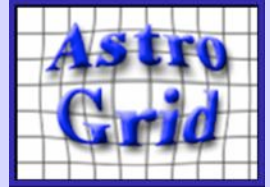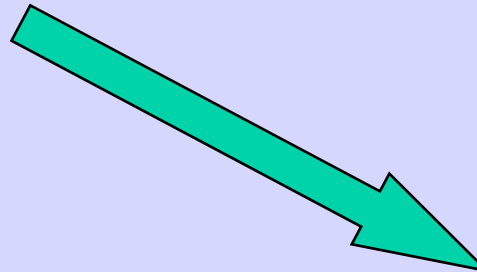# The Virtual Observatory in Action

- VO drivers
- VO vision
- VO progress
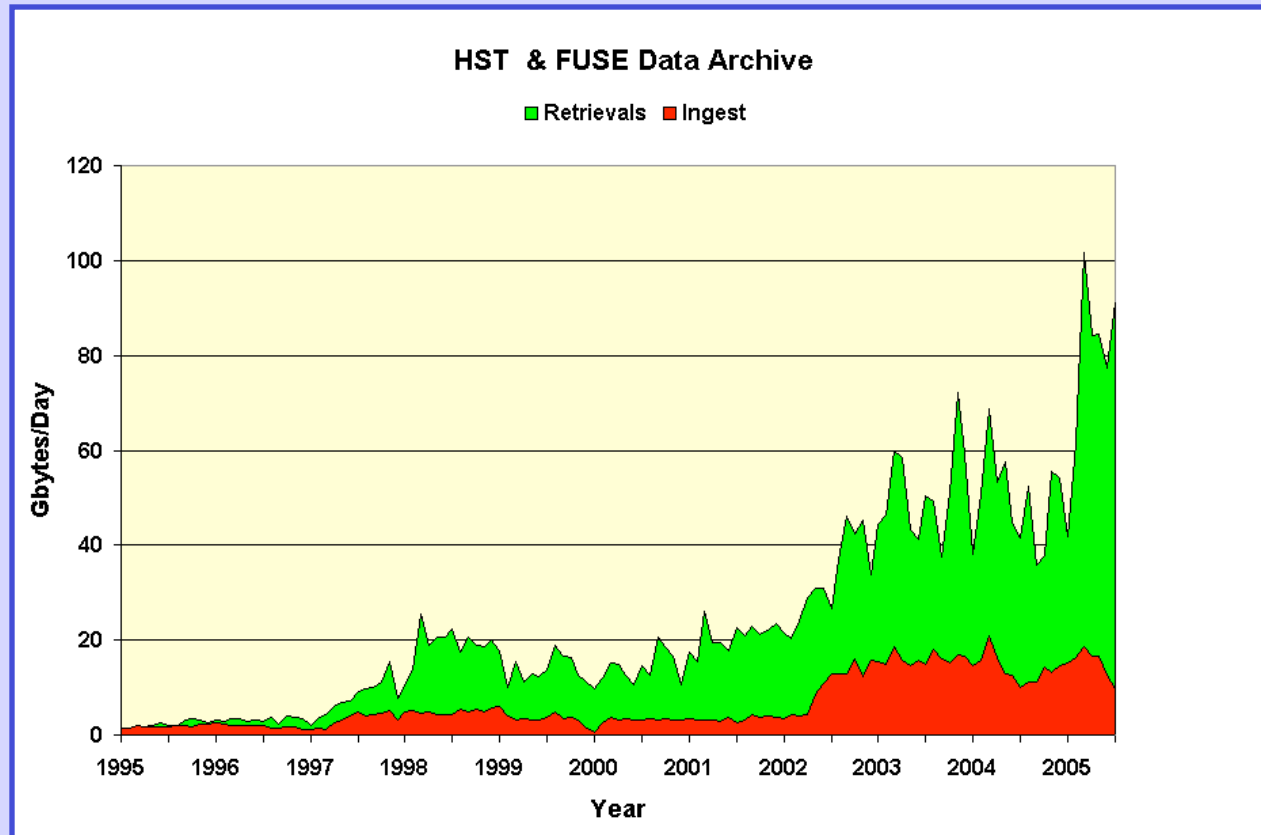
World
AstroGrid
VO Desktop demo

VO drivers : science

science services

- several trends lead to science being done by services from professional data / resource centres

archive re-use

HST & FUSE Data Archive

the archive is the sky

on-line research

- users want on-line availability

- trad
  - download files, analyse at home
  - reduction standardised, analysis home grown

- new
  - analyse in-situ
  - analysis standardised
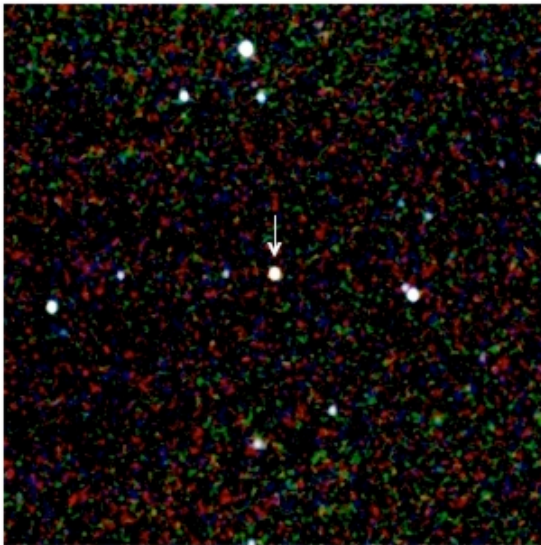
multi-archive research

- most science goals
  require use of multiple data sources ...
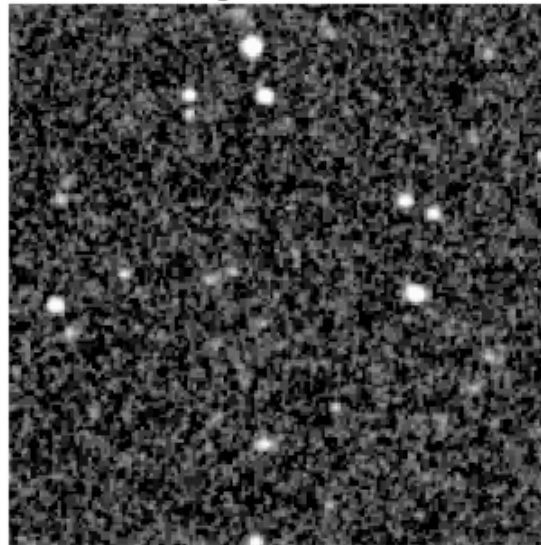
brown dwarfs

2MASS J1146+2230

An L−type dwarf in the constellation Leo

The near−infrared view

The optical view

2MASS Atlas JHK$_s$ Composite Image

Palomar Digitized Sky Survey

2MASS
2 MICRON ALL-SKY SURVEY

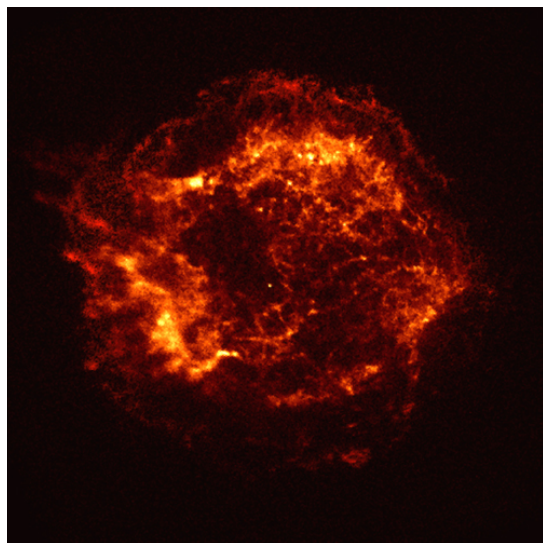J.D. Kirkpatrick (IPAC/Caltech), I.N. Reid (Caltech), R.M. Cutri (IPAC/Caltech),
C.A. Beichman (IPAC/JPL/Caltech), J. Liebert (U of A), M.F. Skrutskie (UMass)

The 2MASS project is a collaboration between the University of Massachusetts and IPAC

multi-λ views of a Supernova Remnant

Shocks seen in the X-ray
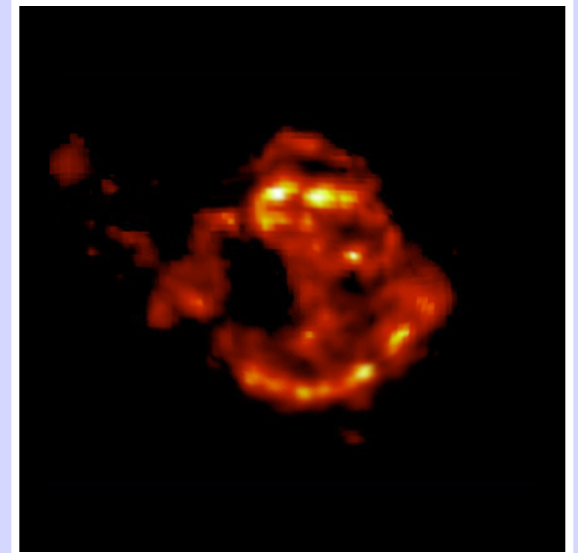
Dust seen in the IR
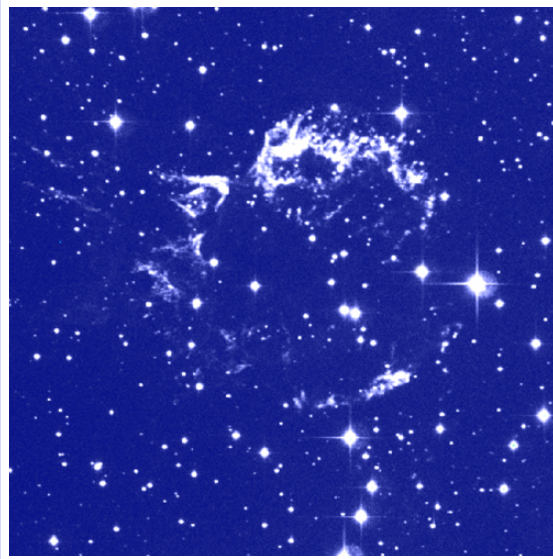
Heavy elements
seen in the optical
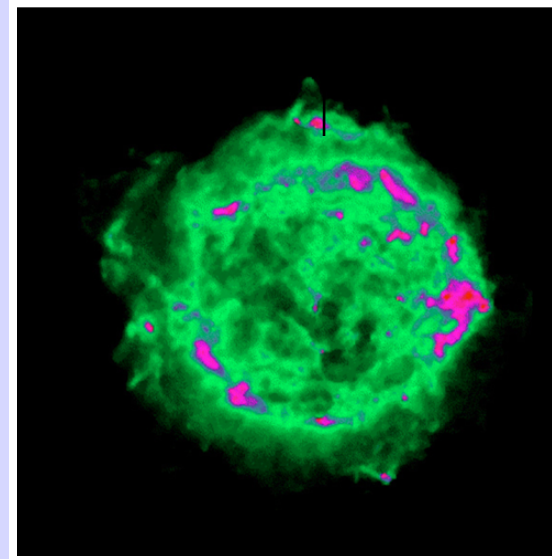
Relativistic electrons
seen in the radio

solar-terrestrial links
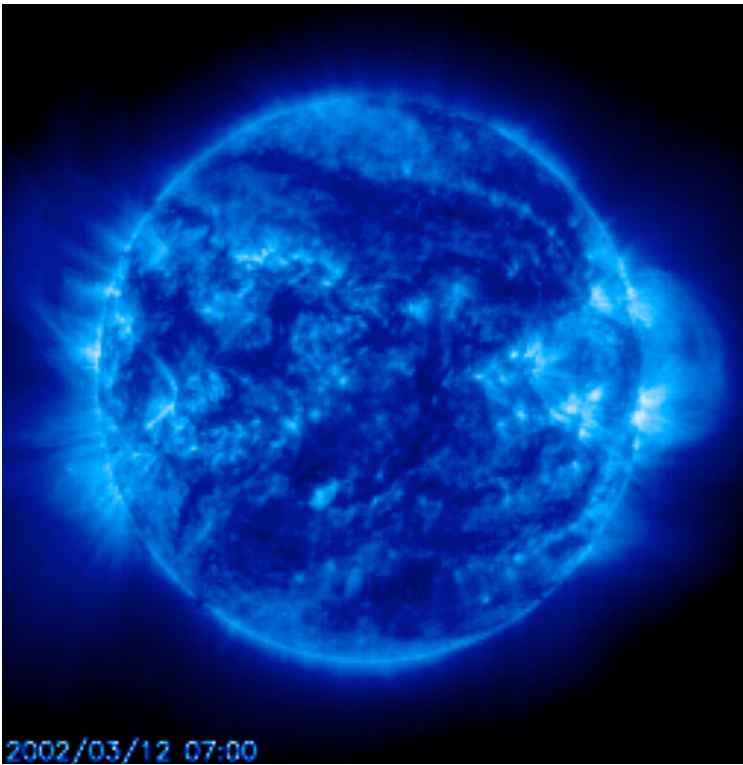
Coronal mass ejection imaged by space-based solar observatory
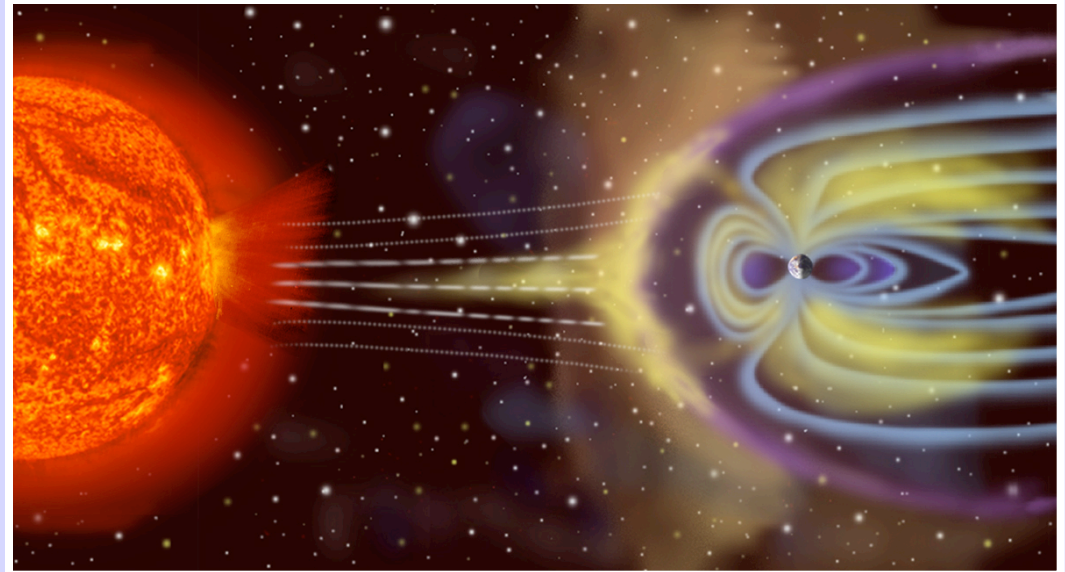
Effect detected hours later by satellites and ground radar

2002/03/12 07:00

large database science

- many goals require large number statistics
  - rare objects : search through billions
    (NEOs, Pop II brown dwarfs, z=7 quasars)
  - weak signal recovery : grav lensing
  - accurate estimation : eg galaxy power spectrum

- often N**2 problems : CPU as well as I/O

- want on-tap search engines and analysis engines

everybody can be a power user

democracy

- facilities in Calcutta
  should be as good as in Caltech

VO drivers : technology

hardware trends

- ## ops, storage, bw : all 1000x/decade
  - can get 1TB IDE = $2K
  - backbones and LANS are Gbps

- ## but device bw 10x/decade
  - real PC disks 10MB/s; fibre channel SCSI poss 100MB/s

- ## and last mile problem remains
  - end-end b/w typically 10Mbps

two bottlenecks

- searching 1TB at 10 MB/s takes a day
  - solved by parallelism
  - but want the engine next to the data
  - and parallel code hard ... ==> people

- transferring 1TB at 10 Mbps takes a week
  - leave it where it is

- ==> data centres provide search service

shift the results not the data

network development

- higher level protocols ==> transparency

- TCP/IP   message exchange
- HTTP   doc sharing (web)
- grid suite   CPU sharing
- XML/SOAP   data exchange

  ==> service paradigm

VO drivers : data growth

archive data rates

- map the sky : 0.1" x 16 bits = 100 TB
- process to find objects : billion row tables
- VISTA 100 TB/yr by 2009
- SKA datacubes 100PB/yr by 2020
- not a technical or financial problem
  - LHC doing 100PB/yr by 2008
- issue is logistic : data management
- need professional data centres

data rich future

- heritage
  - Schmidt, IRAS, Hipparcos
- current hits
  - VLT, SDSS, 2MASS, HST, Chandra, XMM, WMAP, UKIDSS
- coming up :
  - VISTA, ALMA, JWST, Planck, Herschel
- cross fingers :
  - LSST, ELT, Lisa, Darwin,SKA, XEUS, etc.
- plus lots more

- issue is archive *interoperability*
  - need **standards** and **transparent infrastructure**

data access infrastructure is small Δ on huge investment ..

the VO concept

- web all docs in the world inside your PC
- VO all databases in the world inside your PC

whats its not

- not a monolith

- not a warehouse

VO framework

- agreed *standards*
- inter-operable *data collections*
- inter-operable *software modules*

- no central VO-command

- its not a thing
- its a way of life

VO geometry

- not a warehouse

- not a hierarchy

- not a peer-to-peer system

- small set of *service centres*
  and large population of *end users*

what is needed ?

- global standards
- well funded data centres
- working data services
- infrastructure software
- VO aware client tools
- VO aware data mining services

what is needed ?

- global standards   IVOA
- well funded data centres   skin of teeth
- working data services   growing
- infrastructure software   AstroGrid leads
- VO aware client tools   growing
- VO aware data mining services   a few

standards
International VO Alliance

- formal process based on W3C

- key standards agreed
  - formats
  - service metadata
  - data access protocols
  - column semantics
  - s/w interfaces
  - storage addressing

AstroGrid

AstroGrid Project

- UK's contribution to the VO

- Eleven institutions at various times
  - Edinburgh, Cambridge, Leicester, RAL, MSSL, Bristol, Exeter, Manchester, Leeds, Portsmouth, Belfast

- Key partner in Euro-VO
  - AstroGrid, CDS, ESO, ESA + European data centres

- Funded by STFC and EU 2002-2007
  - 2008-9 in balance

- 10-20 FTEs at various times

Aims

- mix and match infrastructure suite
- world leadership in VO technology
- complete working system for UK astronomers

Milestones

- 2002-2004 R&D study; technology development

- 2005-2007 AstroGrid-2
             Construction and prototyping
             AstroGrid Workbench has ~500 registered users

- 2008-2009 AstroGrid-3 : funded but grants not arrived yet...
             Deployed System : VO Desktop
             April 2008 release
             Operations

- 2008 Sept Review of long term status

Key components

- Standards contbns
- Registry implementation
- Virtual Storage : MySpace
- Single Sign-on : Community
- Applications interface : CEA
- Client side middleware : Astro Runtime
- Application interoperability : PLASTIC
- Data Set Access publishers library
- New user tools
- Server side workflow : abandoned
- Client side workflow : adopted Taverna
- Client side workflow : Python scripting library

VODesktop

Concept

- Desktop software calling remote services
- AG core + interoperable tools
- Download installer
- Put in "Start Menu" or similar
- Looks like any other application suite

Not

- Clever web portal
- Server based Web apps
- Server side workflow

  ...... yet

Behind the scenes

- Client side middleware (AR and PLASTIC)
- Core AG services
  - VOSpace, Registry, Sign-on
- Consortium data services with DSA in front
- Rest of world's data services
- Remote app services with CEA wrapper

Core tools

- Resource discovery and bookmarking (VOExp)
- Searching for data (Astroscope)
- Browsing VOSpace (FileExplorer)
- Querying databases (Querybuilder)
- Launching remote apps (Task Launcher)
- Python scripting library

Third party tools that interoperate

- Tables and plotting (Topcat)

- Image manipulation (Aladin)

- Spectral manipulation (SPLAT, VOSpec)

- Graphical workflow building (Taverna)

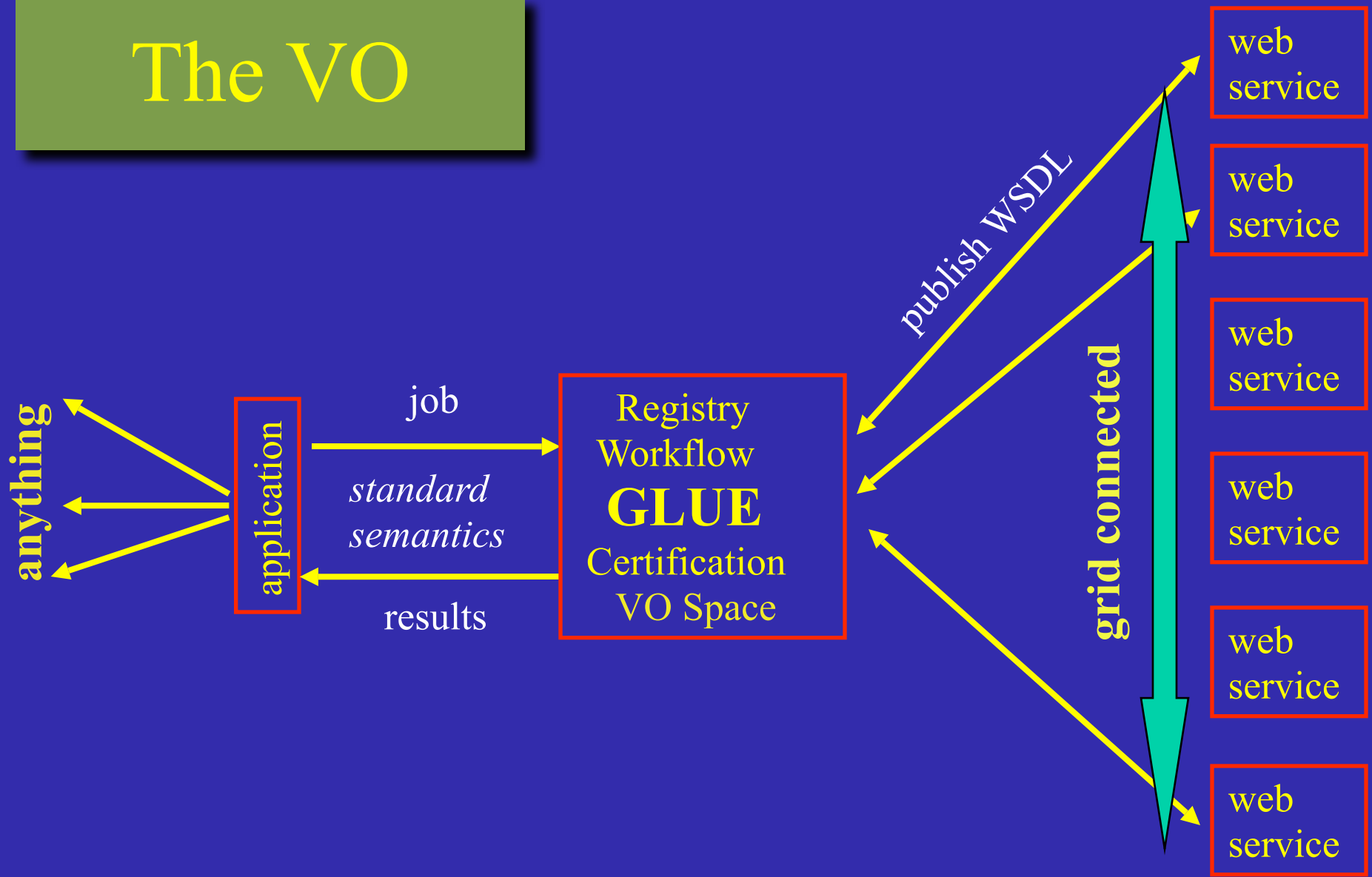 possibly : Visivo, Astroweka, Astroneural

Future

- IDL scripting; any-language scripting
- Run tools from data centre web pages
- All-VOscope
- Helioscope
- Integrate planetarium front end (ViRGO/GSky)
- Proliferation of tools
- Web apps
- Server side workflow engine(s)

Demo

Deleted Scenes

publishing metaphor

- facilities are **authors**
- data centres are **publishers**
- VO portals are **shops**
- end-users are **readers**

- VO infrastructure is **distribution system**.

services

- well funded data centres ?
  - skin of teeth
- working data services
  - growing steadily
  - image libraries; spectrum libraries; catalog searches; SQL queries
- VO portals
  - several effective prototype one-stop-shops
  - Aladin (France), NVO portal (US), AstroGrid workbench (UK)
  - new in April : VO Desktop (AstroGrid/UK)

Infrastructure

- Key software in place from AstroGrid, NVO, JVO, CDS and others
    - Registry (yellow pages)
    - Virtual Storage (VOSpace)
    - Job Execution - workflow
    - API for tools (Astro Runtime)
    - Message protocol for tools (PLASTIC)
    - Single Sign On (Community)

- AstroGrid is the only project that does all of this

VO Tools so far

- ## VO aware versions of old tools
  - eg Aladin, TopCat, Montage

- ## New tools
  - eg VOExplorer, Astroscope, VOSpec, VOPlot
  - AR makes it possible to write your own

- ## Server side applications
  - eg Sextractor, Hyperz, Astroneural, Visivo, WEKA