

# **Automation of VLBI data analysis pipeline**

Leonid Petrov, NASA GSFC, Code 61A

# I. Long-term goal:

- To catch-up GNSS analysis and remove man-in-the-loop
- To shift a work of a data analyst from using a mouse to analyzing pipe-line logs



- To eliminate “analyst noise” and make results reproducible

## II. Current state

- Scheduling VLBI observations — **highly automated**
- Data flow of fs-logs, schedules, submission of L1, L2, L3 — **highly automated**
- Data flow L0 to correlation — **requires manual work**
- Correlation — **requires manual work**
- Visibility analysis — **requires manual work**, recent significant progress
- Group delay preprocessing — **requires manual work**
- Source imaging — **requires manual work**
- Final VLBI data analysis — **moderate level of automation**

### III. Key performance parameters

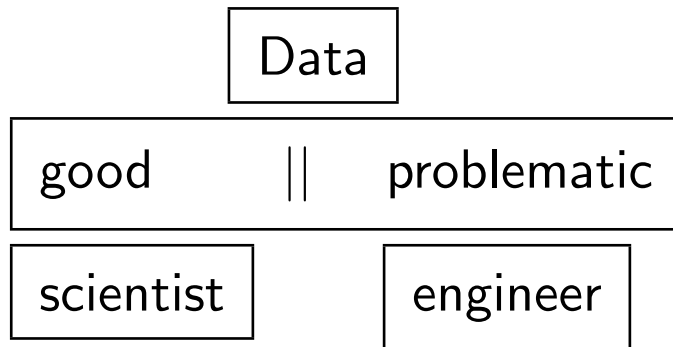
- amount of man-hours to run a given VLBI experiment
- latency of data analysis
- failure rate

## IV. Goals, objectives, and outcome of the development

- identify areas of high ROI
- to reduce the amount of manual work
- to enable fully reproducibility

## V. Approach

- evolutionary approach
- to gradually replace elements/subsystems that require(d) manual work with an automatic solution
- to perform R&D in automatic quality control and correction
- to set a less ambitious goal:



- to develop a testbed

## VI. Required resources

- 1.0 FTE split between a Lead Scientist and a skillful software developer
- A high-end server computer (40K-50K)

## VII. Infusion plan

- overall top-to-down design
- revealing bottlenecks
- setting priorities: what should be automated first
- early adoption of partial automation



## VIII. Issues/Concerns/Risks:

- cultural change: we got used too much to manual processing
- cultural change: *if is not broken, don't fix it* mentality
- finding and training personal will be a challenge
- availability a test-bed data for correlation automation

## IX. Timeframe/Schedule:

- 2 years to complete a basic end-to-end pipeline, not counting time for training
- tuning the pipeline: indefinitely