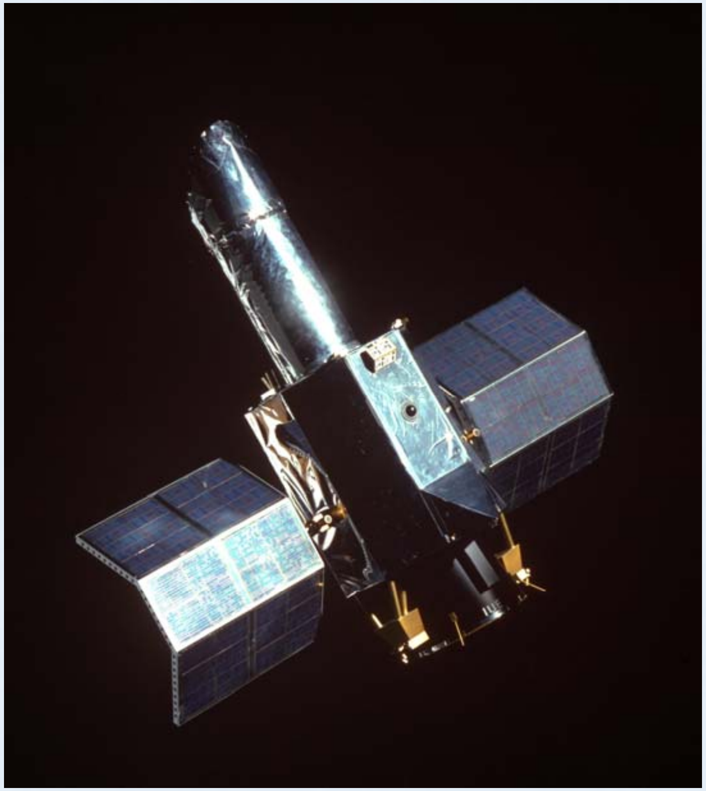
Science Data Reduction Pipelines at NASA
30 Years and counting

Don Lindler
Sigma Space Corporation

January 24, 2007

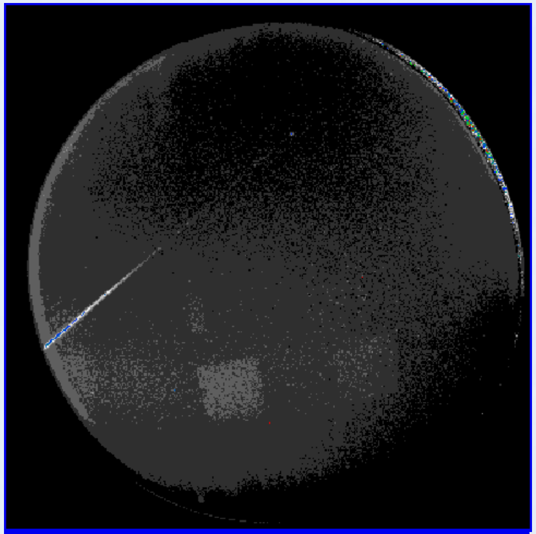
International Ultraviolet Explorer
Joint Project between NASA, ESA, and UK

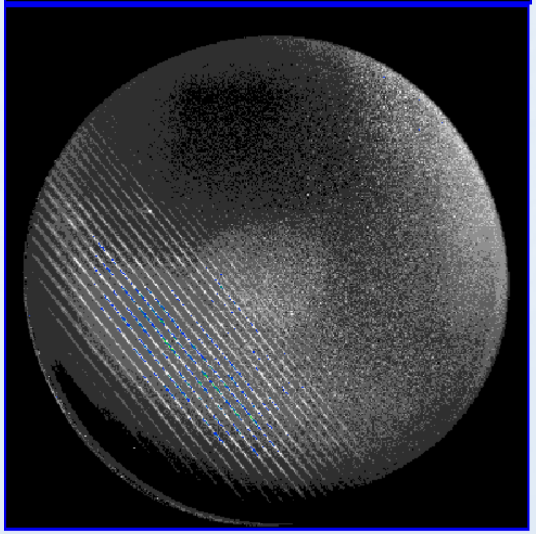
- Launched on January 26, 1978
- Turned off on September 30, 1996
- 18.7 years of operation
- 104470 Spectra
- 9600 astronomical sources



IUE Pipeline and Archive

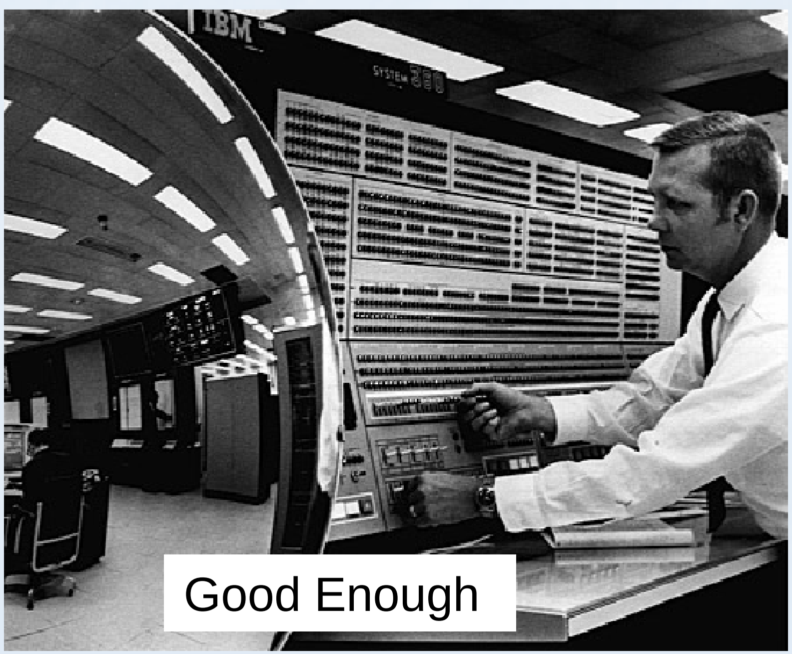
- First astronomical and satellite facility to deliver fully reduced data within 48 hours to the worldwide community of scientists
- the creation of the first worldwide astronomical reduced-data archive delivering 44,000 spectra per year (5 spectra per hour) to astronomers in 31 countries





IUE Pipeline Development

- Utilized NASA/JPL VICAR system (Video Image Communication and Retrieval)
- VICAR ported from IBM mainframe to a Sigma 9 computer.



Good Enough


Much Better

Why choose VICAR?

- Analysis Tools for calibration
- Modules for geometric and photometric correction of images

BIGGEST PROBLEMS

- Expensive transition from IBM to Sigma 9
- Astronomers were not using Sigma 9 computers

Reasons for Success

- Plenty of Funding!
- Modular coding
- Evolution of pipeline algorithms based on the Science and controlled by the users. Frequent meetings (Three Agency and calibration meetings)
- Astronomers were well informed (IUE newsletter)
- Dialog between astronomers and the observatory staff. Astronomers came to the control center for real time observing.

Reasons for Success, continued

- Intermediate outputs supplied to astronomers.
- Data Quality Flags included with output spectra.
- IUE Regional Data Analysis Facilities.

NEWSIPS

New Spectral Imaging Processing Pipeline

- Consistency between archived products during the extended mission.
- After a long study of application executives, it was developed under MIDAS.
- Move from VICAR to FITS format.
- Improvements in S/N, spectral resolution, and absolute sensitivity.
- Again, plenty of funding!

Hubble Space Telescope

- Launched April 24, 1990
- Five initial instruments
  - High Resolution Spectrograph
  - Faint Object Spectrograph
  - High Speed Photometer
  - Faint Object Camera
  - Wide Field and Planetary Camera
- Five Instrument Science Teams with Guaranteed Observing Time

Instrument Team Software Development

- Work began in the late 1970s
- Each team worked independently. Little coordination between teams.
- HST Launch scheduled for 1983
- GHRS and FOS selected to use IDL as the primary analysis tool/programming language.

GHRS system was a major success!

- System used as recently as last year to analyze 20 year old spectral calibration lamp data.
- IDL used for data acquisition, archival, pipeline processing, and analysis
- All data raw data easily search and immediately accessible.
- “On-the-fly” pipeline processing
- Rapid Prototyping
- Propagation of errors.

GHRS continued.

- GHRS approach latter extended to FOS, STIS, ACS, and WFC3 instrument development and to some extent COS.
- Biggest Problem with GHRS system: Formats defined in the pre-FITS era and before ST ScI formats were defined.

HST PODPS (First Try)

Hubble Space Telescope Post Observational Data Processing System

- Part of the Science Operations Ground System (SOGS)
- 1980 to 81: 2 inch thick requirements document written by NASA appointed committee
- 1981: Contract award to TRW
- 1983: First software components delivered for a DEC/VMS based system.

PODPS 1st try continued

- End User was not involved in the design!

1st Try continued.

- SOGS used last-generation programming technology.

  - Supply requirements.

  - Detailed Design.

  - No prototyping.

- Non-modular. Fixing a bug in one part of the program can easily generate a bug somewhere else.
- Very little contact with the instrument teams.

PODPS (2nd Try)

- The Space Telescope Science Institute enters the picture.
- They inherit an unusable system.
- Generation of calibration reference files was not part of the system!
- They negotiate with the instrument teams for a joint development. This time with ST ScI oversight.
- ST ScI Staffing increased from the proposed 100 people to around 400 people at launch.

PODPS (2nd Try) continued

- Pipeline software under SDAS (Science Data Analysis System) running under IRAF.
- Prototyping of software by Instrument Teams
- Problems

  - Machine dependent (DEC/VMS)

  - Machine dependent FITS-like data formats (.HHH and .HHD files)

  - VMS FORTRAN extensions used.

  - Intermediate results not available for all instruments.

PODPS (3rd Try)

- Name changed from SDAS to STSDAS
- First stage: Code no-longer machine dependent
- Second stage: Data Formats no longer machine dependent. Changed to FITS.
- Third stage: Development of OPUS

OPUS

- Operational since 1995.
- Fully distributed for any series of applications.
- Runs multiple instances of multiple processes in multiple pipelines on multiple nodes.
- Easy integration of your own pipeline steps.
- Includes monitoring tools.
- Best of all: The developers of the pipeline modules do not need to know how to use OPUS.
- Complete record of processing included in the output.

OPUS continuing evolution

- Observation Associations
- On-The-Fly calibration
- Automatic generation of calibration reference files (e.g. darks, flats).

OPUS sites

- Hubble Space Telescope (HST).
- Far Ultraviolet Spectroscopic Explorer (FUSE).
- International Gamma-Ray Astrophysics Laboratory (Integral)
- Chandra X-Ray Observatory, AXAF
- BeppoSAX X-ray Observatory
- Spitzer Space Telescope, SIRTF
- The Gemini Observatory

Continuing Problem Areas

- Deciding what enhancements to make when manpower is limited. A well defined decision process is needed with decisions based on the scientific return.

Continued Problem Areas

- Systematic Errors
- Standard Data Quality Flags
  - 8, 16, or 32 bit quantity with each bit representing a different condition
  - Numeric score with each condition representing a different condition. Only the “worst” condition is flagged.
  - Binary
- Analysis software which uses the flags

The Future?

- Model Based Calibration

"Calibration based on instrument models has been demonstrated to provide better accuracy than empirical methods, but in addition it also provides a real understanding of the instrument that enables one to maintain it at maximum performance and quickly diagnose any deviations"

Michael Rosa, ST-ECF

Fig 1: *The European part of the NASA award winning team, the group at ST-ECF: Florian Kerber, Michael Rosa, Paul Bristow (left to right).*

Award winning Astronomers

prefer Model Based Calibration

Conclusion

You work hard, develop new ideas, and improve your instrument calibration. At the end of the day, you go home and have a beer. You repeat this day after day. In the end all that really matters is the beer.