

NASA Earth Science Data Rescue Efforts

Nathan James (1), Jeanne Behnke (1), James E. Johnson (1,2), Asghar Esfandiari (1,2), Emily Zamkoff (1,3), Gary Alcott (1), Atheer F Al-Jazrawi (1,4)

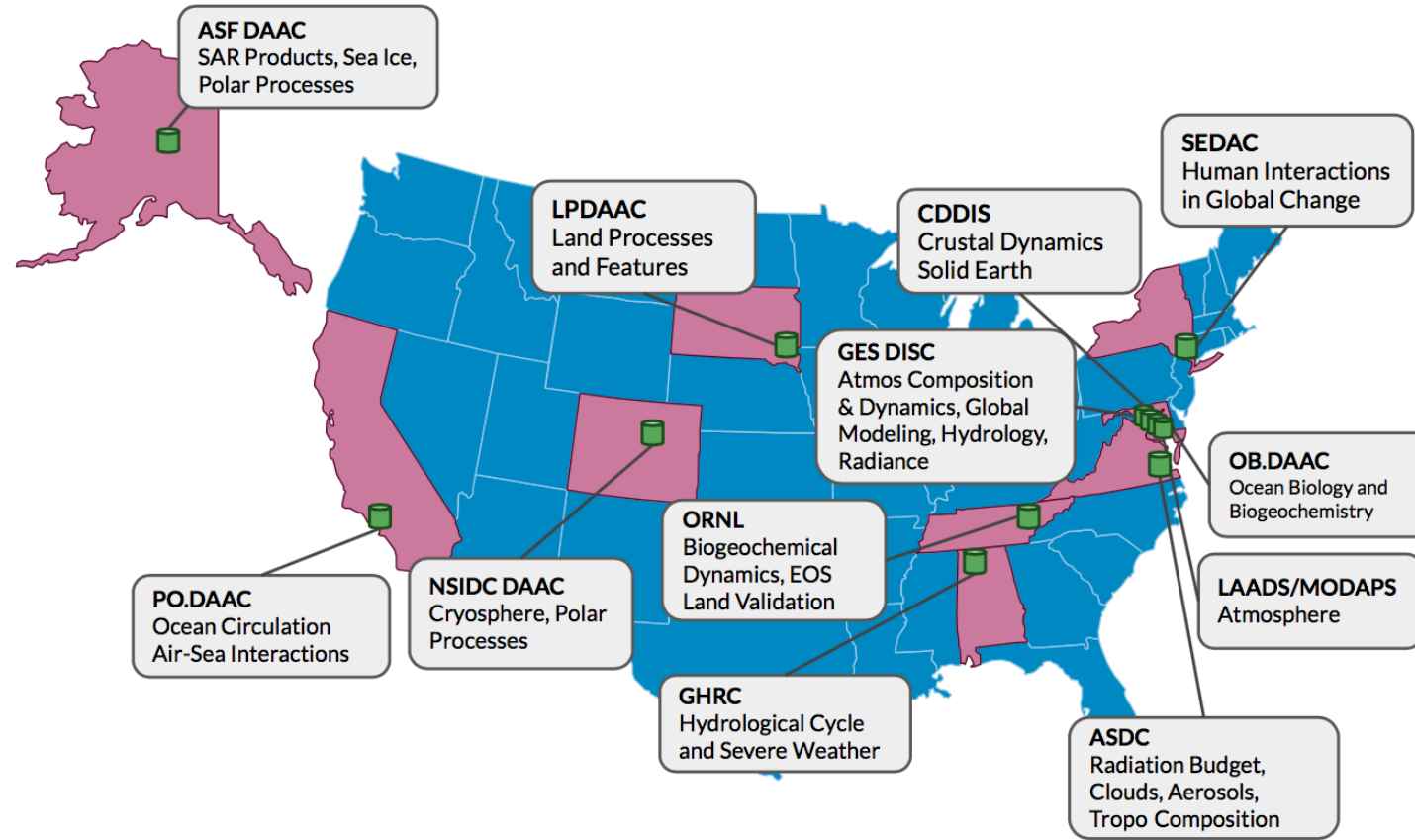
1. ESDIS, NASA GSFC, Greenbelt, MD, United States. 2. ADNET Systems, Inc., Rockville, MD, United States. 3. KBR Wyle Systems, Inc., McLean, VA, United States. 4. Telophase Corporation, Arlington, VA, United States.

Who are we and what do we do?

NASA has been collecting data from Earth remote sensing systems since the inception of the Agency. The data from these sensors are stored in various forms and by many records keepers. When EOSDIS was conceived in the late 1980s, one of the first actions was to separate earth science data from space science and astronomy. The earth science data collection was brought together in 12 active archives for preserving NASA Earth Science data for present and future generations. The EOSDIS data systems focused on new missions starting in the 1990s and today, all 27 PBs of data are online, backed up and accessible worldwide. We continue to work on the data collected prior to the 1990s, a steady effort of recovery and rescue.

Who are we and what do we do?

- NASA EOSDIS is comprised of 12 active archives for preserving NASA Earth Science data for present and future generations.

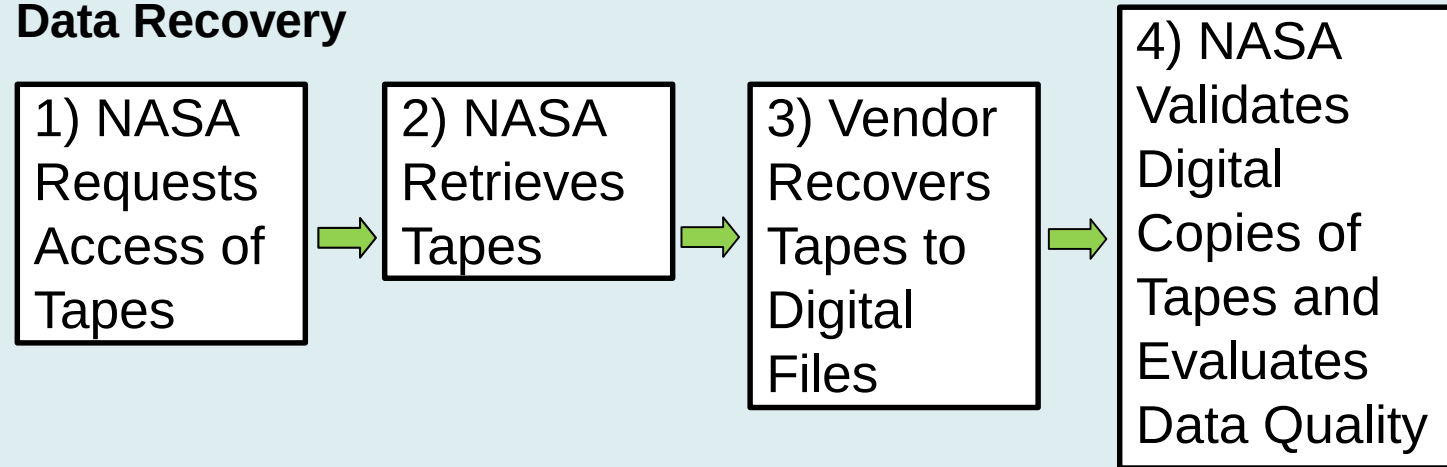


What is the rescue process?

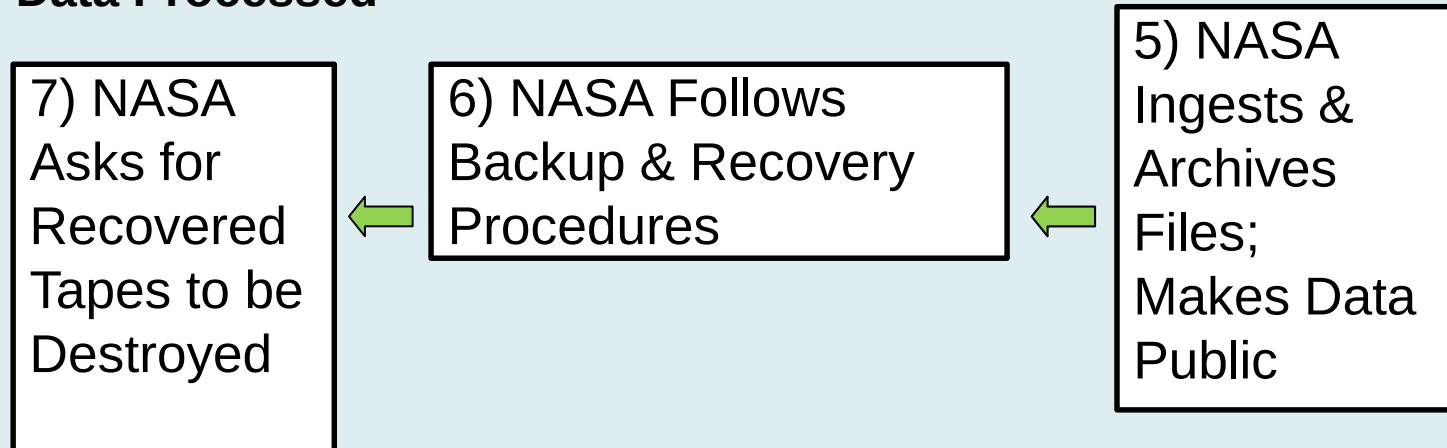
- Currently rescuing NASA data from 1960-1990s that was placed at the Federal Records Center for storage
- Conduct an independent inventory, ship media to the recovery contractor
 - Recovery contractor extracts bits from tape and delivers a corresponding tape image
- For each data set
 - Write specialized code to extract the individual data files from the tape image
 - Validate the science quality of the files including visualization of the data
 - Prepare documentation, archive extracted files, make data available online
- Archive tape images and release original media for destruction

Recovery Process

Data Recovery



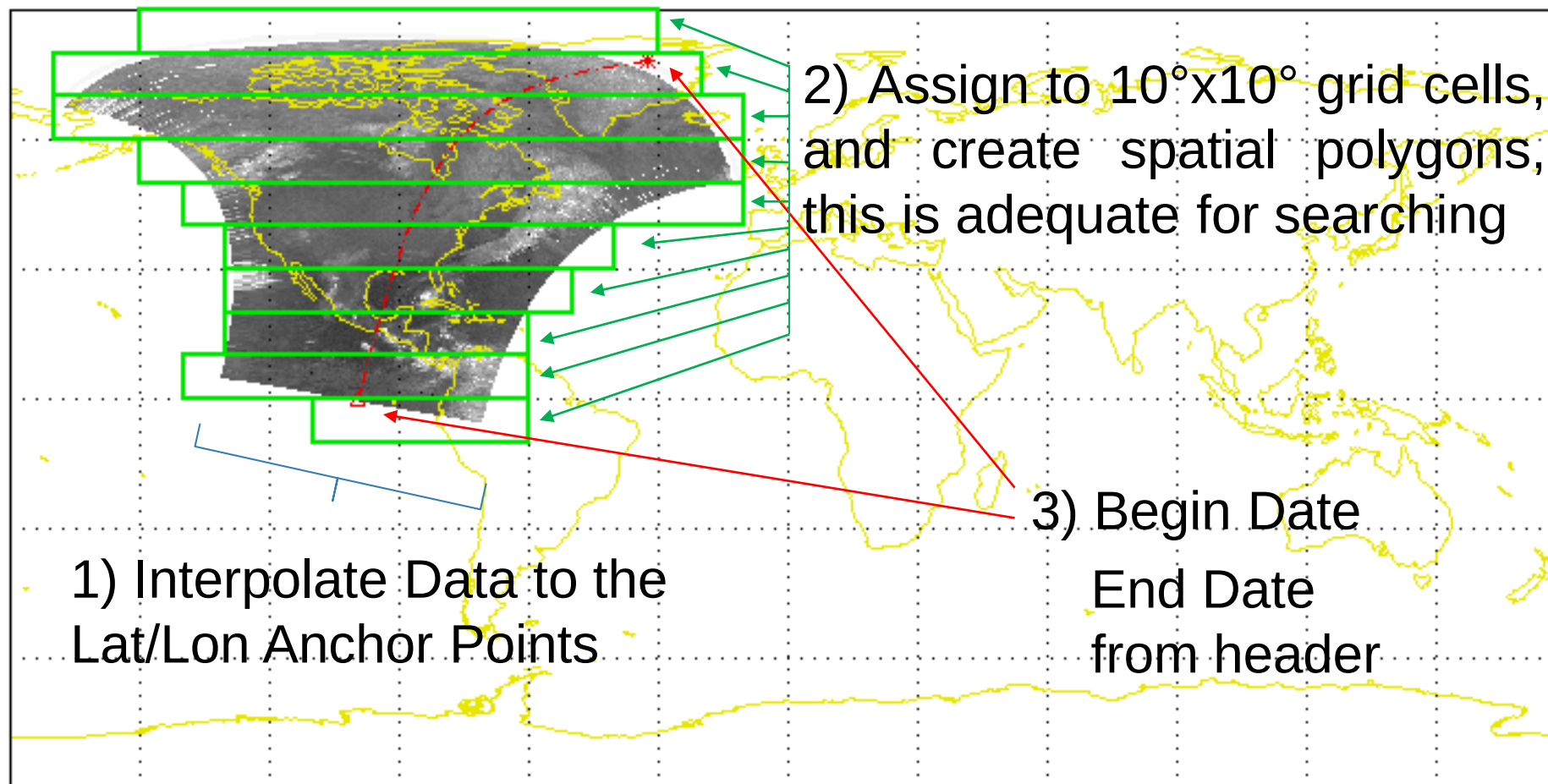
Data Processed



Extract Data Files from Nimbus Tape File

- In the Nimbus era, each experiment team designed their own unique file format, limits software reuse
- No concept of granule level metadata, this has to be extracted from each granule or data file and created new
- Data originally written on outdated IBM-360 machines:
 - use 36-bit or 32-bit words
 - use IBM integer, floats and characters (EBC not ASCII)
- Files have no names, GES DISC creates names based on metadata: experiment, date, orbit and tape number
- Backup tapes must be reviewed individually and compared with primary tape for any missing data files

The File-Level Metadata



What are the challenges?

Attempts to rescue and restore data from past missions often uncover issues involving:

- Bookkeeping/Accounting from original records
 - Data from unrelated missions could be in the same box
 - Missing, illegible or wrong tape labels
- Documentation
 - Lack of useful documents
 - Archived documents don't reflect the final formats and anomalies understood at the time of the mission
 - Non-existent metadata
- Media
 - Fragile, sticky, or corrupt tapes
 - Broken reel hubs
 - Time consuming due to required manual intervention
- Data Processing
 - Determining if tapes are 7-track or 9-track
 - Multiple tape formats in a collection
 - Unknown data file formats
 - Missing or extra end-of-file or end-of-tape markers

How do we meet the challenges?

- Rely on insight from similar data sets already processed
- Rely on information in supporting tape label photos and tape image Comma Separated Values (CSV) files
- Educated guesses
- Tedious examination of bit patterns
- Trial and error
- Patience and persistence

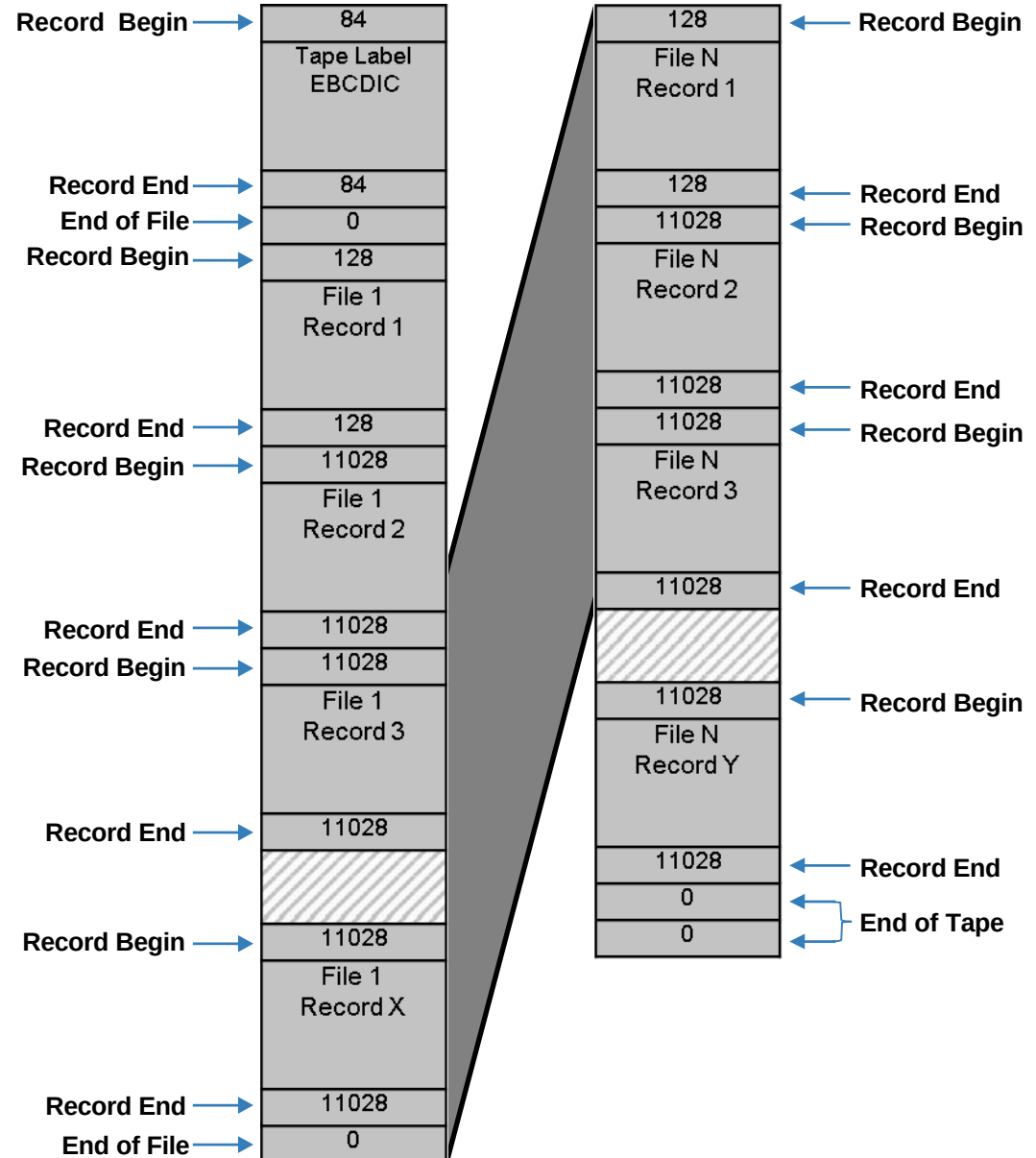
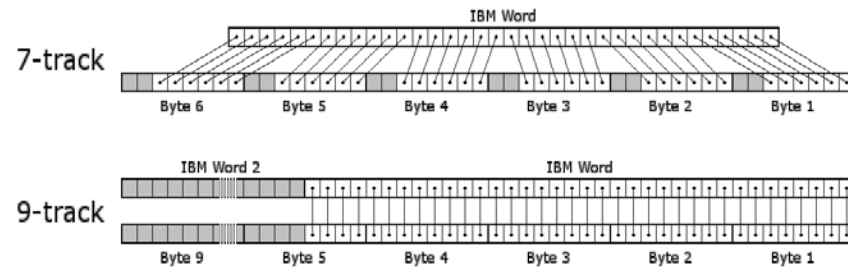
Nimbus Tape Recovery Statistics

- 11973 Nimbus tapes in NASA's data system
- 9378 tapes have been reviewed and processed for data recovery to date
- 106,241 data files have been rescued, validated and made publicly available online.

Nimbus 2 HRIR TAP File Format

TAP Format Header: 32-bit integer
bit 0-30: length of record in bytes
bit 31: 0 = good record, 1 = bad

Reconstructing the original data:



Conclusion

- This is tedious work!
- Important to preserve the data, otherwise lost forever!!!
- No common format makes each product unique
 - *limits software reuse*
- File formats sometimes deviate from documentation
- Corrupted records and data make extraction hard
- Corrupted tapes makes data unrecoverable
- Through a tedious and laborious process, oft-corrupted data are recovered, and incomplete metadata and documentation are reconstructed.
- See <https://disc.gsfc.nasa.gov> for access to the data, documentation, and for more information
- **Reference:** Khayat, M., Kempner, S., "Life Cycle Management Considerations of Remotely Sensed Geospatial Data and Documentation for Long Term Preservation," 2017, <https://ntrs.nasa.gov/search.jsp?R=20160002963>

First data from
Nimbus-1 HRIR
1964/08/29
Orbit 23

