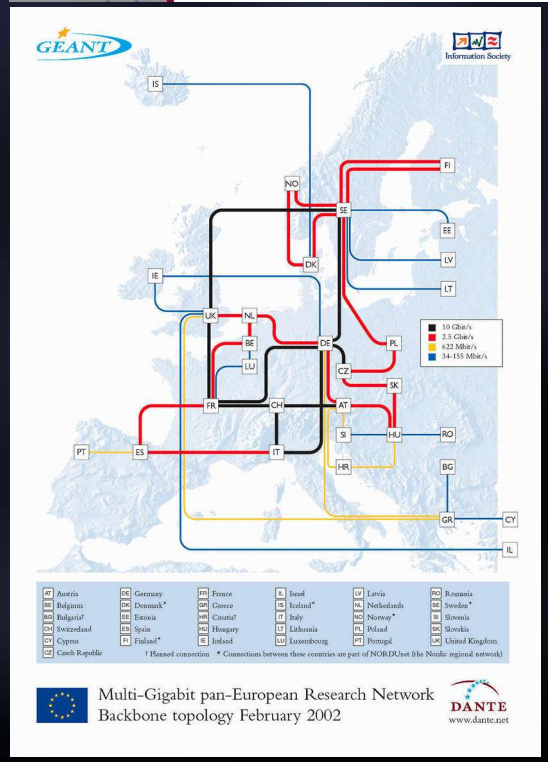
Creating a distributed system with Grids and Web Services

Fast Global Networks

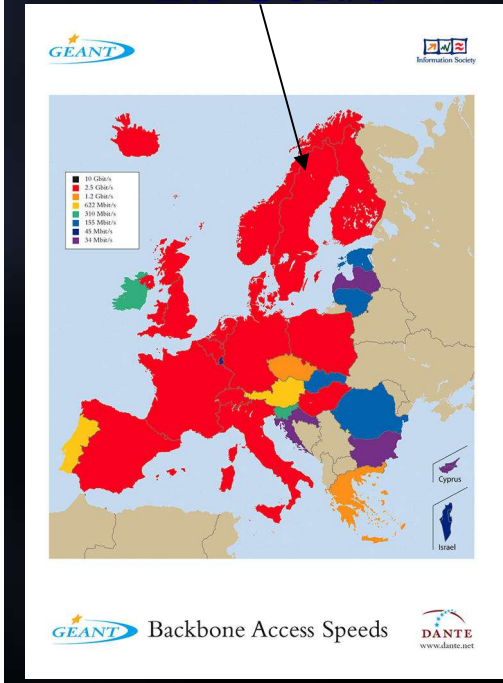


>1 Gbit/sec





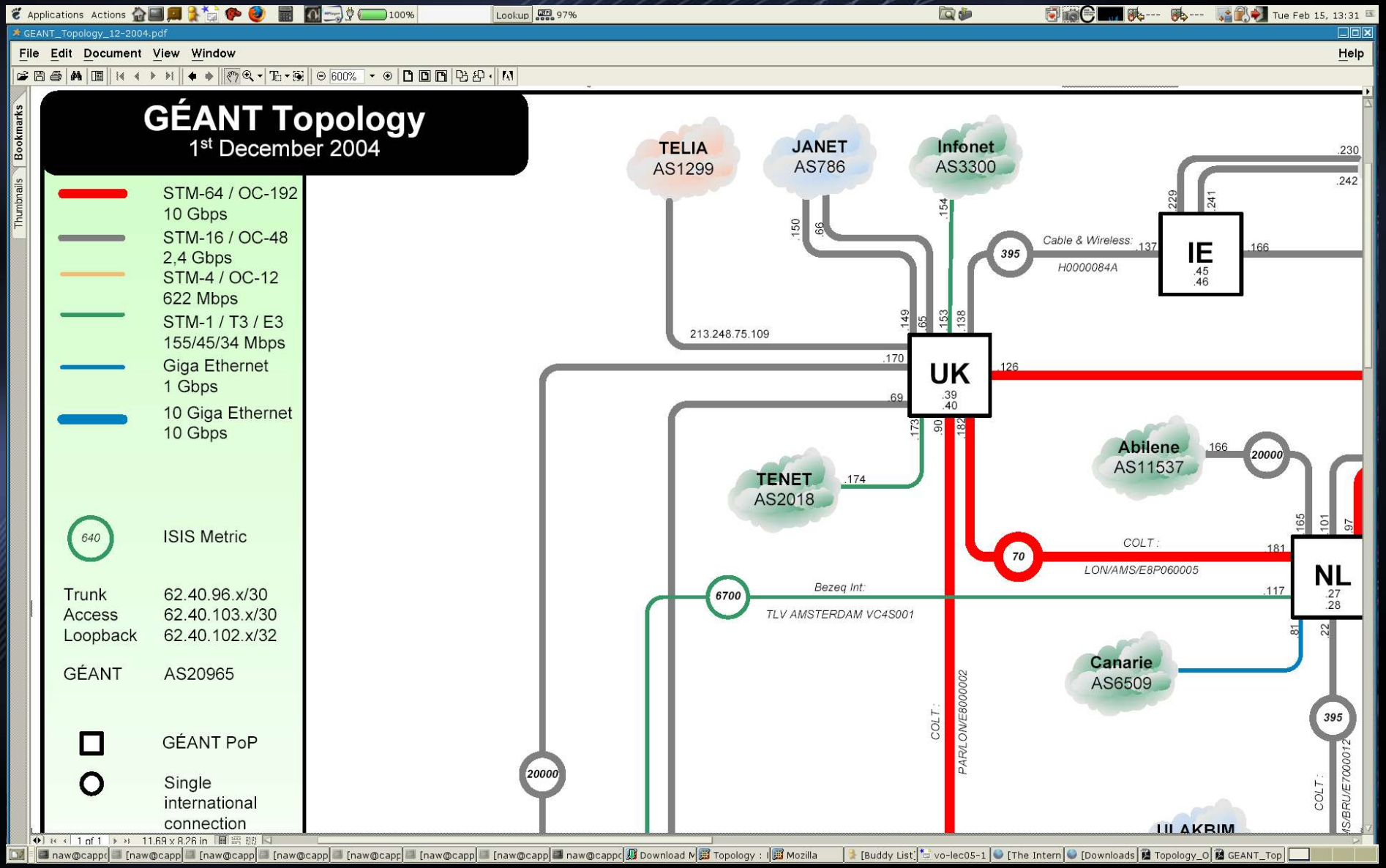
2.5Gbit/s



Geant2 / SuperJanet 5/ UKLight

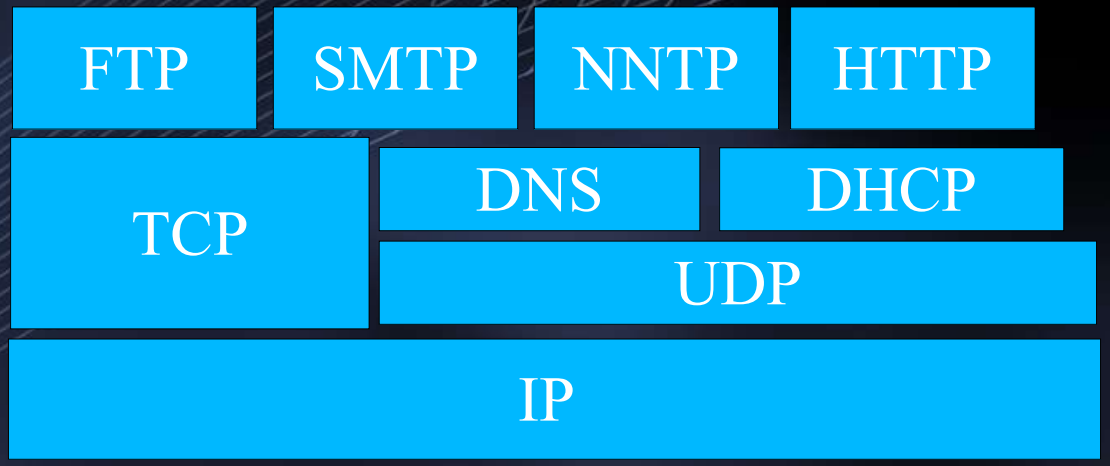
Networks are getting faster, with more cover

New technologies: Lambda-wave optical systems

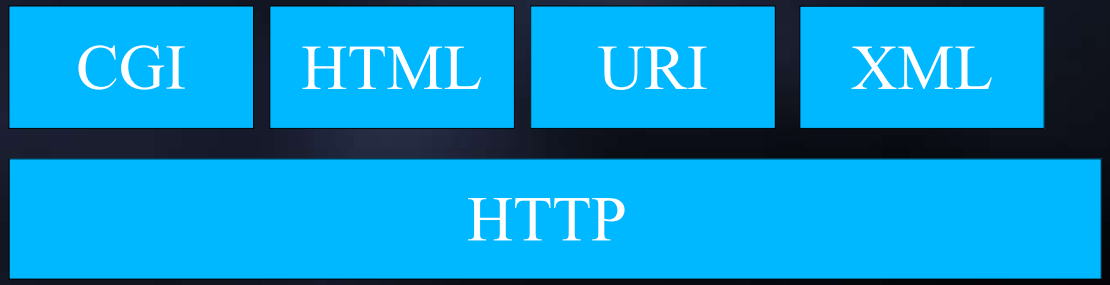


The Internet/ The WWW

Internet core protocols



WWW core protocols



Web services and Grid services add layers of protocols to these

The Grid

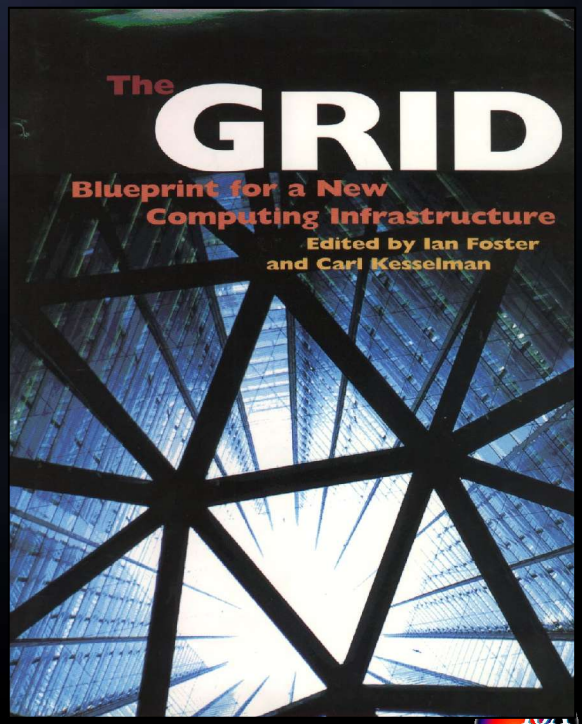
www.globus.org
www.ggf.org

Grid Guru's, Ian Foster / Carl Kesselman:

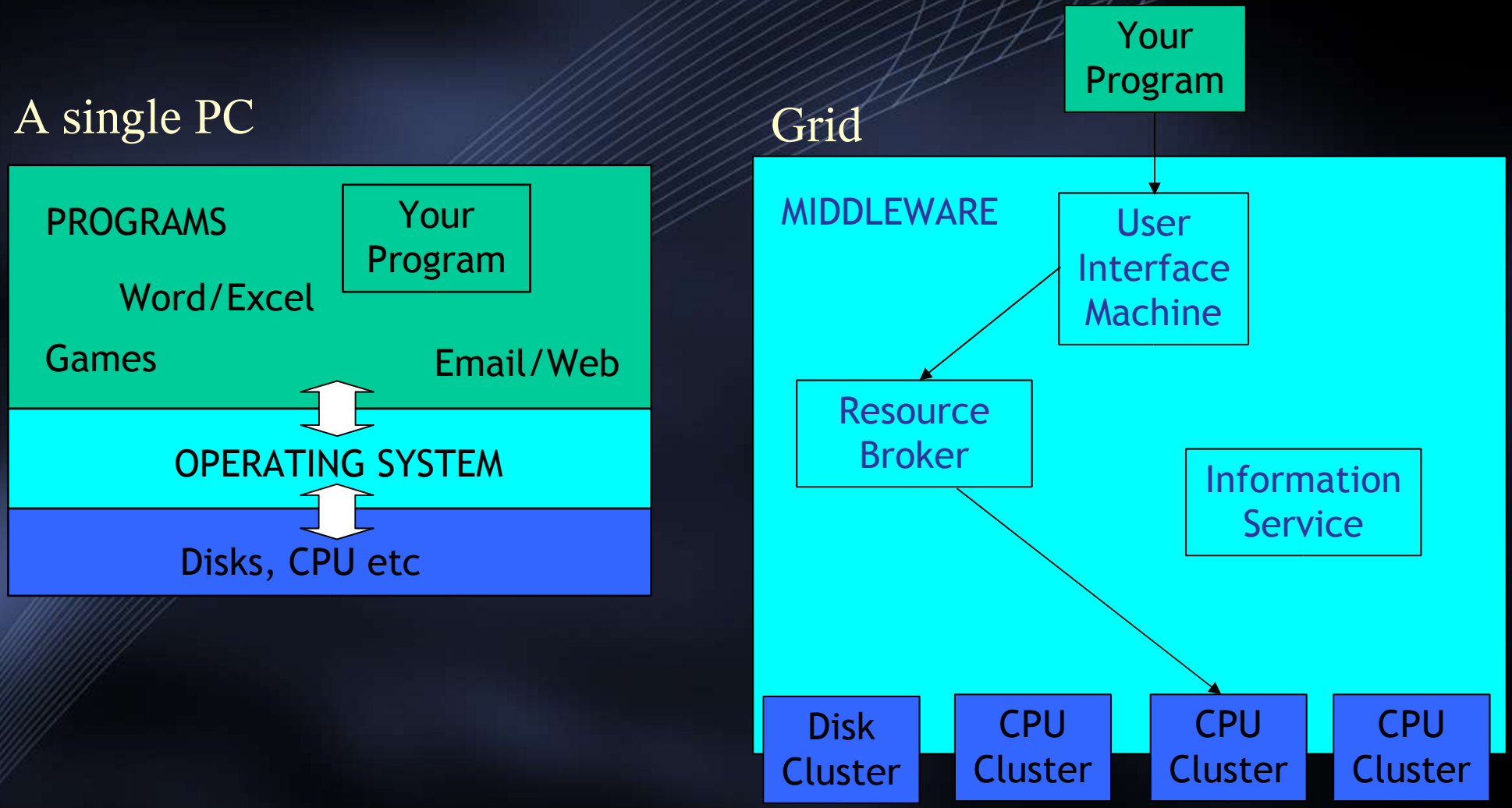
"A computational Grid is a hardware and software infrastructure that provides dependable, consistent, pervasive and inexpensive access to high-end computational capabilities."

- The Grid creates a virtual platform for computation and data manipulation
- c.f. The Internet, a virtual platform for information

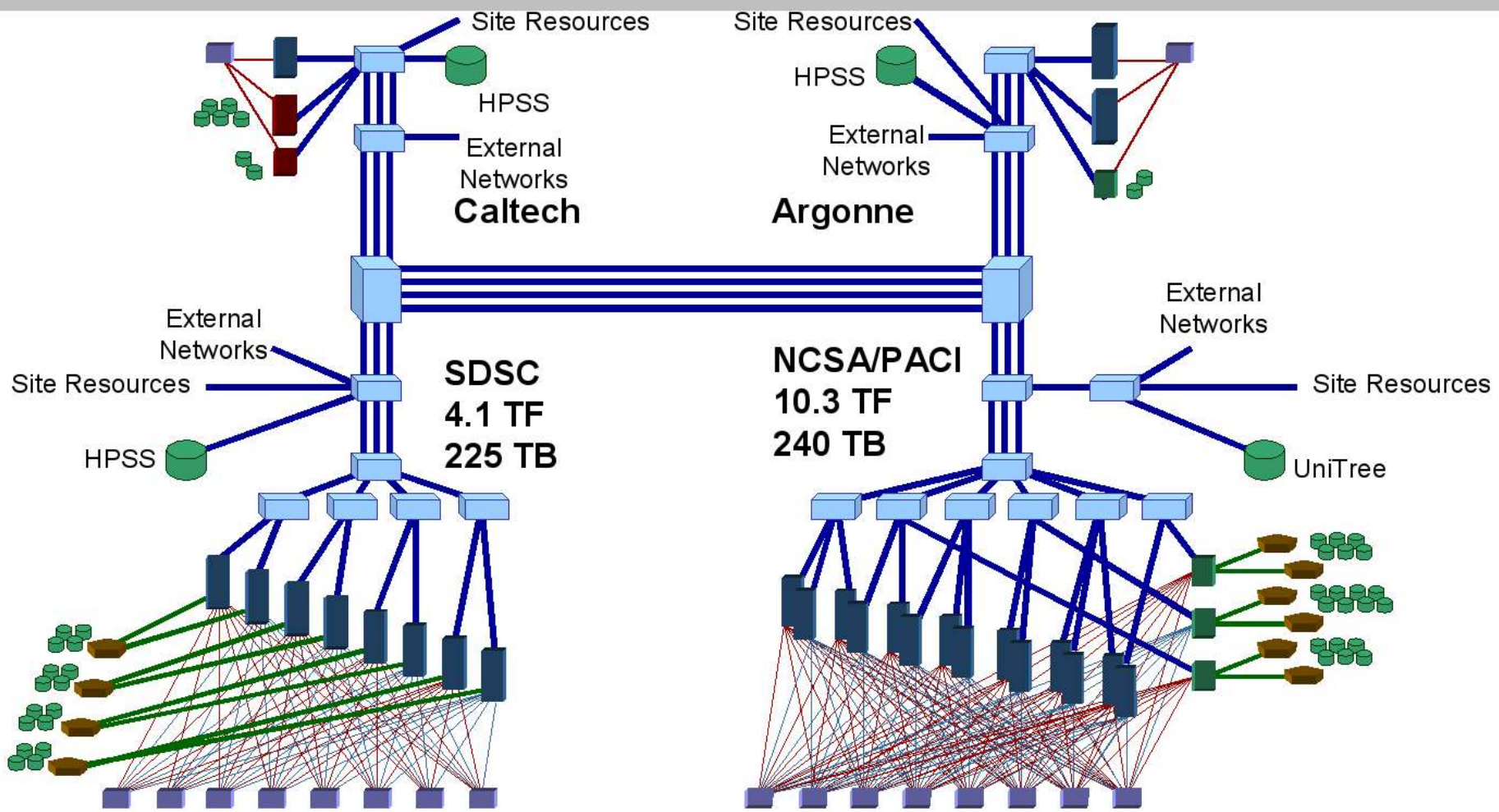
Early history dates to the 1980's
1990s: PVM, MPI, OpenMP: support parallel machines
1995: 1st Grid: I-Way at SC95
Late 1990s: Globus, Legion, Condor, SRB
Late 1990s: Grid Forum -> Global Grid Forum



What is the Grid?:



Overview of Distributed TeraGrid Resources



NVO Summer School Sept 2004

TeraGrid: the USA super grid

Web Services

- Service Orientated Architectures
  - An application architecture within which all functions are defined as independent services with well-defined invocable interfaces which can be called in defined sequences to form scientific processes.
- Core
  - SOAP: simple object access protocol
  - WSDL: web services description language
- Implementations: Apache/AXIS, MS/ .NET (Open Source Mono), Python, Perl, PHP
- WS-I*, WSRF standards: state, security, addressing, management, asynchronous transactions

SOAP/WSDL

SOAP: a lightweight protocol for exchange of information in a decentralized, distributed environment. It is an XML based protocol that consists of three parts: an envelope that defines a framework for describing what is in a message and how to process it, a set of encoding rules for expressing instances of application-defined datatypes, and a convention for representing remote procedure calls and responses.

WSDL is an XML format for describing network services as a set of endpoints operating on messages containing either document-oriented or procedure-oriented information. The operations and messages are described abstractly, and then bound to a concrete network protocol and message format to define an endpoint. Related concrete endpoints are combined into abstract endpoints (services). WSDL is extensible to allow description of endpoints and their messages regardless of what message formats or network protocols are used to communicate, however, the only bindings described in this document describe how to use WSDL in conjunction with SOAP 1.1, HTTP GET/POST, and MIME.

Source: W3C

Grid Services

- Web services handle:
  - Service Description
  - Service Implementation
  - Service Publishing, Discovery and Binding
  - Service Invocation and Execution
  - Stateless services
- Grid services add another layer
  - Secure file transport
  - Programme execution
  - Monitoring services
  - Adds state: allows asynchronous operations
- Open Grid Services Architecture:

Grid/Web Services: Links

- Java: http://java.sun.com/webservices/index.jsp
  - Axis (SOAP implementation): http://ws.apache.org/axis/
  - Globus (GT4, Apr 05): http://www-unix.globus.org/toolkit/
- C#: http://msdn.microsoft.com/vcsharp/
  - Mono: http://www.mono-project.com/about/index.html
  - .Net: http://www.microsoft.com/net/
  - WSRF .NET: http://www.cs.virginia.edu/~gsw2c/wsrf.net.html
- Perl
  - SOAP::Lite http://www.soaplite.com/
  - WSRF::Lite http://www.sve.man.ac.uk/Research/AtoZ/ILCT
- Standards
  - WSRF http://www-106.ibm.com/developerworks/library/ws-resource/
  - SOAP http://www.w3.org/TR/soap/

Grid Services: Links

- Globus and OGSA: http://www.globus.org/ogsa
- Globus and WSRF: http://www.globus.org/wsrf
- OMII: http://www.omii.ac.uk
- EGEE: http://public.eu-egee.org/
  - gLite http://glite.web.cern.ch/glite/

Astro-RG @ Global Grid Forum

interfacing with the grid community

http://forge.gridforum.org/projects/astro-rg

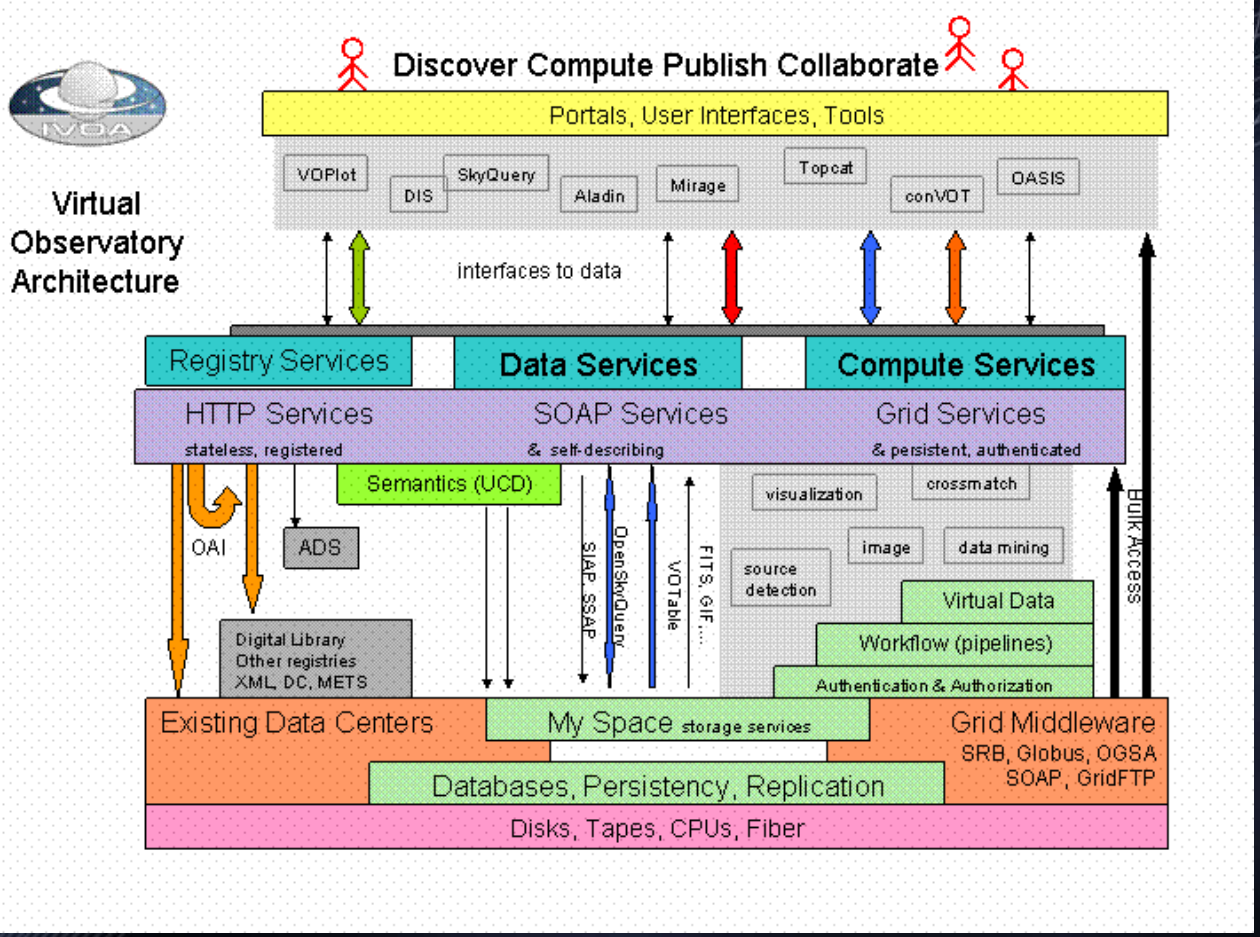
VO Technologies

High level interoperability protocols coupled with Grid Technologies

IVOA: Enabling Interoperability

- The International Virtual Observatory Alliance http://www.ivoa.net
- A global partneship
- Projects represent major global astronomy data providers, e.g. ESO, STScI, NAOJ
- Multi-wavelength
- Multi-site

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

IVOA Working Groups: 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
  - 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'

IVOA: Interest Groups

Aim: issues of importance across working groups

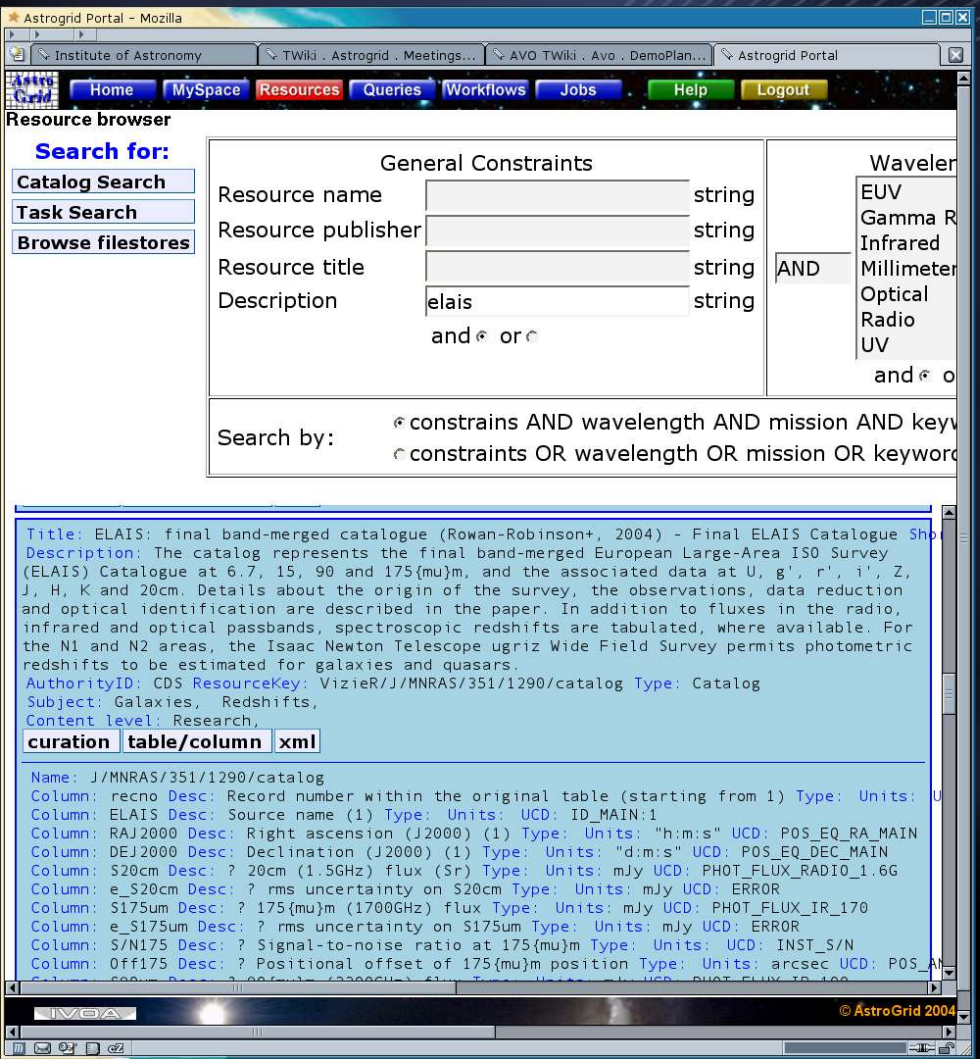
- Applications
  - Interfacing new and legacy apps (e.g. Iraf) to the VO
- RadioVO
  - Issues relevant to 'Radio': e.g. The UV plane, interferometry
- Theory
  - Simulations, mass scale compute
- Data Curation & Preservation
  - Linkages to the Digital Libraries world, a-ph, ADS
- Architecture
  - Fitting it all together
- Networks
- Semantics

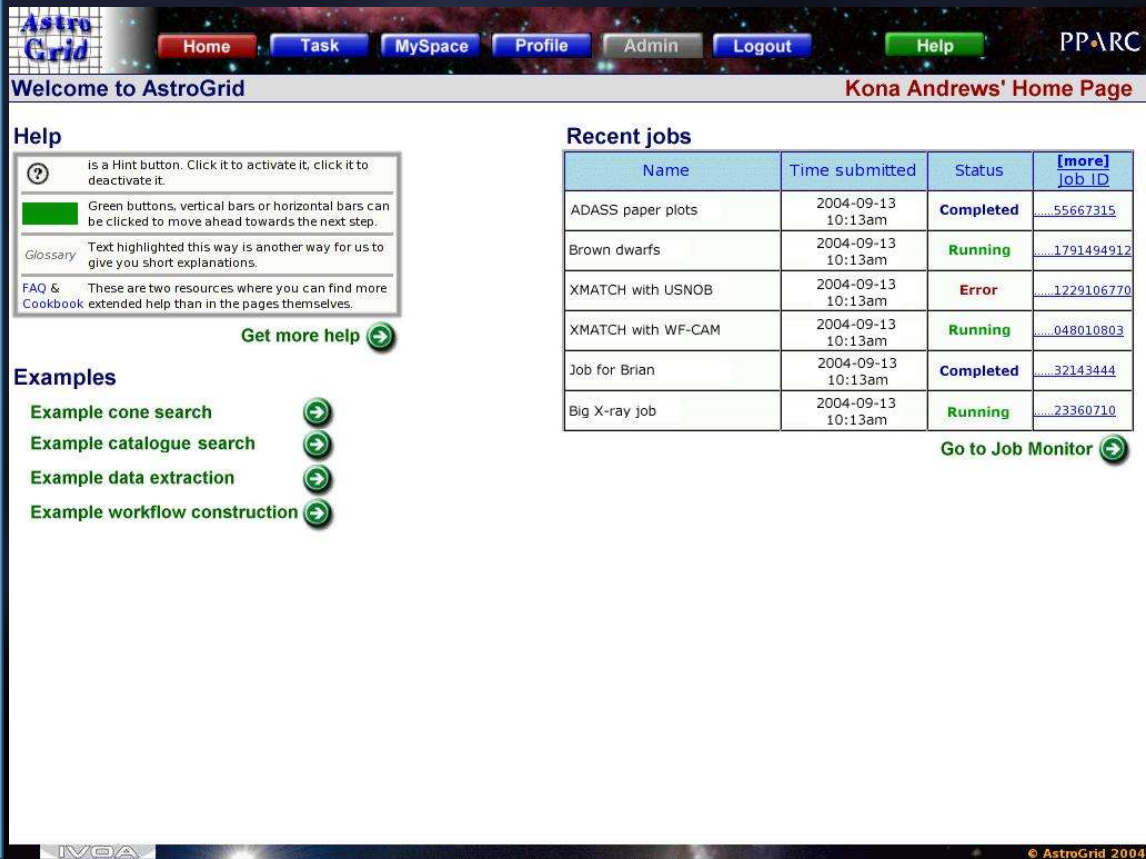
Virtual Observatories

A brief tour: specific examples of use in later lectures

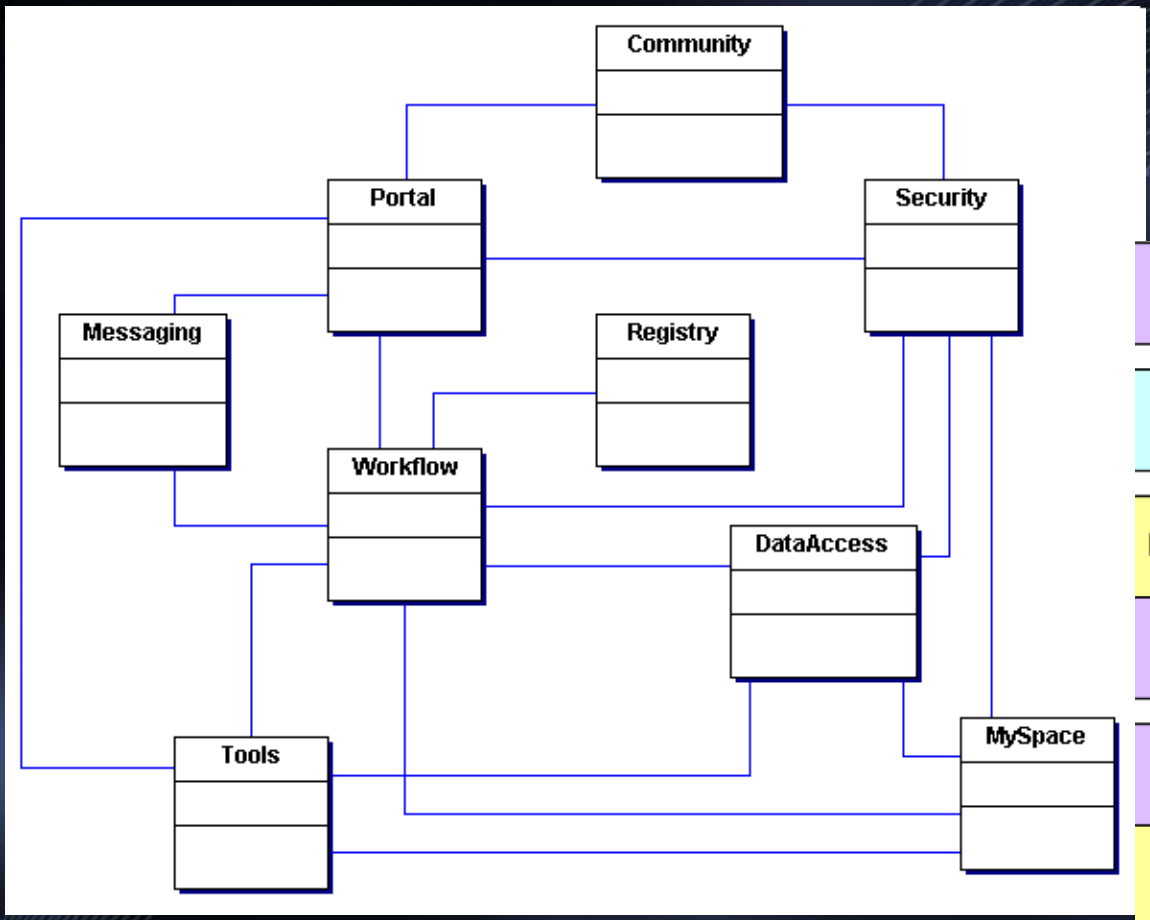
UK-VO: AstroGrid http://www.astrogrid.org

- Web Services based system
- Linkages to main UK data providers
- VO interoperability with external VO's/ Data Providers



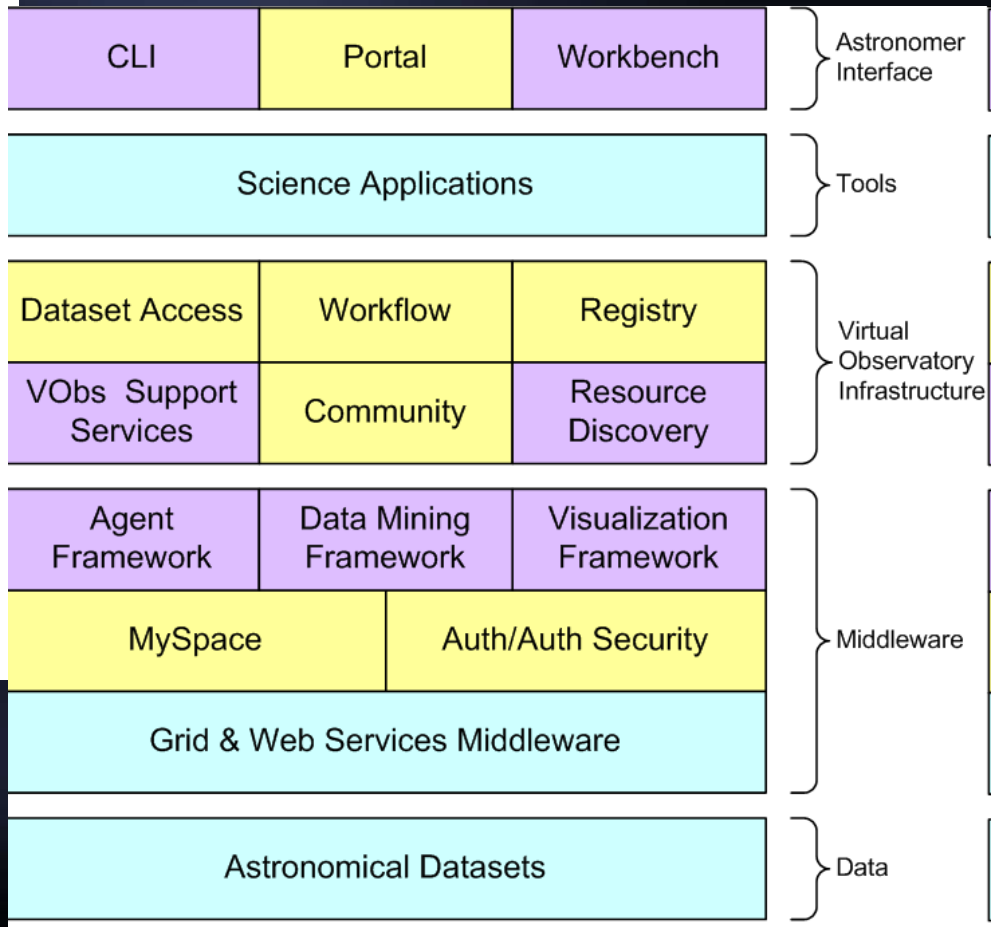


VO Architecture: AstroGrid



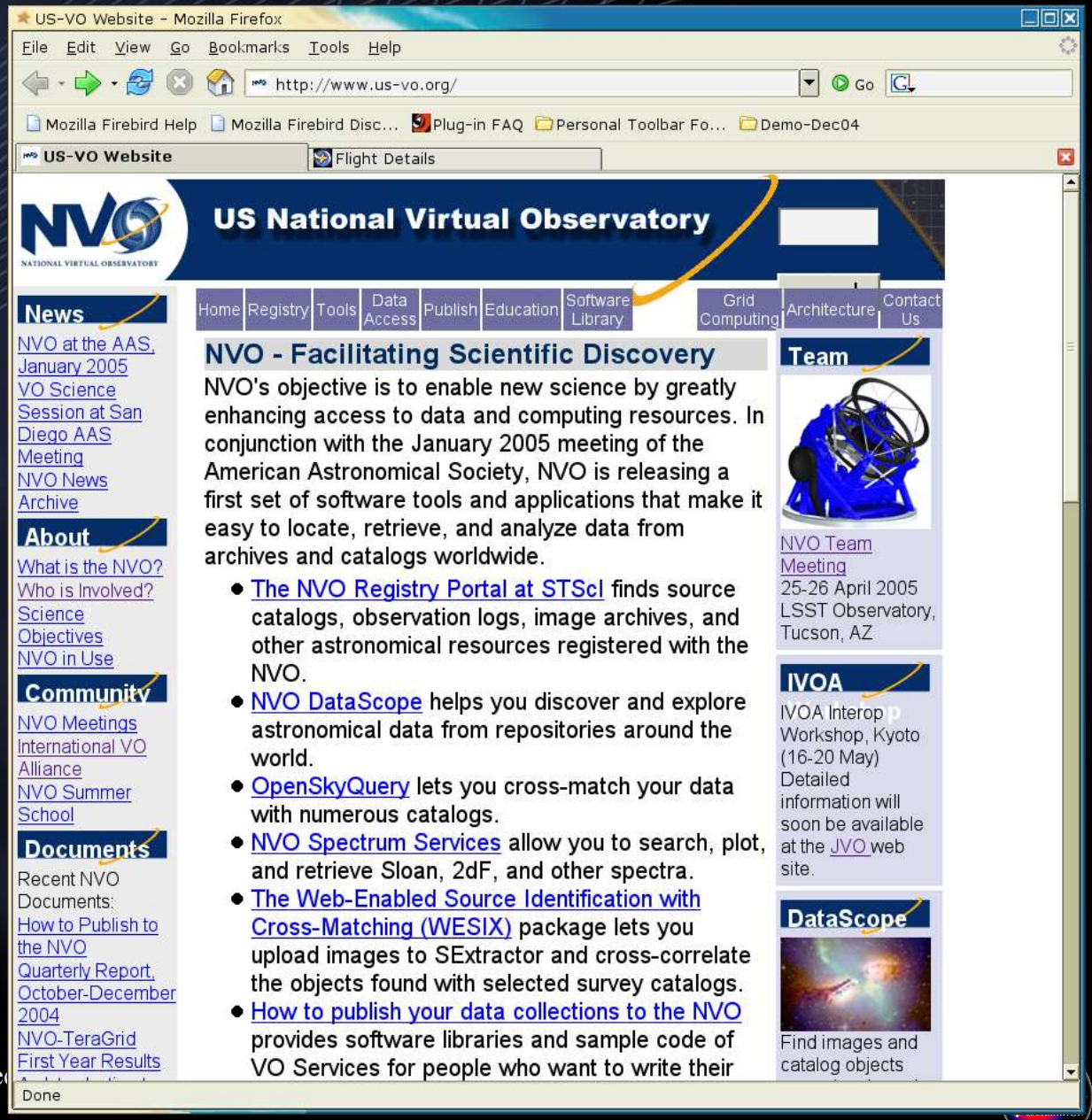
Component based system

Layer diagramme

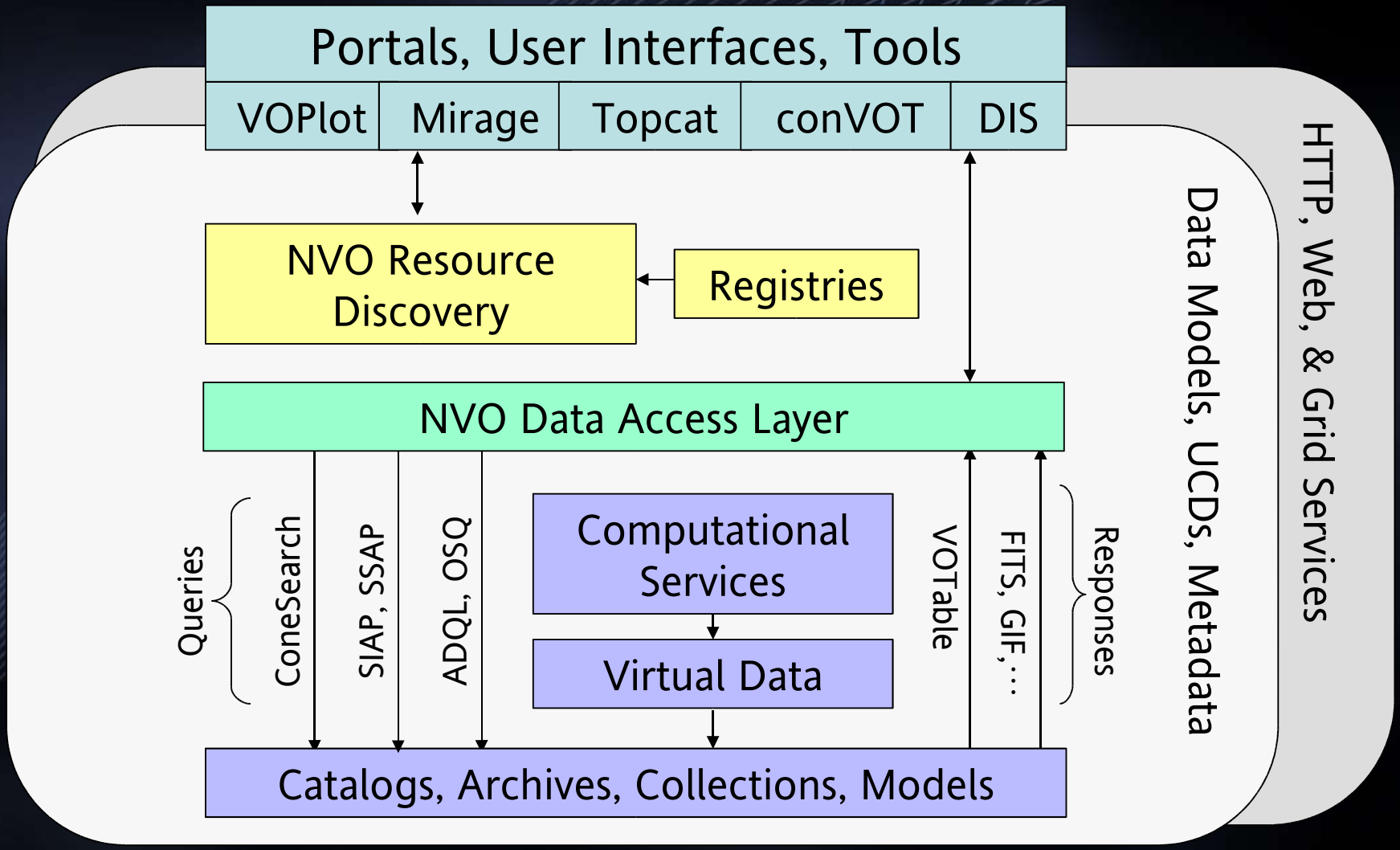


US-VO: NVO http://www.us-vo.org

- 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

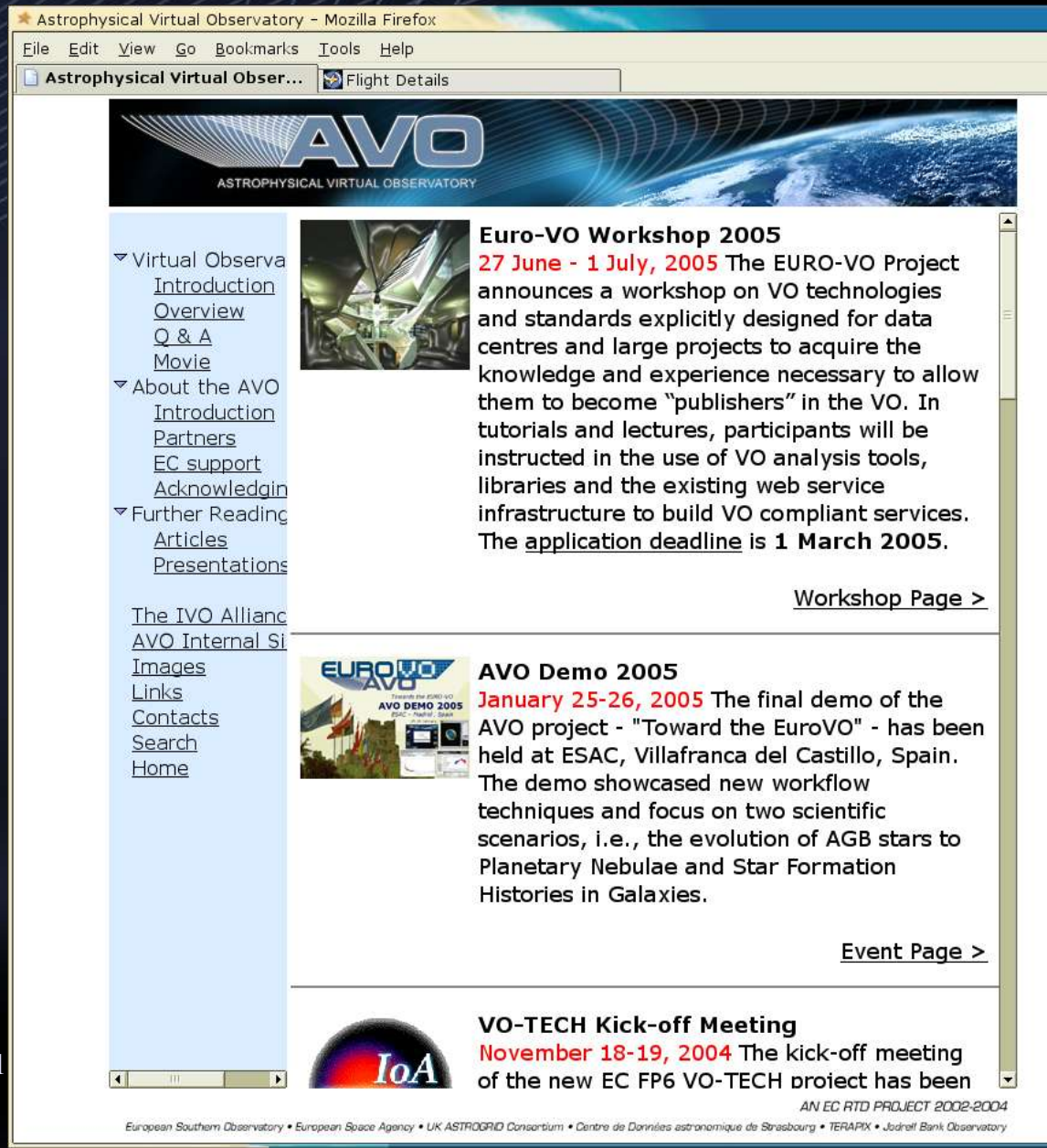


US-VO: Architecture



Euro-VO: AVO http://www.euro-vo.org

- AVO: study programme
- Partners
  - ESO, AstroGrid, ESA/ECF, JBO, CDS, Terapix
- Main outputs
  - Technology Assessments
  - Demo products (to be discussed in a later lecture)
- AVO leads into a larger Euro-VO project
  - Links data centres and missions with technology development & operations

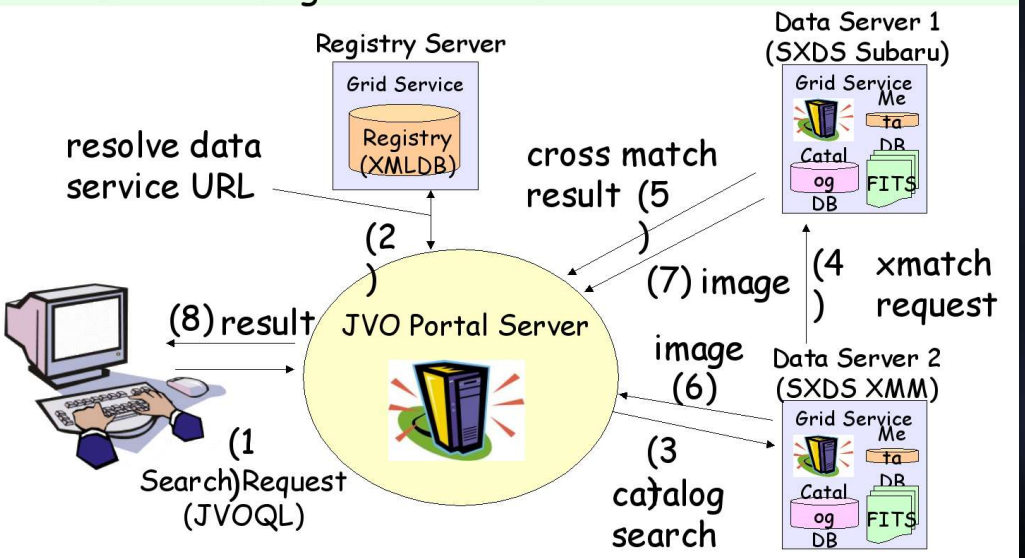


Japan-VO: JVO http://jvo.nao.ac.jp

- Partners
  - NAOJ, JAXA, ICRR, Ochanomizu U, Osaka U, Titech, Fujitsu
  - Data from Suburu, Astro-F, Nobeyama, Alma
- 1st JVO prototype based on Globus Toolkit 2 – 'grid'
- 2nd JVO prototype based on GT3 – grid services, improved performance

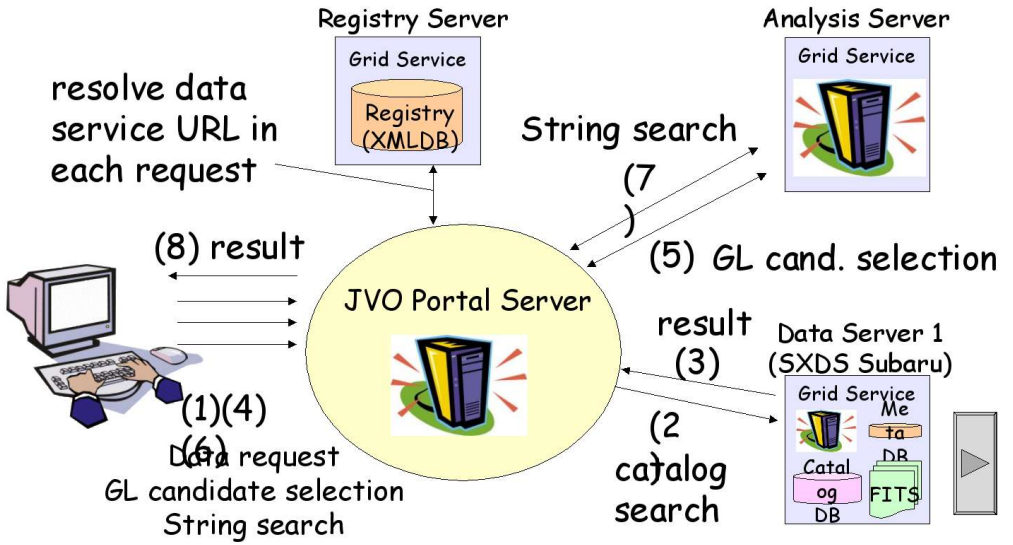
Demo 1: Cross match & Image request

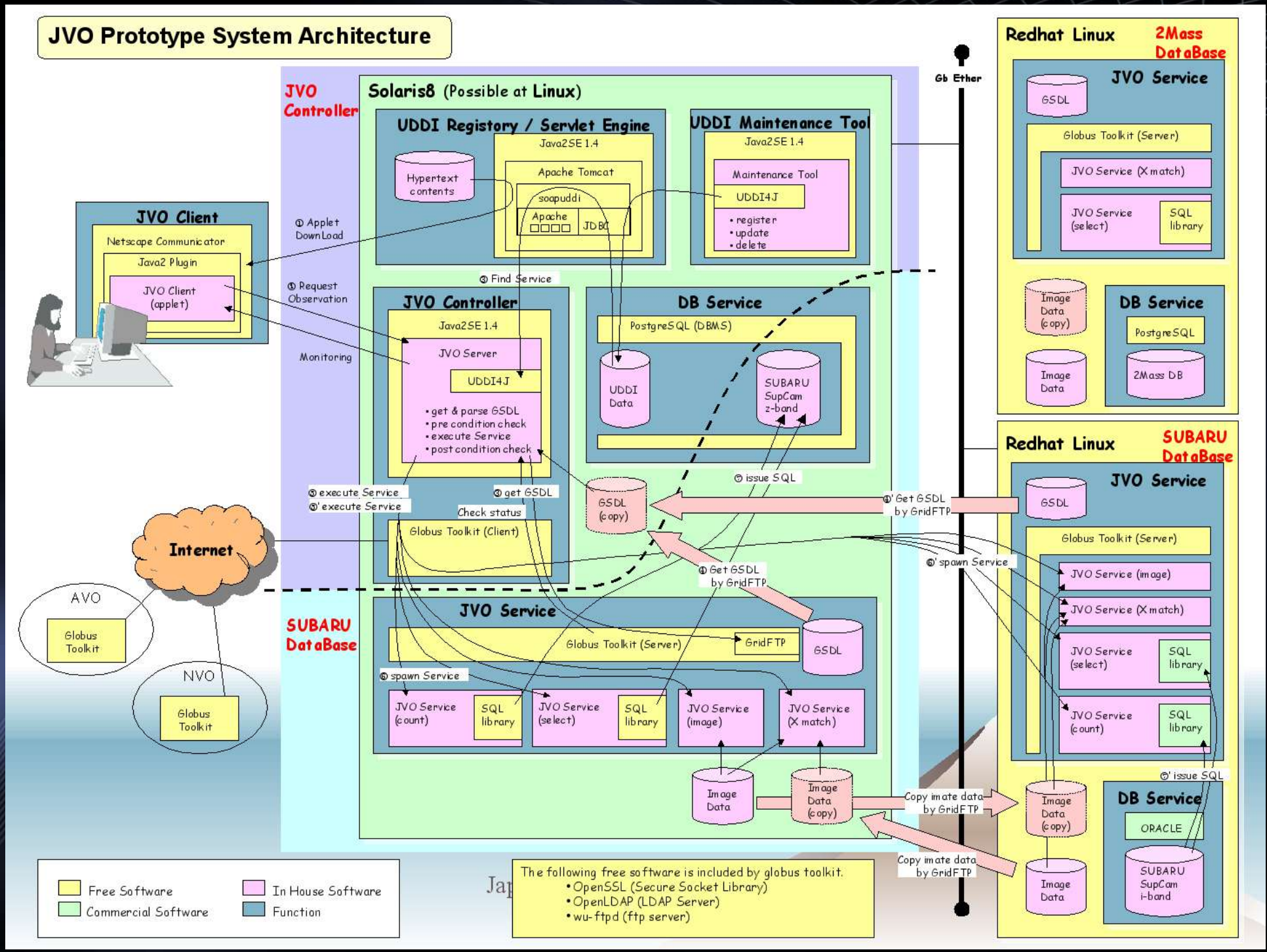
Cross match of the optical and X-ray catalogs of SXDS and image retrievals.



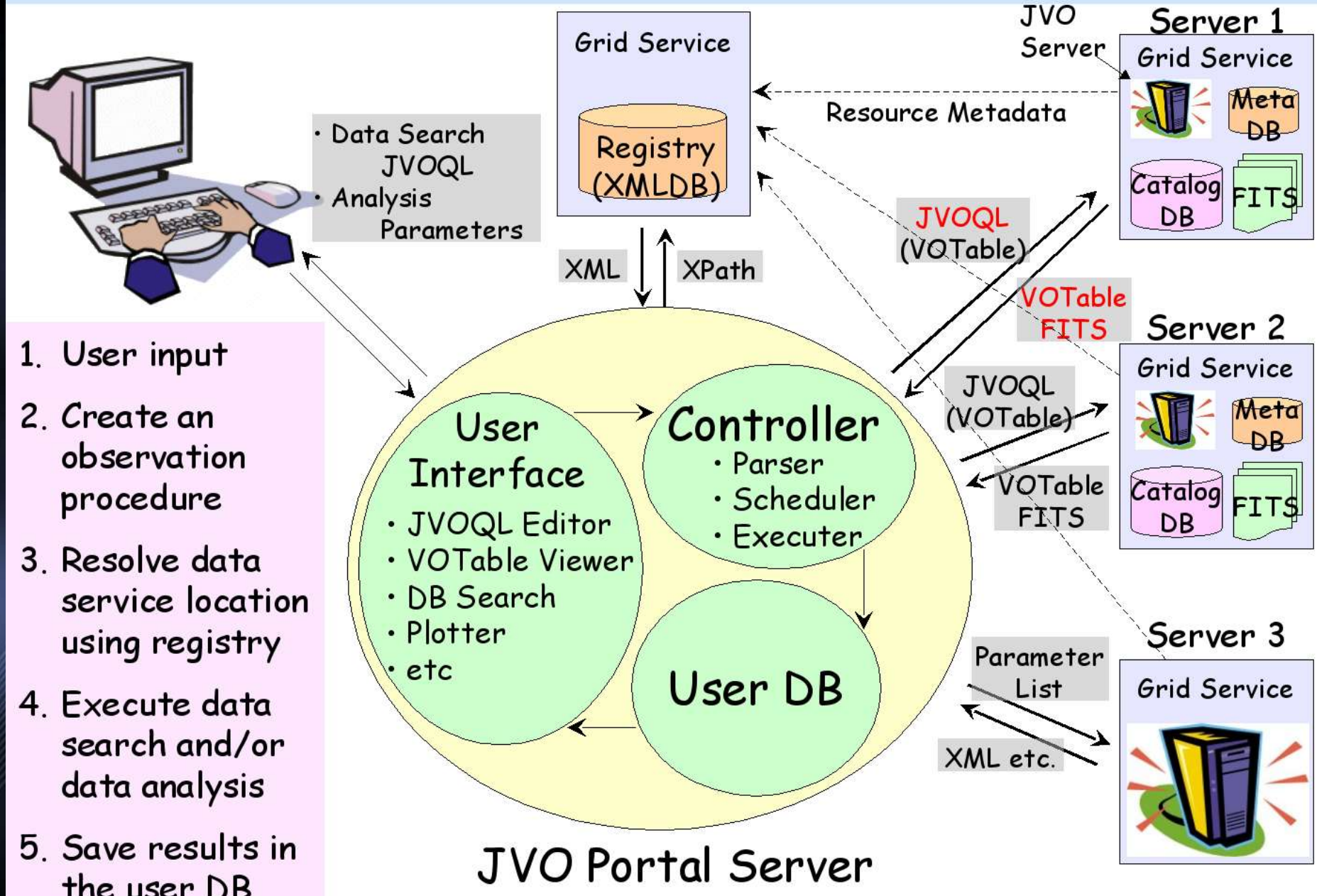
Demo 2: Cosmic String Search

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




JVO Prototype System Architecture
JVO Controller
Solaris8 (Possible at Linux)
UDDI Registory / Servlet Engine
Java2SE 1.4
Hypertext contents
Apache Tomcat
soapuddi
Apache
JDBC
UDDI Maintenance Tool
Java2SE 1.4
Maintenance Tool
UDDI4J
• register
• update
• delete
Gb Ether
Redhat Linux
2Mass DataBase
JVO Service
GSDL
Globus Toolkit (Server)
JVO Service (X match)
JVO Service (select)
SQL library
Image Data (copy)
DB Service
PostgreSQL
Image Data
2Mass DB
JVO Client
Netscape Communicator
Java2 Plugin
JVO Client (applet)
① Applet DownLoad
① Request Observation
Monitoring
② Find Service
JVO Controller
Java2SE 1.4
JVO Server
UDDI4J
• get & parse GSDL
• pre condition check
• execute Service
• post condition check
DB Service
PostgreSQL (DBMS)
UDDI Data
SUBARU SupCam z-band
⑦ issue SQL
⑤ execute Service
⑤' execute Service
③ get GSDL
Check status
Globus Toolkit (Client)
GSDL (copy)
④' Get GSDL by GridFTP
Redhat Linux
SUBARU DataBase
JVO Service
GSDL
Globus Toolkit (Server)
⑥' spawn Service
JVO Service (image)
JVO Service (X match)
JVO Service (select)
SQL library
JVO Service (count)
SQL library
⑦' issue SQL
Internet
AVO
Globus Toolkit
NVO
Globus Toolkit
④ Get GSDL by GridFTP
SUBARU DataBase
JVO Service
Globus Toolkit (Server)
GridFTP
GSDL
⑥ spawn Service
JVO Service (count)
SQL library
JVO Service (select)
SQL library
JVO Service (image)
JVO Service (X match)
Image Data
Image Data (copy)
Copy image data by GridFTP
Image Data (copy)
DB Service
ORACLE
Image Data
SUBARU SupCam i-band
Copy image data by GridFTP
Free Software
Commercial Software
In House Software
Function
The following free software is included by globus toolkit.
• OpenSSL (Secure Socket Library)
• OpenLDAP (LDAP Server)
• wu-ftpd (ftp server)

Architecture of JVO Proto 2



1. User input
2. Create an observation procedure
3. Resolve data service location using registry
4. Execute data search and/or data analysis
5. Save results in the user DB

Other VO projects

- Aus-VO (Australia): http://www.aus-vo.org
- CVO: http://www2.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/cvo/
- China-VO: http://www.china-vo.org
- Draco (Italy): http://wwwas.oat.ts.astro.it/draco/
- GAVO (Germany): http://www.g-vo.org/
- HVO (Hungary): http://hvo.elte.hu/en/
- KVO (Korea): http://kvo.kao.re.kr/
- RVO (Russia): http://www.inasan.rssi.ru/eng/rvo/
- SVO (Spain): http://laeff.esa.es/svo/
- VO-France: http://www.france-vo.org/
- VO-India: http://vo.iucaa.ernet.in/~voi/

Lecture 1: Acknowledgements

- 'Grid' diagramme on Slide 19 Adapted from S. Lloyd, see http://www.gridpp.ac.uk/talks/Inaugural_Lecture.ppt
- Some material on Slide 20 adapted from M. Graham, see NVO Summer School: http://www.us-vo.org/summer-school/proceedings/presentations/WSTechnology.ppt
- Teragrid graphic on slide 21 from Roy Williams, see http://www.us-vo.org/summer-school/proceedings/presentations/Grid.ppt
- The graphic on slide 36 due to Bob Hanisch, see http://www.us-vo.org/summer-school/proceedings/presentations/Introduction.ppt
- The graphics on Slides 5 and 39 are due to Masatoshi Ohishi of the JVO project
- The graphic on Slides 38 & 40 due to Yuji Shirasaki of the JVO project

SOAP Example

Request:

```
<soap:Envelope xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance xmlns:xsd=http
    ://www.w3.org/2001/XMLSchema
    xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/">
    <soap:Body>
        <ComovingLineOfSight xmlns="http://skyservice.pha.jhu.edu">
            <z>float</z>
            <hubble>float</hubble>
            <omega>float</omega>
            <lambda>float</lambda>
        </ComovingLineOfSight>
    </soap:Body>
</soap:Envelope>
```

Response:

```
<soap:Envelope xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema"
    xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/">
    <soap:Body>
        <ComovingLineOfSightResponse xmlns="http://skyservice.pha.jhu.edu">
            <ComovingLineOfSightResult>float</ComovingLineOfSightResult>

        </ComovingLineOfSightResponse>
    </soap:Body>
</soap:Envelope>
```

WSDL Example

```
<definitions>
        <import>*
        <types>
                <schema></schema>*
        </types>
        <message>*
                <part></part>*
        </message>
        <portType>*
                <operation>*
                        <input></input>
                        <output></output>
                        <fault></fault>*
                </operation>
        </portType>
        <binding>*
                <operation>*
                        <input></input>
                        <output></output>
                </operation>
        </binding>
        <service>*
                <port></port>*
        </service>
</definitions>
```

- include other WSDLS
- define datatypes used in <message> elements

- model data exchanged

- a subset of operations supported for an endpoint

    - define input and output messages

- concrete protocol and data format specification for a <portType> element

- identifies actual endpoint for WS