# Tech Session (Day 2 of 3)

## Annenberg Auditorium, Caltech, Pasadena, CA

22 September 2016

**Attendees:**

Johnson, Mafi, Huber, Chen, Law, King, Crichton, Hardman, Stein, Guinness, Semenov, Garcia, McLaughlin, Joyner, Hughes, Gordon, Padams, Nagdimunov, Radulescu, Isbell (by phone), Dan Scholes, Neakrase (phone), Stone, others.

**Tool Overview (Crichton):**

The Tool Working Group is meeting via telecon, first Wednesdays since August; agreed on objectives. Mapping Level 3 requirements to tools, identifying tool gaps, and overseeing registration. Hope to have PDS-wide plan for tool in FY17. Members include Crichton, Guinness, Johnson, King, Padams, Palmer, Lin, and Hardman. Contact via pdstwg@list.jpl.nasa.gov.

Level 3 Tool Requirements include assistance in generating (1.5.1), validating (1.5.2), and submitting (1.5.3) products; documentation for installing, using, and interfacing with each tool (1.5.4); and inspecting file contents (3.3.2), translating among selected formats (3.3.3), translating among coordinate systems (3.3.4), and visualizing data (3.3.5).

It would be nice to have a PDS4 equivalent of NASAVIEW (the PDS3 version is still being downloaded about 150 times per month), a label (or template) generator that uses input from other than a PDS3 label, and a LID uniqueness checker; but there is a Search capability using LID.

**Tool Registry (Hardman):**

The Tool Registry was spawned by a desire to expand the scope of an IPDA tool registry as well as from a request by MC in September 2015. The intent is to include all tools into a single registry. It will be built on top of existing PDS4 components. Product_Service has been updated to support this application (CCB-123).

The initial version is at http://pds-gamma.jpl.nasa.gov/tools/ tool-registry Fifty-seven tools have been registered so far. Updates and new submissions are requested; submissions are reviewed before posting. Submission via Safari do not work (yet).

**Core Tools (Hardman):**

Core tools are used for labeling, validation, and submission. Sources include EN, DNs, MIPL/MGSS, and others.

**Design tools** include Label Creation and Editing (LACE) from NASA Ames, APPS Label Design Tool (LDT) from MGSS, and the PDS4 XML Document editor. The LACE host at Ames will be closed for security reasons in the near future. LACE software is now at EN but Mark Rose continues to make updates (slowly because he has no funding), LDT has been streamlined and may be maintained by EN (if there is PDS funding), and the last has only been proposed by King at the last MC F2F. Neither LACE nor LDT has a large enough user community to justify selection as a core tool; the plan is to test and evaluate them to determine which would be the better selection as a core tool.

The current plan for **generate tools** is to translate XML schemas into a template, which can then be filled out using a command-line program embedded in pipeline processing. Two approaches utilizing the Apache Velocity template engine are being considered — the Generate Tool (which currently translates a PDS3 label to PDS4) and IGPP Docgen (which has been developed by PPI and can use several types of input).

Several DN representatives said they have developed their own 'generate' tools, which (1) confirms that such a tool has been needed for a long time and (2) brings into question the need for a centralized initiative to create a single tool now. King said this underlines the desirability for collaboration within PDS; but he added that JPL rules about shared code have been getting in the way. PPI can share code with JPL; but, if JPL makes modifications, it cannot be shared back. There was disagreement about the importance of these restrictions.

**Validation tools** check labels and products; associated schemas and Schematron files specify syntactic and semantic constraints, but there are additional constraints in the Standards Reference (which have been captured in a Validation Requirements document). The existing Validate Tool validates single products and aggregate products with referential integrity. EN is awaiting final delivery of validation code from NASA Ames. Once in hand, EN will determine what needs to be added.

**Transform Tool** supports image translation among several formats and supports Table translations to CSV; solid examples are needed for testing. The tool transforms the digital data object but not the label, since some of the output formats aren't acceptable PDS4 formats and there would be no LID for the new product.

**Inspection Tools (Hardman):**

NASAVIEW is overdue for a replacement; the new tool should support both PDS3 and PDS4. Requirements will be gathered (e.g., 3.3.2 and 3.3.5) and there will be preliminary design/development of a more general **Inspect Tool** in FY17. For requirements, there will be a look back at TBTOOL and OAL in addition to NASAVIEW; SBN has also developed a PDS4 Viewer, which will likely be relevant. The Inspect Tool will be handed off to a new hire next week; development will probably be in Python and there should be a development plan by Thanksgiving. RMS has a library of low-level Python code that could be incorporated.

**Tool Issues (Crichton):**

1. Transformations
2. Inspect Tool (high priority)
3. Integrate software efforts
4. Capture NASA Ames software
5. Shift to service
6. Easier way to request tools/services

**ISIS and PDS4 (Isbell):**

ISIS was developed as a cartographic/spectrographic product many years ago; the current version is ISIS3. A PDS4 ingest capability is being developed; Chris expects this to become available as missions require support. Code is written in C++.

**Igpp.docgen (King):**

This software is based on earlier work by IMG; it is based on Apache Velocity and can read PDS3 labels, text files, spreadsheets, and CDF metadata. Apache Veloicty Template Language (VTL) is text plus Apache directives (mark-ups). The output can be text files, including well-formed PDS3 labels or XML. It is available at http://release.igpp.ucla.edu/igpp/docgen.

**MIMIC (King):**

MIMIC manages a collection of files that have profiles (time stamp, size, checksum, etc.). It is used to create mirrors of file collections. It can be restarted if interrupted; it uses checksums to ensure fidelity, transfers data with parallel tasks, and caches inventory to improve efficiency. rsync is an alternative but it generates a complete file list with checksums for each run; this is significant overhead. There may be problems if the bundle/collection hierarchy is not mimicked by the directory structure (including secondary members). Some (but not all!) synchronizations can be accomplished in 15 minutes.

**XML Display (King):**

You can define a style sheet to go with your XML; with that you can customize display on your browser. It requires addition of a processing directive before the root document tag in each label. Todd loads the style sheet into the same directory as his XML label. There may be other ways to reference and access the style sheets; implementation may depend on the browser. In an example (pds4.xsl, which is on the meeting web site). Todd can click on a <local_identifier_reference> and the browser jumps to the referenced object.

**On-Line Archiving Facility (OLAF) (Stone):**

OLAF is a self-service facility for archiving (http://sbntools.psi.edu/olaf/); it has been used for over 10 years and many users can go pretty far in the process before SBN personnel need to intervene. It is used by smaller data providers who do not have pipelines. By design, it makes the same kinds of decisions for similar products, which are limited to tables. OLAF is presently in transition between PDS3 and PDS4.

**A General PDS4 Data Tool (Nagdimunov):**

Lev has developed a PDS4 viewer (http://sbndev.astro.umd.edu/wiki/PDS4_Viewerwhich shows the label in a format similar to Todd's, though it appears to be more compact. Plots of table data and images can be displayed. Todd said he has tried using it on PPI data; some products were quite large, and he was pleased with the results. Why not call this the Inspect Tool?

Lev has looked at the more general problem of reading and displaying data. He was asked to maintain a program which read data files; he encourages programmers to 'write to the standards' rather than reading data. On the other hand, he wants software to be used by scientists (not other programmers). Python 2.6+ or 3.3+ with optional NumPy seems to be a good choice. He is not worried by large data objects (there are 7 GB images in PDS), but users usually want to study only small portions of these. He has been challenged by fields which have arrays as elements, discontiguous fields (apparently both in MAVEN data), unbounded ASCII integers (that would map to more than 64 bits in binary) and reals, collapsed text fields, and delimited tables which have no length limits.

Hughes asked whether Lev is asking for changes in standards/policies or whether he is simply noting ambiguities. Lev did not give a clear answer, but several in the audience picked up the discussion … with no resolution. Steve thanked Lev for bringing up the issues.

**PDS4 Label Design Tool (Padams):**

LDT is a web-based label tool that does not require knowledge of XML; a team of interns is reworking the previous version of the software. It allows the user to create a specific label for an existing product; it could then be edited to create a template if there were many similar objects needing labels. Go to https://pdsimg-services.jpl.nasa.gov/ldt/ for the current software, which Jordan demonstrated.

**Geoscience Tools (Scholes, Stein):**

**Archive Management Tool**: AMS tracks the integrity of the GEO holdings; it is used in-house monthly. There is a data base of archive directory and file contents which can be compared against current contents; the backup holdings are checked every six months. Only a few errors have been found — during early runs, when there was accidental corruption.

**Label Generation Tool**: This tool merges product metadata in spreadsheet form with templates to produce PDS4 labels; it is being used to move Brown ReLab spectral data into PDS4. Some

ReLab data were ingested into PDS3, but there were integrity issues, and it has been decided to start over.

**Image Processing Library**: This library is used by GEO personnel for PDS3 image processing functions, such as stretching.

**Discussion (Crichton):**

King said the Registry is a good idea; but many DN personnel would like to install potentially useful tools on local systems, exercise them, and share experiences. Without this, developers will continue their old habits of developing software locally, often replicating efforts at other nodes. Crichton replied that there is the intention of calling Tech Sessions yearly to stimulate collaboration. King said he doesn't care about the formal design process for tools; he just wants to use the tools and see how they work. For some functions, Todd would like to 'pick one' and move forward, if we're going to work together.