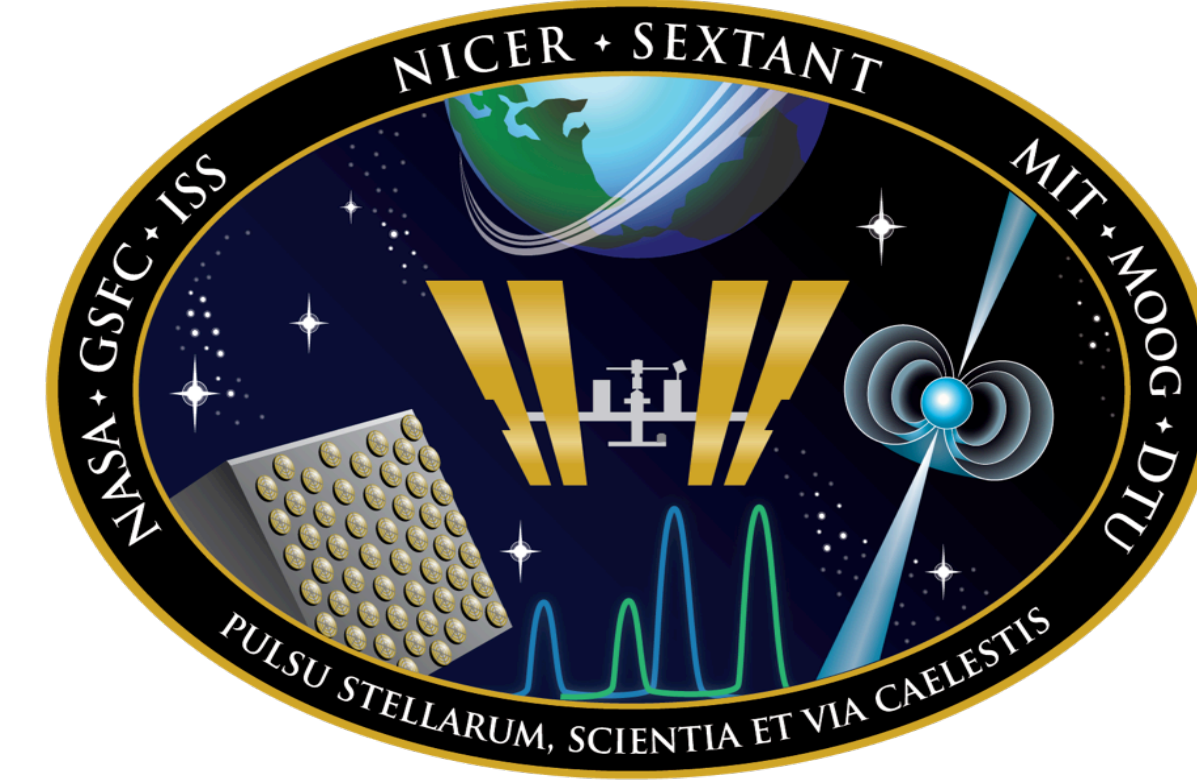
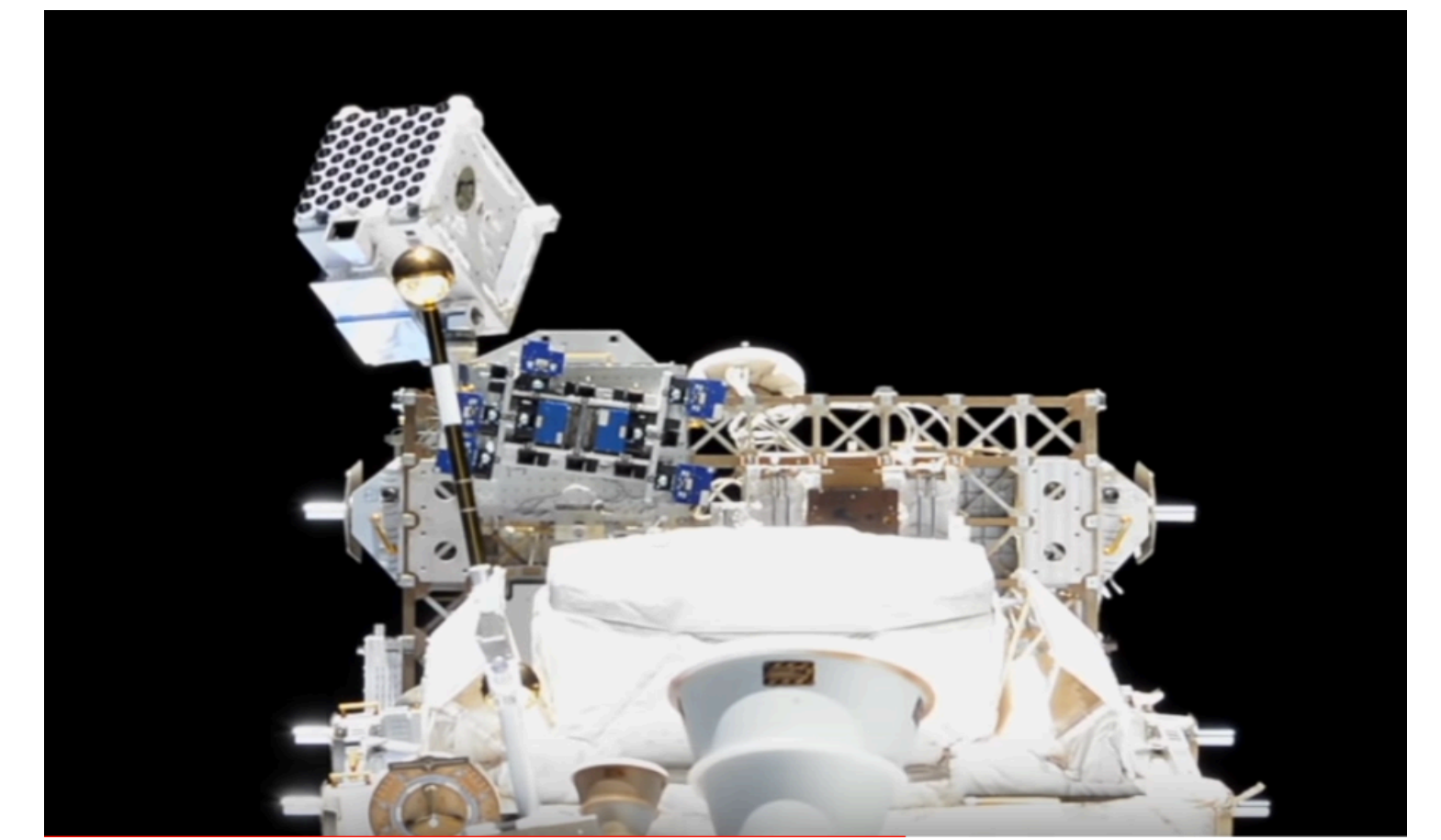


NICER Science Data Pipeline



Abstract

NICER (Neutron star Interior Composition Explorer) is a NASA Explorer payload mounted on the International Space Station. Launched in June 2017, it is designed to measure the composition of neutron stars by observing them at X-ray wavelengths. Data stream from the mission, through the science processing pipeline, and on to the High Energy Astrophysics Archive. We describe the different steps in the pipeline, from ingesting the data to processing to storage, the different flows of data, and how the status of the data is tracked throughout the entire cycle. We use a combination of Python, Perl, and shell scripts, along with a relational database, to move the data through the pipeline. The HEASoft software package is used to process the data. The data are moved to long term storage at the HEASARC, where interested parties may retrieve them for further study.



Ingest

Raw data

The input telemetry comes to the pipeline in the form of CCSDS (Consultative Committee for Space Data Systems) packet files. These files contain all payload data, including X-ray counts and NICER pointing (RA and Dec). These files list the pointing, at regular time intervals, updated every 0.1 seconds. Data for a single sequence could be entirely contained in one CCSDS file, or it could span multiple files. A single CCSDS file could contain data about multiple sequences.

Three different types of telemetry files are delivered, at a cadence of: every 15 minutes, every 4 hours, and once per day. It takes an expected 36 hours to receive all of the data for a particular day.

Ingest

Input telemetry files are delivered to a "hot directory" on the NICER machine which is constantly checked for new additions. Once a file downloads, it is moved into the final ingest directory.

The tool `pktrouter_nicer` retrieves the time-index information from the CCSDS files, and records the data in the Ingest Database.

Trend monitoring is performed on a daily basis. The pipeline checks the database for recent deliveries. It [re]runs the daily script for days for which new data has been delivered. One daily trend product is a detailed NICER observatory pointing history, which is used to reconstruct the target observation timeline. The software verifies that there are no large gaps in the input pointing history, which would either need new data to fill in, or need to be reconstructed from the uploaded observatory schedule, before passing the sequences to the next step in the pipeline.

Processing

The data processing scripts are run daily, after the trend monitoring has completed. Sequences are marked as 'ready to process' when new data arrive that matched their time ranges. These sequences are then added to a queue for the processing script. Each of the cores on the machine may be leveraged in the processing. Plans are underway to add additional machines to the processing capabilities.

Once the processing is complete, the script then determines from the database if the sequence should be delivered to the encrypted archive or the public archive, and marks each as such.

Science Tools

The NICER science data is processed from Level0 (raw) to Level1 (FITS format, uncalibrated), to Level2 (calibrated and screened) and Level3 (high level browse products). Basic tools are available in the public multi-mission HEASoft (v6.25) package*. High level science processing scripts are maintained within the pipeline infrastructure.



Here we are!

Archive

HEASARC

Once the observation sequences are processed, they are sent to the High Energy Astrophysics Science Archive Research Center (HEASARC). The pipeline uses DTS to send the data.

The delivery information is recorded in the Observation Database, including delivery date, encryption status, and receipt confirmation from the archive.

The Browse interface for the NICER archive data can be found linked from the following URL:
<https://heasarc.gsfc.nasa.gov/W3Browse/nicer/nicermastr.html>

Encrypted data that are still within their validation period are sent to a private section of the archive. Data are typically validated and available to the public, unencrypted, two weeks after the observation.

Pipeline

A dataset is an observation segment, or sequence, identified in the pipeline and tracked through the database by the ObservationID. A single sequence contains the data for each snapshot of the target on a day, which may not be continuous in time.

Parts of the code structure of the pipeline are adapted from previous missions. An established pipeline has evolved from being used with Suzaku, Hitomi, and NuSTAR, and will be used with future missions as well. NICER took portions of this pipeline that adapted well, and adjusted certain parts to better fit the specific needs of NICER, and created new pipeline code where necessary.

Results

As of November 2018, the NICER Observatory has delivered to the HEASARC archive:

- 8,617 validated and public sequences
- 939 encrypted sequences in their two-week validation phase

A total of 43,530,791 seconds have been processed.

NICER has observed 274 X-ray targets, and has accumulated a total photon count of 28 billion photons.

Databases

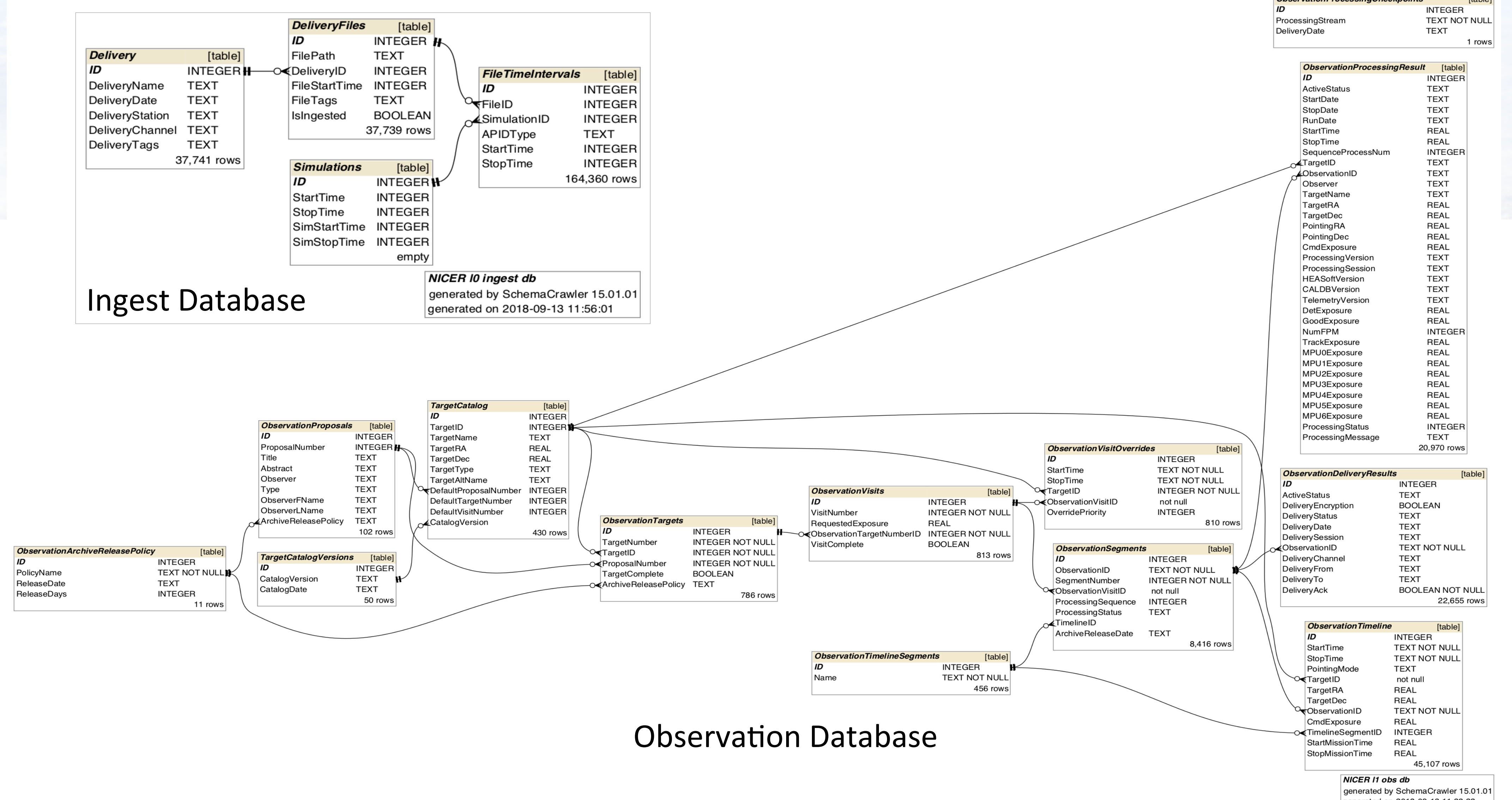
At each step of the pipeline process, information is recorded in one of the databases.

Ingest Database

The ingest database tracks information about the incoming data files. It records basic data such as the file name, file path, and delivery time. It also records the start and stop times of the data inside the file. These times are used in the daily trend monitoring script to compare against the timelines of sequences, to see which sequences have received new data and should be processed.

Observation Database

This database records as-flown information about each observation segment. It records facts about each of the science targets and their associated science proposals. It tracks each snapshot that makes up an observation in a detailed observing timeline. The current status of a sequence as it goes through the pipeline is recorded in the ObservationSegments table. It states whether the sequence is waiting to be fully ingested, ready to go through the data processing script, ready to deliver, or has been delivered, and to which part of the archive. The ObservationProcessingResult table holds data about each round of processing of a sequence, including target information, and start and stop times. In the ObservationDeliveryResults table, information about the delivery of the data to the archive is recorded, such as delivery date and encryption status.



Kristin Rutkowski (Adnet Systems, NASA/GSFC)
Craig Markwardt (NASA/GSFC)
NICER Science Team
www.nasa.gov/nicer
www.heasarc.gsfc.nasa.gov/docs/nicer/
svs.gsfc.nasa.gov/Gallery/NICER.html
nicerhelp@bigbang.gsfc.nasa.gov

References

CCSDS Space Packet Protocol <<https://public.ccsds.org/Pubs/133x0b1c2.pdf>>
DTS <heasarc.gsfc.nasa.gov/dts/>
FITS <<https://fits.gsfc.nasa.gov/>>
*For more information on the HEASOFT software package, see Poster P5.9

Software Used

- Shell, Perl, Python scripts
- Compiled code
- SQLite database
- HEASOFT*
- Data Transfer System (DTS)
- git