HST Archive & Data Distribution

Herb Kennedy
Archive Branch
OUTLINE

• Who we are
• What we do
• What’s in the Archive
• What’s new in the Archive
• What’s coming in the Archive
• HST Data flow
Who We Are??

Currently there are two branches in the Operations & Data Management Division (ODM) responsible for data in the Archive.
Who We Are?? (cont.)

• OPUS Branch
  - Al Holm    Branch Chief

  Sid Parsons, Rich Arquilla, Tracy Ellis, Dorothy Fraquelli,
  Forrest Hamilton, Debbie Kenny, Mark Kochte, John Scott
Who We Are?? (cont.)

- Archive Branch
  - Rick White Branch Chief
    - Faith Abney Archive Systems Analyst
    - Sara Anderson Archive Specialist
    - Alberto Conti GALEX Database Engineer
    - Mike Corbin Archive Scientist
    - Anthony Houston Archive Specialist
    - Andrea Jackson DADS Operations
    - Herb Kennedy Data Archive Operations Supervisor
    - Tim Kimball Senior Software Engineer
    - Karen Levay Archive Systems Analyst
    - Myron Smith Archive Scientist
    - Rachel Somerville Archive Scientist
    - Randy Thompson Archive Systems Analyst
    - Bernie Shiao GALEX Database Engineer
The Archive Branch:

- The Archive Branch is responsible for the operation of the “Multimission Archive at STScI” (MAST)
  - Currently represents data from 13 satellite missions, of which four (HST, FUSE, CHIPS, GALEX) are active.
- We provide data delivery services and technical support to users
- We ensure the scientific integrity of the data
- We are responsible for the operation of the “Data Archive and Distribution System” (DADS).
  - This system represents the hardware and the software required to archive, distribute, and populate the databases for HST and FUSE.
WHATS IN THE ARCHIVE?

- Total size of the HST Archive as of Jan 1 2004:

17.11 Terabytes

Note: This is a virtual number. For active HST instruments the archive only stores the Raw files in compressed format and calibrates the data upon retrieval using “On-The-Fly Reprocessing” (OTFR) software. This process uses the POD files and the calibration reference files to generate the best possible data when retrieved.
WHAT’S IN THE ARCHIVE (cont.)

• MAST Represents data from 13 satellite missions:
• Active missions currently archived:
  - HST (launched 1990)
  - FUSE (launched 1999)
  - CHIPS (launched 2003)
  - GALEX (launched 2003)
• Legacy missions in the archive:
  - EUVE, IUE, ASTRO (HUT, UIT, WUPPE), ORFEUS (BEFS, IMAPS, TUES), Copernicus
• Total size of the archive including MAST Missions: 26.13 terabytes
WHAT’S IN THE ARCHIVE (cont.)

MAST Holdings (GB)
(Jan 1 2004)
WHAT’S IN THE ARCHIVE (cont.)

- Data ingest:
  - 16 GB per day (average)
- Data Distribution:
  - 50 GB per day (average)
WHATS NEW IN THE ARCHIVE

Hardware:

• Installed EMC Symmetrix 8830
  - Spinning disk storage
  - 32 TB of storage space (8 TB available for the Data Depot)
  - Will eventually take the place of jukeboxes
    (jukeboxes will be used as redundant archive)
  - Currently in the process of migrating all archive data to EMC.
    Process will be completed early this year.
  - All data will be online/faster response time
EMC Symmetrix 8830
WHATS NEW IN THE ARCHIVE (cont.)

Hardware:

• Installed SUN FIRE 15K ENTERPRISE CLASS SERVER
  - Replaces the old mixture of VMS, UNIX, and Tru64 clustered machines
  - 68 UltraSparc III CPU’s, 272 GB RAM, partitioned into 7 virtual machines
  - Contains development, test and operations domains for DADS and OPUS along with the corresponding databases.
WHAT’S NEW IN THE ARCHIVE (cont.)

Software:

• Distribution redesigned software v10.2
  - Installed December 2003
  - Replaced distribution SW running on VMS & UNIX
  - Allows for greater operator control, queue manipulation, and prioritization of user requests in the system
  - Secure FTP delivery mode
  - Electronic distribution, Media distribution, On-The-Fly-Reprocessing, (OTFR) and jukeboxes are all controlled from a single GUI.
<table>
<thead>
<tr>
<th>Timestamp</th>
<th>Request ID</th>
<th>Priority</th>
<th>Delivery Mode</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004-01-12T19:48:29</td>
<td>04011219482746762clarke</td>
<td>NORMAL</td>
<td>SFTP</td>
<td>running</td>
</tr>
<tr>
<td>2004-01-12T18:59:18</td>
<td>04011218591746761baly</td>
<td>NORMAL</td>
<td>SFTP</td>
<td>running</td>
</tr>
<tr>
<td>2004-01-10T22:49:55</td>
<td>04011022495246735macri</td>
<td>NORMAL</td>
<td>SFTP</td>
<td>running</td>
</tr>
<tr>
<td>2004-01-10T03:17:02</td>
<td>04011003163650719jll</td>
<td>NORMAL</td>
<td>SFTP</td>
<td>running</td>
</tr>
<tr>
<td>2004-01-10T02:28:21</td>
<td>04011002273646718harligan</td>
<td>NORMAL</td>
<td>SFTP</td>
<td>running</td>
</tr>
<tr>
<td>2004-01-09T20:38:56</td>
<td>040109223646621armaloi</td>
<td>NORMAL</td>
<td>CFRM</td>
<td>running</td>
</tr>
<tr>
<td>2004-01-08T22:32:04</td>
<td>040108222046620saznin</td>
<td>NORMAL</td>
<td>CFRM</td>
<td>running</td>
</tr>
<tr>
<td>2004-01-08T22:32:02</td>
<td>04010822202646620saznin</td>
<td>NORMAL</td>
<td>CFRM</td>
<td>running</td>
</tr>
<tr>
<td>2004-01-08T22:20:39</td>
<td>04010822202646609armaloi</td>
<td>NORMAL</td>
<td>CFRM</td>
<td>running</td>
</tr>
<tr>
<td>2004-01-08T22:20:15</td>
<td>04010822202646607armaloi</td>
<td>NORMAL</td>
<td>CFRM</td>
<td>running</td>
</tr>
<tr>
<td>2004-01-06T19:56:30</td>
<td>04010619562046513armaloi</td>
<td>NORMAL</td>
<td>CFRM</td>
<td>running</td>
</tr>
<tr>
<td>2004-01-06T15:30:58</td>
<td>0401052109454654iprathini</td>
<td>NORMAL</td>
<td>DVD</td>
<td>running</td>
</tr>
<tr>
<td>2004-01-08T04:42:00</td>
<td>04010517101546409kalakk</td>
<td>NORMAL</td>
<td>FtpStaging</td>
<td>running</td>
</tr>
<tr>
<td>2004-01-04T18:41:48</td>
<td>0401041841464670haojing</td>
<td>NORMAL</td>
<td>SFTP</td>
<td>moved to trouble</td>
</tr>
<tr>
<td>2004-01-02T17:54:49</td>
<td>04010217544946776winhenick</td>
<td>NORMAL</td>
<td>FTP</td>
<td>moved to trouble</td>
</tr>
<tr>
<td>2004-01-02T17:52:40</td>
<td>0401021752364675winhenick</td>
<td>NORMAL</td>
<td>FTP</td>
<td>moved to trouble</td>
</tr>
<tr>
<td>2004-01-02T17:50:16</td>
<td>0401021750174674winhenick</td>
<td>NORMAL</td>
<td>FTP</td>
<td>moved to trouble</td>
</tr>
<tr>
<td>2004-01-02T17:50:16</td>
<td>0401021750174674winhenick</td>
<td>NORMAL</td>
<td>FTP</td>
<td>moved to trouble</td>
</tr>
<tr>
<td>2004-01-02T06:28:28</td>
<td>04011206285656764apelrili</td>
<td>NORMAL</td>
<td>FTP</td>
<td>moved to trouble</td>
</tr>
<tr>
<td>2004-01-01T06:28:28</td>
<td>04011206285656764apelrili</td>
<td>NORMAL</td>
<td>FTP</td>
<td>moved to trouble</td>
</tr>
<tr>
<td>2004-01-01T06:28:35</td>
<td>04011206285656764apelrili</td>
<td>NORMAL</td>
<td>FTP</td>
<td>moved to trouble</td>
</tr>
<tr>
<td>2004-01-01T06:28:35</td>
<td>04011206285656764apelrili</td>
<td>NORMAL</td>
<td>FTP</td>
<td>moved to trouble</td>
</tr>
<tr>
<td>2004-01-01T01:23:55</td>
<td>04011001234046816kessen</td>
<td>NORMAL</td>
<td>SFTP</td>
<td>moved to trouble</td>
</tr>
<tr>
<td>2004-01-01T01:20:12</td>
<td>0401100120184246714coberltv</td>
<td>NORMAL</td>
<td>SFTP</td>
<td>moved to trouble</td>
</tr>
<tr>
<td>2004-01-01T06:59:36</td>
<td>04010619592546557ramaloi</td>
<td>NORMAL</td>
<td>CFRM</td>
<td>moved to trouble</td>
</tr>
<tr>
<td>2004-01-06T20:46:32</td>
<td>04011320305846842gregg</td>
<td>NORMAL</td>
<td>FTP</td>
<td>committed to BIGI</td>
</tr>
<tr>
<td>2004-01-06T20:53:11</td>
<td>040112052544678daries</td>
<td>NORMAL</td>
<td>FTP</td>
<td>committed to BIGI</td>
</tr>
<tr>
<td>2004-01-11T00:33:14</td>
<td>0401110032426736ro</td>
<td>NORMAL</td>
<td>DVD</td>
<td>committed to BIGI</td>
</tr>
<tr>
<td>2004-01-09T00:11:14</td>
<td>0401090003846844jro</td>
<td>NORMAL</td>
<td>DVD</td>
<td>committed to BIGI</td>
</tr>
<tr>
<td>2004-01-05T23:57:10</td>
<td>0401052356374678jro</td>
<td>NORMAL</td>
<td>DVD</td>
<td>committed to BIGI</td>
</tr>
<tr>
<td>2004-01-05T23:57:10</td>
<td>0401052356374678jro</td>
<td>NORMAL</td>
<td>DVD</td>
<td>committed to BIGI</td>
</tr>
<tr>
<td>2004-01-03T21:03:35</td>
<td>04011321032946846daries</td>
<td>NORMAL</td>
<td>FTP</td>
<td>committed</td>
</tr>
<tr>
<td>2004-01-03T20:44:34</td>
<td>0401132044324685prucesteve</td>
<td>NORMAL</td>
<td>FTP</td>
<td>committed</td>
</tr>
<tr>
<td>2004-01-03T20:40:36</td>
<td>040113204030560448key</td>
<td>NORMAL</td>
<td>FtpStaging</td>
<td>committed</td>
</tr>
</tbody>
</table>
WHAT’S COMING IN THE ARCHIVE

Software:

- Ingest redesign software
  - Moves current DADS ingest running on VMS to SUN FIRE 15K platform.
  - Uses OPUS pipeline processing
  - OPUS/DADS GUI
  - Installation scheduled for late spring 2004
HST DATA FLOW

INGEST PIPELINE:

- OPUS receives the raw data from PACOR
- OPUS generates an archive request
- DADS SW reads the request and copies files to a staging (EMC)
- Validates the dataset to ensure that all files and header keywords are present.
- Copies the data to optical disk platters (primary & safestore)
- Copies the data to the Data Depot on EMC
- Populates the operational database (Catlog) with header keywords
INGEST PIPELINE (cont.)

- Returns a response to OPUS if ingest is successful
- Returns a response to DADS for re-ingest or investigation if unsuccessful
- Continually replicate the operational database (Catlog) to the browse database (Zeppo)
- Data are available to the user after replication via StarView or the MAST web interface.
DISTRIBUTION PIPELINE:

- Users are registered and obtain an account on-line.
- Users query the browse database via the MAST World wide web or Starview
- Users list datasets and submit requests to DADS
  - Request only the dataset files they require.
  - The science data is processed through OTFR
  - Request the data delivery option
    1) Data SFTP or FTP’ed directly to user’s computer
    2) Data copied to a host computer (STDATU) and the user pulls the data via FTP
    3) Data copied to hard media (CD, DVD) and shipped
### StarView version 7.22

#### Enter qualifications for: Quick Search

<table>
<thead>
<tr>
<th>Label</th>
<th>Qualification (click cell to edit)</th>
<th>Get Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset Name</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dec</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radius (degrees)</td>
<td>0:10</td>
<td></td>
</tr>
<tr>
<td>PI (last name)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposal ID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flag</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target Name</td>
<td>M81</td>
<td></td>
</tr>
<tr>
<td>Instrument</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Results for: Quick Search

- **Proposal ID:** 9454
- **Release Date:** 2003-03-12 14:05:03
- **Target Description:** GALAXY; NUCLEUS; LINER
- **Instrument:** ACS
- **Config:** ACS/HRC
- **Start Time:** 2003-03-12 08:14:25.3
- **Flag:** NORMAL

<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>RA</th>
<th>Dec</th>
<th>Instrument</th>
<th>Flag</th>
<th>Apertures</th>
<th>Central Wavelength</th>
</tr>
</thead>
<tbody>
<tr>
<td>X0KK0101T</td>
<td>+09.55 33.11</td>
<td>+69 03 55.49</td>
<td>FOC</td>
<td>NORMAL</td>
<td>512x512</td>
<td></td>
</tr>
<tr>
<td>X0KK0102T</td>
<td>+09.55 33.11</td>
<td>+69 03 55.49</td>
<td>FOC</td>
<td>NORMAL</td>
<td>512x512</td>
<td></td>
</tr>
<tr>
<td>X0KK0201T</td>
<td>+09.55 33.11</td>
<td>+69 03 55.49</td>
<td>FOC</td>
<td>NORMAL</td>
<td>512x512</td>
<td></td>
</tr>
<tr>
<td>J6FF03011</td>
<td>+09.55 33.17</td>
<td>+69 03 55.11</td>
<td>ACS</td>
<td>NORMAL</td>
<td>HRC</td>
<td>2727.024</td>
</tr>
<tr>
<td>J6FF03021</td>
<td>+09.55 33.17</td>
<td>+69 03 55.11</td>
<td>ACS</td>
<td>NORMAL</td>
<td>HRC</td>
<td>3356.902</td>
</tr>
<tr>
<td>J6FF02011</td>
<td>+09.55 33.17</td>
<td>+69 03 55.11</td>
<td>ACS</td>
<td>NORMAL</td>
<td>HRC</td>
<td>2714.109</td>
</tr>
<tr>
<td>X0KK02011</td>
<td>+09.55 33.11</td>
<td>+69 03 55.44</td>
<td>ACS</td>
<td>NORMAL</td>
<td>HRC</td>
<td>3034.748</td>
</tr>
</tbody>
</table>

- **Proposal Abstract**
- **Instrument Info**
- **Calibration Info**
HST DATA FLOW (cont.)

DISTRIBUTION PIPELINE (cont.)

• OTFR requested data are processed using the latest calibration reference files and software available for the current instruments
• Data requested for legacy instruments will return data calibrated at the time of ingest
• The DADS system also identifies and retains requests that fail FTP delivery or OTFR processing for operator investigation
• The system automatically notifies the user via email when the request is submitted and when it is complete.
Subject: MAST Data Retrieval Request Received: kennedy46854
Date: Tue, 13 Jan 2004 16:58:57 -0500 (EST)
From: archive@stsci.edu
To: kennedy@stsci.edu

Your MAST Data Retrieval Request was received Tue Jan 13 21:58:57 GMT+00:00 2004
Your request has been accepted, and assigned ID: kennedy46854
Status may be monitored at:
   http://archive.stsci.edu/cgi-bin/reqstat?reqnum=kennedy46854

Your files will be delivered to the staging area on host: archive.stsci.edu
Your directory will be: kennedy46854
When your request is complete, you will receive email notification,
and you can use anonymous ftp to retrieve your data.

Datasets to be retrieved for this request:
u21y0d02t

Please contact the archive hotseat (archive@stsci.edu) or
410-338-4547 with any questions. Thank you.
Status for kennedy46854 is

- Request ID: 04011321585646854kennedy
- Tracking ID: kennedy46854
- State: complete-succeeded
- Start Time: 2004-01-13T22:01:02 UT
- OTF Processing Progress: 100.0%
- Delivery Mode: FtpStaging
- Files Delivered: 6

The DADS Queue currently contains 23 requests waiting to run.
The average time for a request to complete is 6 hours 1 minutes 51 seconds +/- 3 hours 13 minutes 23 seconds.
Subject: Retrieval status for request kennedy46854
Date: Tue, 13 Jan 2004 17:10:52 -0500 (EST)
From: archive@stsci.edu
To: kennedy@stsci.edu

Your request kennedy46854
was completed at Tue Jan 13 22:10:52 GMT+00:00 2004
REQUEST DELIVERY STATUS: SUCCESS

Your data has been shipped to
dads_ops@stdatu.stsci.edu:/stage/kennedy/kennedy46854
Total size of delivered files: 10609920 bytes

Delivery status of individual files

u21y0d02t_d0f.fits with status: success (5054.0625 kb)
u21y0d02t_q0f.fits with status: success (5054.0625 kb)
u21y0d02t_q1f.fits with status: success (101.25 kb)
u21y0d02t_shf.fits with status: success (36.5625 kb)
u21y0d02t_trl.fits with status: success (14.0625 kb)
u21y0d02t_x0f.fits with status: success (101.25 kb)
The Archive Hotseat handles all archive related questions and concerns.

A copy of HST data is also delivered to the following institutions:
- CADC  Canadian Astronomical Data Center
- ECF    European Coordinating Facility
- NAOJ   National Astronomical Observatory of Japan
Questions??????