Cosmic Variability —
Mass Data Management and Information Technology
Challenges of Astronomical Databases

Wen-Ping Chen (陳文屏)
Institute of Astronomy
National Central University, Taiwan
2006 October 24 @ CODATA Beijing
Challenges of Astron. Observations

• Sensitivity --- farther, fainter, older
• Resolution --- clearer (angular), finer (spectral), …

Next Frontier in Astrophysics

• Celestial objects vary in brightness (minor planets, stars, AGNs, gravitational lenses, etc).

• Time domain has not been much exploited.

Lesson on Project Management

• Software cost should not be sneezed at, especially if the data are to be publicly available

• Rule of Thumb --- $1 hardware; $1 software; $3-10 informatics (i.e., databases)
The **BATC** (Beijing-Arizona-Taipei-Connecticut) project, a multi-wavelength sky survey which involves institutes in China, US and Taiwan, was initiated in early 1990s by a group of Chinese astronomers including Fang Lizhi, then exiled in the US.

The project, after the initial hardware, management, and communication challenges, has collected a total of 500-600 GB worth of imaging data, and now enters its scientific production peak (totaling > 100 SCI papers) after more than a decade of operations.

It takes mutual trust to build up a collaboration!
Outline

**TAOS**
Taiwan-America Occultation Survey

Fully operational; in Taiwan

**Pan-STARRS**
Panoramic Survey Telescope and Rapid Response System

Being constructed; in Hawaii, USA
Geographical Vantage:
- Many high mountains
- West Pacific
- Low latitude

→ time-domain astrophysics
LULIN OBSERVATORY

鹿林天文台

Elevated to 2862m; above the inversion layer

... seen from Yusan (Jade Mt; 玉山) 4000- m
LELIS

LOT

Lulin Observatory

SLT

TAOS

窄波段巡天計畫

中美掩星計畫 3號望遠鏡

超輕型望遠鏡

中美掩星計畫 4號望遠鏡

控制中心 1m望遠鏡

中美掩星計畫 1,2號望遠鏡
Scientific Activities at Lulin

- **Time Variability**
  - Part of global network
  - Mass data processing
  - Data mining

- **TAOS** (Taiwan, USA, Korea)

- **LELIS** (NCU)

- **Taiwan Oscillation Network** (NTHU)

- **Sprite** (NCKU)

- **Taiwan Earth-Shine Network** (NTHU)

- **Atmospheric Experiments**

So far, discoveries of 10+ supernovae, 150+ asteroids ...
So far there are more than 1000 objects found beyond the orbit of Neptune, Pluto being one of the largest. These are seen by reflected sunlight, so only the largest can be detected by large telescopes.
The **TAOS** (Taiwan-America Occultation Survey) project, a novel telescope array set up by groups from Taiwan, US and Korea, began routine observations in early 2005 and has the potential to give clue to the formation and evolution of our Solar System.

Comet nuclei too faint to be detected by direct imaging may be “seen” when they move in front of a background star --- a stellar occultation event.
Project Overview

- A census of the small objects in the outer solar-system
- An array of wide-field telescopes (D=50 cm, f/1.9, FOV=3 sq. deg) to monitor brightness changes of ~1,000 stars at 5 Hz rate
- Looking for a ‘blink’ of starlight (occultation) when an object (> 2 km) moves in front of a distant star
- **Frequency of events ➔ population of “interveners”**
- Data rate a few 100 GB per night; only “interesting” data downloaded via the dedicated microwave E1 connection
- Real-time data analysis (light curves, rank statistics)
- Requiring coincidence detection of the same event by all telescopes to guard against false positives
Typical CCD imaging

Every star, together with surrounding skies, get exposure at the same time

TAOS data

Integrate for 200 ms and then read out 32 rows of pixels, with the shutter remains open

The sequence continues, so each star appears as a series of dots ‘zipper’

⇒ ‘Fake’ neighboring stars and skies!
TAOS Telescopes

Lulin Observatory
Central Taiwan
altitude=2862 m

TAOS is the only one of its kind in the world to conduct a census of small (1-2 km size) icy bodies at the outer reach of the solar system.

With a special data acquisition and a non-parametric statistical analysis scheme

100 GB/night
Adaptive Aperture Photometry Pipeline

- Fast
- Moderately accurate
- Compensate for image motion

Diagram:

1. **stare**
2. **pixel coordinates**
   - $(x, y, ra, dec, mag, id, hms+dms)$
3. **pre-AAP**
   - optimal aperture size
4. **AAP**
   - $(t0, t1, flux, ferr, x, y, seq)$
   - 69 bytes

**Compressed zipper**
- decompressing...
- decompressed zipper
- smooth & statistics
- optimal aperture size
- ap-size text file
- pixel-coord text file
- pixel-coord text file
Sample Output of the TAOS Photometric Light Curves

<table>
<thead>
<tr>
<th>beg. time</th>
<th>end time</th>
<th>count</th>
<th>ct err</th>
<th>x</th>
<th>y</th>
<th>rb</th>
</tr>
</thead>
<tbody>
<tr>
<td>66293.663414</td>
<td>66293.818818</td>
<td>26656.42</td>
<td>141.08</td>
<td>478.73</td>
<td>49.79</td>
<td>2</td>
</tr>
<tr>
<td>66293.913422</td>
<td>66294.068817</td>
<td>26787.13</td>
<td>137.69</td>
<td>478.86</td>
<td>125.79</td>
<td>2</td>
</tr>
<tr>
<td>66294.163444</td>
<td>66294.318814</td>
<td>26650.18</td>
<td>142.46</td>
<td>478.72</td>
<td>201.90</td>
<td>2</td>
</tr>
<tr>
<td>66294.413452</td>
<td>66294.568816</td>
<td>26329.72</td>
<td>138.72</td>
<td>478.86</td>
<td>277.74</td>
<td>2</td>
</tr>
<tr>
<td>66294.663449</td>
<td>66294.818813</td>
<td>26454.31</td>
<td>140.94</td>
<td>478.79</td>
<td>353.83</td>
<td>2</td>
</tr>
<tr>
<td>66294.913452</td>
<td>66295.068814</td>
<td>27486.91</td>
<td>145.05</td>
<td>478.93</td>
<td>429.89</td>
<td>2</td>
</tr>
<tr>
<td>66295.163458</td>
<td>66295.318812</td>
<td>27318.67</td>
<td>140.27</td>
<td>478.77</td>
<td>505.80</td>
<td>2</td>
</tr>
<tr>
<td>66295.413483</td>
<td>66295.568820</td>
<td>26646.11</td>
<td>140.48</td>
<td>478.61</td>
<td>581.82</td>
<td>2</td>
</tr>
<tr>
<td>66295.664491</td>
<td>66295.818816</td>
<td>25601.26</td>
<td>146.22</td>
<td>478.41</td>
<td>657.70</td>
<td>2</td>
</tr>
<tr>
<td>66295.913628</td>
<td>66296.068814</td>
<td>26560.96</td>
<td>143.56</td>
<td>478.57</td>
<td>733.74</td>
<td>2</td>
</tr>
<tr>
<td>66296.163496</td>
<td>66296.318812</td>
<td>26050.15</td>
<td>145.18</td>
<td>478.49</td>
<td>809.87</td>
<td>2</td>
</tr>
<tr>
<td>66296.413508</td>
<td>66296.568819</td>
<td>26172.10</td>
<td>148.19</td>
<td>478.33</td>
<td>885.90</td>
<td>2</td>
</tr>
<tr>
<td>66296.663500</td>
<td>66296.818816</td>
<td>25504.24</td>
<td>149.49</td>
<td>478.31</td>
<td>961.89</td>
<td>2</td>
</tr>
<tr>
<td>66296.913505</td>
<td>66297.068813</td>
<td>25484.46</td>
<td>139.52</td>
<td>478.32</td>
<td>1037.92</td>
<td>2</td>
</tr>
<tr>
<td>66297.163527</td>
<td>66297.318812</td>
<td>26705.86</td>
<td>146.23</td>
<td>478.26</td>
<td>1114.00</td>
<td>2</td>
</tr>
<tr>
<td>66297.413534</td>
<td>66297.568816</td>
<td>25456.52</td>
<td>153.24</td>
<td>478.25</td>
<td>1190.05</td>
<td>2</td>
</tr>
<tr>
<td>66297.663539</td>
<td>66297.818815</td>
<td>24907.12</td>
<td>144.34</td>
<td>478.47</td>
<td>1266.08</td>
<td>2</td>
</tr>
<tr>
<td>66297.913528</td>
<td>66298.068814</td>
<td>25864.63</td>
<td>140.92</td>
<td>478.66</td>
<td>1342.22</td>
<td>2</td>
</tr>
</tbody>
</table>
• Pipeline Flow from image-taking to archival
• Arrows represent the flow of messages between components
2004 February 21
TAOS detected the occultation event of HIP 079407, $m_V=8.8$ mag) by (51) Nemausa ($m_V=11.9$)
Prediction by Isao Sato (左藤勳)
D~150 km

\[ \Delta t \sim 6.25 \pm 0.50 \text{ s} \]
The 1 m telescope at Lulin also detected the same event with traditional CCD imaging.

\[ t_{\text{exp}} = 1 \text{ s} \]
\[ t_{\text{ro}} < 2 \text{ s} \]

By A. Chen
2006 Feb 06 three TAOS telescopes detected a suspected occultation of TYC 076200961 ($m_V \sim 11.83$) by (286) Iclea ($m_V \sim 14.0$ mag, $D \sim 97$ km)
Panoramic Survey Telescope And Rapid Response System
Project Overview

- All-sky survey \((3\pi)\)
- Frequent revisit (cadence 4-7 days)

Wide-Field Imaging
Short Duty Cycle
Efficient Operations

- An array of 4 telescopes, located in Hawaii, each of D=1.8 m, equipped with a 1.4 gigapixel camera of an Orthogonal Transfer Array CCD detector (=40 cm square focal plane) \(\Rightarrow\) 7 square-degree FOV with 0.26” pixels

- Detection of moving, transient, and variable celestial objects down to very faint limits
- Cumulate very deep sky images several TB/night
The Telescopes
The Detector

- Independently addressable orthogonal transfer CCDs (cells)
  - Reducing cost by increasing yield
  - Fast readout: Gigapixels in 2 s
  - On-Chip guiding
  - Minimizing effects of bright stars
  - Compensating for image motion
The Site(s)

Site for Prototype telescope (PS1) --- Haleakala High Altitude Observatory (Maui)
Eventual Mauna Kea site for full Pan-STARRS
Each raw image from a single Pan-STARRS camera will contain 2 Gbytes (2 bytes per pixel). In the full survey mode, typical exposures last 30 seconds, so the raw data rate is several terabytes per night for the full telescope system. The amount of data produced by Pan-STARRS is so large that it will not be practical to archive every image. Software techniques are therefore being developed to extract the important information from the images, while allowing less crucial information to be discarded.

Data storage requirement ~0.5 Pb in year 1
The Data Flow

Subsystems

✓ TEL – Telescopes
✓ CAM – Cameras
✓ OTIS – Observatory, Telescope & Instrument Software
✓ IPP - Image Processing Pipeline
✓ MOPS – Moving Object Processing Software
✓ PSPS – Published Science Data Products
Summit Process Flow Diagram

Telescope System
- Dome
- Telescopes
- Shutters

Camera System
- Cameras
  - Detector Host
  - Detector Data Storage

OTIS
- TCS
- Pan-STARRS Telescope Scheduler
- Observation Sequencer
- OTIS DA
- Observation Tool
- OTIS Weather Server
- Weather Forecast Servers
- Meteorological Stations
- Other External Data Servers
- Survey Mission

Legend:
- PS Subsystem
- PS Hardware
- Survey Mission
- Data Collection
- PS Process
- External System
- Pixel data
- Metadata
- Commands
PSPS Overview
# Database Sizing Justification

<table>
<thead>
<tr>
<th>Object DB</th>
<th>P2</th>
<th>P4Σ</th>
<th>P4Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>obj/deg</td>
<td>2.7E+04</td>
<td>1.1E+05</td>
<td>2.0E+05</td>
</tr>
<tr>
<td>deg/fpa</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
</tr>
<tr>
<td>FPA/night</td>
<td>3000.00</td>
<td>750.00</td>
<td>750.00</td>
</tr>
<tr>
<td>nights/year</td>
<td>250.00</td>
<td>250.00</td>
<td>250.00</td>
</tr>
<tr>
<td>bytes/obj</td>
<td>64.00</td>
<td>64.00</td>
<td>64.00</td>
</tr>
<tr>
<td>DB OH</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
</tr>
<tr>
<td>Years</td>
<td>10.00</td>
<td>10.00</td>
<td>10.00</td>
</tr>
<tr>
<td>PB</td>
<td><strong>0.36</strong></td>
<td><strong>0.36</strong></td>
<td><strong>0.67</strong></td>
</tr>
</tbody>
</table>

### Data Product

<table>
<thead>
<tr>
<th>Data Product</th>
<th>Size (PB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static Sky Img</td>
<td>1.51</td>
</tr>
<tr>
<td>Object Data</td>
<td>1.43</td>
</tr>
<tr>
<td>Cum. Sky Cat.</td>
<td>0.19</td>
</tr>
<tr>
<td>Metadata</td>
<td>0.04</td>
</tr>
<tr>
<td>Postage Stamps</td>
<td>0.01</td>
</tr>
<tr>
<td>MOPS</td>
<td>0.0021</td>
</tr>
<tr>
<td>Filtered Trans.</td>
<td>0.000001</td>
</tr>
<tr>
<td><strong>Total (PB)</strong></td>
<td><strong>3.19</strong></td>
</tr>
</tbody>
</table>

### Cum. Sky Catalog

<table>
<thead>
<tr>
<th>deg</th>
<th>3.0E+04</th>
</tr>
</thead>
<tbody>
<tr>
<td>obj/deg</td>
<td>4.3E+05</td>
</tr>
<tr>
<td>filters</td>
<td>6.00</td>
</tr>
<tr>
<td>bytes/obj</td>
<td>300.00</td>
</tr>
<tr>
<td>Compress</td>
<td>1.00</td>
</tr>
<tr>
<td>DB OH</td>
<td>4.00</td>
</tr>
<tr>
<td>Copies</td>
<td>2.00</td>
</tr>
<tr>
<td>PB</td>
<td><strong>0.19</strong></td>
</tr>
</tbody>
</table>

### Static Sky Images

<table>
<thead>
<tr>
<th>deg</th>
<th>3.0E+04</th>
</tr>
</thead>
<tbody>
<tr>
<td>pix/deg</td>
<td>3.2E+08</td>
</tr>
<tr>
<td>filters</td>
<td>6.00</td>
</tr>
<tr>
<td>bytes / pix</td>
<td>7.20</td>
</tr>
<tr>
<td>Compress</td>
<td>0.40</td>
</tr>
<tr>
<td>DB OH</td>
<td>1.00</td>
</tr>
<tr>
<td>Copies</td>
<td>9.00</td>
</tr>
<tr>
<td>PB</td>
<td><strong>1.51</strong></td>
</tr>
</tbody>
</table>
Conclusions

- Time-domain astrophysics as a niche science
- Astronomers demanding to push the IT forefronts
  - Telescope/Detector technology → larger, finer observations
  - Rapid cadence → huge data volume
- Data processing, analysis, storage, archival, distribution
  ($1 \text{ hardware, } $1 \text{ software, } $3-10 \text{ DB})
- Need to involve software engineers, IT managers, statisticians … from the very beginning of a project to design the experiment