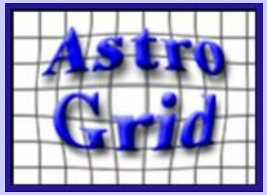
The Virtual Observatory
what it is and where it came from

- VO drivers
- VO vision
- VO progress

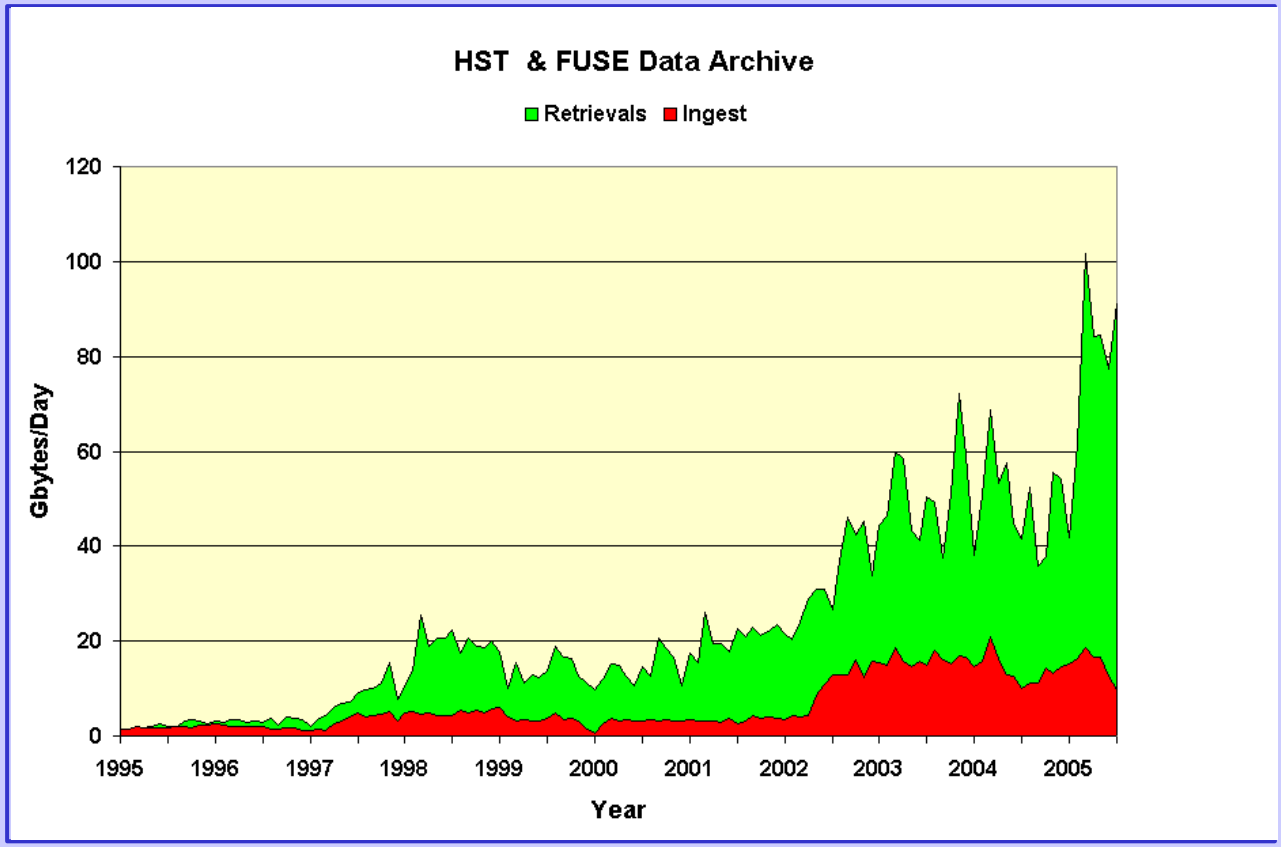


VO drivers : science

science services

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

archive re-use



- processed data (catalogs etc) ==> primary sources (SDSS, UKIDSS, etc)

on-line research

- users increasing assume on-line availability
- trad
  - download files, analyse at home
  - reduction standardised, analysis home grown
- new
  - analyse in-situ
  - analysis standardised

multi-archive research

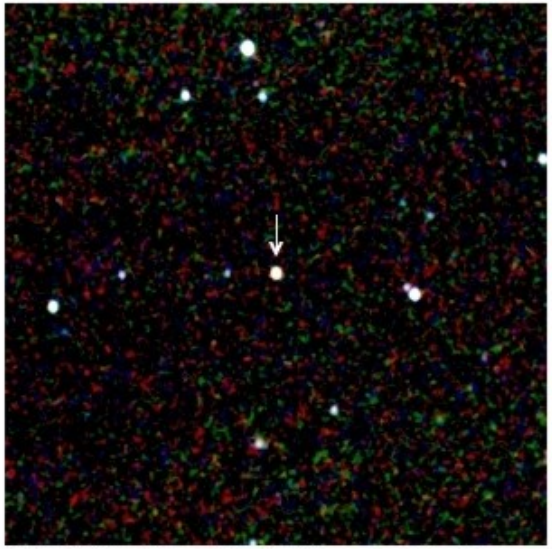
- most science goals
  require use of multiple data sources ...

brown dwarfs

2MASS J1146+2230

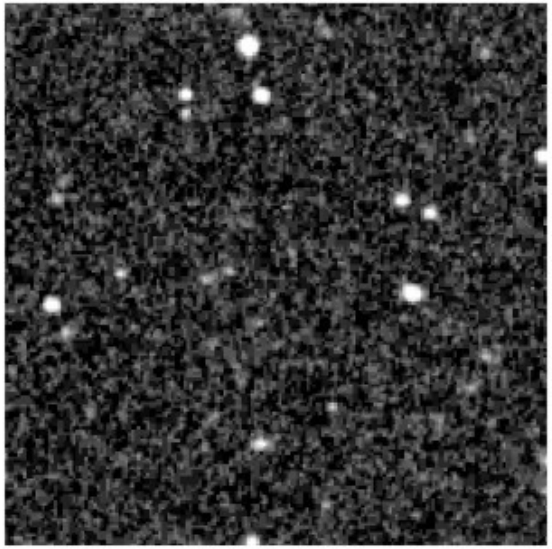
An L–type dwarf in the constellation Leo

The near–infrared view



2MASS Atlas JHK$_s$ Composite Image

The optical view

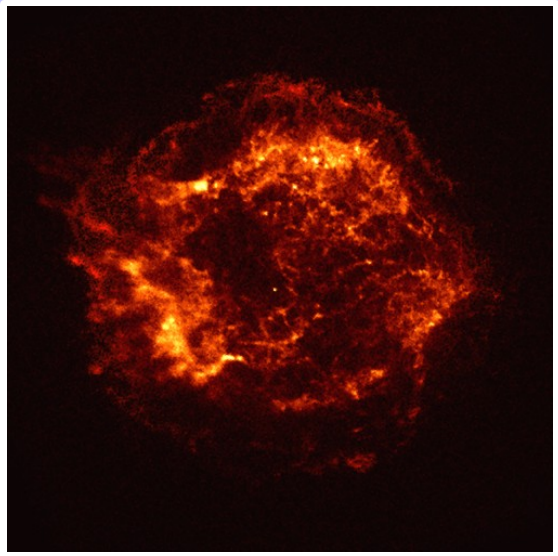


Palomar Digitized Sky Survey

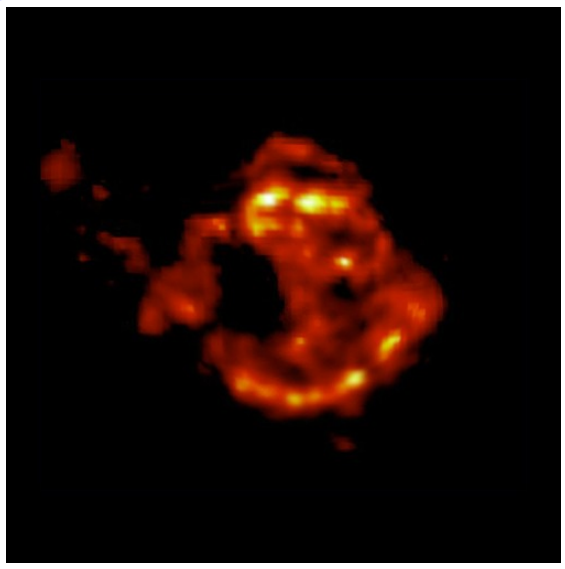
J.D. Kirkpatrick (IPAC/Caltech), I.N. Reid (Caltech), R.M. Cutri (IPAC/Caltech),
C.A. Beichman (IPAC/JPL/Caltech), J. Liebert (U of A), M.F. Skrutskie (UMass)

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

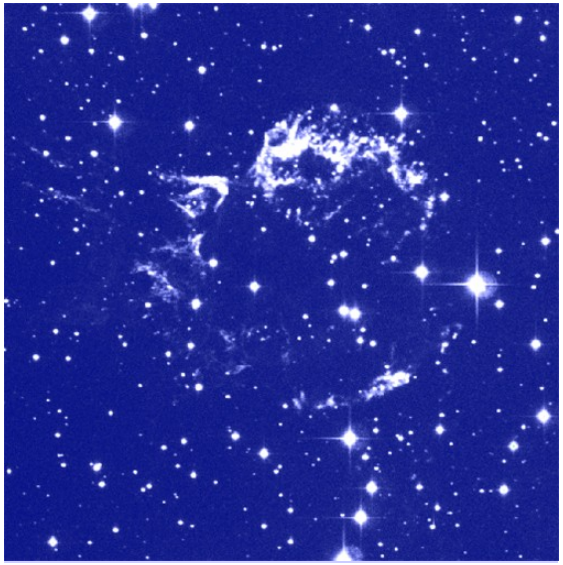
multi-λ views of a Supernova Remnant



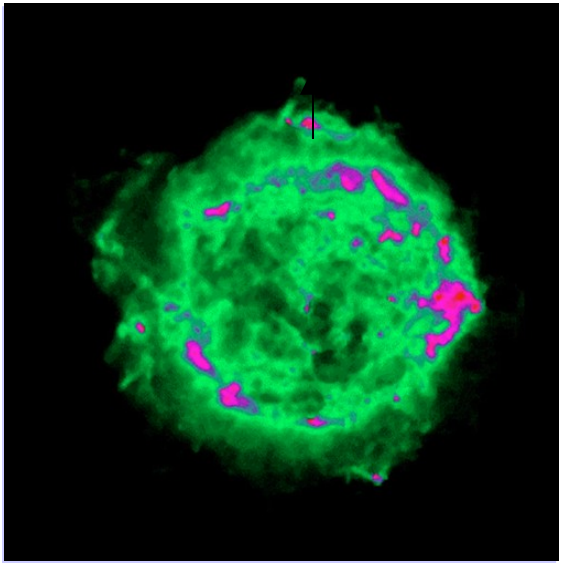
Shocks seen in the X-ray



Dust seen in the IR

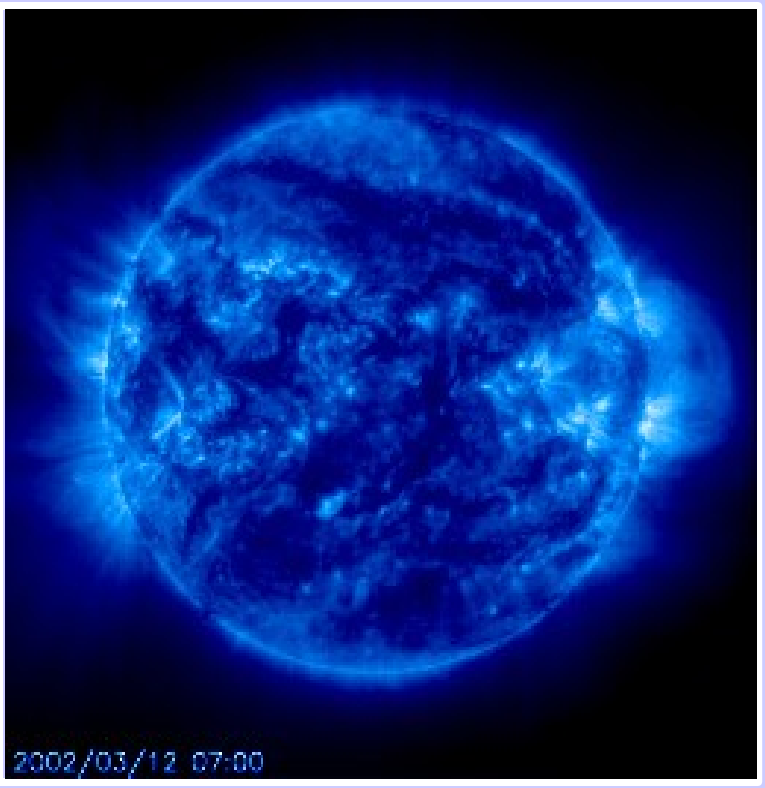


Heavy elements
seen in the optical

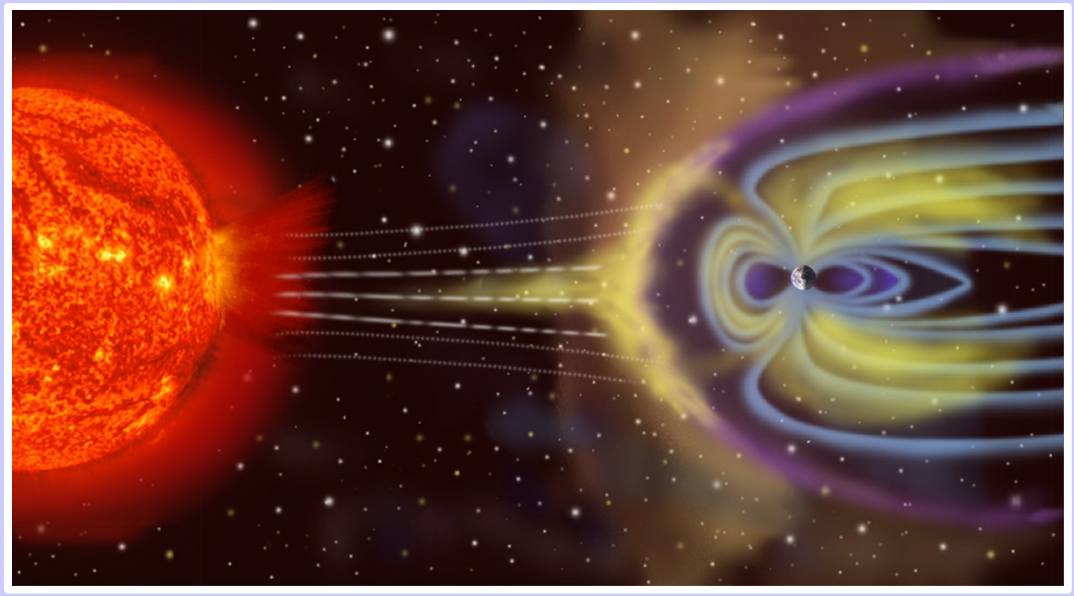


Relativistic electrons
seen in the radio

solar-terrestrial links



Coronal mass ejection imaged by space-based solar observatory



Effect detected hours later by satellites and ground radar

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 = $3K
  - 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
  - shift the results not the data
- ==> data centres provide search service

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 2007
- SKA datacubes 100PB/yr by 2020
- not a technical or financial problem
  - LHC doing 100PB/yr by 2007
- 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

data access infrastructure is small Δ on huge investment ..

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

VO Vision

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

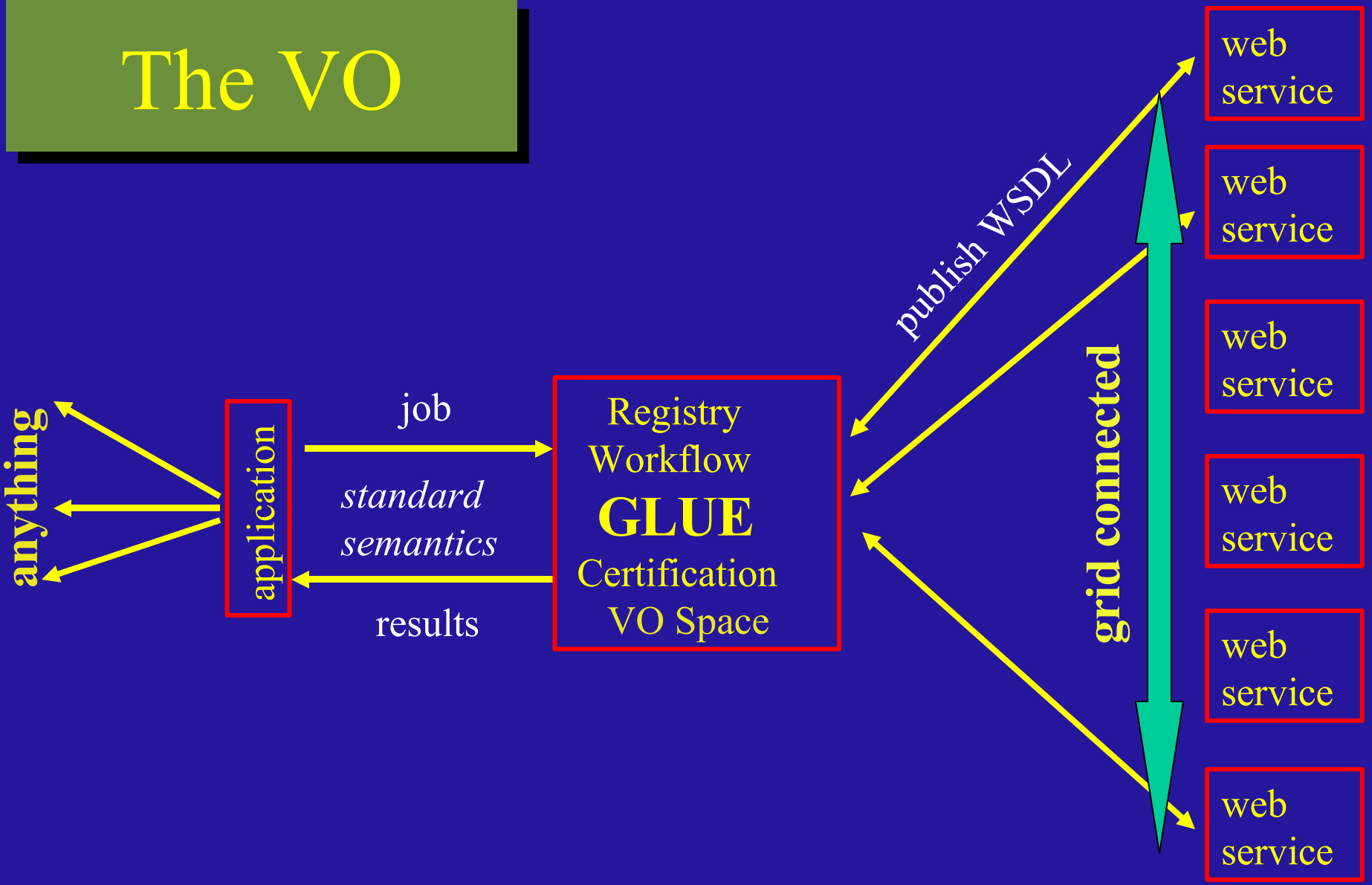
- its not a thing
- its a way of life

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*

The VO


anything
application
job
standard
semantics
results
Registry
Workflow
GLUE
Certification
VO Space
publish WSDL
grid connected
web service
web service
web service
web service
web service
web service

publishing metaphor

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

- VO infrastructure is **distribution system**.

VO progress

what is needed ?

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

standards
International VO Alliance

- formal process based on W3C
- key standards agreed
  - formats
  - service metadata
  - data access protocols
  - column semantics
  - s/w interfaces

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 one-stop-shops
  - SkyQuery, Aladin, NVO portal, AstroGrid workbench
- VO aware data mining services
  - little so far

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)
- Key next step
  - Single Sign On (Community)

VO Tools so far

- VO aware versions of old tools
  - eg Aladin, TopCat, Montage
- New tools
  - eg DataScope, Astroscope, VOSpec, VOPlot
  - AR makes it possible to write your own
- Server side applications
  - eg Sextractor, Hyperz, Astroneural, Visivo, WEKA

scoresheet

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

following development projects 2001-2006
all this is well underway (except No. 2 ? ....)
*==> **time to make VO a working reality***

FIN