Operational Recommendations for Capturing History and Infusing Data Science (ORCHIDS)

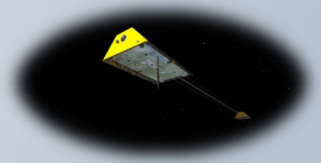
Preparing your mission data for future analysis

Jack Lightholder
Lukas Mandrake
James Montgomery
Sarah Myers
Kevin Shannon
Brian Kahovec

Jet Propulsion Laboratory
California Institute of Technology

© 2020 California Institute of Technology. *Government sponsorship acknowledged.* Published by The Aerospace Corporation with permission.

Motivation: Trapped Mission Data, Limited Analysis



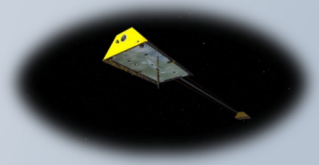
Operating Spacecraft

Science + Engineering

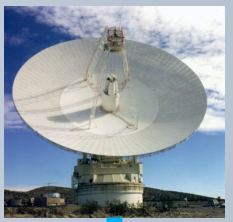
Ground Station

- Spacecraft downlink both Science and Engineering Data
- Traditional Operations treat these flows very differently

Motivation: Trapped Mission Data, Limited Analysis



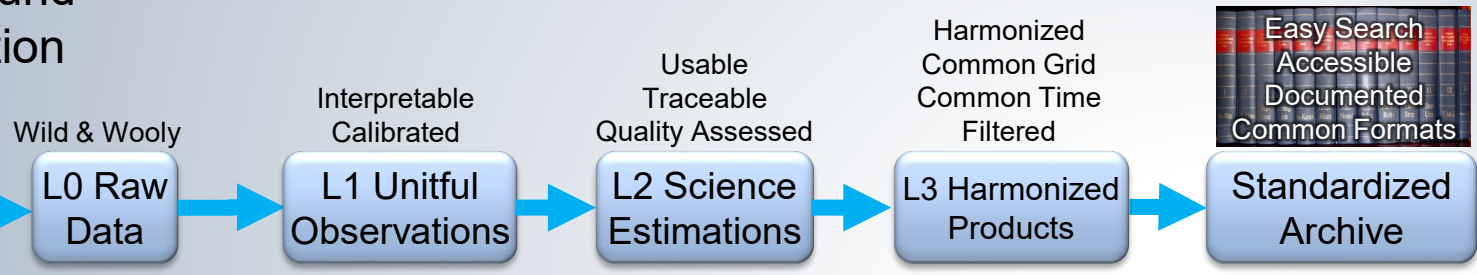
Operating Spacecraft



Ground Station

Science Data

- Science data is destined for exploration
- L0-L3 processing prepares data for analysis
- Standardized archives (DAAC's) serve data
- Search abilities permit researcher access
- Quality assessment and metadata captured
- Excellent document is assumed and required

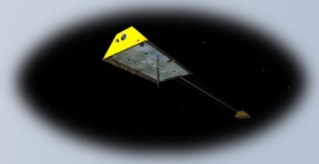


Science Pipeline is Healthy & Optimized for Analysis

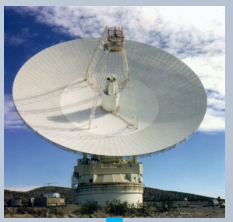
(because that's the point of science data)

JPL

Motivation: Trapped Mission Data, Limited Analysis



Operating Spacecraft



Ground Station

- Engineering data is for “expert consumption only”
- Shallow or emergency analysis assumed
- Wake the person who understands this stuff!
- New personnel overwhelmed with missing details
- Even basic analysis is stymied, let alone data-driven science

Science Data

Wild & Wooly
L0 Raw Data

Interpretable
Calibrated
L1 Unitful Observations

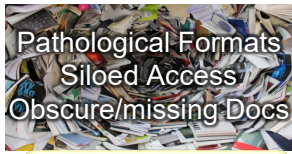
Usable
Traceable
Quality Assessed
L2 Science Estimations

Harmonized
Common Grid
Common Time
Filtered
L3 Harmonized Products

Standardized Archive

Eng. Data

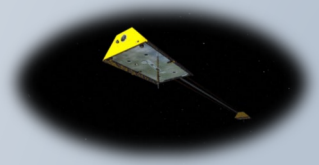
Wild & Wooly
“L0” Raw Data

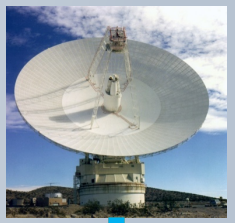


Nonstandard Archive

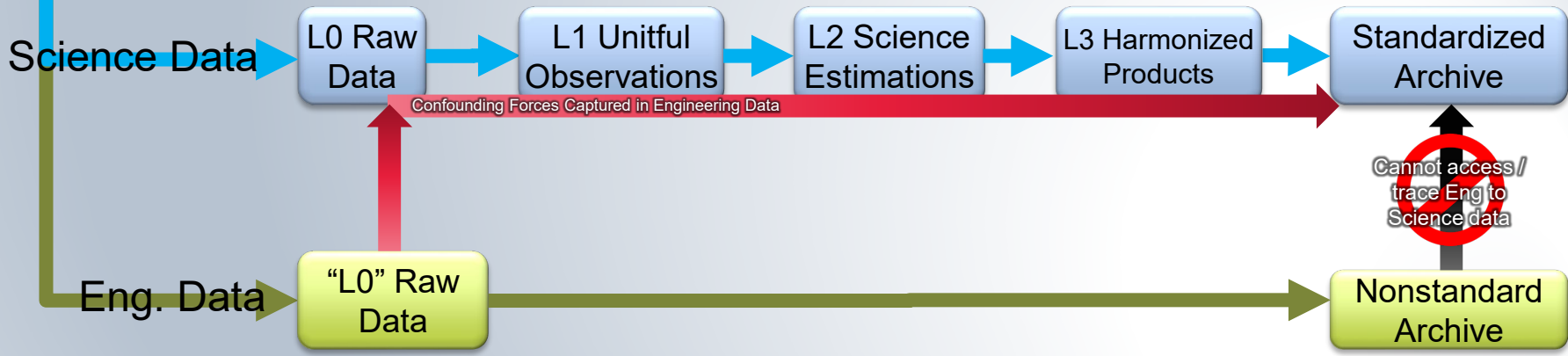
Motivation: Science Anomaly Situation

For GRACE missions, spacecraft IS the instrument!

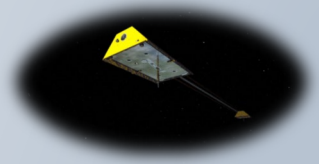




- (this really happens!)
- Scientists trace L2/3 anomalies back to L0
- Issue is on instrument! Linked to spacecraft itself
- No worries… scientists will just look at Eng data
- What do you mean it can't be easily harmonized?
- What do you mean I can't easily make graphs?
- Advanced analysis thwarted… spin up Ops effort
- False dichotomy between Science/Eng data maintained



Motivation: Ops Anomaly Situation



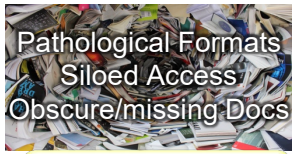
Panic! Streetlight effect forces minimum effort paths

- Look at the eng data we CAN get at
- Check only the most likely relationships
- Confirmational analysis only… exploration hindered
- System interactions are hardest to explore

Eng. Data

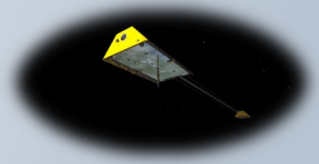
Wild & Wooly

"L0" Raw Data



Nonstandard Archive

Motivation: Multi-Mission Comparison



Try to compare data from multiple missions or ATLO to Flight

- Totally different archive, formats
- Mature missions no longer have original personnel!
- Requires research project just to harmonize
- Neither archive captured sufficient info for cross-comparison
- May not actually be possible within reasonable time/cost

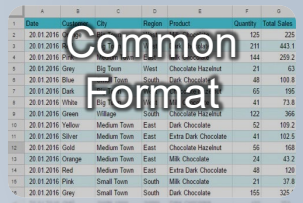
Part of solution already known!

NASA's Analysis Ready Data (ARD) concept: the data we all want & need

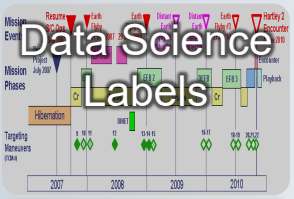
Analyst can just start work

Easily select and interpret data

Event Logs, Known Issues
Faulty Channels, Design Intentions,
Expert Knowledge



Any program can read
No special loaders



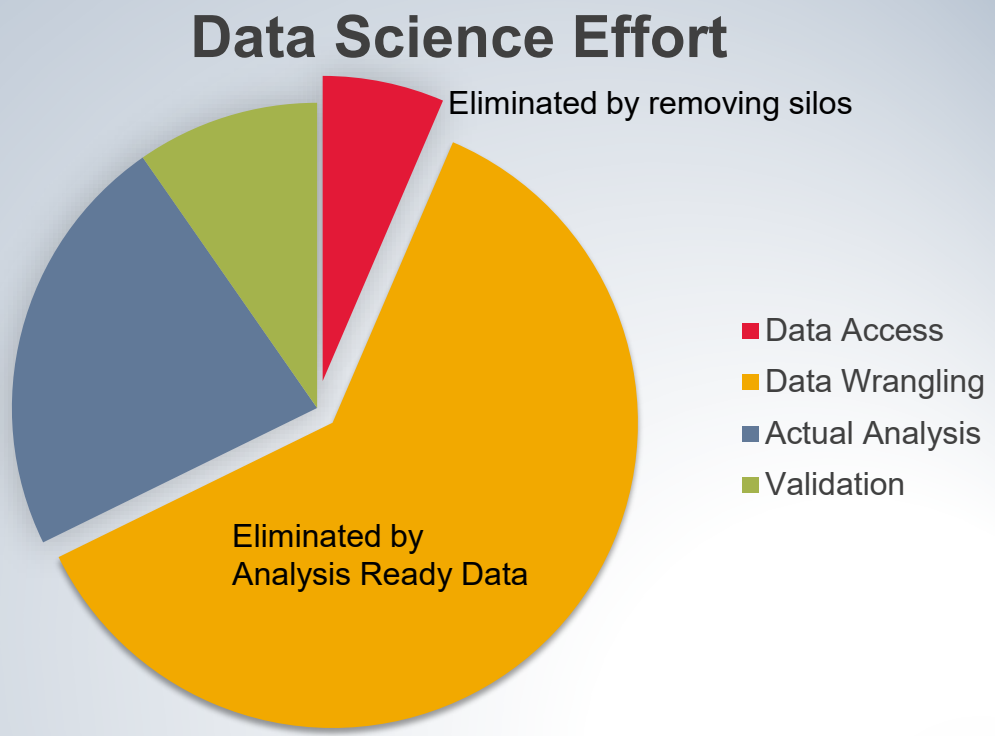
Timelines, events, channel information stored as proper labels

Simple, searchable, standard, scalable

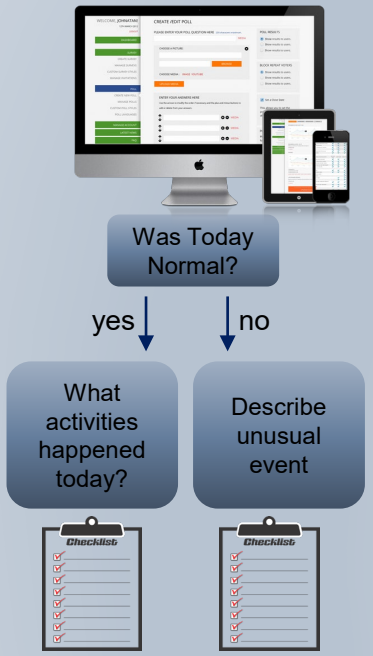
Removing Barriers

Data access, format conversion, time-alignment and missing data point management consumes > 60% of time spent on analysis projects.

Analysts must chase SME's to overcome incomplete / esoteric documentation.

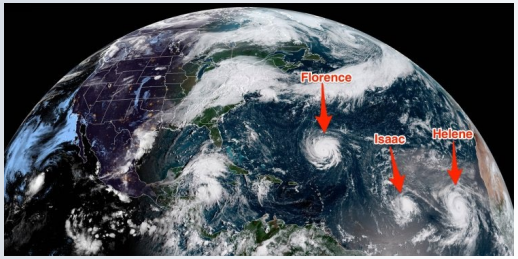


Continuous Knowledge Capture Tools – Observational Science in Ops!



- Simple, easy, high-level daily questions
- Gives overall daily context as labels
- Ever-growing list of standard categories
- Machine-readable database

- Time range annotation (labels)
- What happened when, with categories & notes
- Important patterns & relationships between channels

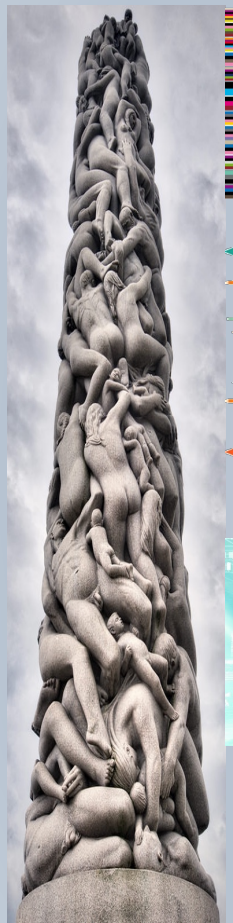


- Anomalies should be treated like Hurricanes
- Each has formal name, trajectory, details
- Start & stop time clearly delineated
- Characteristic features specifically highlighted
- Minimize freeform text!

Delivered solutions – Parallel pipeline for zero disruption

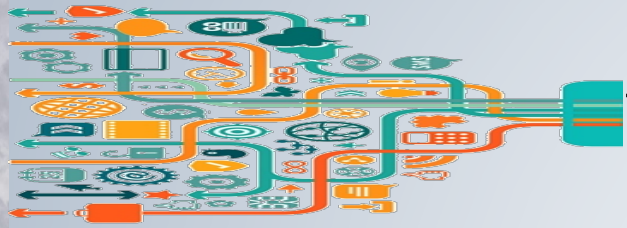
New Pipeline

Established Ops Workflow
(it's made of people!)



Translation Library

- Converts current formats to Data Science formats
- Decodes Ops-specific datatypes (e.g. on-change)



Data Alignment / Harmonization System

- Common time bins, densification (L3 equivalent)
- Merges Metadata

Analytic Database

- On-demand data regridding, filtration
- On-demand data computation results
- On-demand featurization for DS use cases

Development Lessons Learned

Data Access
- Data source APIs optimally need to provide time series for downstream use

Extensibility
- Isolate file format (mission) specific code
- Fast data point inserts for batch and continuous modes of operation

Data Harmonization
- Consider the frame of reference
- Beware of features whose data types evolve over time
- Revisit repeated value time series

Analysis Ready Data
- Store sparse and query to meet the needs of the user
  - Binning rate, interpolation strategies, etc.
- Capture accompanying metadata early in spacecraft design
  - Utilize flight rate of occurrence to prioritize creating metadata for active missions

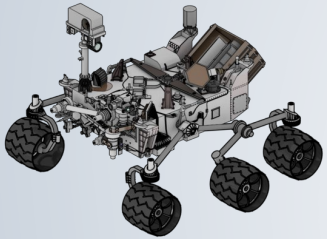
Tools Enabled by New Pipeline

Fast Exploration of ARD

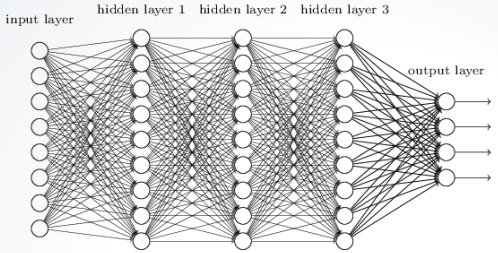
Anomaly detection / Focus of Attention

Interactive Data Dashboarding

Expert Labeling Tools



Multi-Subsystem Trending Tools



Machine Learning Model Infusion

Jet Propulsion Laboratory
California Institute of Technology