NATIONAL VIRTUAL OBSERVATORY

THE BASIS FOR THE VIRTUAL OBSERVATORY: HISTORY & CONTEXT

ETHAN J. SCHREIER

AUI, STSCI & JHU

IAU Joint Discussion 08

LARGE TELESCOPES & VIRTUAL OBSERVATORY: VISIONS FOR THE FUTURE

13-26 July 2003, Sydney, Australia

--Astronomy and Astrophysics in the New Millennium, National Research Council

National Science Foundation, Information Technology Research Program, Cooperative Agreement AST0122449

The Basis for the Virtual Observatory: History and Context

- 1960s-70s: Prehistory - computers enter astronomy
- 1970s-80s: Early years - digital data & user facilities
- 1980s: Adolescence - astronomy as big science & IT innovator
- 1990s: Adulthood - established astronomy data systems;
  internet/web in popular use & mass market s/w

- Now: *VO's time has come*
  major surveys, large telescopes, new age of astronomy discovery with massive multi-dimensional databases

from punch cards to petabytes in 50 years

- Future: *(senility??)* - further expansion, an operational VO, coordination with other disciplines
  - *An interplanetary virtual observatory ?*

Prehistory of the Virtual Observatory

- 1960s - 70s: early use of computers in astronomy
  - operating telescopes & taking data
  - electronic detectors introduced in astronomy (eg PMTs => CCDs)
  - use of digital data
    - ⇨ beginning of computer-based data analysis
    - ⇨ digitizing existing tables, catalogs,… ca 1970
  - space astronomy missions:
    - computer-based testing and integration
    - astronomy & planetary missions: data in telemetry

Prehistory of the Virtual Observatory
(cont)

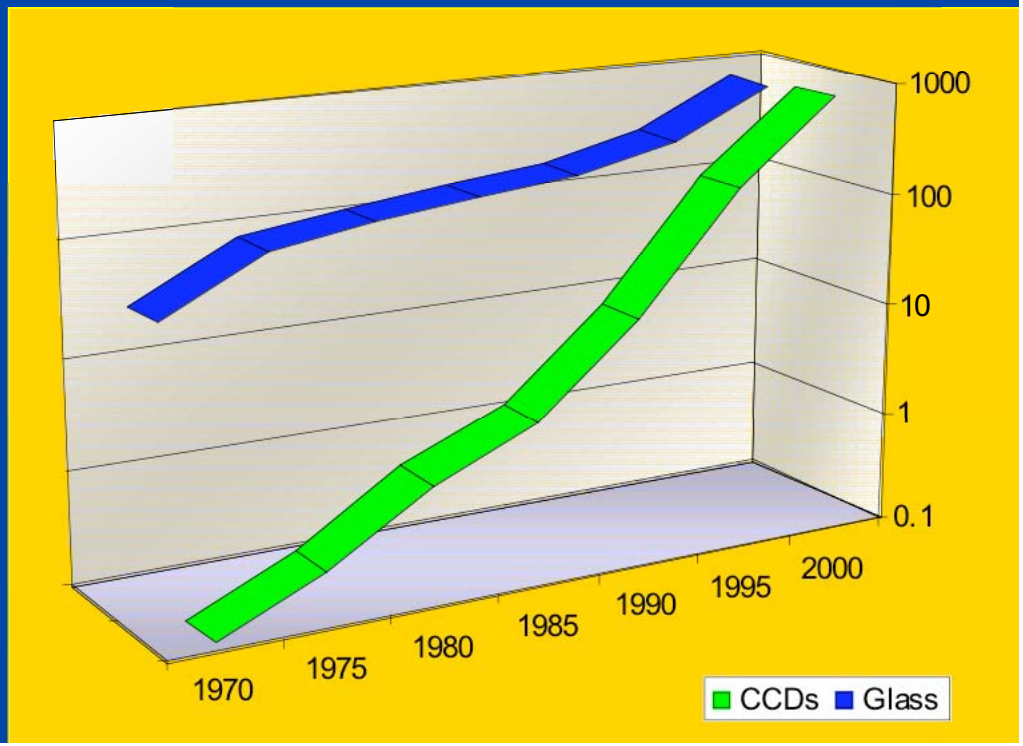
Establishment of early data centers

- photographic plate collections - measuring engines *(!);* occasionally digitized
  - e.g. Harvard plate stacks (since 1800s) still not completely digitized
- presence of plates led to digital star catalogs
- electronic data centers, e.g.
  - SAO - SAOCAT/SKYMAP, used for UHURU in 1970
  - CDS (1972) - SIMBAD stellar data base
  - ADC (1978) - based on CDS model

Early Years - the underpinnings

Numerous dramatic changes by late 1970s - 1980s

- Electronic detectors the norm => required electronic archiving
- Evolution of facility-class observing facilities
  - expansion of ground-based optical: NOAO, ESO, CFHT…
  - NRAO/VLA - radio
  - NASA's Einstein Observatory - X-ray



Early Years - the underpinnings (cont)

Digital data became widespread (although small volume) but lacked "data management" and preservation

- CODMAC – informed by need to preserve planetary data
  - found "data management" missing management
  - systems approach needed
  - but mainly
    - ➢ *user involvement/control needed at all stages of development lifecycle (… before data became information became knowledge)*
    - ➢ *incompatible with normal NASA-type procurement process*

Early Years - the underpinnings (cont)

Attempts to make astronomy facilities useful for other than initial observer, or specialized user, *e.g.*

- Mission-oriented centers to support observers, e.g. Einstein Observatory in 1978 w/20% GO time
- CDS evolved in 1983 to "…*collect, homogenize, distribute, preserve astronomical information for the usage of the whole astronomy community*"
- STScI established in 1981 to support *all* users of first large optical facility in space (launched in 1990)
  - ⇨ *With HST, astronomy became big science, with resources + visibility for user support (including non-optical astronomers)*

User facilities => need for standards, user-oriented analysis systems

- Development/adoption of FITS
- Development of facility data analysis systems (*e.g.* AIPS, IRAF, STSDAS); special purpose systems (*e.g.* FOCAS, MIDAS,…)

Early Years - the underpinnings (cont)

- Growing coordination among different astronomy sub-disciplines: multi-wavelength astronomy
  - mission/wavelength-specific user facilities & science-oriented, value-added systems - e.g. SIMBAD, NED, IPAC, SKYVIEW,…
- Beginning of wide-spread networking (*e.g.* ARPAnet, DECnet..); recognition of power of networks combined with archives and common data analysis tools
- Potential of a distributed *astronomy data system*

  ⇨ *NASA-sponsored workshops about and astronomical data system (~1988)*

  ⇨ *Attempts at an ADS (NASA), ESIS (ESA), …*

Early Years - the underpinnings (cont)

Initial ADS not a smashing success, but:

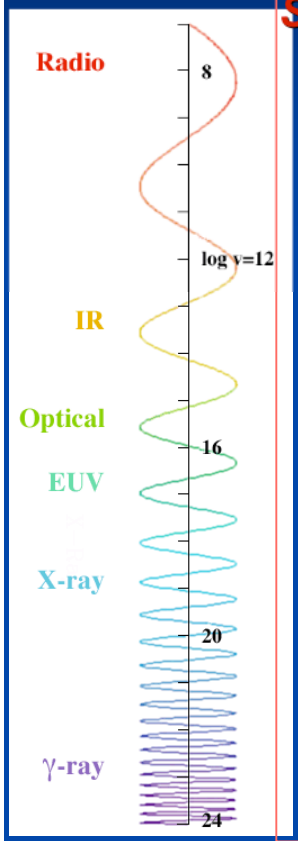
- Produced enormously successful bibliographic service
- Formalized links among major data centers
- Provided lessons for future astronomy systems, *e.g.* VO
  - *timing is critical* - leverage developing technology & minimize custom software: use industry standards, open & freely available software & tools,
  - Concentrate on interoperability and metadata standards, don't force internal database standards
  - Utilize distributed development and consensus, not overly-centralized management
  - Add value & insure buy-in by users (not just IT professionals)

The 1990s

Large-scale research using archived data

- Major end-to-end mission data systems, *e.g.*
  - HST, VLT, CXC, GEMINI, SIRTF, NRAO….
- Major discipline data centers **, *e.g.***
  - HEASARC, MAST, IRSA, SDSS, NRAO…
  - ESO, STECF, CDS, CADC …
  - ,…

A NASA PERSPECTIVE



Mission Science Centers

MAP
SWAS
Planck
Herschel
ISO
SIRTF
SOFIA
2MASS
HST
GALEX
FUSE
Chandra
XMM-Newton
BeppoSAX
RXTE/Swift
INTEGRAL
GLAST

Wavelength Focused Science Archive Centers

NSSDC/ADF
IRSA
MAST
HEASARC

Permanent Archive

NSSDC

Integrating Services

NED
ads NASA
SkyView
SIMBAD/CDS
ADC

Astronomy and Physics Community
General Public
Education Community

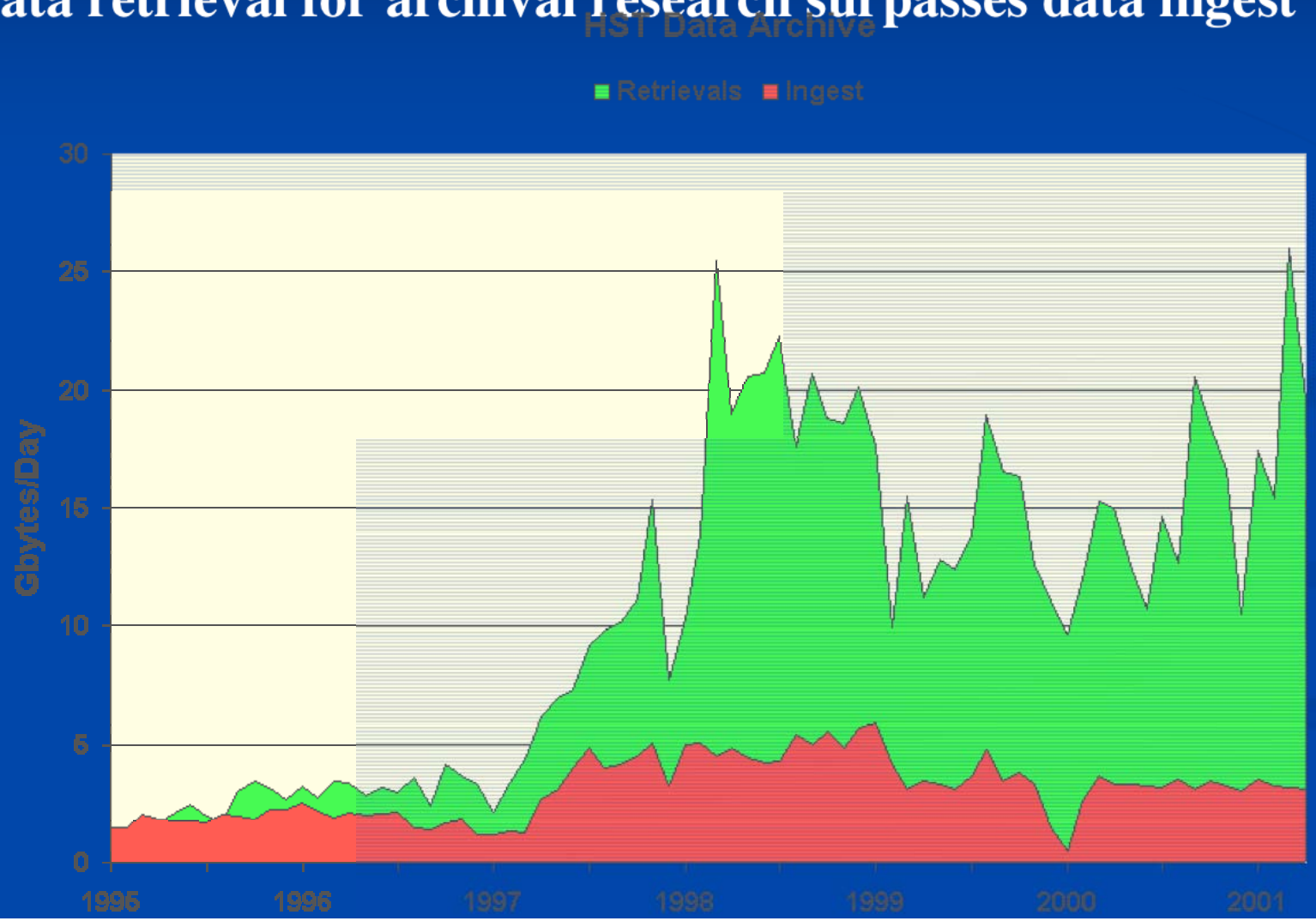
The 1990s:

Large-scale research using archived data

- Major end-to-end mission data systems, *e.g.*
  - HST, ESO, CXC, GEMINI, SIRTF, NRAO….
- Major discipline data centers **, *e.g.***
  - HEASARC, MAST, IRSA, SDSS, NRAO…
  - ESO, STECF, CDS, …
  - CADC, ANTF,…
- Creation of large datasets for archival research - SURVEYS
  - GSS/DSS/GSC-II
  - 2MASS
  - HDF
  - SDSS
  - 2df, GOODS, 6df, DEEP, MACHO, DPOSS, COBE, MAP, NVSS, FIRST, GALEX ...
  - Observatories committing >~ 20% of time to survey teams ……

The 1990s
(cont)

Data retrieval for archival research surpasses data ingest



The Decade of the Virtual Observatory

"…the National Virtual Observatory (NVO) is the committee's top priority among the small initiatives. The NVO will provide a "virtual sky" based on the enormous data sets being created now and the even larger ones proposed for the future. It will enable a new mode of research for professional astronomers and will provide to the public an unparalleled opportunity for education and discovery."

Astronomy and Astrophysics in the New Millennium, NRC/NAS 2000

The Decade of the Virtual Observatory cont

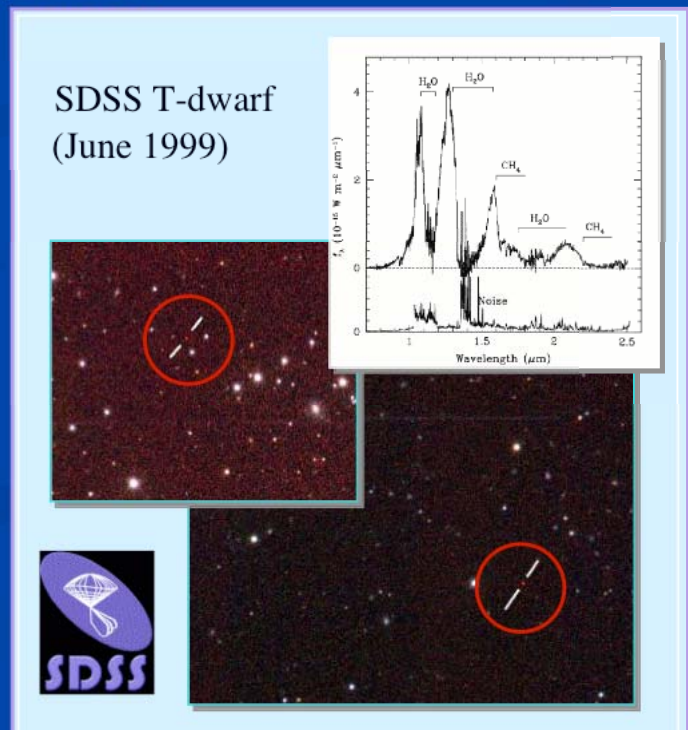
VO would provide new discovery space

Simultaneous access to multiple archives at multiple wavelengths

- Access to analysis tools, super-computing resources

  => data mining

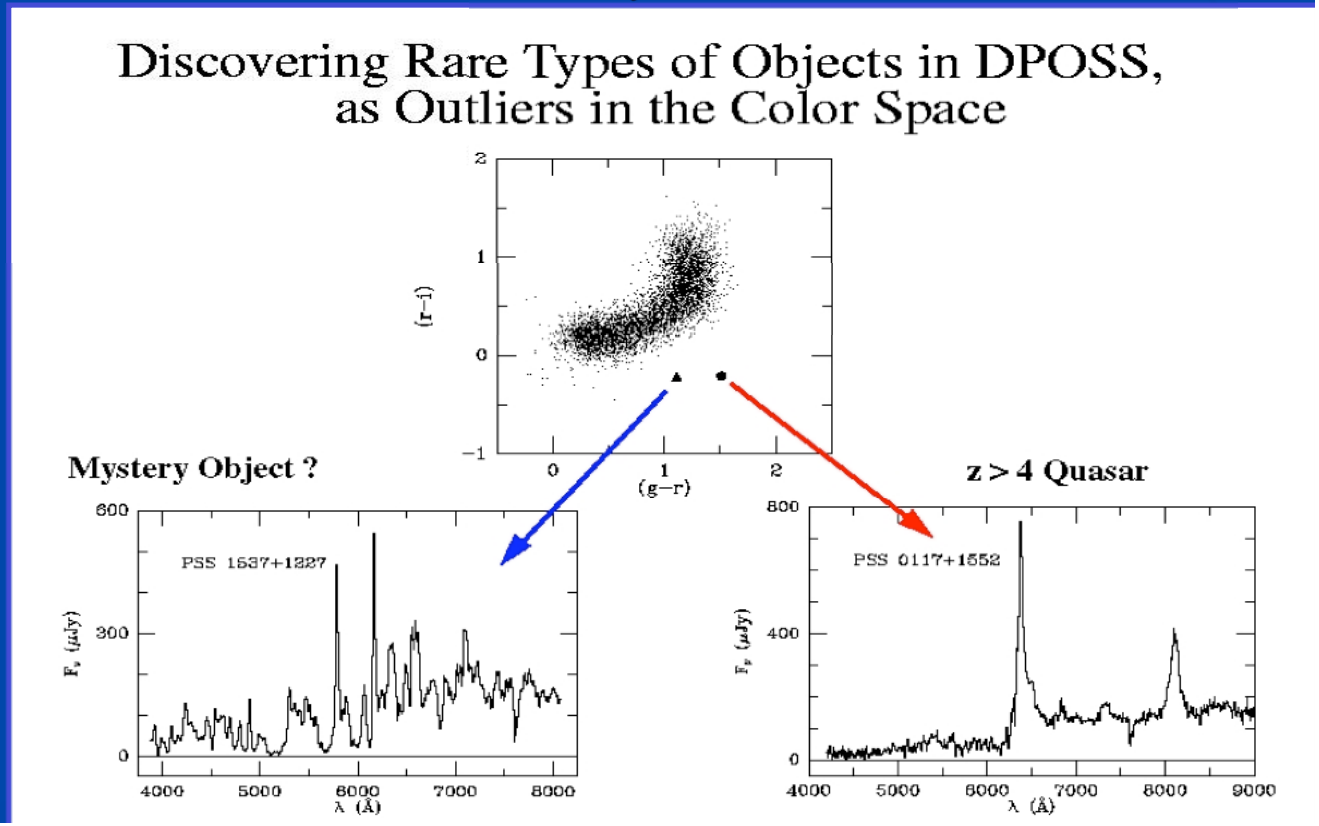
- Confront data with simulations and theory



The Decade of the Virtual Observatory
cont

VO would enable *statistical astronomy* - massive correlations in multiple dimensions: across wavelengths, across time,…

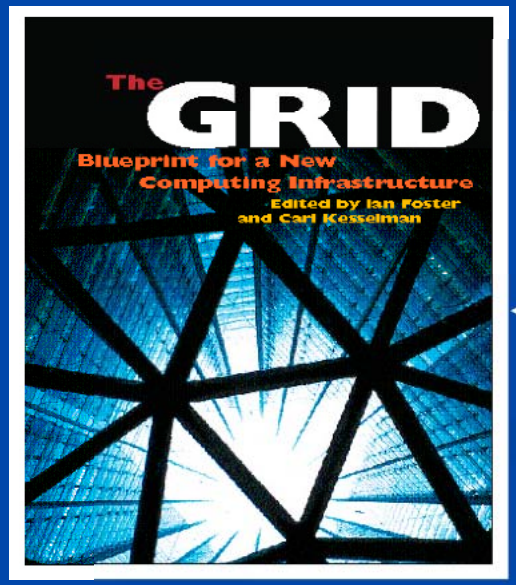
outliers => ultimate serendipitous discovery



The Decade of the Virtual Observatory
…more than just astronomy

"…Internet computing and Grid technologies promise to change the way we tackle complex problems. They will enable large-scale aggregation and sharing of computational, data and other resources across institutional boundaries. And harnessing these new technologies effectively will transform scientific disciplines ranging from high-energy physics to the life sciences."

Dr. Ian Foster, GLOBUS Project



First Initiatives

A US *Interim Steering Committee* decided to pursue NVO through the NSF Information Technology Research program in fall 2000

- Large astronomy/IT/CS collaboration: 42 collaborators, 16 organizations, 8 international liaisons
- Nine education partners
- Leveraging on similar efforts such as GriPhyN, many IT developments

Goals: create the necessary information technology framework

demonstrate utility with science prototypes

NVO ITR Proposal Team

	A. Szalay	*JHU*	
	P. Messina	*Caltech*	
C. Alcock	*U. Penn.*	M. Livny	*U. Wis.*
K. Borne	*ADC/Raytheon*	C. Lonsdale	*IPAC*
T. Cornwell	*NRAO*	T. McGlynn	*HEASARC/USRA*
D. DeYoung	*NOAO*	A. Moore	*CMU*
G. Fabbiano	*SAO*	R. Moore	*SDSC/UCSD*
A. Goodman	*Harvard*	J . Pier	*USNO*
J. Gray	*Microsoft*	R. Plante	*NCSA/UIUC*
R. Hanisch	*STScI*	T. Prince	*Caltech*
G. Helou	*IPAC*	E. Schreier	*STScI*
S. Kent	*FNAL*	N. White	*GSFC*
C. Kesselman	*USC*	R. Williams	*Caltech*

Collaborators

B. Berriman	IRSA	J. Good	IPAC
R. Brissenden	SAO	I. Griffin	STScI
R. Brunner	Caltech	B. Madore	NED/IPAC
C. Cheung	GSFC	J. Mazzarella	NED/IPAC
K. Cook	LLNL	B. McLean	STScI
A. Connolly	U. Pitt.	M. Postman	STScI
D. Curkendall	JPL	A. Rots	SAO
G. Djorgovski	Caltech	S. Strom	NOAO
I. Foster	U. Chicago	A. Thakar	JHU
R. Gal	JHU	D. Tody	NOAO

International Liaisons

P. Benvenuti	ST-ECF/ESA	F. Murtagh	U. Belfast
D. Durand	CADC	R. Norris	ATNF
F. Genova	CDS	S. Okamura	U. Tokyo
A. Lawrence	ROE	P. Quinn	ESO

Education Partners

Association of Science-Technology Centers	International Planetarium Society
National Air and Space Museum	Silicon Graphics (Digital Planetarium)
Spitz (Electric Sky)	Maryland Space Grant Consortium
Gettysburg College (Project CLEA)	UC Berkeley (CSE@SSL)
American Museum of Natural History	

Toward an International Virtual Observatory

VO is by necessity international

- Astronomy is big *and* international science
  - International collaborations
  - Requires multi-wavelength approach, international resources & data
- Data volumes keep increasing (doubling times < 12 months) and cannot be replicated indefinitely

NVO followed rapidly by
Astrovirtel, Opticon, AVO, Astrogrid…

Toward an International Virtual Observatory

Astrogrid	VO United Kingdom
Aus-VO	Australian Virtual Observatory
AVO	Astrophysical Virtual Observatory
China-VO	Virtual Observatory of China
CVO	Canadian Virtual Observatory
FVO	French Virtual Observatory
GAVO	German Astrophysical Virtual Observatory
IDGAR	VO Italy
JVO	Japanese Virtual Observatory
KVO	Korean Virtual Observatory
NVO	National Virtual Observatory
RVO	Russian Virtual Observatory
VO India	Virtual Observatory India

The proof of concept is in the diverse local implementations & IVO's global coordination by consensus

Toward an International Virtual Observatory

Astrogrid	VO United Kingdom
Aus-VO	Australian Virtual Observatory
AVO	Astrophysical Virtual Observatory
China-VO	Virtual Observatory of China
CVO	
FVO	
GAVO	
IDGAR	
JVO	Japanese Virtual Observatory
KVO	Korean Virtual Observatory
NVO	
RVO	
VO India	

nb. The philosophy of the VO + net => would be a struggle to keep VO *from being* international

International liaison & voluntary coordination are evolving naturally

The proof of concept is in the diverse local implementations & IVO's global coordination by consensus

Toward an International Virtual Observatory

The International Virtual Observatory Alliance

A Mission and Roadmap Statement
2002-2005

Peter Quinn, Bob Hanisch and Andy Lawrence
On the behalf of the ASTROGRID, AVO and NVO Projects
June 10 2002

International Virtual Observatory: Where we are today

Harvesting the results of a "thirty year trend toward …collectivization and democratization" in astronomy

(with apologies to Andy Lawrence, and despite the current state of the world)

Goal is to create a *digital sky* comprising *all* relevant archives

- make distributed archives speak the same language - appear as one
  - searchable by the same tools
  - accessible through a uniform interface
- linked to real observatories
- linked to computational resources and services
- linked to theoretical capabilities

International Virtual Observatory: What we are doing today

- Continue to develop the framework
- Continue and advance international IVO collaboration via consensus - an alliance of the willing
  - e.g. IVOA, IAU Commission 5 for validation ??
- Demonstrate power of VO via science demonstrations
  - *Must be useful to astronomers, not just CS/IT*
- Start defining the operational era of VO

International Virtual Observatory: The Future

The major new international facilities, *e.g.*
ALMA, JWST, VISTA, LSST, SKA, GSMT, ELT/OWL(sic), HST2…

- will involve major collaborations
- will require intensive computation, on-line data processing & data mining
- will carry out coordinated surveys & generate major new archives

The Virtual Observatory

- will involve major collaborations
- will require intensive computation, data processing & mining techniques
- will facilitate the scientific exploitation of coordinated surveys

=> synergism of new observing facilities and the VO

=> they must coordinate and evolve together

=> observatories will be nodes of the VO

Is Astronomy Alone?

some final remarks

- high energy physics: is concentrating on grid computing to meet the computational problem of analysing petabytes/year… *is very relevant*
- surveillance communities probably solved many problems related to mining results from multiple data bases… *but is not much good to us*
- earth science: fleets of spacecraft producing data at 10-100x astronomy

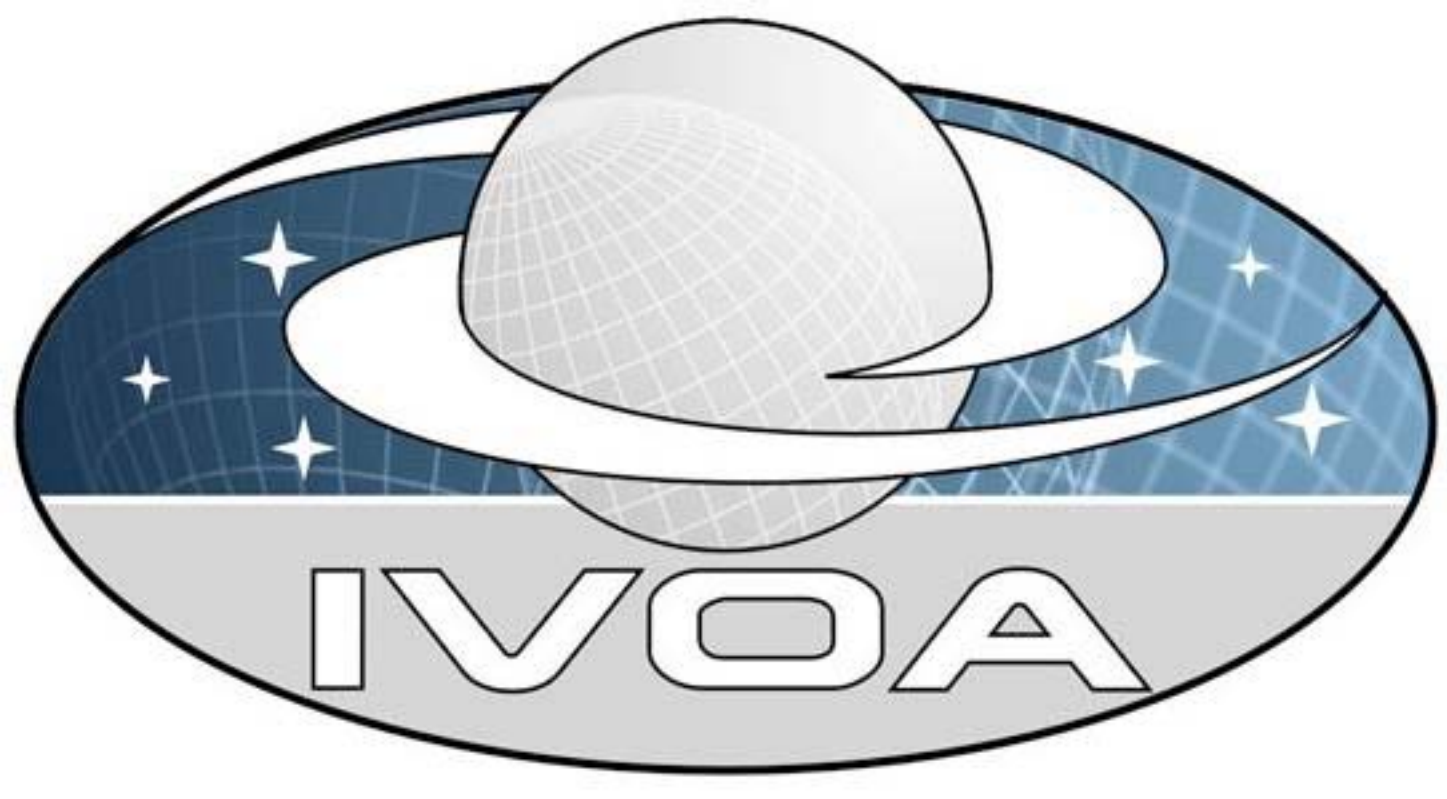
  EOSDIS after a decade of calls for a distributed system, has now done so: dealing with terabytes of data per day & creating a federation of nodes

  - Astronomy data holdings catching up , community will have to start watching
  - Earth science may become frontier of real-world IT experiments for science
  - Larry Smarr, who started at Harvard as an astronomer, then led NCSA, is now chair of NASA's earth science advisory council

VO/EOS may have related growth paths ⇨ *space nodes in the VO???*

"…A plan filed with the Internet Engineering Task Force (5/01) would make it possible to create a communication network between spacecraft, asteroids, satellites, and planets. The project is supported by NASA and the Defense Advanced Research Projects Agency." www.ietf.org/internet-drafts/draft-irtf-ipnrg-arch-00.txt,

The Virtual Observatory is on its way



FLTs & Virtual Observatory need and complement each other