

START TO END OF LINE ON 11/20/92

## **Some Comments on the PGS Toolkit Study Report**

Ed Masuoka

MODIS SDST/Code 920

November 19, 1992

Here are my comments on the toolkit report. I look forward to hearing more about the toolkit in early December. I've attached a copy of Tom Goff's comments in case you haven't already received them. I've used your section headings and titles to help you located where the comments refer to. Please send me email at emasuoka@ltpsun or call on 6-7608 if you have questions.

### **3.3.1 Production Control and Scheduling Tools**

While the PGS processing is largely event and data driven, I believe that there is a time driven aspect as well. For example, should there be a requirement to produce a MODIS product every 24 hours, then at some point say after 20 hours the MODIS processor must run even without the necessary data for all products.

The toolkit should include tools that will enable programs to save their processing environment (files, control structures and data structures maintained in memory) and processing history in the event that they must suspend processing. Dr. Robert Evans from the MODIS Ocean Team has suggested that this should be handled by a SQL-based tool that will facilitate storage and retrieval of processing history and complex control sequences that will be needed either: in the event that a process must be suspended due to non-receipt of required data or to rapidly implement changes in processing that are the result of regional changes in the oceanic conditions.

### **3.3.1.2 Initialization Tools**

Given that Run IDs, Stream IDs and Product IDs will be inserted into a standard product's metadata, we would like to have tools which allow science processing software to recover this information from the stored metadata to facilitate reprocessing. I agree with Tom Goff's comments about including both forward pointers (to the algorithms that used the product as input) and backward pointers (to the products and the algorithm that were needed to create it) in the metadata.

Tools for determining what resources are available, requesting the resources and reserving memory and other resources for the duration of a program or processing stream would be very useful in MODIS processing.

### **3.3.1.2 Termination Tools**

We definitely need to be able to pass resources between programs within a production stream and keep track of and save the resources if the stream is interrupted, in particular data structures in memory and temporary disk files. .

### **3.3.2 I/O Tools**

Direct access to a given record or block of records is highly desirable for the standpoint of efficiency and should be supported in the tool kit. I want to retrieve and update block x without touching all x-1 blocks before it.

Won't RAM based mass storage or RAM cache buffers be transparent to the user, i.e. appear like very fast disks even without the tool kit?

### **3.3.2.2 Production Stream File I/O Tools**

We plan to keep as much data in memory or RAM disk as possible during the production stream. Sharing these data structures with other teams could bog down the PGS unless we write them out as disk files.

I assume that these stream files exist online on the PGS only and are never stored in the DADS. I am also assuming these files would be kept around for a relatively short time (less than a few days).

### **3.3.2.3 Product and Ancillary Data Access Tools**

Who is conducting the data structures study and when can we hear more about it?

What is the schedule for delivery of the IMS and DADS toolkits?

#### **3.3.2.3.2 Level 1-4 Product Access Tools**

How do the MODIS investigators and MODIS Support Teams influence the decision process on Level 1-4 standard structures?

#### **3.3.2.3.3 Metadata and IMS Access Tool**

"Tools for querying IMS or DADS for the availability and content of data products may also be provided" - please change "may also be provided" to "will be provided". These tools will be important when we are reprocessing or querying the IMS to see if a data set we require for our processing from another instrument has been produced.

#### **3.3.2.3.11 Error Output Tools**

Tom Goff has suggested a processing log file in which the MODIS data processing programs write status messages, errors and so on. This file could be examined in the event of unexplained errors in the processing or archiving of MODIS products. One would still want a separate status log file, which holds information strictly on the completed data sets.

#### **3.3.2.3.13 Data Validation Graphics Output Tools**

I would make this a low priority unless it is clear that the IMS toolkit will not provide visualization and graphics capabilities.

#### **3.3.2.3.14.1 Digital Terrain Elevation Data Access Tools**

You will need to run these tools only once or twice for a given topographic data set to get the interpolated surface to use for elevation corrections during production processing.

#### **3.3.2.3.15 Tools to Access Common Physical Constants**

I assume these tools will be implemented as constants in include files.

#### **3.3.3 Math&Modeling Support Tools**

The mathematical operations mentioned in this section should be handled by commercial math and stat libraries. I think the phrase "is likely to be provided by" should be changes to "should be provided by" to emphasize our commitment to use COTS software, such as IMSL.

### **3.3.3.1 Math and Statistics Libraries**

I prefer IMSL libraries having used them over the years. They are used in the LAS image processing software package developed by Code 930 and EDC.

Will you present results of the survey of EOS science investigators regarding their preference for a given math and stat libraries at the December presentation?

### **3.3.3.3 High Performance Processing Tools**

In this section, you refer to "the toolkit" which implies to me one toolkit. However, given the likelihood of PGSs with different architectures aren't we really talking about as many as seven PGS toolkits, each of which contains tools that take advantage of the specific architecture dependent features of the PGS running in a given DAAC.

In your second paragraph you point out that the SCFs may require special tools to emulate the PGS architecture-specific functions. Given the variety of choices for computing systems at an individual SCF, one may be faced with developing code to emulate these functions for each of the seven PGS on a number of different vendor platforms (SUN, HP, DEC, SGI and IBM to some of the systems already in place in MODIS SCFs).

### **3.3.6.1 Remote Database Access Tools**

I think this is a useful tool set provided it is used ahead of the main production stream to stage the latest data to near-line or online mass storage prior to processing.

### **3.3.6.2 DAAC Specific Tools**

See comment under 3.3.3.3 High Performance Processing Tools section.

### **3.3.6.3 Debugging Tools**

If programmers from SCFs have access to the PGS or a PGS-clone like the one in the MODIS Team Leader's Computing Facility via Internet, then it may be more efficient for them to use the full suite of interactive debugging tools on the target PGS rather than a subset of these functions in a toolkit.