



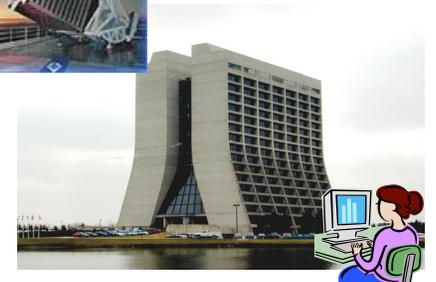
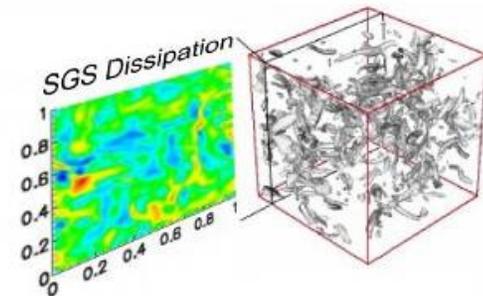
SkyServer: Understanding the Universe

Alex Szalay & Jordan Raddick, JHU
Jim Gray, Microsoft Research



The Evolution of Science

- Thousand years ago:
science was empirical
describing natural phenomena
- Last few hundred years:
theoretical branch
using models, generalizations
- Last few decades:
a computational branch
simulating complex phenomena
- Today:
data exploration (eScience)
synthesizing theory, experiment and computation with advanced data management and statistics



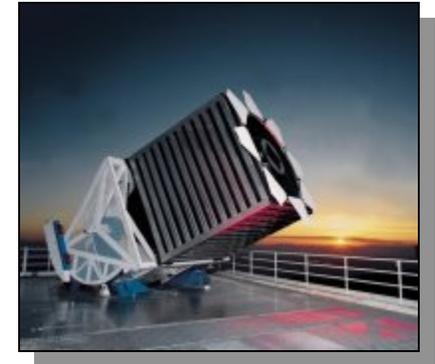
Why Is Astronomy Special?

- Especially attractive for the wide public
- It has no commercial value
 - *No privacy concerns, freely share results with others*
 - *Great for experimenting with algorithms*
- It is real and well documented
 - *High-dimensional (with confidence intervals)*
 - *Spatial, temporal*
- Diverse and distributed
 - *Many different instruments from many different places and many different times*
- The questions are interesting
- There is a lot of it (soon petabytes)



Background

- The Sloan Digital Sky Survey (SDSS)
The “Cosmic Genome Project”
 - *5 color images of _ of the sky*
 - *Pictures of 300 million celestial objects*
 - *Distances to the closest 1 million galaxies*
- JHU: build the public archive for the SDSS
- Lots of debate who the archive is for
 - *“power users”*
 - *“astronomers”*
 - *“students and amateurs”*
 - *“wide public”*
- Interesting challenge in digital publishing
 - *We have to publish first in order to analyze*



What is SkyServer?

- SkyServer is an educational website
- Access to an underlying multiterabyte database
- More than 50 hours of educational exercises
- Background on astronomy
- Tutorials and documentation
- Searchable web pages
- Interactive visual tools for data exploration
- Prototype eScience lab

<http://skyserver.sdss.org/>



History

- Database
 - *Started with Objectivity/DB*
 - *Switched to MS SQL Server (Szalay, Gray, Kunszt) as a “hobby”, using the MS Terraserver (Gray,Barclay) as a basis*
- Web Interface
 - *Started as an experiment, built in one week, to showcase at an Intel meeting for higher education in SF (Jan 2001)*
 - *Goal: integrate pixel space with catalog space (Terraserver)*
 - *Initial test pages by a JHU undergraduate (Blair Lanier)*
 - *Color images built by Szalay, using custom code and Photoshop*

EDR: Early Data Release

- SDSS Early Data Release (June 6, 2001)
- 100 GB catalogs, few hundred square degrees
- SkyServer aimed solely at public outreach
- Built in 2 weeks by Szalay and Gray (20 hour days)
- Web site design by Szalay
- Content writing done by Stephen Landy
- Hardware donated by Compaq
- Highly interactive, using browser independent DHTML (“browser hell”)

SkyServer Goals

- Provide easy, visual access to exciting new data
 - *“hot off the press”*
- Illustrate that advanced content does not mean a cumbersome interface
- Understand new ways of publishing scientific data
- Target audience
 - *Advanced high-school students*
 - *amateur astronomers*
 - *wide public*
- Multilingual capabilities built in from the start
 - *Heavy use of stylesheets, language branches*

Student Projects

- Developed by Jordan Raddick over the first year
 - *Science writer with good physics background*
 - *Collaborating with high school teachers (Rob Sparks @Fermilab)*
- Various exercise levels
 - *Advanced, Basic, Quiz*
- Teacher and student registration
 - *Teacher guides developed*
- Rosa Gonzales (Mexico) used in a summer program
 - *Developed an exercise in Spanish, then translated to English*

Tutorials and Guides

- Developed by Jordan and Postdocs
 - *How to use Excel*
 - *How to use a database (guide to SQL)*
 - *Expert advice on SQL*
- Automated on-line documentation
 - *Ani Thakar, Roy Gal*
 - *Database information, Glossary, Algorithms*
 - *Searchable Help*
 - *All stored in the DB, and generated on the fly*

DR1: Data Release 1

- The first main data release of SDSS (May 2003)
- 1.1TB of catalogs, linked to 6TB of low level data
- SkyServer has undergone a major facelift
 - *New graphic design by Curtis Wong, Asta Roseway (MS)*
 - *Modified stylesheets and embedded scripts only*
 - *Web site translated in 2 days*
- New visual tools using Web Services
 - *Szalay, Gray, Maria Nieto-SantiSteban*
- API's published
- Formal helpdesk in place
- Created MySkyServer
 - *0.65GB laptop version*



DR2: Data Release 2

- Live in March 15, 2004, with 2.2 TB of catalogs
- Only incremental changes in interface
- Web site under source control
- Color images dramatically improved
- New translations under way
 - *Japanese, French, German, Spanish, Hungarian*
- Tools overhauled
 - *now embraced by professional astronomers*
- Enormously increased traffic
- Moving to 3-way web front end + 3 DB servers
- Collaborative tools: MyDB with group access

Metrics

- All traffic logged
 - *Web logs, SQL logs, Web service logs*
 - *Will reach 40M hits next few days*
- Clearly stated privacy policy
 - *Aggregate traffic visible to every one*
 - *IP addresses queryable*
 - *All registered user information is secure*
 - *Contents of SQL queries logged, but not visible*
 - *Used for optimizing DB design*
- Several research projects by CS groups
 - *Workflow analysis (optimal caching, virtual data expt)*

Support

- Immense amount of volunteer time
- SDSS Project provided the DB development effort
- Fermilab did data processing, supported Rob Sparks
- Close links to iVDGL/GriPhyN
 - *Raddick works closely with M. Campanelli*
- Direct financial contributions:

– <i>NSF SGER</i>	\$ 50K
– <i>NASA IDEAS grant</i>	\$ 50K
– <i>NASA MD Space grant Consortium</i>	\$ 5K
– <i>Howard Hughes Inst.</i>	\$ 20K
– <i>Microsoft Research</i>	\$ 100K
– <i>Compaq/HP</i>	\$ 50K
– <i>Intel</i>	\$ 100K



Collaborations

- Gary Greenberg (Northwestern U)
 - *Collaboratory => recent joint NSF grant*
- Carl Pennypacker (Berkeley)
 - *Hands On Universe -- link to on-line student telescopes*
 - *Astrophysics Journal for High School Students*
- Western Maryland College
 - *New modules developed*
- National Virtual Observatory
 - *Standard Web Service interfaces for discovering and delivering educational content*
- Lots of volunteer help from teachers
 - *Feedack about content level*

Visibility

- Presentations and demo booths at various meetings
 - *NSF Astronomy Open Day*
 - *AAS meetings*
 - *Physics teacher conferences*
- Summer courses (R.Kron @ Yerkes Observatory)
- Papers published
 - *Physics Teacher, Sky and Telescope*
 - *Various conference proceedings, posters*
- Slashdot, Microsoft Research
- 10 mirror sites over the world (“sneakernet”)
 - *Canada, Japan, Germany, India, China, Australia, UK, Hungary*

Summary

- Amazing journey, happened gradually
- Very enthusiastic welcome by teachers and public
- Now embraced by professional astronomers
- Clear that we need lot of help from educators
- Like a “folk tale” of the 21st century:
many individuals and institutions added their contributions, and the whole is much more than the sum of the parts
- Expect real discoveries to be made by high-school students