The US National Virtual Observatory

David De Young
and the USNVO Collaboration

Trends in Astrophysical Data

- Astrophysical data is growing exponentially
  - Doubling every year (Moore's Law):
    both data sizes and number of data sets

- Computational resources scale the same way
  - Constant funding levels will keep up with the data

- Main problem is the software component
  - Currently components are not reused
  - Software costs are an increasingly larger fraction
  - Aggregate costs are growing exponentially

Discoveries

- ## When and where are discoveries made?
  - Always at the edges and boundaries
  - Going deeper, using more wavelength bands
  - Physicists make many measurements and discard most; Astronomers make many measurements and find discovery in their entirety and combination

- ## Metcalfe's law
  - Utility of computer networks grows as the number of possible connections: $O(N^2)$

- ## VO: Federation of N archives
  - Possibilities for new discoveries grow as $O(N^2)$

- ## Current sky surveys have proven this
  - Very early discoveries from SDSS, 2MASS, DPOSS

Data Publishing Roles

Roles	Traditional	Emerging
Authors	Scientists	Collaborations
Publishers	Journals	Project www site
Curators	Libraries	Bigger Archives
Consumers	Scientists	Scientists
	read->analyze	query-> analyze

Changing Patterns

- Exponential growth
  - Data will be never centralized

- More responsibility on projects
  - Becoming Publishers and Curators
  - Larger fraction of budget spent on software
  - Lot of development duplicated, wasted

- More standards are needed
  - Easier data interchange, fewer tools

- More templates are needed
  - Individuals develop less software

Evolving Standards

- Astrophysics has a good track record

- FITS: universally used to share low level data
  - Individual images, tables, files
- But: new industry standards emerging
  - XML, SOAP
- Requirements of modern data exchange:
  - More dynamic (streams, queries)
  - Merging heterogeneous sources

Accessing Data: Today

1. Locate data from user supplied source
2. Download and study documentation
3. Identify necessary data components
4. Copy data to local machine
5. Read and filter data locally
6. Perform the analysis locally

Time Consuming and Inefficient

Accessing Data: Soon

Phase 1

1. Auto-discovery of data, and documentation
2. Study documentation
3. Filter (query) data from remote source
4. Analyze incoming data stream directly

Phase 2

- Perform even analysis remotely,
  close to the data source

Remote Resources

Today

- Accessing remote data:
  - WWW, FTP
  - Data formatted in certain ways
    - HTML, FITS

- Accessing remote computing:
  - Hard configured local area clusters
  - Remote supercomputers
  - Need to move data to the computing
  - Available resources do not always match problem

Emerging New Concepts

- Standardizing distributed data
  - Web Services, supported on all platforms
  - Custom configure remote data dynamically
  - XML: Extensible Markup Language
  - SOAP: Simple Object Access Protocol
  - WSDL: Web Services Description Language

- Standardizing distributed computing
  - Grid Services
  - Custom configure remote computing dynamically
  - Build your own remote computer, and discard
  - Virtual Data: new data sets on demand

A Response to These Trends

- THE VIRTUAL OBSERVATORY CONCEPT

- Characteristics
  - Distributed
  - Science Driven
  - Integrated With Information Technology
  - Broad Based Community Support
  - Builds on Existing Infrastructure

The US National Virtual Observatory

- National Academy of Sciences "Decadal Survey" recommended NVO as highest priority small (<$100M) project

  "Several small initiatives recommended by the committee span both ground and space. The first among them—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 …"

  —Astronomy and Astrophysics in the New Millennium, p. 14

The USNVO Initiative

- ## ORIGINS
  - VO White Paper (Alcock, Prince, Szalay): Jun1999
  - First NVO Workshop (JHU): Nov 1999
  - Formation of Initial Working Groups (Science, Management, Technical): Nov 1999
  - Formation of Interim Steering Cte: Feb 2000
  - Second NVO Workshop (NOAO): Feb 2000
  - Presentations to NASA and NSF: May 2000
  - First Major NVO Meeting (CIT): Jun 2000
 - Submission of Proposal to NSF: May 2001

Project Team

- NSF ITR project, "Building the Framework for the National Virtual Observatory" is a collaboration of 17 funded and 3 unfunded organizations
    - Astronomy data centers
    - National observatories
    - Supercomputer centers
    - University departments
    - Computer science/information technology specialists
- PI and project director: Alex Szalay (JHU)
- CoPI: Roy Williams (Caltech/CACR)
- $10M award for five-year period, beginning 1 Nov 01

Proposal Team

	A. Szalay	*JHU*	
	R. Williams	*Caltech*	
C. Alcock	*U. Penn.*	M. Livny	*U. Wis.*
K. Borne	*ADC/Raytheon*	C. Lonsdale	*IPAC*
T. Cornwell	*NRAO*	T. McGlynn	*HEASARC/USRA*
D. De Young	*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*		

Project Management
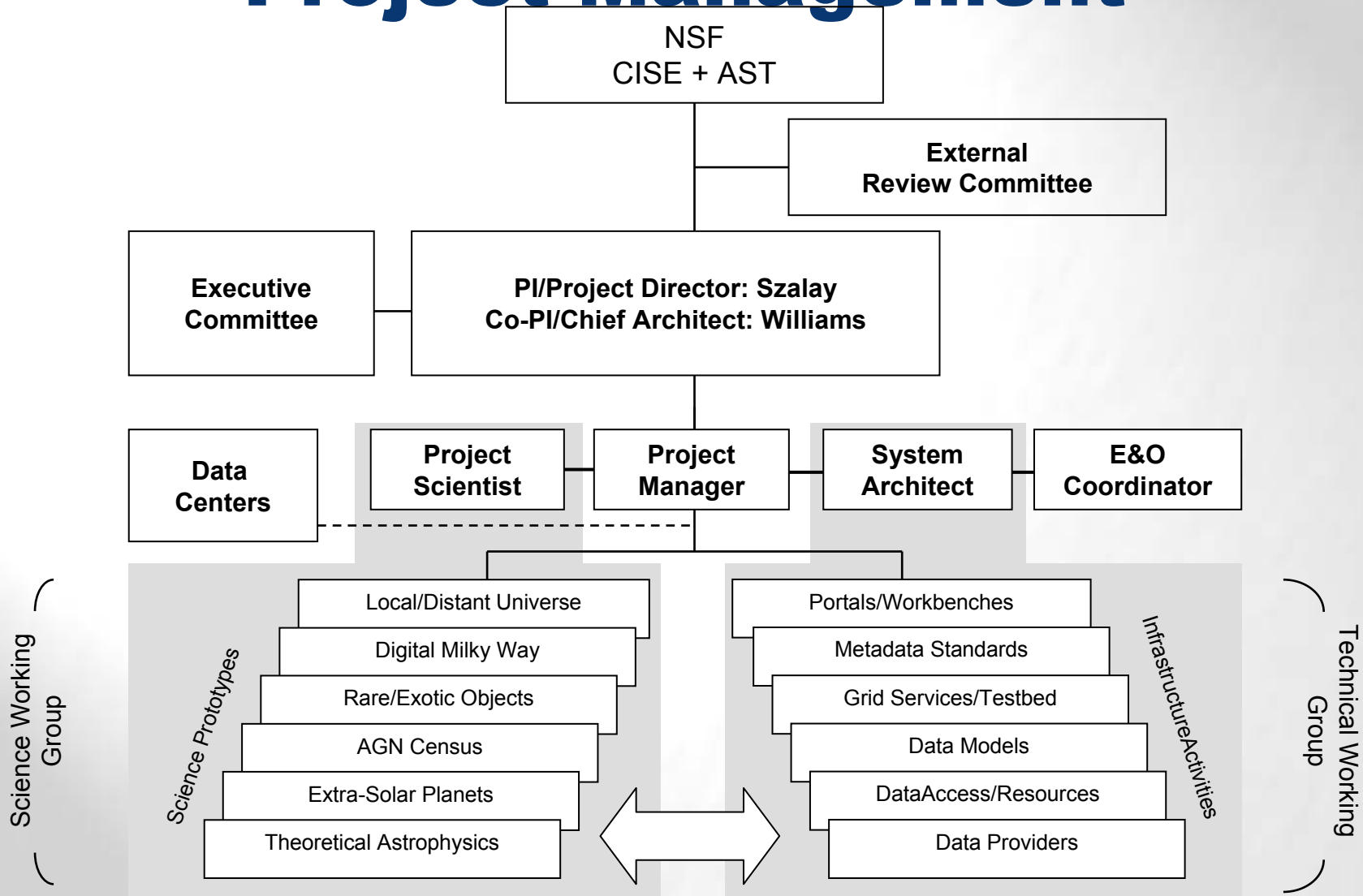
Team Organization

- Executive Committee
  - A. Szalay, R. Williams, R. Hanisch (PM), D. De Young (PS), R. Moore (SA), G. Helou, E. Schreier

- Education & Outreach
  - M. Voit, Coordinator

- First Working Groups established
  - Metadata (R. Plante/NCSA)
  - Systems (R. Moore/UCSD)
  - Science (D. De Young/NOAO)

- Project teams established for initial science demonstrations
  - GRB follow-up (T. McGlynn/HEASARC)
  - Brown dwarf search (B. Berriman/IPAC)
  - Cluster galaxy morphologies (R. Plante/NCSA)

Education & Outreach

- Integral part of project
- Emphasis is on development of partnerships
- Initiated with a workshop this summer at STScI (July 11-12)
  - Understand requirements on NVO services from perspective of formal education, informal education, commercial/corporate, and public outreach content developers

Education/Outreach 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
American Museum of Natural History	

Management Plan

- Formal management plan delivered to NSF in January 02

- 11 major work breakdown categories, with sub-elements to three levels

- All level-two technical WBS areas have designated lead who is responsible for tasks and schedule within that area

Building the Framework for the National Virtual Observatory

NSF Cooperative Agreement AST0122449

Management Plan

December 2001

The challenge of building a framework to enable the National Virtual Observatory will be met with a management structure that supports distributed research and development. We take optimal advantage of the domain expertise already resident in the organizations supporting the existing archival systems, sky surveys, and source catalogs of the astronomy community and meld this diversity with state-of-the-art information technology. Our structure ensures strong communication and coordination among the distributed, multi-disciplinary, heterogeneous resources, with accountability to both the community and the funding agency. It ensures that astronomy needs drive technology development.

Work Breakdown Structure

WBS	Year 1	Year 2	Year 3	Year 4	Year 5	Total
1 Project Management	2.35	2.65	2.55	2.55	2.55	12.65
2 Data Models	1.90	2.90	1.55	0.70	0.55	7.60
3 Metadata Standards	3.15	4.70	1.85	1.00	0.85	11.55
4 Systems Architecture	1.75	1.75	1.05	0.45	0.45	5.45
5 Data Access/Resource Layer	1.95	2.15	2.10	1.85	1.35	9.40
6 NVO Services	1.90	2.80	1.90	1.05	0.50	8.15
7 Service/Data Provider I&I	0.25	0.50	0.65	0.75	0.80	2.95
8 Portals and Workbenches	1.10	1.70	1.50	1.75	1.60	7.65
9 Test-Bed	0.85	2.05	2.15	2.35	2.70	10.10
10 Science Prototypes	1.85	2.45	2.35	2.80	3.00	12.45
11 Outreach and Education	0.20	1.00	1.00	1.00	1.00	4.20
	17.25	24.65	18.65	16.25	15.35	92.15
In-Kind Contributions	5.20	6.50	6.40	5.75	5.70	29.55

Milestones

- Nov 2001 – Jan 2002: Established project structure

- May 2002: Defined initial science demos

- June 13, 2002: Formed International VO Alliance

- Nov 15, 2002: Internal testing of science demos

- January 2003: Initial science demonstrations (AAS)

- August 2003: Intermediate NVO science demos (IAU)

Reporting and Communication

- Formal Quarterly and Annual Reports to NSF; copied to NASA

- Informal monthly reports to project manager

- Biweekly project status telecons with level-two WBS leaders

- Weekly Executive Committee telecons

- Weekly or biweekly working group telecons (Metadata, Systems, Science)

- Archived e-mail exploders for all working groups and management discussions

NVO: How Will It Work?

- Define commonly used small services
- Build higher level toolboxes/portals on top
- Do not build `everything for everybody'
- Use the "90-10" rule:
  - Define the standards and interfaces
  - Build the framework
  - Build the 10% of services that are used by 90% of the community
  - Let the users build the rest from the components

Development Approach

- First year: emphasize prototyping and experimentation, leading to real demos but not necessarily production-level software or system

  - Many IT tools now available; extensive evaluation through prototypes necessary to refine choices
  - Set up framework for more formal software management (baseline, test, revision control) for a distributed development effort in year 2

- NSF ITR project is not expected to define and "deliver" the entire NVO

Critical Issues

- Science demonstrations
  - Identified, scoped and scheduled

- Service registry issues
  - Needs international coordination (Garching)

- User interface issues
  - Need to retrofit existing portals

- EPO requirements
  - Impact on metadata standards

Role of Science Prototypes

- Keep focus on user- and science needs

- Identify most common services

- Verify standardization efforts

- Encourage data providers to participate

- Demonstrate to community that NVO tools will
  - arrive soon
  - will be useful for everybody
  - can evolve incrementally

- First science demos planned for January 2003

Initial Science Prototypes

- Brown-Dwarf search
  - Distributed query across several archives
  - Correlations with non-detections
  - Example of typical NVO search

- Gamma-Ray burst
  - Event follow up service
  - Exercise in standards compliance/interoperabilty

- Galaxy evolution in clusters
  - On-the-fly image analysis and pattern recognition
  - Exercise in grid computing

International Collaboration

- European initiatives underway
  - Astrophysical Virtual Observatory funded by European Commission (€3.3 million, three years)
  - AstroGrid, funded by UK e-science program (£5 million, three years)
- Other international efforts starting:
  - Canada (C$4M recently approved), India, Japan, Chile, Germany, Russia, Australia
- International VO roadmap announced at Garching VO conference, 10 June 2002
- International VO Alliance formed, 13 June 2002
- Regular telecons among NVO, AVO, and AstroGrid leadership
- Frequent technical contacts among partners
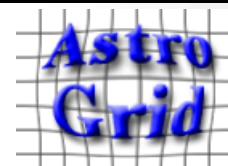
IVOA Participants

- **AVO**
  - P. Quinn (co-chair)
  - B. Pirenne
  - K. Gorski
  - F. Genova
  - P. Benvenuti
- **AstroGrid**
  - A. Lawrence
  - N. Walton (sec'y)
  - T. Linde
- **Russian VO**
  - O. Malkov
  - V. Vitkovskij
- **Canadian VO**
  - David Schade

- **NVO**
  - A. Szalay
  - R. Williams (tech coord.)
  - R. Hanisch (chair)
  - R. Moore
  - D. De Young
  - G. Helou
  - E. Schreier
  - G. Djorgovski
- **Australia**
  - R. Norris
- **German AVO**
  - W. Voges
- **India VO**
  - Ajit Khembavi

International Standards

- Active collaboration among NVO, AVO, and AstroGrid on VOTable

  - V1.0 released on April 15

  - Basis for testing metadata models, exchange protocols, encoding mechanisms

- Continued development of FITS standard

  - World Coordinate System definitions

    - Framework definition

    - Celestial coordinates

    - Spectral dispersion relations

    - Distortion functions

    - Time

Summary

- NSF ITR NVO project is one of four major and numerous other small VO-related initiatives now underway world-wide

- NVO is adopting, adapting, or developing necessary technology as derived from science requirements

- Project management approach seems to be working based on the first six months experience

- NVO project is dealing with many of the management challenges that will face the ultimate VO organization

http://us-vo.org