

Real Web Services

Jim Gray

Microsoft Research

455 Market St, SF, CA, 94105

Gray@Microsoft.com

<http://Research.Microsoft.com/~Gray>

Talk at

OSI Soft, Wednesday, 14 May 2003

Slides at

<http://Research.Microsoft.com/~Gray/talks>

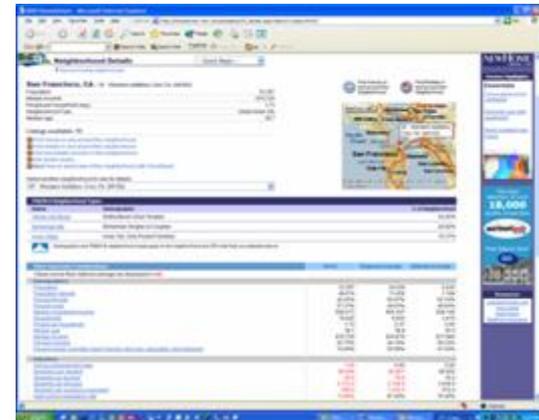
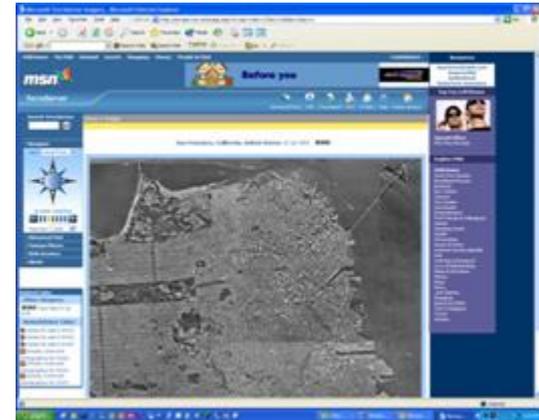
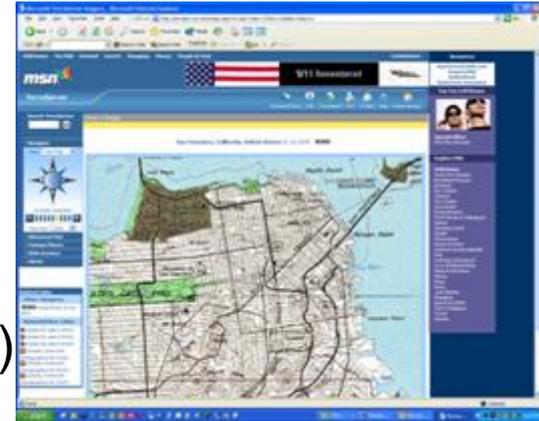
Outline

- **Theme:**
 - You can now easily publish data**
 - So, the issue now is what to publish?**
 - What do you want to share with others?**
- TerraServer becomes a WebService
- Astronomy: Data Mining Web Services
- What possible relevance to you?

TerraServer

TerraService.net

- A photo of the United States
 - 1 meter resolution (photographic/topographic)
 - USGS data
 - Some demographic data (BestPlaces.net)
 - Home sales data
 - Linked to Encarta Encyclopedia
- 15 TB raw, 6 TB cooked (grows 10GB/w)
- Point, Pan, zoom interface
- Among top 1,000 websites
 - 40k visitors/day
 - 4M queries/day
 - 3 B page views (in 5 years)
- All in an SQL database



TerraServer Becomes a Web Service

TerraServer.net -> TerraService.Net

- Web server is for people.
- Web Service is for programs
 - The end of screen scraping
 - No faking a URL:
pass real parameters.
 - No parsing the answer:
data formatted into your
address space.
- Hundreds of users but a
specific example:
 - US Department of Agriculture



Before you buy.

TerraServer

Advanced Find | Famous Places | Web Services | About | Home Advisor

GetAreaFromPt

```
public AreaBoundingBox GetAreaFromPt(LonLatPt center, Theme theme, Scale scale, int displayPxWidth, int displayPxHeight)
```

The `GetAreaFromPt` method returns the tile meta-data for a Geographic rectangle. Use this call to identify the tiles required to construct an image of a specific size and resolution with a known center point. The `GetAreaFromPt` is typically called in applications that want to control the size of the display area. Using `GetAreaFromPt`, the caller controls the specific Longitude and Latitude point to be displayed in the center. The **LonLatPt** point parameter identifies the Geographic center of the rectangle of interest. The **Theme** and **Scale** input parameters identify the type imagery and resolution of interest. The `displayPxWidth` and `displayPxHeight` parameters identify the size of image the caller intends to create. The `GetAreaFromPt` method computes the TerraServer tiles that overlap the corners of your image area in the resolution specified in the **Scale** parameter.

Sample Call:

```
TerraService ts = new TerraService();
LonLatPt center;
center.Lon = -122.5;
center.Lat = 37.5;
try {
    AreaBoundingBox abb = ts.GetAreaFromPt(center, Theme.Photo, Scale.Scale16m,
}
catch { // Catch exceptions thrown by GetAreaFromPt, e.g. bad theme, bad
```

And now.. 6 slides from the “customer” who built a portal using TerraService

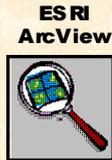


Business Applications Need Data

Vision: *One Stop Shopping to Data Anywhere, Anytime, Anyplace*

Strategic Business Applications

Public Access to Service Center Data



Customer Service Toolkit



Web Based Application



Services



- One stop Shopping
- Site Location
- Data Selection
- Data Extraction (*cookie cutting*) for vector, raster, and tabular
- Component Architecture

- Data Formatting including reprojection and Mr. Sid compression
- Data Packaging
- Data Delivery including FTP, CD, and immediate download
- Public and Internal Security

- Standards Enforcement
- Automated Retrieval under program control
- Compatibility with FGDC and Open GIS Standards
- COTS or GOTS based
- Print Map

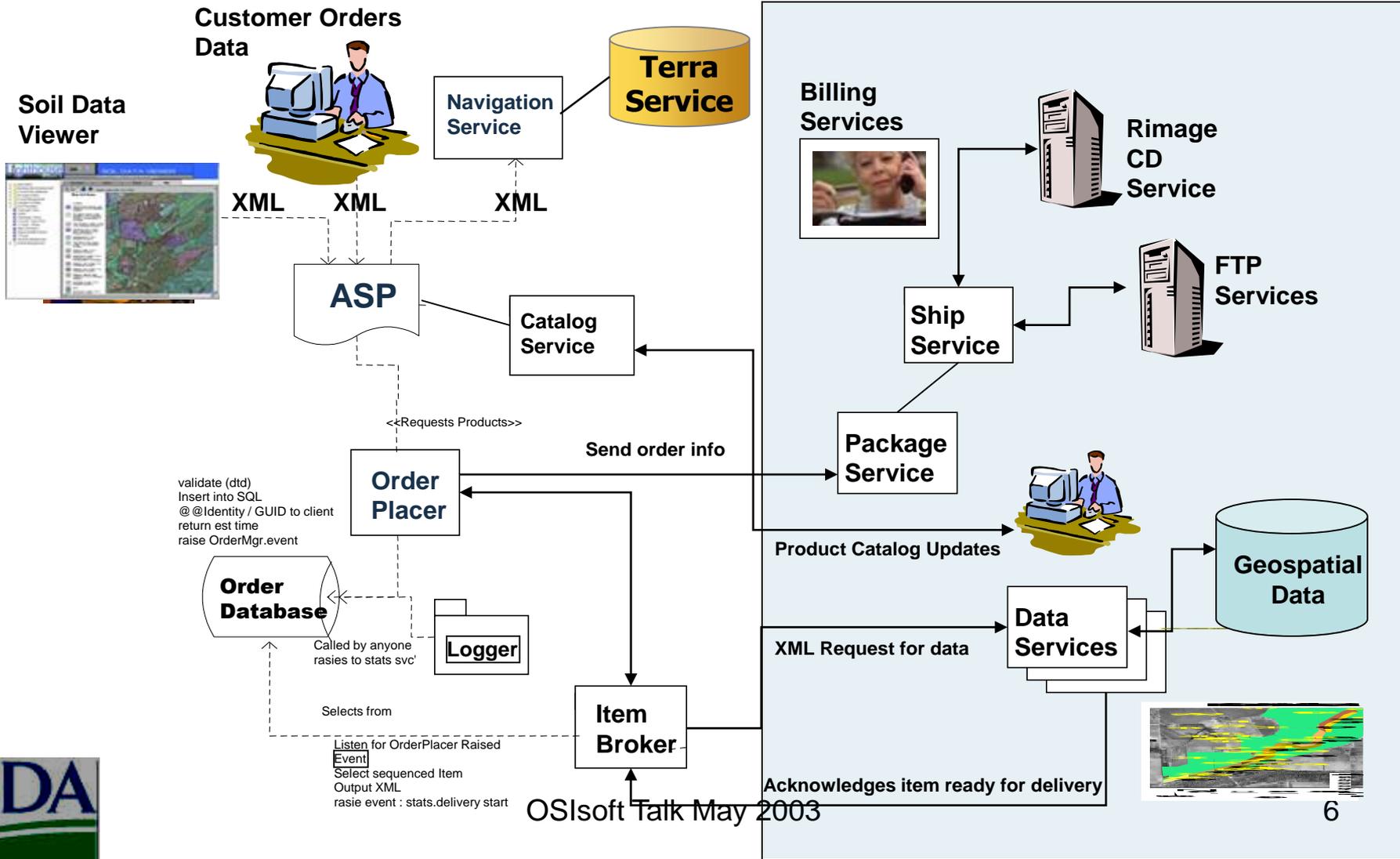
Data Marts & Warehouses



Data Gateway Functional Overview

ITC - Fort Collins, Colorado

NCGC - Fort Worth, Texas



Order Process

Lighthouse returns confirmation

The screenshot shows the 'Geospatial Data Gateway' website. The navigation bar includes 'Home', 'About', 'News', 'Help', and 'Contact'. The date is 'Apr 19 | Thu' and the text 'NRCS + FSA + RD' is visible. The main content area is titled 'Order Confirmation' and is divided into several sections:

- Data Format:**

Projection	Geographic (Lat/Long) NAD83
Extent	Standard
Compression	Zip
Image	
Vector	Shape File
- Shipping Info:**

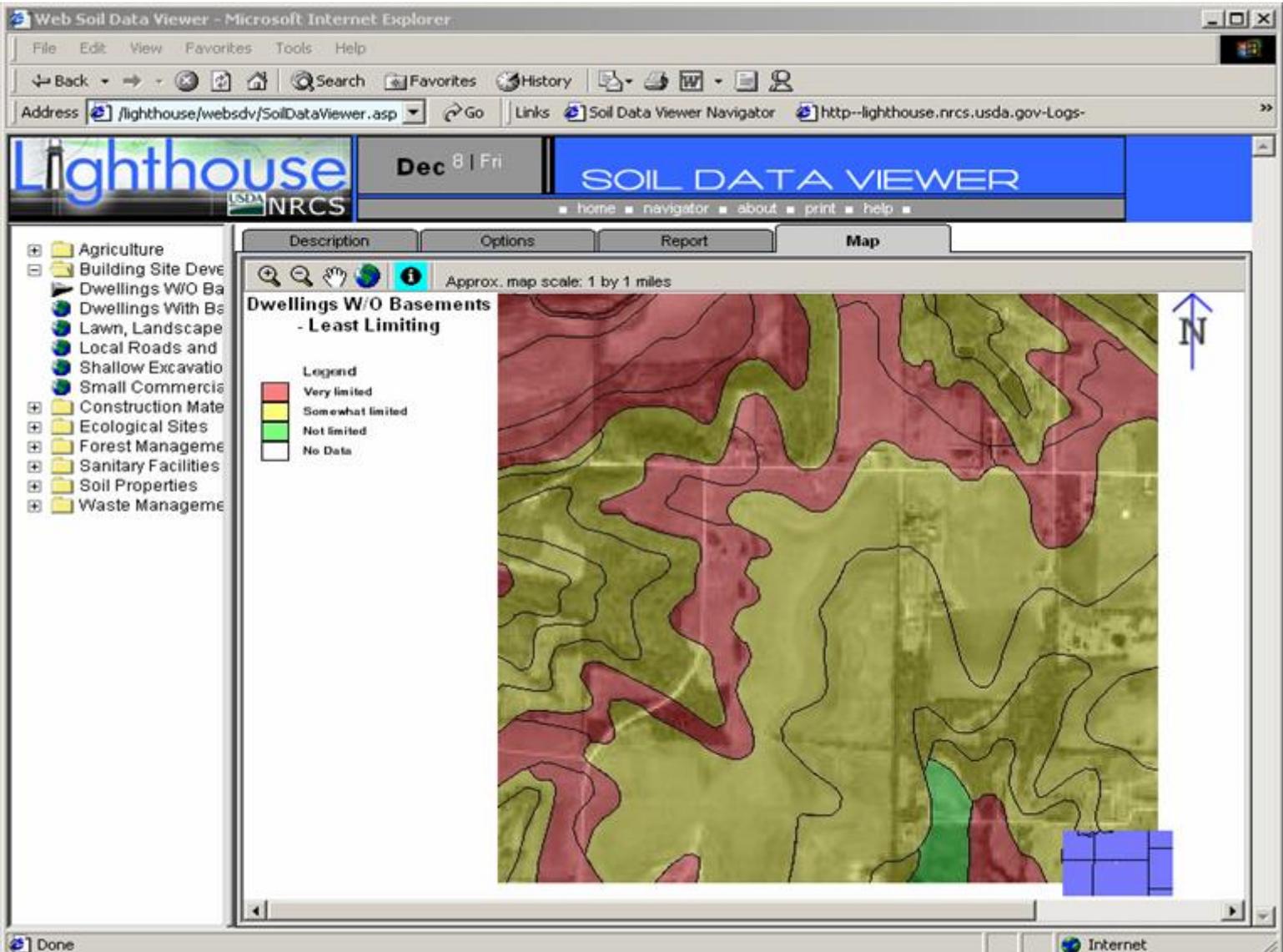
Name	Wendall Oaks		
Organization	USDA - NRCS		
Email	woaks@itc.nrcs.usda.gov		
Address	2525 Redwing Road		
City, ST, Zip	Fort Collins, Colorado 80526	Ship via	FTP
Phone	970 202-9900	Fax	970 202-9901
- Products Ordered:**

Item	Description
1	Common Land Units

On the left side, a yellow box with 'S5' and 'Step 5' is displayed. Below it, the 'Instructions' section reads: 'Recheck current information. If any errors are present you may return to any previous step to correct. If the information is correct, the order may be submitted for processing.' At the bottom left, there is a 'Place Order' button.

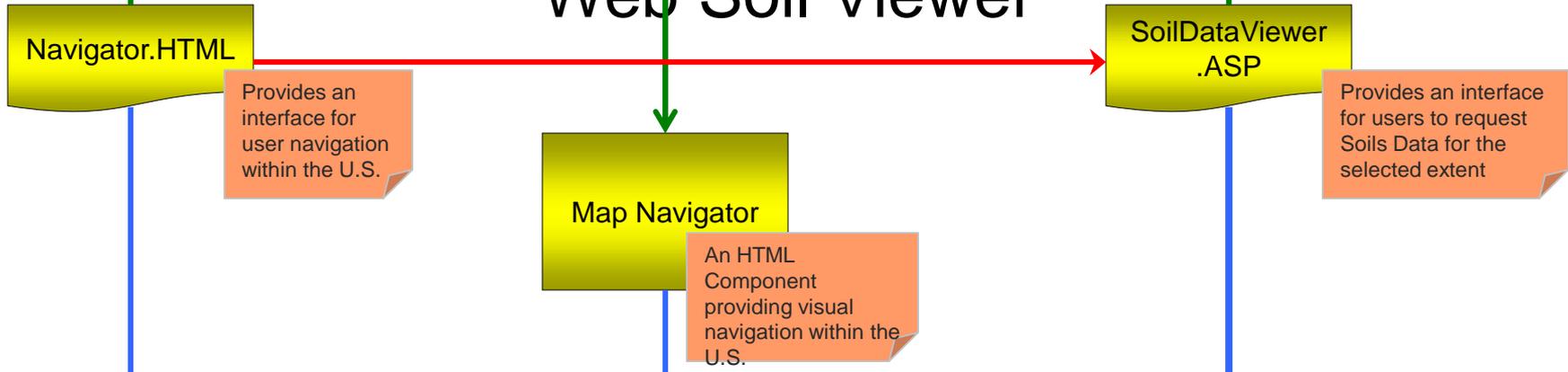
Custom End Product

Soil Interpretation Map

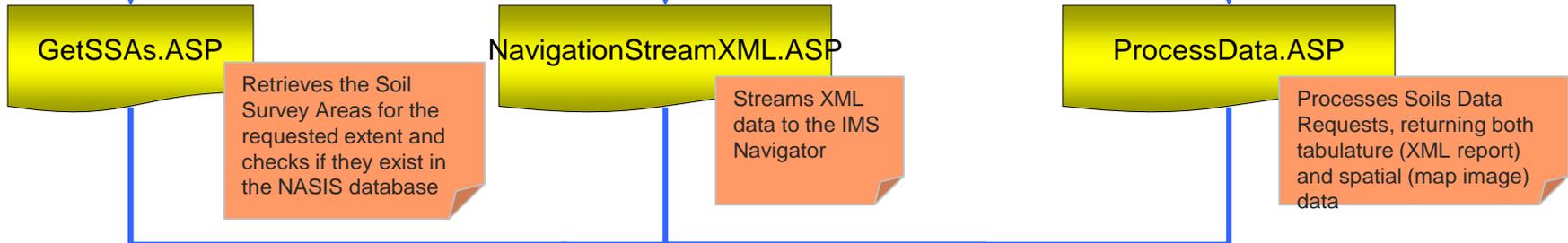


Client - Internet Explorer

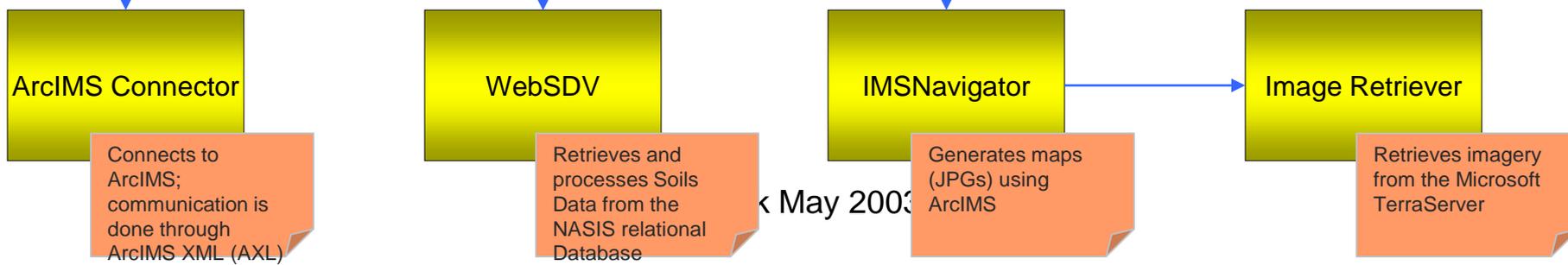
Web Soil Viewer



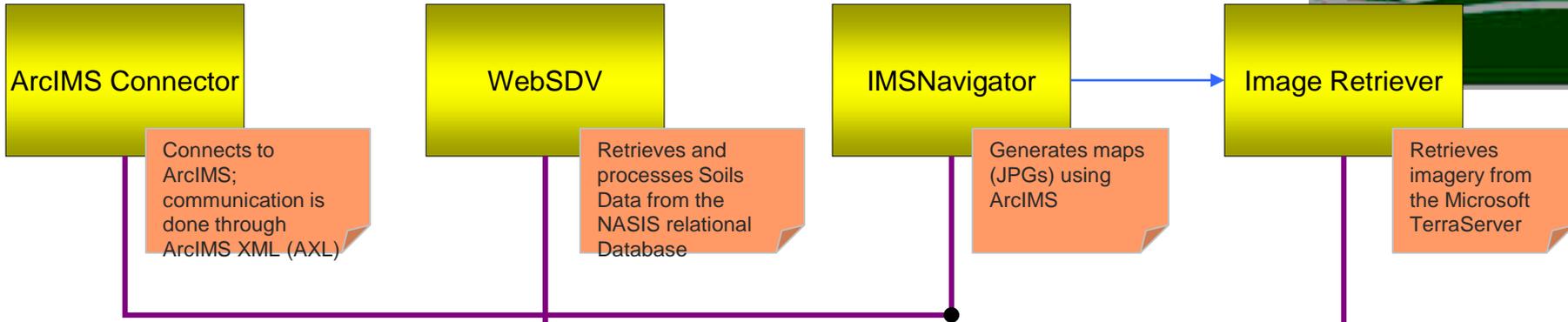
Web Server - Internet Information Services



Web Server - COM+ Applications



Web Server - COM+ Applications



Database Server - ESRI Spatial Data Server



Database Server - Microsoft SQL Server



Microsoft Terraserver



Brief tour of TerraService

Outline

- Theme:
 - You can now easily publish data
 - So, the issue now is what to publish?
 - What do you want to share with others?
- TerraServer becomes a WebService
- **Astronomy: Data Mining Web Services**
- What possible relevance to you?

SkyServer

SkyServer.SDSS.org

- Like the TerraServer, but looking the other way: a picture of $\frac{1}{4}$ of the universe
- Pixels + Data Mining
- Astronomers get about 400 attributes for each “object”
- Get Spectrograms for 1% of the objects

SkyServer Object Explorer - Microsoft Internet Explorer

SkyServer Object Explorer

SDSS

Search by
ObjId
Ra,dec
5-part SDSS
Plate-Fiber
SpecObjId

Summary

PhotoObj
Field
Frame
PhotoZ
Neighbors
Navigate
FITS

SpecObj
SpecLine
SpecLineIndex
XCredShift
ELredShift
Spectrum
Plate
FITS

NED search
Virtual Sky

Save in Notes
Show Notes

Print Page

SDSS J121755.52+002623.87

GALAXY ra=184.481364, dec=0.4399658, ObjId = 2255030989160697

status	TARGET PRIMARY OK_STRIPE OK_SCANLINE PSEGMENT RESOLVED OK_RUN GOOD SET
flags	BINNED1 SATURATED INTERP COSMIC_RAY NOPETRO NODEBLEND CHILD BLENDED
PrimTarget	TARGET_GALAXY_BIG TARGET_GALAXY
SecTarget	



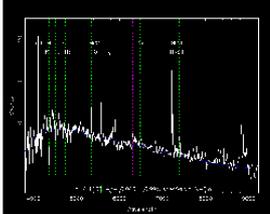
run	752
rerun	8
camcol	5
field	273
obj	249
rowc	1128.2
colc	282.6
parentid	2255030989160695
nchild	0

u	g	r	i	z	reddening_r	petroRad_r
17.57	15.88	15.52	15.21	15.43	0.07	25.108

fiberMag_r	petroMag_r	devMag_r	expMag_r	psfMag_r	modelMag_r
19.61	16.21	15.59	17.27	19.60	15.59

SpecObjId= 81006758046203904

plate	mjd	fiberid	z	zErr	zConf	specClass	ra	dec	fiberMag_r	objId
287	52023	631	0.100	0.00006	9.93E-1	GALAXY	184.48137	0.43999	19.57	2255030989160697



zStatus	XCORR_EMLINE
zWarning	
PrimTarget	TARGET_GALAXY
SecTarget	
eClass	2.293E-1
emZ	0.100
emConf	6.468E-1
xcZ	0.100
xcConf	9.93E-1

Cross-identifications

OSIsoft Talk Ma

Why Astronomy Data?

- **It has no commercial value**

- No privacy concerns
- Can freely share results with others
- Great for experimenting with algorithms

- **It is real and well documented**

- **High-dimensional data** (with confidence intervals)
- **Spatial data**
- **Temporal data**

- **Many different instruments from many different places and many different times**

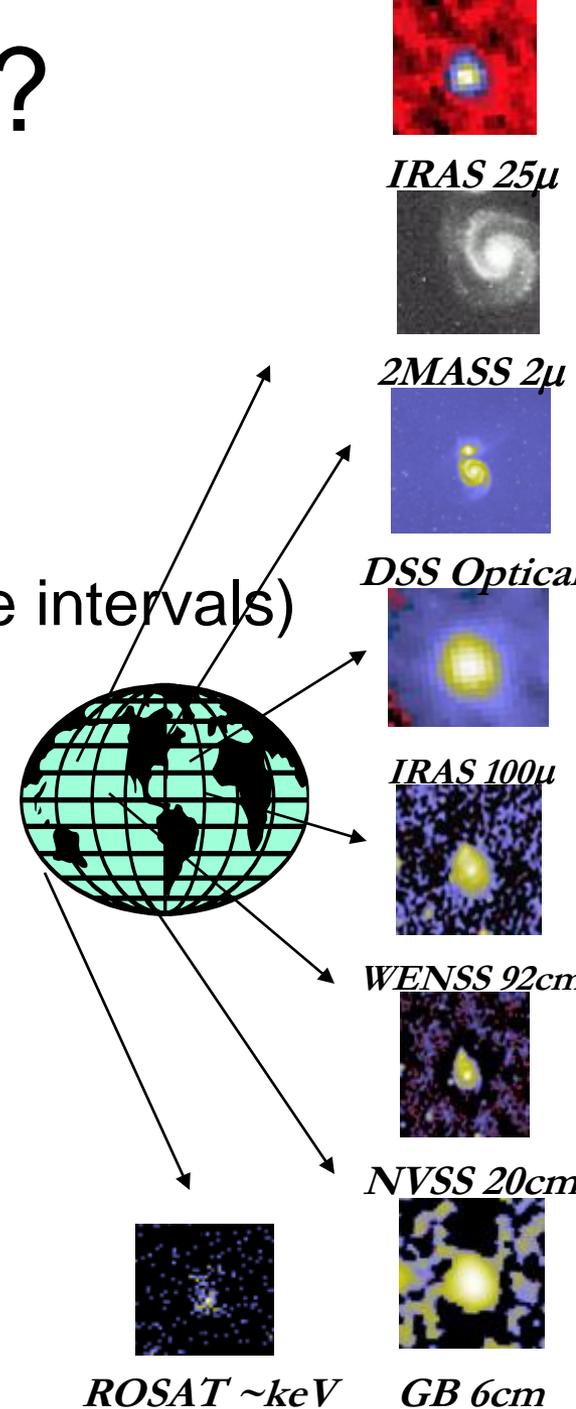
- **Federation is a goal**

- **The questions are interesting**

- How did the universe form?

- **There is a lot of it (petabytes)**

OSIsoft Talk May 2003



Demo of SkyServer

- Shows standard web server
- Pixel/image data
- Point and click
- Explore one object
- Explore sets of objects (data mining)

Virtual Observatory

<http://www.astro.caltech.edu/nvoconf/>

<http://www.voforum.org/>

Premise: Most data is (or could be online)

So, the Internet is the world's best telescope:

- It has data on every part of the sky
- In every measured spectral band: optical, x-ray, radio..
- As deep as the best instruments (2 years ago).

– It is up when you are up.

The “seeing” is always great

(no working at night, no clouds no moons no..).

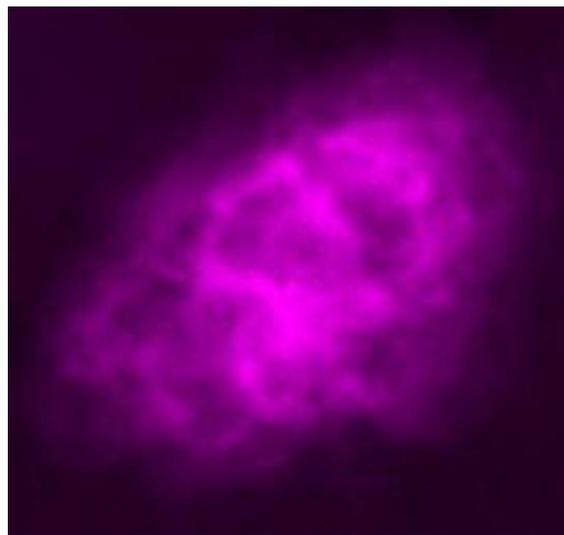
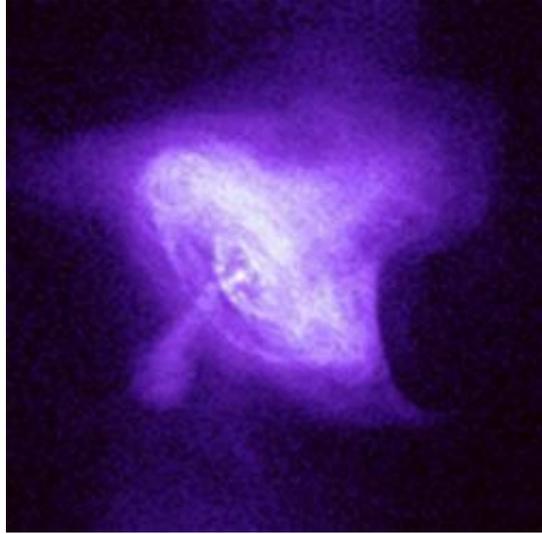
– It's a smart telescope:

links objects and data to literature on them.



Time and Spectral Dimensions

The Multiwavelength Crab Nebulae



X-ray,
optical,
infrared, and
radio

views of the nearby
Crab Nebula, which is
now in a state of
chaotic expansion after
a supernova explosion
first sighted in 1054
A.D. by Chinese
Astronomers.

Data Federations of Web Services

- Massive datasets live near their owners:
 - Near the instrument's software pipeline
 - Near the applications
 - Near data knowledge and curation
 - Super Computer centers become Super Data Centers
- Each Archive publishes a web service
 - Schema: documents the data
 - Methods on objects (queries)
- Scientists get “personalized” extracts
- Uniform access to multiple Archives **Federation**
 - A common global schema

Grid and Web Services Synergy

- I believe the Grid will be many web services share data (computrons are free)
- IETF standards Provide
 - Naming
 - Authorization / Security / Privacy
 - Distributed Objects
 - Discovery, Definition, Invocation, Object Model
 - Higher level services: workflow, transactions, DB,...
- Synergy: commercial Internet & Grid



Web Services: The Key?

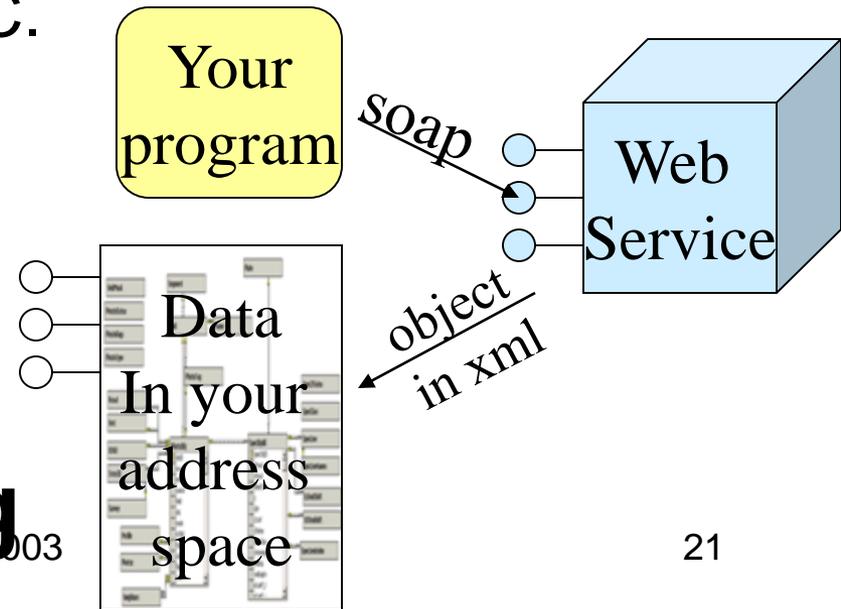
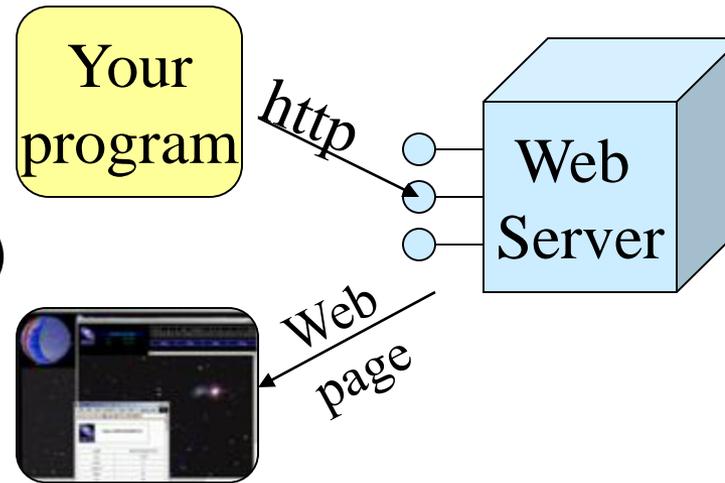
- **Web SERVER:**

- Given a url + parameters
- Returns a web page (often dynamic)

- **Web SERVICE:**

- Given a XML document (soap msg)
- Returns an XML document
- Tools make this look like an RPC.
 - $F(x,y,z)$ returns (u, v, w)
- Distributed objects for the web.
- + naming, discovery, security,...

- **Internet-scale distributed computing**



SkyQuery: a prototype

- **Defining Astronomy Objects and Methods.**
- **Federated 3 Web Services** (fermilab/sdss, jhu/first, Cal Tech/dposs)
multi-survey cross-match
Distributed query optimization (T. Malik, T. Budavari, Alex Szalay @ JHU)

<http://SkyQuery.net/>

- **My first web service** (cutout + annotated SDSS images) **online**
 - <http://skyservice.pha.jhu.edu/devel/ImgCutout/chart.asp>
- **WWT is a great Web Services (.Net) application**
 - Federating heterogeneous data sources.
 - Cooperating organizations
 - An Information At Your Fingertips challenge.

SkyNode Basic Web Services

- Metadata information about resources
 - Waveband
 - Sky coverage
 - Translation of names to universal dictionary (UCD)
- Simple search patterns on the resources
 - Cone Search
 - Image mosaic
 - Unit conversions
- Simple filtering, counting, histogramming
- On-the-fly recalibrations

Portals: Higher Level Services

- Built on Atomic Services
- Perform more complex tasks
- Examples
 - Automated resource discovery
 - Cross-identifications
 - Photometric redshifts
 - Outlier detections
 - Visualization facilities
- Goal:
 - Build custom portals in days from existing building blocks (like today in IRAF or IDL)

Demo of Image Cutout Service

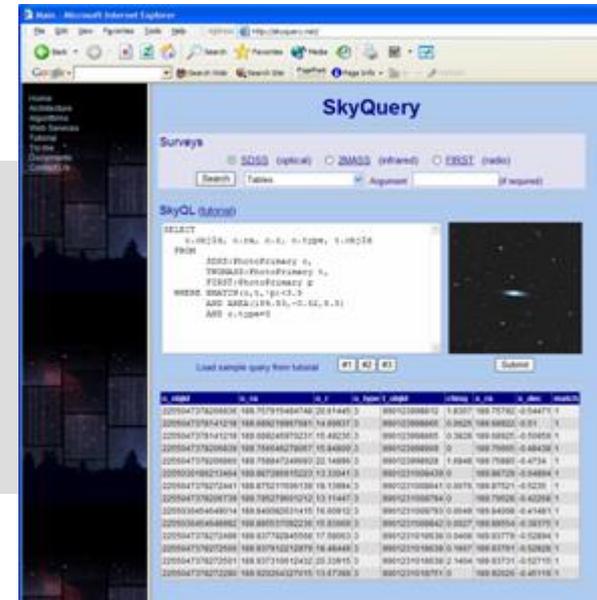
- Shows image cutout
- Show project and debugging project
- Show hello World
- Show “theAnswer” method

SkyQuery (<http://skyquery.net/>)

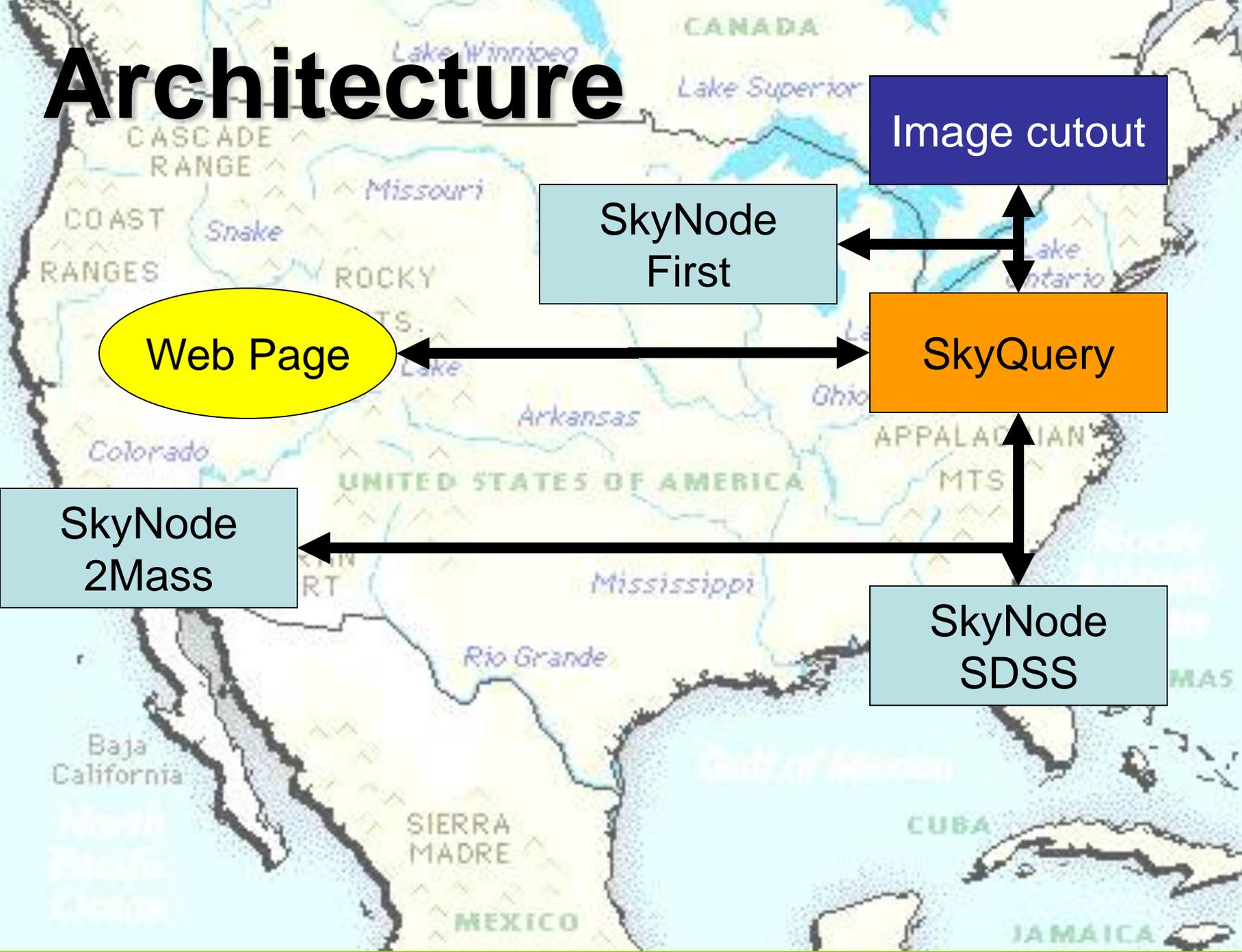
- Distributed Query tool using a set of services
- Feasibility study, built in 6 weeks from scratch
 - Tanu Malik (JHU CS grad student)
 - Tamas Budavari (JHU astro postdoc)
- Implemented in C# and .NET
- Allows queries like:

```
SELECT o.objId, o.r, o.type, t.objId
FROM SDSS:PhotoPrimary o,
      TWOMASS:PhotoPrimary t
WHERE XMATCH(o,t)<3.5
      AND AREA(181.3,-0.76,6.5)
      AND o.type=3 and (o.I - t.m_j)>2
```

OSIssoft Talk May 2003



Architecture



Demo of SkyQuey

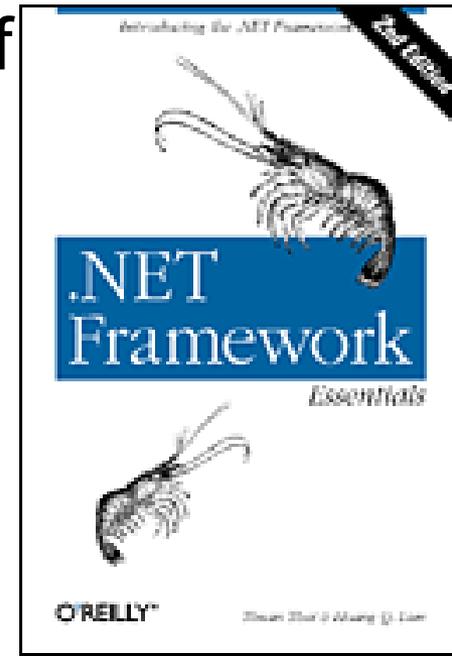
- Show a few queries just to prove it works.

Outline

- Theme:
 - You can now easily publish data
 - So, the issue now is what to publish?
 - What do you want to share with others?
- TerraServer becomes a WebService
- Astronomy: Data Mining Web Services
- **What possible relevance to you?**

Summary So Far

- Some real web services deployed today
- Easy to build & deploy
- Services publish data, Portals unify it
- Tools really work!
- I'm using C# and foundation classes of VisualStudio, a great! Tool
- A nice book explaining the ideas:
[\(.Net Framework Essentials, Thai, Lam isbn 0-596-00302-1\)](#)



Possible Relevance to You

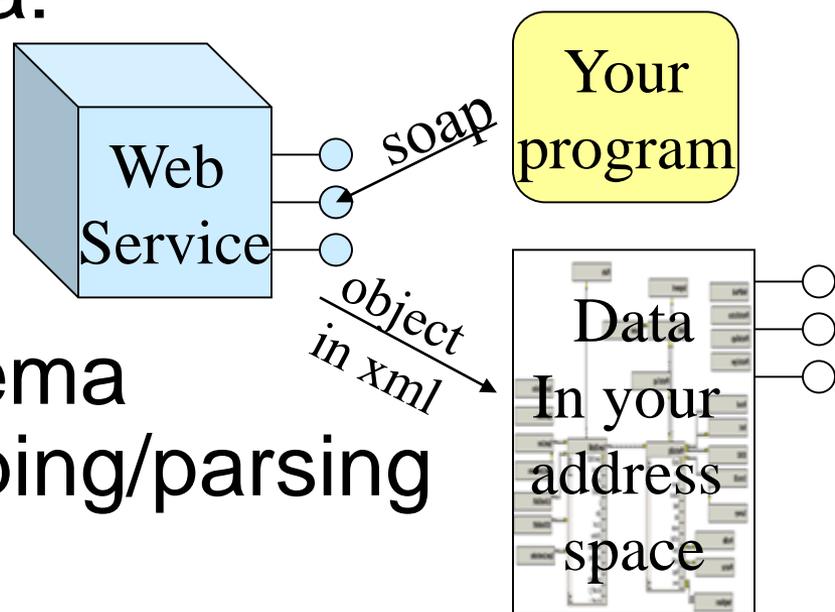
- This web service stuff is REAL

- If you have a class,
It is a way to publish data:

Internet
Intranet

- It is a way to find data
data comes with schema
no more screen scraping/parsing

- Business model unclear
– Your ideas go here.



Outline

- Theme:
 - You can now easily publish data
 - So, the issue now is what to publish?
 - What do you want to share with others?
- TerraServer becomes a WebService
- Astronomy: Data Mining Web Services
- **What possible relevance to you?**