

# Simplifying the Web User's Interface to Massive Data Sets

Roberta Allsman

The Australian National University

Supercomputer Facility

[robyn.allsman@anu.edu.au](mailto:robyn.allsman@anu.edu.au)

# Mass Storage Environment

## Storage Tek

1 – Powderhorn 9310 Automated Cartridge system

300 TB capacity

8 – Timberline 9840 tape drives

20 GB/tape

2 – Redwood tape drives

50 GB/tape

LSCI Storage and Archive Manager File System

SAM-FS

# Overview

- Motivation
- User Community
- Requirements
- Process Overview
- Security Model
- Future Directions

# Motivation

Help data store users  
with the largest archives  
to access their data effectively  
without  
taking on their role of project data manager.

# MACHO Project

## Search for Dark Matter in the Galactic Bulge

- 7 TB image data, 70 MB/image
- 1 TB photometry data
- Non-proprietary

## Purpose

- Make data available to the worldwide astronomy community

# Wide Field Imager (WFI) Project

Observatory management of data taken using the MSSSO  
WFI device

- Over 10 TB during 10 year device life; 100 MB/image
- Mixed proprietary and non-proprietary data with continuous roll into non-proprietary status

Purpose

- Provide proprietary data to project via web
- Provide non-proprietary data to worldwide community

# Client Requirement: A Non-Invasive Web Site

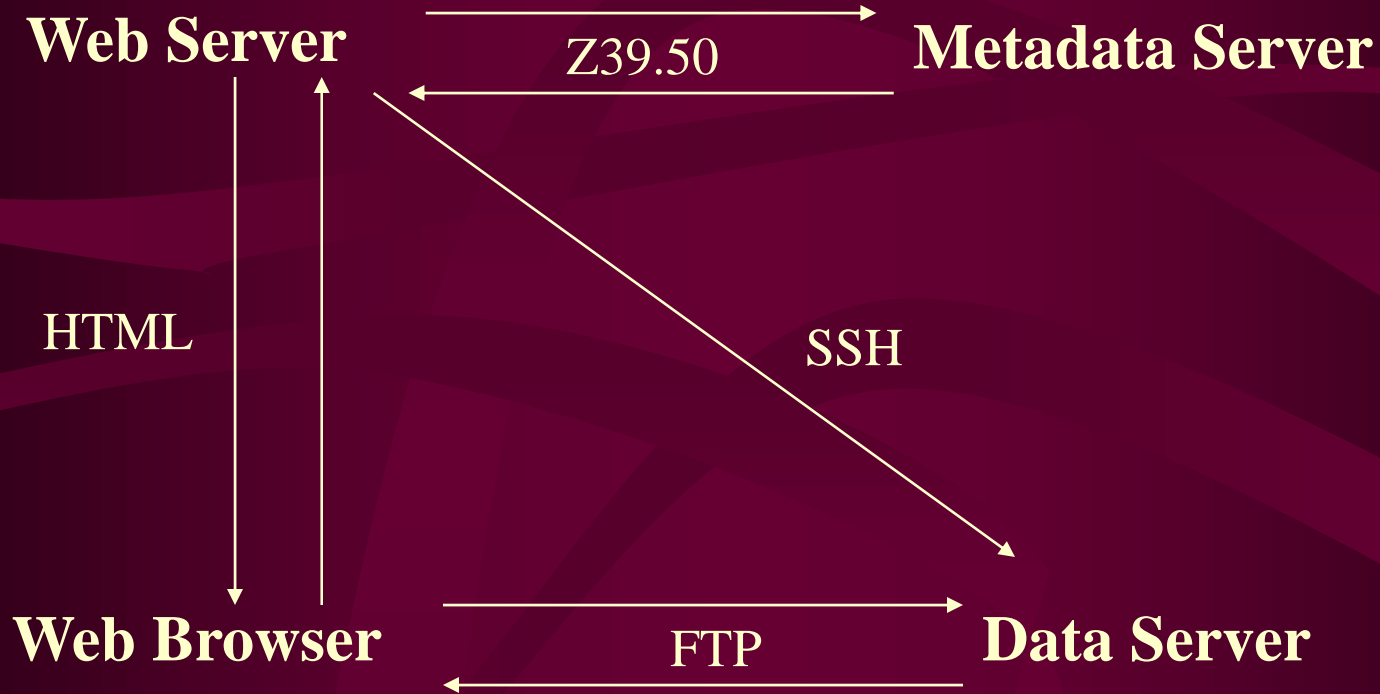
- No need to alter web browser configuration
- No account required for access to non-proprietary data
- No account required on user's local system for delivery of data
- Proprietary data restricted to designated project members

# Additional Considerations

- Data delivery efficiency since data being transferred internationally
- Security implications of intermixed proprietary and non-proprietary data
- Simple system which the data owner maintains
- Minimal software costs



# Process Overview



# Division of Responsibility

## ANUSF

- Manages data server
- Provides single account for data management

## Project

- Manages web server and metadata (MD) server
- Maintains MD catalog and data within data store

## Web User

- Accesses data via his standard web browser
- Only needs web account if proprietary data wanted

# Search Form

## MACHO Project: Variable Star Catalog Retrieval

[Startup](#) | [Help](#) | [Error Recovery](#)

Click options below to expose or remove input fields.

[Classification](#)

[Variability Index](#)

Sample: Units [ HMS/DMS/arcmin ] Equinox [ J2000 ]  
RA [ 5:1:15.2 ] Dec [ -69:25:59.5 ] Search Radius [ 1 ]

[Location](#)

Sample: Units [ radians ] Equinox [ J2000 ]  
RA [ 1.31447 ] Dec [ -1.21183 ] Search Radius [ .0003 ]

Sample: Units [ degrees ] Equinox [ J2000 ]  
RA [ 75 ] Dec [ -69 ] Search Radius [ .1 ]

[Units](#)

[Equinox](#)

[RA](#)

[Dec](#)

[Search Radius](#)

[Average Magnitude](#)

[Average Amplitude](#)

[Average Period](#)

[Review and edit search query.](#)

Display at most  [records per page.](#)

# Search Result

## MACHO Project: Image Search Result

Your query was (Field=1) and (ObsDate>19930327) and (ObsDate<19930401).

**7 hits**

[Help](#) | [New Search](#) | [Previous](#) | [Next](#)

Unix batch retrieval script from selected images.

<u>Summary</u>		<u>Field</u>	<u>Domain</u>	<u>RA</u>	<u>Dec</u>	<u>ObsId</u>	<u>Date</u>	<u>Fetch</u>	<u>Batch</u>
<a href="#">Short</a>	<a href="#">Full</a>	1	LMC	5:5:21.3	-69:5:13	4596	19930328	<a href="#">Now</a>	<input type="checkbox"/>
<a href="#">Short</a>	<a href="#">Full</a>	1	LMC	5:5:21.3	-69:5:13	4619	19930328	<a href="#">Now</a>	<input type="checkbox"/>
<a href="#">Short</a>	<a href="#">Full</a>	1	LMC	5:5:21.3	-69:5:13	4634	19930329	<a href="#">Now</a>	<input type="checkbox"/>
<a href="#">Short</a>	<a href="#">Full</a>	1	LMC	5:5:21.3	-69:5:13	4658	19930329	<a href="#">Now</a>	<input type="checkbox"/>
<a href="#">Short</a>	<a href="#">Full</a>	1	LMC	5:5:21.3	-69:5:13	4690	19930330	<a href="#">Now</a>	<input type="checkbox"/>
<a href="#">Short</a>	<a href="#">Full</a>	1	LMC	5:5:21.3	-69:5:13	4716	19930330	<a href="#">Now</a>	<input type="checkbox"/>
<a href="#">Short</a>	<a href="#">Full</a>	1	LMC	5:5:21.3	-69:5:13	4764	19930331	<a href="#">Now</a>	<input type="checkbox"/>

# Authentication

occurs when the identity of the user is determined by virtue of userid/password validation.

The requester's process is then granted access to all data and processes available to the authenticated userid.

# Authorization

is the process whereby an authenticated user is determined to have legitimate access to a specific datum.

# Authentication: Web Accounts

- All users requiring access to proprietary data must acquire a web account from the Observatory.
- The web account may be used for authentication into multiple projects.
- The Project PI is authorized to update the list of authorized users of his project's data.

# Authentication: the authtoken

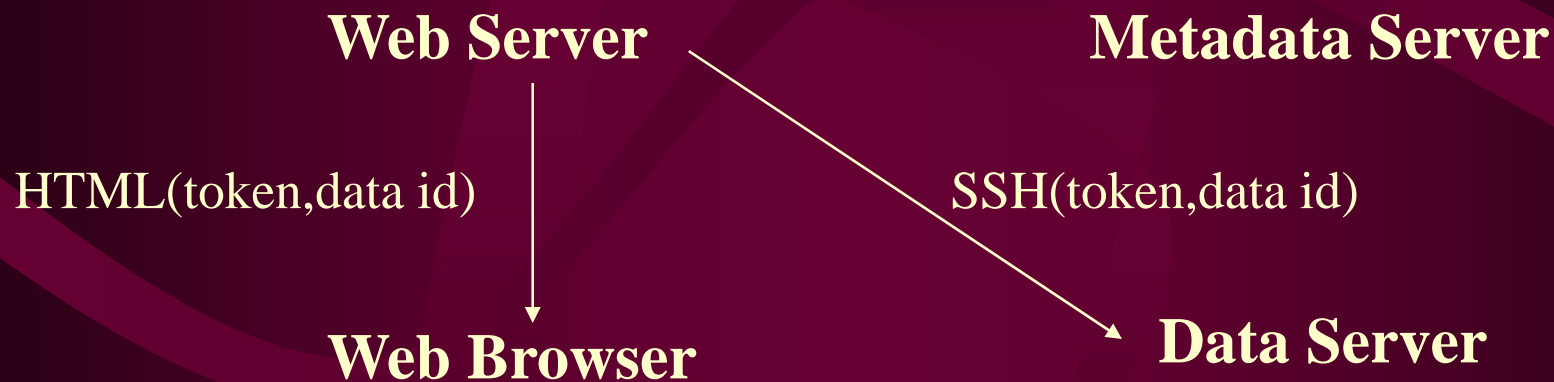
- The search form's authentication uses APACHE 'htpasswd' with 'anonymous' user support.
- A unique authtoken is created upon authentication.
- The authtoken is based on APACHE 'mod\_unique'
  - with a slight encoding change to enable use within an 'ftp' URL.



# Authorization

If access to a specific datum is authorized, the authtoken and data identifier are

- forwarded to the data server and
- included in the search response



# Authtoken Inclusion

## MACHO Project: Image Search Result

Your query was (Field=1) and (ObsDate>19930327) and (ObsDate<19930401).

7 hits

[Help](#) | [New Search](#) | [Previous](#) | [Next](#)

Unix batch retrieval script from selected images.

<u>Summary</u>		<u>Field</u>	<u>Domain</u>	<u>RA</u>	<u>Dec</u>	<u>ObsId</u>	<u>Date</u>	<u>Fetch</u>	<u>Batch</u>
<a href="#">Short</a>	<a href="#">Full</a>	1	LMC	5:5:21.3	-69:5:13	4596	19930328	<a href="#">Now</a>	<input type="checkbox"/>
<a href="#">Short</a>	<a href="#">Full</a>	1	LMC	5:5:21.3	-69:5:13	4619	19930328	<a href="#">Now</a>	<input type="checkbox"/>
<a href="#">Short</a>	<a href="#">Full</a>	1	LMC	5:5:21.3	-69:5:13	4634	19930329	<a href="#">Now</a>	<input type="checkbox"/>
<a href="#">Short</a>	<a href="#">Full</a>	1	LMC	5:5:21.3	-69:5:13	4658	19930329	<a href="#">Now</a>	<input type="checkbox"/>
<a href="#">Short</a>	<a href="#">Full</a>	1	LMC	5:5:21.3	-69:5:13	4690	19930330	<a href="#">Now</a>	<input type="checkbox"/>
<a href="#">Short</a>	<a href="#">Full</a>	1	LMC	5:5:21.3	-69:5:13	4716	19930330	<a href="#">Now</a>	<input type="checkbox"/>
<a href="#">Short</a>	<a href="#">Full</a>	1	LMC	5:5:21.3	-69:5:13	4764	19930331	<a href="#">Now</a>	<input type="checkbox"/>

# Simplifying Security Restrictions

## **No Proprietary Data**

- No authentication check prior to search form display
- No authorization check during hit summary creation
- No authorization check on data retrieval

## **Only Proprietary Data**

- Authentication check prior to search form display
- No authorization check on hit summary creation
- Authorization check on data retrieval

# Providing Value Added Analysis

## MACHO Project: Variable Star Search Result

Your query was (Field=1) and (Tile=3319) and (Seqn=10:12).

### 3 hits

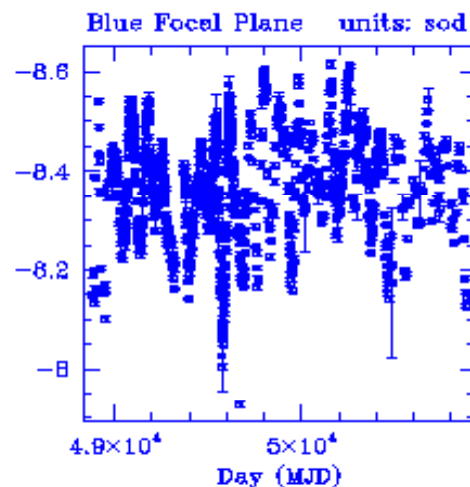
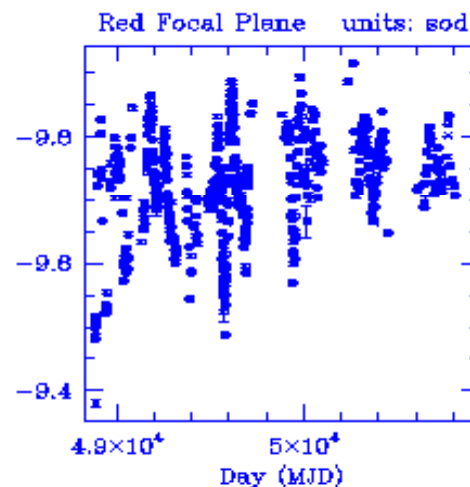
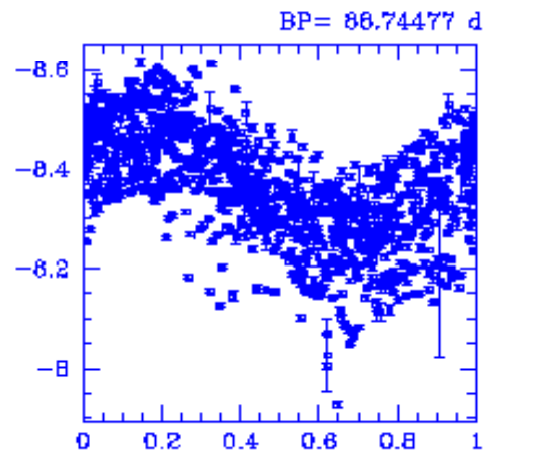
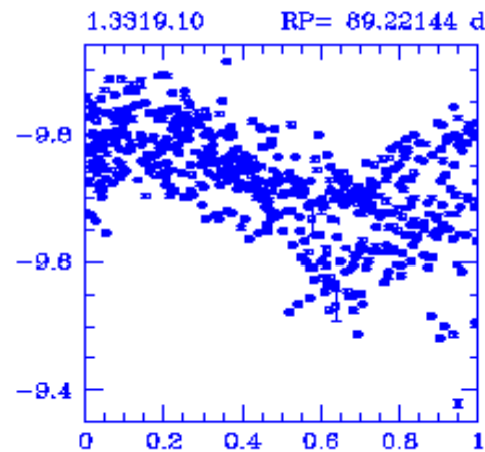
[Help](#) | [New Search](#) | [Template Images](#) | [Previous](#) | [Next](#) | [ftp Bundle](#)

<a href="#">Light Curve</a>	<a href="#">Field. Tile.Seqn</a>	<a href="#">Location (J2000)</a>	<a href="#">Variability Index</a>	<a href="#">Classification</a>	<a href="#"># Obs</a>	<a href="#">Obs w/2 Pts</a>	<a href="#">Focal Plane</a>	<a href="#"># Pts</a>	<a href="#">Period (Days)</a>	<a href="#">Magnitude Ave (K-C)</a>	<a href="#">Amplitude Ave</a>	<a href="#">Sup RSA</a>
<a href="#">View</a>	1.3319.10	<a href="#">RA</a> 5:1:15.2640	24.11	none	1077	490	<b>r</b>	497	89.2214	15.702 V	0.226	3.11
Bundle <input type="checkbox"/>		<a href="#">Dec</a> -69:25:59.5200					<b>b</b>	1041	88.7447	14.084 R	0.265	3.75
<a href="#">View</a>	1.3319.11	<a href="#">RA</a> 5:1:16.2000	1.64	none	1077	541	<b>r</b>	548	0.1226	16.966 V	0.052	1.06
Bundle <input type="checkbox"/>		<a href="#">Dec</a> -69:25:41.8800					<b>b</b>	1056	0.3314	17.112 R	0.053	1.14
<a href="#">View</a>	1.3319.12	<a href="#">RA</a> 5:1:16.8480	1.6	none	1077	548	<b>r</b>	555	0.166	16.958 V	0.124	1.49
Bundle <input type="checkbox"/>		<a href="#">Dec</a> -69:26:2.4000					<b>b</b>	1043	1.0962	17.206 R	0.13	1.36

# Value Added Analysis

## MACHO Project: Lightcurve Browser

[Help](#)



[Zoom & Show input parameters](#)

X axis (Day MJD)

Y axis (Red Focal)

Y axis (Blue Focal)

Zoom In on  X  Y  XY

Zoom Out on  X  Y  XY

Show

Extract

Extract

Units

Add phasing on redraw

MACHO Star: 1.3319.10 RA: 05 01 15.253 DEC: -69 25 59.44 Observations: 1077

# Future Work

- Different metadata server
- Enabling Virtual Observatory access
- Converting to a Secure APACHE server