

## **PDS Face to Face Technical Session hosted by EN at JPL Pasadena, Sept 24 and Sept 25, 2008**

### **Attendees**

Atmospheres (Lyle Huber, Joni Johnson), Engineering (Sean Hardman, Paul Ramirez, Steve Hughes, Emily Law, Ron Joyner, Dan Crichton, Elizabeth Rye, Julie Wang, Betty Sword, Richard Chen, Mike Cayanan, John Ho), Imaging (Chris Isbell, Patty Garcia, Alice Stanboli, Rafael Alanis, John Diehl, Bob Sucharski, Chris Braadshaw, Jeff Anderson), Geosciences (Ed Guinness, Susie Slavney, Tom Stein), NAIF (Boris Semenov), PPI (Todd King, Steve Joy, Bill Harris), Rings (Mitch Gordon), Program Office (Al Schultz, Mike Martin), Small Bodies (Anne Raugh)

### **Day 1 – Sept 24, 2008 Notes**

Dan Crichton, opening presentation: “PDS Technical Session”: pds-tech-session-20080923.pdf

- Purpose – Ensure PDS-4 meets node needs, understand gaps in plan, address Ray Arvidson’s questions, begin to form high level architecture. Collaborative session. Not a design review, but input to design. “Implementation agnostic.”
- Key goals from PDS2010 Vision Statement: simplified, rigorous standards, adaptable tools, on-line services, reliable, scalable infrastructure, etc.
- Dan – need extensible system; not determine specifics which might be budget-dependent; challenge in communicating an architecture: providing a useful view for stakeholder. Selected Zachman framework – trying to come up with MC view.
- Rather than worrying about specific functions at this point, Dan wants to make sure we design an underlying infrastructure flexible enough to support whatever specific functions we later determine are appropriate.
- Susie – Ask what are the artifacts, output, accomplishments to bring back to node management.
- Anne – raise concern about assumptions made in Ray Arvidson’s questions; Ed: states that the MC pretty much bought off on these assumptions.
- Todd– note that Dan’s infrastructure allows us to leverage off of capabilities and contributions of the various nodes.
- Lyle – suggest outcome to recommend to the MC depends on budget.
- Anne – question about ordering of the vision.

Mitch Gordon, Data Architecture (DA) Team presentation and discussion: “PDS 4 Data Architecture Part I (1 &2)”: PDS4\_DataArch\_A\_part\_I-1.pdf; PDS4\_DataArch\_A\_part\_I-2.pdf

- Mitch exercises – attempt to reach consensus among nodes
- Unresolved fundamental issues/ Strawman Information Model:
  - Exercise 1: Prioritize customers/drivers for DA

- Things that were “missing” from Mitch’s chart:
  - Jeff – accommodating new contemporary formats.
  - Rodney Heyd – handling large data volumes (felt to be more system than data architecture; scalability in terms of capacity)
  - Todd – customer is HQ and managers – need reporting capabilities.
  - Chris – locating and identifying data; making sure users can search for and find things in the catalog (granularity?)
  - Ron – subsetting of data (Mitch – related to granularity)
  - Ron – absence of tools (sort of sub-setted under multiple)
  - Paul – tracking (related to Todd’s point on reporting) understanding life cycle of PDS data
  - Patty – reporting and assessment (Todd reporting thing)
  - Susie – data needs to be easy to find from outside PDS as well as in (Google example)
  - Elizabeth – data integrity, quality assurance (tension with taking complex (but simple to use) onboard S/C formats and converting them to simple archive formats – potential for introduction of errors)
  - Steve – easily (in a technical sense) extensible standards
  - Todd – many of these aren’t drivers, they’re the things that we’re trying to accommodate
  - Lyle – adaptable
- Discussion and prioritization of Mitch’s A through Q list
- Final overall rankings: (see also: PDS 4 Data Architecture Reprise.pdf, slide 4)
  - a) 72
  - b) 36
  - c) 39
  - d) 29
  - e) 63
  - f) 26
  - g) 19
  - h) n/a (same as “c”)
  - i) 37
  - j) 14
  - k) 15
  - l) 11
  - m) 22
  - n) 13
  - o) 35
  - p) 7
  - q) 2
- Exercise 2: Weigh characteristics of “Consensus Extreme” versus “Distributed Extreme”. See slides 6-11 of PDS4\_DataArch\_A\_part\_I-1.pdf for aspects of each “extreme”.  
See slide 5, PDS 4 Data Architecture Reprise.pdf for results.

- Action Item – provide comments and input to result spreadsheets.

Sean Hardman, System Architecture Team presentation: “Preliminary PDS 2010 System Architecture Specification”, pds2010\_system\_arch\_spec.pdf

- Action Item – provide comments / input to Architecture scope questions after given more thought – “we want a rigid core PDS and local extensions”.
  - Q1 – information model has multiple sections, for now, focus on observational data, administrative services data (deliveries, reporting, tracking) need work.
    - question - should model include operational information like deliveries, reviews, orders, housekeeping, etc.?
  - Q2 - Subnodes are part of DNs, should they be integrated into the system, or interface to the system?
  - Q3 – Grandfather existing projects – probably not, need negotiation case by case basis, migration vs bridging?
- PDS, NVO interoperability - how much to expose our services interface?
- Action Item – provide additional comments input to Architecture Principles
  - Statement that no data released to public until after peer review is inconsistent with some mission and node policies (slide 16).
  - Data accessibility is not represented in the implications of data stewardship principle (slide 16).
  - Dan suggested that Ed write up recommendation for MC to consider the priority of long term preservation and system usability.
  - Data corruption, natural disaster, et al., could disrupt nominal ops, but need to recover quickly – can’t completely avoid disruption. (slide 17).
  - Reword and clarify the intent of “control the technical diversity of system across the nodes” (slide 18).
  - Change “a data dictionary” to “a data dictionary model”. (slide 21).
  - Not all requirements are initiated by the MC, but approved by the MC – change “levied” to “approved” (slide 30).
  - Nodes do not necessary want leading technology , but stable technology, should there be a principle/guideline for EN to monitor and perform technology investigation (particularly emphasized by Anne).
  - Software (e.g. compression), documentation be addressed as part of the data stewardship.

Steve Hughes, Data Architecture Team presentation: “PDS Data Architecture, Developmental Approach”: PDS4\_DataArch\_080923\_B.pdf

- Steve - would like input from the team if the WG has covered all the bases in terms of the model.
- Todd – voice there’s fuzziness between logical and physical model.
- Ed – ask where are the requirements captured for the data architecture.

- slide 6 - discussion of relationships among PDS and external entities: PDS “references” = passive, PDS “adopts” = active, PDS “assimilates” = grab standards but then don’t care about their further development.
- Also slide 6, “heliospherics” should be “heliophysics”.

Mitch Gordon, Data Architecture Team presentation and discussion: PDS Data Architecture Part II (1 and 2)”: PDS4\_DataArch\_A\_part\_II-1.pdf, PDS4\_DataArch\_A\_part\_II-2.pdf

- Mitch - expect what comes out of the meeting is consensus of the tech group of what issues need to be sorted out by the MC, what guidance is needed from the MC, the sense of approach and direction WG is going in. Need to be able to say what we hope MC meant by “fewer, simpler formats”.
- Implications of “pruning PDS3”, or “Fresh Start”.
- Fresh Start Straw Man; Minimalist Approach; four simple Base Structures, supported Abstract Classes and User Classes .
- Complex example - ISIS cubes not part of encoded stream base. what presented was simply an example of an approach to add structure to PDS3, keep in mind to distinguish archival format vs processing format.
- Lyle – identification and documentation of why, as well as how much and what data are in the base structures, identification of tools to accommodate this structure, need to analyze cost and benefits of migrating the data.
- Elizabeth – raise concern, not convinced straw man structure sufficient to address SPICE and a repeating record structure (e.g. variable length records – Dick Simpson’s records); model needs to be robust enough to handle “basket case” data (good examples are GEO’s TES and Viking IRTM data – ask Susie for details).
- Elizabeth - will / should decoding an encoded stream base ~~will~~ render one of the other base structures?
- Mitch - point out that PDS has to change policy of not modifying the raw data if trying to meet the drivers providing simple formats and simplifying the base structures
- Mitch - need guidance from MC on what should be driver(s) (i.e. is current user community or long-term archive the higher priority?)
- Lyle – also need to worry about backward compatibility .
- Ed - Separating the underlying (byte-level) structure and objects is necessary.
- Part2-II, slide 11 “heterogeneous” should be “homogeneous”.

## Day 2 – Sept 25, 2008 Notes

Chris Bradshaw presentation, “PDS Imaging Node Management Portal”:  
PDS\_IN\_Management\_Portal\_lg.pdf

- Address data integrity and reporting requirements, leverage Open Source S/W, scalable
- Use Samhain for data integrity functions (inventory files, checksum, logging), not much differences in comparing performance against md5deep.
- Use AWStats for web stats.
- Backup management, Doc & Collaboration, Reporting are under development.

Rodney Heyd presentation, “HiRISE JPEG2000 Experience”: HiRISE JPEG2000 Experience.pdf

- Benefits: Allows for GIS metadata, multiple resolutions in single file, more efficient network transfers, lossy or non-lossy compression (example: 29.6 Gb uncompressed:1.79 non-lossy compressed).
- Issues: Limited client s/w supporting jpip, client s/w doesn't support all JPEG2000 capabilities (geotiff metadata).

Sean Hardman presentation, “PDS Tools and Platforms”: pds\_tools\_platforms.pdf

- Planned deployments of upgrades to VTool, LTDTTool and NASAView targeted between now and end of calendar year.
- Need to update the EN project pages .
- Elizabeth - wants her data dictionary tool (slide 2 – Sean suggested not likely to happen until PDS4).
- Tom Stein - volunteer to provide a 64 bits machine for C-base application, and/or testing (slide 3).
- Survey of current platform support across the nodes (Anne, Lyle say Solaris 10 okay).
- Need to upgrade to Java 1.5 (per Sun plan to end Java 1.4). There were no objections by attendees of Tech Session.

Sean Hardman presentation, “System Architecture: Core Services and Definition” (starts on slide 32 of “pds2010\_system\_arch\_spec.pdf”)

- Six stakeholders (NASA Program, MC, System Engineer, Data Engineer/Software Engineer, Data Provider, Data Consumer) with unique viewpoints, views and concerns.
- (Slide 35) Anne -The MC is also concerned about how the system addresses the needs of the mission/data providers (and what's involved in migration to a new system). Sean - Resource allocation is also another MC concern.
- Views mapped to Zachman Framework, chosen to model What, How and Where aspects.
- By F2F MC in November, have further identified services and their provisions.
- Slide 49 discussion:

- Elizabeth – does registry just contain identifier information or location info as well? Sean – just identifier now, could include location info later. (slide 49)
- Ed: where are the services hosted (e.g. ingest)? (Do you envision all data ingested coming through some kind of single application?), Dan: map services to central vs distributed paradigm, or both. Partitioning of the provision has yet to come. Sean: this is the “grand” or “ideal” view, some of it might be centralized (ex. catalog ingestion) other parts node specific. Lyle: services only represent half of what we do, before ingest, there’s a lot of work needed before products are finished for ingestion. (The diagram is missing all the things we do to the data we receive from data providers before ingestion (like validation, generation of documents, etc. This diagram works once the product is finished, but we need another diagram describing the processing of getting to the point where the product is completed.) Sean: Services have API. Dan: Services are web/network based applications. The preparation of products may be tools segment that’s not shown. But will be addressed.
- General questions about where security should go.
- Action Item : add another diagram to address Lyle’s comment above
- Sean: provide more detailed description, more detailed functional view, more structure focus on application architecture for F2F in Nov (Nov. 21-22).
- Implementation governance : more about change control
- Mitch - would like to have some kind of service/approach to give heads up to other nodes about certain development for more coordination (such as the IMG’s management portal).
- Dan - need more views of the system (Should more internal functions (standards, node responsibilities) be shown somewhere on the diagram? Sean: probably, should maybe be included in operator part of the diagram. Elizabeth: Software? Sean: could be included in Document box.)
- Dan - suggest interim tech telecon before Nov. MC? Todd: get updates out and circulate them via email. Sean: Would definitely like feedback on services diagram. (slide 68).
- Action Item: updated materials to tech members by end of October.

Mitch Gordon, Day2 Discussion, “PDS4 Data Architecture Reprise”: PDS 4 Data Architecture Reprise.pdf

- Todd: why is PDS4 straw man better than what we have for PDS3?
- Elizabeth: “Data storage and interpretation” implies for archive, right? (I.e. other formats permissible at node) Anne: yes.
- Tech group needs to define fewer and simpler formats/structures and propose to the MC. The MC just wants to make sure that users and data providers are not complaining about their complexity. (Mitch is trying to focus on what MC meant by “fewer, simpler formats”.) Todd disagrees that there’s any ambiguity here. Susie says we should stop worrying about what the MC was thinking and try to tell them what we think it means.
- Jeff - add “ with rigid meta data standards” to “ a few, simple structures for data storage and interpretation”.

- Todd – does not see a benefit of the statement.
- Mike - MRO HiRISE, how the change of these structures will affect them?
- Todd - may be that we don't find any of our formats meeting long-term requirements, then we need to define new formats, Anne: it depends on the drivers and the whole point of the exercise.
- 1 node 1 vote for the change of statement .
- Mitch - need input from the tech group to revise the exercise 1 category list (duplication, etc), next time around will be 1 node per vote. Data Archive Group will re-issue with explanatory paragraph (relevant to whole system, not just data). Based on outcome, prepare recommendations to MC what drivers /characteristics should be included in PDS4.

Dan Crichton - WRAP-UP: tech-session-wrap-up.pdf

- Capture:
  - Decisions made
  - Items for the MC
  - Action Items for Tech Group
  - Recommendations to Working Groups