SF23A-0044

Phases of Science Software Development

(Do Phases of the Moon Affect Phases of Science Software Development?)

Albert J. Fleig, PITA Analytic Sciences, 8705 Burning Tree Rd, Bethesda, MD 20817, fleig@comcast.net

Mike Linda SAIC, 7501 Forbes Blvd. Suite 103, Seabrook, MD 20706, Mike.Linda@gsfc.nasa.gov

Curt Tilmes NASA Goddard Space Flight Center, Code 922, Greenbelt MD 20771, Curt.Tilmes@nasa.gov

Concept Development

Four to seven years before launch of the instrument into space.

Mathematical description of the algorithm concept, typically for a perfect instrument, with perfect calibration.

Algorithm Theoretical Basis Document, ATBD, may be required for external review during this phase.

Portions of the algorithm may be coded and tested with sample data.

Input, output, and metadata formats not considered.

Programming and processing system development staff should be in place by end of this phase.

Note that the proposal and the budget for the entire life of the project, including resource estimates for staff and processing, are due long before this phase is complete.

Nominally Complete Preliminary Algorithm

Should be completed about **two years before launch** for "at-launch" products.

Reads properly formatted input data from predecessor algorithms.

Writes properly formatted output data.

Includes required metadata for processing and archival.

Runs with test input data, provides a first indication of ultimate processing resources (e.g. flops, memory, I/O).

Requires a reasonable approximation of the at-launch physics be coded.

Requires that predecessor output format has been defined.

Requires that conventions about output format and metadata have been defined.

May require agreement with other Project participants about variable names and units.

Requires a source of input test data with physically reasonable values and formatted in a reasonable approximation of the at-launch format.

Requires support systems such as a test processing system, programming standards, delivery guides, and configuration tracking be in place.

Pre-Launch Test Version

Tests start **12-15 months before launch** and this version must be available then.

Runs automatically in the intended operational processing system.

Ingests all needed input and ancillary data in the nominal at-launch form.

Produces output with all of the required metadata to support downstream processing and the data distribution and archival system.

This is a good time to define, code and test downstream or Level 3 code for presenting the results of the algorithm. The physics may still be in work, but it is now possible to decide how to grid, bin, and display the initial results.

Includes code for all important aspects of at-launch algorithm including checking input data and generating output data and metadata in the correct format. Includes quality and error flags for output products.

Requires agreement on file naming conventions and process initiation methods.

Requires a substantial amount of input test data (and associated ancillary data) in the anticipated at-launch format.

Requires a scaleable implementation of the at-launch processing system be implemented and available for these tests.

Requires participation in seemingly endless pre-launch test planning, evaluation, and report meetings.

Launch-Ready Algorithm

Often required to be in place **3 months before nominal launch.**

Corrects problems found during the pre-launch tests. Also incorporates changes in input data resulting from problems found in the pre-launch tests.

Incorporates updates to the core physics as algorithm development and understanding of anticipated instrument performance continues.

Includes checking error flags and quality flags as defined for the launch-ready input products that are used by this algorithm.

Has been tested successfully with real input data collected during pre-launch tests. (This is a test of input, output, and error handling routines, not of the core physics.)

Requires that the at-launch processing system, staff, and procedures be in place and functioning well.

May be updated several times as launch slips, but needs to be complete in case it does not.

Tools for reading, plotting, and investigating the intended output products should be available during this phase.

Launch and Early Operations

Typically **3-6 months from launch until nominal operation** begins and up to a year or so before instrument characterization is adequate.

Spacecraft and instrument are operated in engineering deployment and early test modes.

Instrument gains, temperatures, and voltages may be non-nominal.

There are likely to be "undocumented features" (blunders) in input data sets. Instrument operation and calibration are often not as expected.

Errors may also be found, and need to be corrected, in the code for the product under discussion.

Substantial effort will be needed to determine whether unanticipated results stem from real features in the phenomena being measured, instrument performance problems, calibration algorithm artifacts, or unexpected features of the processing algorithm.

There is often a period of rapid change to the processing algorithm as actual performance of the instrument and real observation of the phenomena to be measured are obtained.

Rapid changes in the instrument operation, calibration, and processing algorithm often result in multiple versions of the output product being available. Configuration management and a method for determining when changes should be implemented are required.

First Release of Data

The first release of data to people outside the Project often occurs **9-15 months after launch.**

Once the major pre-launch blunders have been found and corrected, instrument characterization and algorithm changes stabilized, and the data reprocessed from the beginning, it is useful to release the data to a wider audience with two goals:

Allow for wider inspection and identification of possible product artifacts.

Allow users to become familiar with data content and format and to develop their own analysis programs.

While this first release of data may support some investigation of the phenomena being measured, it is often recommended that publication be deferred until a validated data set is available.

Validated Data Release

Typically **9 months to two years after launch.**

Produced by reprocessing the released data period with consistent versions of the input data and processing algorithms.

Accompanied by a validation report describing the accuracy and precision of the data, the methods used for determining this, and any known problems in the data.

This data set is intended for use in studies and publications subject to the limitations stated in the validation report.

Ongoing validation effort is required for the entire lifetime of the data set. The validation report covers only data for the date range specified in the report. It is understandable to hope that data produced after the report will continue to be equally as valid, but this is often not the case. Instruments degrade with time. Situations not observed during the validation period occur. There may be substantial changes in the observing regime (e.g. volcanic eruptions producing long lasting aerosol changes.)

Even though the data has been validated, algorithm development is likely to continue and instrument characterization to continue. Periodically it will be decided that the enhancements produce data that is sufficiently improved that reprocessing the entire data set with the best current knowledge is warranted.

Final Processing

Often **6 to ten years after launch.**

At some point the instrument will fail or the flight project will come to an end. This often leads to one final reprocessing of the data incorporating everything that is known about how to improve the data.

At the completion of this "final" reprocessing, care should be taken to archive all of the production algorithms and all of the programs required to produce the tables used in the algorithms.

It may also be necessary to find a home for archiving and distributing the data once the project is dissolved.

Post-final Processing

Ten to fifty years later.

If the measured parameter remains of interest and similar instruments continue to be developed and flown, it is likely that there will be further improvements in retrieval technology during follow-on projects. Especially in cases where long term trends of a parameter are of interest, this may lead to a desire to produce a consistent long term climate data set using the best current techniques with data from both current and predecessor instruments. An example is the Total Ozone Measuring Spectrometer instrument, TOMS, first flown on Nimbus 7 in 1978 with subsequent versions flown on multiple spacecrafts. Twenty three years of this data was just reprocessed in 2004. Measurements at wavelengths similar to those of TOMS are currently being made with the OMI instrument and will be made with the OMPS instruments to be flown in the next decade. It is likely that there will be further reprocessing of all of this data as algorithms continue to evolve.

People retire or leave the project. Computer systems and languages evolve. Institutional memory fades with time. It becomes increasingly difficult to remember all the details of how a data set was made. At the same time, computing technology improves and what was once a daunting data processing task becomes relatively trivial. Reprocessing twenty three years of TOMS data only took a week on our new system once the algorithms were successfully ported. The importance of archiving complete information and documentation of both how the processing was done and why certain decisions were made was brought home by this effort.

This work was carried out under NASA contract number NAS5-00220.