

Day 2, 01: Objectives and Overview of Tools/Tool Working Group Report

- This afternoon will be for questions about using tools
- Formation of tool working group recently

Charter:

- Level 3 requirements: making sure requirements are up to date
- Are we capturing what exists and how do people get to it?
- What are the gaps? How to prioritize fixing these gaps?
- When we have overlapping tools, what do we do? Input, collection, culling
- Making sure nodes have input about requirements, functionality
- Supporting testing, giving feedback

Logistics:

- pdswg@list.jpl.nasa.gov
- Monthly telecom, attempting to create traceability, mapping, gap analysis

Work Plan

- Not just NASA version: what is international version of tool registry?-IPDA
- PDS-wide planning for tools in FY17

Level 3 Tools Reqs

- generation, validation, submission, documentation
- design vs. generation
- Services: inspection, translation, visualization

Gap Analysis

- EN and other nodes
- Sharing tools: bring this up this afternoon, how to make this more effective in PDS?

Status

- Google Sheet gap analysis:
 - 46 tools: some not in registry
 - planning category gap
 - submission and peer review categories missing: is there a common tool that can be developed?
 - Design, peer review, analysis/reader requirements missing

- Roadmap discussions to address some of the gaps
- PDS4 version of NASAView: how o maintain 2 versions of a tool? PDS3/PDS4
- Functionality needed: LIDS correctness, integrity

Near Term Actions

- How to store TWG notes? Google Sheets, Redmine
- Discussion at Mgr council meeting

Discussion: Is it possible to search by LID?

- A: Yes, you can search by LID. Want tailored interface, though.
- Bundle identifier requirements, services vs. desktop applications
 - Web-based?
- We should have a web-interface that we can just enter tool and track it. What is the way to capture requests?
 - Can be brought up at next meeting
 - Sean's JIRA: can enter tools issues at <https://pds-jira.jpl.nasa.gov/secure/Dashboard.jspa> (might have to request for permission)

Day 2, 02: Tool Registry

[Map](#)

Overview

- want to expand IPDA Tool Registry, want to make it less static, more easily searchable
 - new registry uses Info Model
 - better integration of information
- Question: how will search criterion be monitored?
 - Will be discussed tomorrow, is a goal, but not tool registry related

Architecture

- Separate search interface for registry: takes you to online services area
- Migrated existing tools into product labels, haven't done resources yet (there are a lot)

Status

- Not on public site yet, on PDS gamma

- **Need better descriptions, registration: please do this Nodes!**
 - review current list of tools and submit updates or new entries
 - Need to make this a formal action item
 - Need to go through registry: weeding and entering based on public usability/interest

Search Interface Demo: <http://pds-gamma.jpl.nasa.gov/tools/tool-registry/>

- Categories of search related to Information Model
 - Category might be changed by Sean based on how well it fits in category
- Search tools are the biggest category
- Is PDS version important for something like a search tool? Does it matter?
 - Not really. But, can't select PDS3 AND PDS4 at the moment because of IM
- When you select tool, you get detailed info, link to tool, requirements, description, category
- Submitting new tool:
 - Tool is desktop-based, service is web-based
 - Required fields are red: some fields need to be optimized, can't be submitted in Safari
- Question: When fixing PDS3/PDS4 section, can you also add N/A?
 - One of the issues I have is that when PDS3/PDS4 is selected, it goes into the registry on the backend. The selection of PDS3 or PDS4 will be optional to narrow if you want, but you won't have to select one or the other
- Product label: minimal info
 - Adds internal reference to which node the tool came from

Next Steps:

- Resolve current issues: Safari, URL, edit entries
- Allows you to enter release date info: might not actually be important, have to update often
- Verifying that there is a representative set of tools to release to the public site
 - Is there a timeframe for this release?
 - By next TWG meeting, two weeks. Operational by mgmt council meeting
- **BUG:** Firefox windows 10: When clicking on tool, it is linking to other tools
 - timing issue with search interface
- Tying tools to missions?
 - this is done on the back end
 - Titles: MSL vs Mars Science Laboratory, will yield different results
 - will be updated in the backend search

Day 2, 03: Core Tools: Label design, Label Generation, Validation, and Transformation

[Map](#)

Overview

- provide tools that satisfy 1.5 x and 3.3x requirements
- Emphasis on Open source software
 - transformation tool does good job of images from PDS3, but doesn't support every PDS3 revision

Lifecycle

- Focusing on core

Approach: develop set of common core functions

- various libraries present through PDS3 and PDS4 iterations
- Architecture: Label functions and data functions built over open sources and PDS community
- Design: early in project, focused on use of Oxygen/Eclipse, now we have other software being developed-how do we determine what is best going forward?

Design Tools

- LACE
 - no funding to maintain this
- APPS Label Design Tool: Jordan will present later
 - will have to decide between LACE and APPS at some point
- PDS4 XML Document Editor- Todd King
- Question: What is Mark's obligation to do this?
 - He doesn't have funding, Ames has no more money for this. He's doing it out of niceness
 - It's not a sure thing
- None of them have a large enough community to justify selection
- Money is there for testing and evaluation of functionality, will allow feedback
- By EN taking control of LACE, it can be made more accessible
- Question: For LACE: it should be available by November?
 - I hope. Hoping to get an eval done, then web-based and desktop available. Need to get all of the software first
 - When I do get LACE, it's probably not updated to the latest version of the model

Generation

- Mostly command-line tools that could be integrated into a mission pipeline
- You have to put a template together to use the tools

- Generate Tool
- IDPP DocGen
- Talk about merging these two tools into one: low priority, need more motivation/need to make this work
- Other nodes have written other tools, this shows that maybe nodes needed a generate tool
- Question: Are the other tools in Velocity templates?
 - Maybe? Knowledge for writing velocity templates can transfer
 - Compare advantages and disadvantages between different generate node tools
- Would you care if your tool was replaced? Java-based, python?
 - Is this the best use of resources?
 - This illustrates why open-source and collaborative environment is so important
 - Issues of resources, necessity, licensing, lack of access to source code
- Comment: You will evaluate design tools, but lack the motivation to do this with generate tools. Please do this with generate tools too. **Want to have something to give data providers.**
 - Answer: less visibility, less demand for generate tools. We can get it moving forward, though.

Validation

- use XML to do bulk of validating labels
- Validation Requirements Document: additional constraints from PDS Standards Ref.
- What it does now: validates single products and aggregate products
- Waiting on final delivery of the code: have a lot, but not all
 - now have single validation tool for PDS3/PDS4
 - Will give a desktop version of the volume validator
 - have not seen, don't know much about its functionality
 - Might be downtime on validator (few weeks), but it can be brought back with Mike's help
 - Want to have this in the build, but clock is ticking, need review of requirements
- Question: Core library API not available: Can you make an example of how to validate just one label and share it?
 - Yes. I have it. Changes in versions prevented me from releasing it, but I can
- Question: Has anyone written their own validation tool? Besides Ron?
 - No.
- Question: PDF-A document validation in a bundle?

Transformation Tools

- Transforming product labels and dat to/from PDS4 formats
- Only available command-line tool, but looking to make available as service
 - what kinds of reqs, hosting, deployment questions
- We need solid examples of well-labeled products with data to run through tools- for testing and for users
 - consistent and updated examples from discipline nodes
- Question: PDS4 → CSV, is this available for PDS3 too?
 - not in this tool, but inspect tool does. In the works, planned

- Question: What transformations are you planning to support in the future?
 - Up to the nodes. Need to prioritize a list of those transformations, tool working group will be looking at this
- Question: Is this something I can set up to do batch jobs? Does it give me PDS4 label?
 - Have to write script to do that. Making Velocity template to do create PDS4 label from PDS3
- Question: FITS Image in PDS4 label?
 - like Array 2D image, FITS images are somewhat supported now, but we had issues with it before.
- Question: CDF, FITS, HDFI: Any plan to have array transform to these?
 - Not yet, but we can add your input to TWG
 - Action: create a list, talk to TWG
- IPDA review comments: documentation issues, a couple bugs -will be fixed in this build
 - Greater need for documentation overall, they are working though the most obvious
- Suggestion: tool documentation is very terse, maybe beef it up to help guide
- Array 3D spectrum with more than one band to JPEG, or GIF, what would it do?
 - you just get one band
 - Scaling factors are being applied now
- table delimited to CSV, what is it doing?
 - not a lot. you'll always have commas
- Question: Who writes this documentation? SEE: transform tool, operation, known issues, PDS4 image issues
 - Sean writes some of it. Will look at this further, needs further attention
 - Share this feedback as you're reading it

Next Steps

- Evaluating design tools
- Merging, writing generate tools
- Evaluate Validate tool
- Prioritize additional transformations

Other core tools: package tools, tracking service?

- packaging —> transport service: will talk about tomorrow

Day 2, 04: Inspect Tool Plans

Map

Overview:

- Working on replacement of NASAView, long overdue
 - most popular tool, most fixes necessary
- New tool will support PDS3/PDS4
- Evaluate Lev's tool, see if they can bring in his functionality and integrate into new tool
 - brought someone on (start next week), will be in Python, most likely

Requirements Gathering

- satisfy existing requirements going forward
 - Look at Lev's requirements
- list of tool requirements and what it should support, then get feedback from node community
- What will new hire do if you're still gathering reqs?
 - help gather reqs, hoped to be further along

Design

- Initial thoughts on design: prototypes with AMES
 - browse products tool from collections, bundles,, or data set, with thumbnails
 - tool recognizing what it is, transforms available, start visualizing tables, something like that
- Question: there is use for tool that looks at single product
 - That functionality will not be excluded
- Desktop and browser-based application

Development

- Looking at developing in Python
 - Lev has hesitations
 - past apps, libraries in Java-how to reconcile with Python?
- Are you aware of Rings Node Github Python Library?
 - Yes.
- Question: (slide 6) What does desktop application mean to you?
 - i.e. NASAView, essentially a GUI running on desktop

- Question: Python/Java- have you thought about which libraries you might make available publicly? Commercial software support?
 - Will borrow and share with other nodes internally, making source code available publicly will be tricky, libraries not complete enough

Next Steps:

- Have reqs set by Thanksgiving, 2016
- Have progress report for mgmt council meeting
- code in Dec.
- Once code is out, will be more opportunity for feedback

Comment: With Transform/Translate tools, this tool is most helpful in getting communities to like PDS more

Comment: Many users want tool to bring data into their own environments effortlessly —> commercial applications

- want data more quickly, not ending to download, figure out how to view
- Python library with methods —> pipelining into their own storage
- Reader library in tool registry?
 - Yes. But you can't find tools registry.

Day 2, 05: Tool Takeaways, discussion

Map

- 1) Transformations and prioritization of those
- 2) Inspect tool
- 3) Integrate generation tools, or summarize, understand
 - Those who made their own tool: do you want to have it in comparison?
- 4) Capture all Ames software
- 5) Shift to open source structure
- 6) It would be nice to have a quick, easy way to suggest tools and submit needs
 - not JIRA
 - web form, timed response
- 7) Software development environments
 - commons area

- Identify process for code sharing, communication/tracking about requests
- Should PDS think about broader Github type space, environment for sharing?
- Is it possible to write a generate tool? The work goes into setting up the template, how to plug the numbers in. Where is the common tool?
 - there is single template tool
 - The tool can be common through the processor.: i.e. PDS labels, extract image metadata, can access types of data we typically label
 - This is complicated problem
- Need to have JPL make code available to the rest of the community
 - Todd hosting Github repository, we can contribute
- If JPL employee writes code and contributes to an open-source project, are there stings attached?
 - Has to be signed off to the source i.e. Apache
 - Greater movement in direction to contribute to open source at JPL and in PDS
- Does it make sense for us to piggyback on existing Github or make our own?
 - Easier to just use Github, headache to manage
 - JPL has internal Github for lab, NASA has project on Github
 - PDS project on big NASA Github?
- In process of generating new context bundle—soon
- Tool for peer-review?
 - in TWG Google Doc

Day 2, 06: Node-specific Tool Presentations

[Map](#)

ISIS and PDS4

- ISIS3 & PDS4: not using PDS4 libraries
- ISIS is working on ingest application
- looking to do PDS4 export in the long run

igpp-dogen

- MAVEN uses this to generate labels
- document generator based on Apache Velocity
- Output can be text files
 - release pages for software
- What is Apache Velocity?
 - Velocity Template Language, allows content of docs to be parameterized
 - Is like it's own programming language

- variables are objects and have properties and methods
 - velocity loops: can iterate over variables with multiple values
- standard java libraries can be integrated, helps with creating PDS templates
- once you design pattern and get data entered, you can generate PDS4 labels quickly and in a bulk way
- Question: couldn't provide full path to data files I wanted to use?
 - can't have colon in path, is separator: maybe we can update this?
- Question: We want a link to download?
 - On first slide, submitted to tool registry
- Self-contained package; can download and run pretty much immediately, pretty easy to get running
 - download, bundle, to directory
- What are the important formats to include going forward?

Mimic

Tool used to copy and sync archive

- manages collection of files with profiles
 - can determine what's new and what's changed
- Why a new tool? other tools are not designed for syncing archives
 - Mimic manages inventories, so it will only copy what is new, not all of the files
- Cached on remote site and locally, doesn't need to do checksums against all items, quicker running times
 - must be collaborative, Mimic works better when operating on both sides
 - Cache the checksums, not running checksums every time
- Mirror: to make someplace copy of archive/data set
- Bundling Collection: from PDS, can create bundle in Mimic
- Planned: mimic as user tool
 - generate inventory for transfer
 - originally designed system-to-system, but users might be able to benefit
- When you clone a bundle, do you assume it mimics the product tree?
 - is a Mimic bundle is not the same as a PDS4 bundle. Mimic has multiple folders with Mimic config files, inventories
- How much data is being inventoried and synchronized from archive?
 - 20TB in 20 min, if no changes
- Does it work the other way around? Can you do rate limiting?
 - Not built-in, you would have to do it at different level

Formatted Views of XML

Ways to make it easier to view in a browser

- attach XML stylesheet to render nicer in browser
 - some restrictions based on browser
- Do I need PDS stylesheet for each directory?
 - can have relative path info, don't know if browser can determine multiple

- stylesheets
 - has to be colocated
- need additional statement in every file identifying stylesheet

OLAF: Online Archiving Facility

Self-service portal for submitting products to the PDS

- in use at SBN for 10+ years
- Akin to the “turbo-tax” of PDS
 - shields users from inner workings of PDS
- Used by mostly smaller data providers, smaller amounts of data
- Only produces tables currently, other formats in dev
- User logs in, uploads, will generate bundle, collections, products and docs —> peer review
- Transitioning from PDS3—> PDS4

Question: LADEE LDEX issues —> did resolutions get pushed back into OLAF?

- yes

DEMO:

- User logs in, fills out fields
 - supported at collection-level and product level
- To list of data types accepted
 - tables are the furthest along, can describe more details about table
- description of how tables are formatted, column descriptions
- Question: requires 4tran (?) format or pos eX (?) ? **[not sure of file formats here, needs review]**
 - 4tran now, could make it accept both
- Created dataset and loaded product to it, described product, its fields and the metadata
- Also has admin interface
 - currently data is tied to peer-review, but is going to be changing

Development

- better usability, supporting more data dictionaries, more flexibility, access by other nodes

Question: What happens if you have multiple tables?

- OLAF supports batch upload capabilities

Question: Can do one bundle per one collection, but what about multiple?

- one data collection per bundle currently, but want to expand that

When might version be available for wider use?

- Possibly a few months, if you're willing to ride out bumps

Question: mostly for one-off data sets, or will work for accumulating data sets?

- Mostly used for one-off. We would have to consider plans for accumulating data sets
- Seems really useful for DAPS
 - easy tool to point individual providers to
 - makes decisions and adds info-can be customized
- SBN wants to make tools useful for those unfamiliar with PDS, don't need to know all of the standards
- Used OLAF for Apollo restorations and it worked well, can be good for PDS4

PDS4 Viewer

(World's quickest) Demo:

- label view similar to Todd's style guides
- table viewer: you can plot, change colors, parameters, type of table
- Image viewer: like DS9, can scale, set parameters, colors

Comment: Good Inspector tool, can it be the inspector?

- need it to work on PDS3 and PDS4, have to replace NASAView

Trying to write general PDS4 data tool (mostly the reader, not the viewer)

- writing to standards is difficult, allowing all PDS4 possibilities
 - don't write to examples, look at the standards, see what you need

Project start

- was partly for reviewers
 - does label actually describe data structure?
- IDL is proprietary, reviewers and users had complaints
 - here's python

Users

- scientist, non-programmer
 - make it simple for user to download and install into python
 - what version of python to run? Both 2 and 3
- Dependencies: NumPy is optional, LXML better than library that comes with python, SciPy

Ram usage is not concern for SBN, but is on PDS-wide scale, this tools is developed for the PDS as a whole

- memory mapping: don't want to read all the memory at once because of memory limitations, but this isn't supported

Labels and Group Fields

- arrays described in group fields: had to re-write code to support this because it is allowed

Unbounded numeric types: can have ASCII integer greater than 64 bits

- in Python, a big integer will take 24 bits to store it, because it is an object—> 24x memory multiplier
 - have workarounds

Reading labels

- collapsed text fields: defined by IM, not by XML, this matters when writing software

SEE for PDS4 Viewer: http://smbdev.astro.umd.edu/wiki/pds4_viewer

Question: Statement about deciding if we want discontinuous fields?

- this software does support the standard but it makes software more difficult
 - not likely revisiting this

Comment: As a developer, areas of the model are under question-- is this what was intended?

- do we want to revisit ambiguous aspects of the model?
- Needs to be discussed: make a list

PDS4 Label Design Tool

Upgrade:

- web interface for creating pds4 labels
- Use IM model questionnaire
 - as a user, you never have to look at or know XML
- Produces valid PDS4 label, but only validates using schema, not value validation
- Warning: tool is in beta
- SEE: <https://pdsimg-services.jpl.nasa.gov/ldt>

Demo:

- selecting product type, product information, walks through different areas
 - all items being driven by the model
- Discipline dictionaries are present
 - some bugs and disabled content still
- After getting through node-level items, you can add mission specific information and

then exports valid xml label

- Can leave some of the fields blank, as a way to just get a template for the label, then you can manually fill them out
- Can have pull-down of items, or auto-populates based on schematron files, can be added as long as its in JSON
- Question: will the source code be provided so there can be a self-hosted version?
 - Yes, I think we can. Costine pushing for making this open
- Want to abstract away the name of some of the fields
 - Ask questions about type rather than just say what the value title is
 - Want to use this for DAP-like proposals

Plans

- Fix bugs, validation updates,
- scientist-centered interface: the terminology is still too PDS-specific
 - Questionnaire-like interface (Turbotax)
 - Advanced user option-can fill out form directly
- Question: You have authentication service- tie into LDAP security service?
 - can be done, not currently
- Question: how long before it's fairly complete? beyond product observational
 - We need to get this out before the beginning of next year because of all the DAPS, next 3 months

Geoscience Node Tools

Archive Mgmt System: track integrity of archival holdings at directory and file level

- database holds baseline catalog of archive: file and directory level
- Run monthly validation scans
 - will summarize scans and report through GUI, can sort
- Question: how often do you find corruptions, or changes?
 - mostly when it was first implemented, not often
- Uses command-line interface
- Needs to add bundle for PDS4

PDS4 Label Generation Tool: merge metadata in spreadsheet form with templates to produce PDS4 product label files

- Field corresponds to column in spreadsheet
- simple config file: helpful for PDART
- Question: migration or is it new?
 - this is new
- Can I get my hands on it?
 - Yep. Not extensive documentation, but you can use it.

Image Processing Library: internal for Node personnel (analyst's notebook), image processing

- Needs to bring in XML parser to make functional for PDS4
- Customized linear stretch of single- and 3- band images
- Allows for overlaying user parameters on the image

Day 2, 07: Wrap up

Map

Lyle has peer-review tool, allows a few fields for commenting

- Good for forcing people to categorize their comments

Discussion: How do we achieve better sharing and community support?

- tool registry is one way to raise awareness
- want more collaboration and use each other's tools
- Continued tech sessions for sharing opportunities
 - if we don't meet regularly, we become divided
- EN focus on predictably, but as nodes continue to develop, other focus may be necessary
- How do we track issues for tool? Maintenance?
 - as a community, we have to decide what we want to do and exactly where we can go to locate information
- Cross-PDS tool dev environments?
- Challenge: not wanting to over-impose, but allow sharing
- Want to know about what people are working on, even if it's not finished
- Value of actually meeting in-person
- TWG can help with the importance of commons area

Tomorrow: 8:30-noon: focus on services