Appendix A: PDE IRB Sub-Committee Details

Archiving (storing data, formatting, metadata requirements, documentation, redundancies and gaps)

Summary of Sub-Committee Objectives:

The Archiving Sub-Committee will analyze and assess the current state of the PDE archiving processes and archive locations to provide findings and recommendations on how PSD data is effectively being archived. This Sub-Committee will focus on whether data are appropriately being prepared for archiving, if the right resources are available to take data from its source to an archive, and whether an appropriate archive is available to the community. This Sub-Committee will tackle topics related to:

- Standard formatting (Standard data types used in the community?)
    - Are the current data standards working for the community that works with the data?
    - Do all the archives and databases within the PDE work effectively together?
- Documentation/metadata (Sufficient to enable going from zero to functional?)
- Submission tools (Sufficiently enforce format/metadata standards?)
- Improvements/upgrades (Upgradable as processing, etc., improves?)
- Long-term curation (Long-term maintenance of archived data, models, etc.)
- Redundancies and Gaps (multiple archives, centralized resource, etc.)
    - What PDE archives and databases include the same data and, if so, what value does that add to the PDE?
    - Is there a PDE element for all the data the community is interested in accessing?

PDE elements included

- PDS archives
    - Discipline node (DN) data holdings: Atmospheres, Cartography & Imaging Sciences, Geosciences, Planetary Plasma Interactions, Ring Moon Systems, Small Bodies; also NAIF support node
    - DN Data Node holdings: THEMIS, HiRISE, LROC
    - Registry Catalog
- NSSDCA
    - Spectral data archives and collections: HITRAN/Harvard, RELAB/Brown Univ. (coming to PDS Geo), TIR/ASU, VIS-NIR/USGS, CRISM spectral library (PDS Geo)
- Sample archives and databases
  NOTE: The initial PDE IRB will NOT review sample curation of returned samples, meteorites, IDPs, or those collected at terrestrial analogs. While it is understood that data on these samples may be contained in some elements of the PDE, it has been determined that in order to prevent too much expansion of the scope of the PDE IRB,

and the expertise required, this subject area will be excluded. It is PSD's intention that the issues in this area (such as the lack of an archiving system for terrestrial analog data) will be assessed later and fully, intentionally, integrated into the PDE. As such, these elements will NOT be prioritized for review by the PDE IRB:

- Astromaterials Acquisition and Curation Office JSC
- Meteorite samples
- Asteroid samples: Hayabusa/Itokawa samples, OSIRIS-REx/Bennu samples
- Genesis, Stardust/Wild 2, Cosmic Dust, Microparticles, etc.

However, these elements will be prioritized for review by the PDE IRB:

- JSC Curation online catalogs, such as:
  - Lunar sample catalog: Apollo rock sample image archive, Apollo Samples and Photo catalog, Lunar Sample Compendium
  - Antarctic Meteorite Database
- AstroDB (inc Moon DC; format derived from/related to PetDB, Columbia Univ.), MoonDB/Astromat

- Data archive tools
  - PDS/PDS4: Generate, OLAF, PLAID/APPS, Validate, Transform, PDS4 JParser, GDAL, MakeLabels, Local Data Dictionary Tool (LDDTool), Docgen – create PDS4 labels from PDS3 labels – PPI, Mimic – stream data, synchronize servers - PPI
  - HAPI (Heliophysics API) – stream time series – Heliophysics Division

2. Searching

discoverability of data, open access, translation.

Summary of Sub-Committee Objectives:

The Searching Subcommittee will analyze and assess how easy it is for a community user to navigate the PDE and whether there are improvements that could be made to facilitate this navigation. This Subcommittee will provide findings and recommendations on how open the PDE and its elements are to a variety of user types (e.g. established researchers, citizen scientists). This Sub-Committee will tackle topics related to:

- Open Access (Effective open data policy? User-friendly access to datasets?)
    - Does PSD have an effective open data policy that is implemented across PDE elements? How would compliance with such a policy be enforced?
- Discoverability (Easy to find datasets? Search by metadata? Synthesized?)
- Usability (Datasets are “science-ready” / “analysis-ready” for end users?)
- Translation (Enhanced data usability by translating into other formats?)
    - Does the PDE enhance data usability by providing resources for users to train themselves or others to become more sophisticated data users?

PDE elements included

- Archives (see Archiving Sub-Committee elements)
- Tools (see Archiving and Utilization Sub-Committees elements)
- General Public Data Access
    - Planetary Photojournal

Utilization

(tools for analyzing data, skill development, open software, software development)

Summary of Sub-Committee Objectives

The Utilization Sub-Committee will assess and analyze whether the infrastructure of the PDE is effective in allowing the community of users to work with the data and if the appropriate resources are available to facilitate this work. Additionally, this Sub-Committee will explore whether the program infrastructure to support tool and software development is sufficient to keep the PDE evolving and stable. This Sub-Committee will tackle topics related to:

- Open Software (Effective open software policy?)
    - What definitions and language are appropriate to differentiate needs and policies as they relate to software?
- Discoverability (Easy to find/use software tools?)
- Development (Support for software development, esp. early career?)
    - Are citizen scientists supported by our development of tools and software to utilize the PDE?
    - How do elements developed in the PDART program move to be key elements of the PDE with long-term, stable resource support?
- Data Skill Development (Skills for providing foundational products?)

PDE elements included

- Community data visualization, search, and access/retrieval tools
    - PDS
        - Analyst's Notebook (Geo), Atlas (CIS-JPL), Annex (CIS-USGS), Ferret (SBN), Marsviewer (CIS-JPL), Small Bodies Image Browser (SBN), ODEs (Geo), OPUS (RMS), Viewmaster (RMS), UPC,PILOT (CIS-USGS), Map-a-Planet (CIS-USGS), Photojournal (CIS-JPL), PDS4 Viewer (SBN), NASAView (EN), Cosmographia (NAIF), Spectral Library (Geo), Aspera high-speed data transfer (Geo), WebGeoCalc (NAIF), CATCH (SBN), MPC live DB distribution (SBN), MPChecker (SBN - MPC), MPC DB query tool (SBN - MPC), Splash – time series data analysis and display – PPI, PNG Walks PPI (Iowa), VISTA PPI
    - USGS
        - Planetary Nomenclature (on behalf of IAU); Working groups for planetary bodies, coord syst, etc.
        - Geologic Mapping & GIS: MRCTR and GIS data and tools; apatial data access initiatives; community training events
        - IAA Spatial Data Infrastructure: Astropedia; Geospatial data metadata, cataloguing, access initiatives; ISIS3 planetary cartography software; Geodesy program (Coordinate systems, reference frames, IAU

international coordination and working groups); “Cartography” program and data products; community training events
    - JMars/Moon, etc.
        - Geospatial data access, tools, visualizations, processing
        - Community training events
    - Treks
        - Geospatial data access, tools, visualizations
        - Community training events
    - Quickmap, ACT (Erick Malaret)
    - RPIFs
        - Hardcopy, unique data collections, books, maps, etc.
        - Digital data access, tools
        - Community outreach and training events
    - Small Body Mapping Tool
    - Planetary Data search tools for/within Astrophysics Division:
        - ADS (often used to find PSD products, including PDS holdings)
        - IRSA NEOWISE image viewer & photometry search (IPAC, Caltech)
    - Autoplot – University of Iowa
    - Topcat – Virtual Observatory Alliance
- Community data processing tools
    - Ames Stereo Pipeline
    - GDAL
    - MRCTR GIS lab and tools
    - Projection on the Web (POW) tool at USGS
    - Mission tools (sometimes shared within teams only): HiView, HiRISE, Univ AZ; LunaServ, LROC, ASU
    - Quickmap MESSENGER/LROC/M3, ACT
    - Titan Swatch Viewer (TSV), Cassini team
    - Community toolkits and collaborations: OpenPlanetary, PlanetaryPy, SpicePy, etc.
    - PDS: SPICE Toolkit, WebGeocalc (NAIF), Others
    - AI4Mars
- Planetary Spatial Data Infrastructure (PSDI) elements
    - Foundational products
        - Missions
        - Community derived products
        - Geodesy program
            - GSFC elements
            - Missions
    - Policies
    - Standards
        - PDS
        - FGDC
        - ISO

        - IPDA collaborations, tools, services
        - International coordination with IAU, national space agencies such as ESA/PSA, JAXA,ISRO, CNRS, IVOA etc.
    - Working groups, collaborations
- NASA Derived Data Products, Software
    - PDART
    - DAPs
    - Other
- NASA GitHub
- NASA Planetary Github
- PDS tools and databases
- USGS Astrogeology ISIS3
- NASA software
    - From missions
    - From DAPs

4. Mining & Automation

(searching and using large datasets, high-end computing, cloud computing, ML/AI)

Summary of Sub-Committee Objectives

The Mining and Automation Sub-Committee will analyze and assess how emerging computing techniques, resources, and services support the PDE and PDE users. In particular, this Sub-Committee will focus on current and future best practices for a data ecosystem that should be brought to fruition for the PDE to meet current and emerging user needs and improve the PDE's effectiveness in serving the community. This Sub-Committee will tackle topics related to:

- Access / Cloud Computing (Commercial services meet needs? Plans for future?)
    - What would be the advantages and disadvantages of moving PDE elements to the cloud environment?
- Analysis / HEC (Sufficient resources provided to community? Reasonable application process? Resources up-to-date?)
    - Do the current HEC resources meet the computing and processing needs of the community?
- Automation / ML (How is this being used/coordinated in PSD? Plans for future? What can/should the PDE do that best enables users to use PDE with AI and ML methods?)
    - Example. Are the current investments in automation and streamlining sufficient to see maximal gains?

PDE elements included

- High-end computing (HEC) facilities
    - HECC/NAS, ARC: Data Portal, ECCO, HELIO, QUAIL
    - NCCS, GSFC
    - Others?
- Cloud computing resources (NASA, commercial)
    - Amazon AWS
    - Microsoft Azure
    - Nebula IaaS/Paas/SaaS/DBaaS, ARC
    - Google Cloud
- PDS data services
    - PDS API
    - Data in Cloud?
- AI and machine Learning
    - JPL Machine Learning and Instrument Autonomy Group
    - JPL Artificial Intelligence Group
    - AI4Mars

5. Inter-relational

(relationships among the ecosystem elements)

Summary of Sub-Committee Objectives

The Inter-relational Sub-Committee will analyze and assess the stakeholder and management elements of the PDE. This Sub-Committee will provide findings and recommendations on possible ways to make the PDE more cohesive and clearly connected. This Sub-Committee will provide insight into potential advantages and disadvantages of management and communications structures and identify gaps that should be filled in order to boost the collaborative nature of the PDE. This Sub-Committee will tackle topics related to:

- Ecosystem governance
- Ecosystem cybersecurity
- Discoverability of data (SC2) relies on searches/filters based on metadata (SC1)
- Software/models (SC3) requires datasets (SC2), should these be stored in the same place (SC1)?
- Community outreach
    - How can we discover how well the community is being served by the PDE?

PDE elements included

- Project offices and program management: PDS, NSSDCA, PDCO (MPC, Harvard CfA/SAO; CNEOs, JPL; IAWN, UMD) AMMOS, JPL
- Mission-critical/Mission-enabling activities: JPL HORIZONS system, JPL NAIF/SPICE, Telescopes (incl radar? Goldstone? DSNs? TDRS?), Planetary Spatial Data Infrastructures, community outreach and training efforts / skill development (see also panel 1), training by PDS for PDS, training by PDS for community (including SPICE training classes), outreach events at national meetings, meetings (domestic and international)
- Planetary Data Workshops
- PSIDA
- Mission team events: by and for teams, for public and data users, tools and tutorials (including SPICE tutorials and self-training packages)
- Stakeholders: NASA HQ, Other NASA divisions, NSF, NASA Missions, International missions, Individual data providers, PDS, planetary data users, PDS Data and Archive Working Groups, MAPSIT, CAPTEM, IPDA, Cartography Working Groups, general public (educators, students, citizen scientists)
- Relationships with and among stakeholders: communications, support, feedback from user groups, surveys, etc.