Engineering and PDS 2010 Project Overview

Dan Crichton

March 22-24, 2010

Agenda

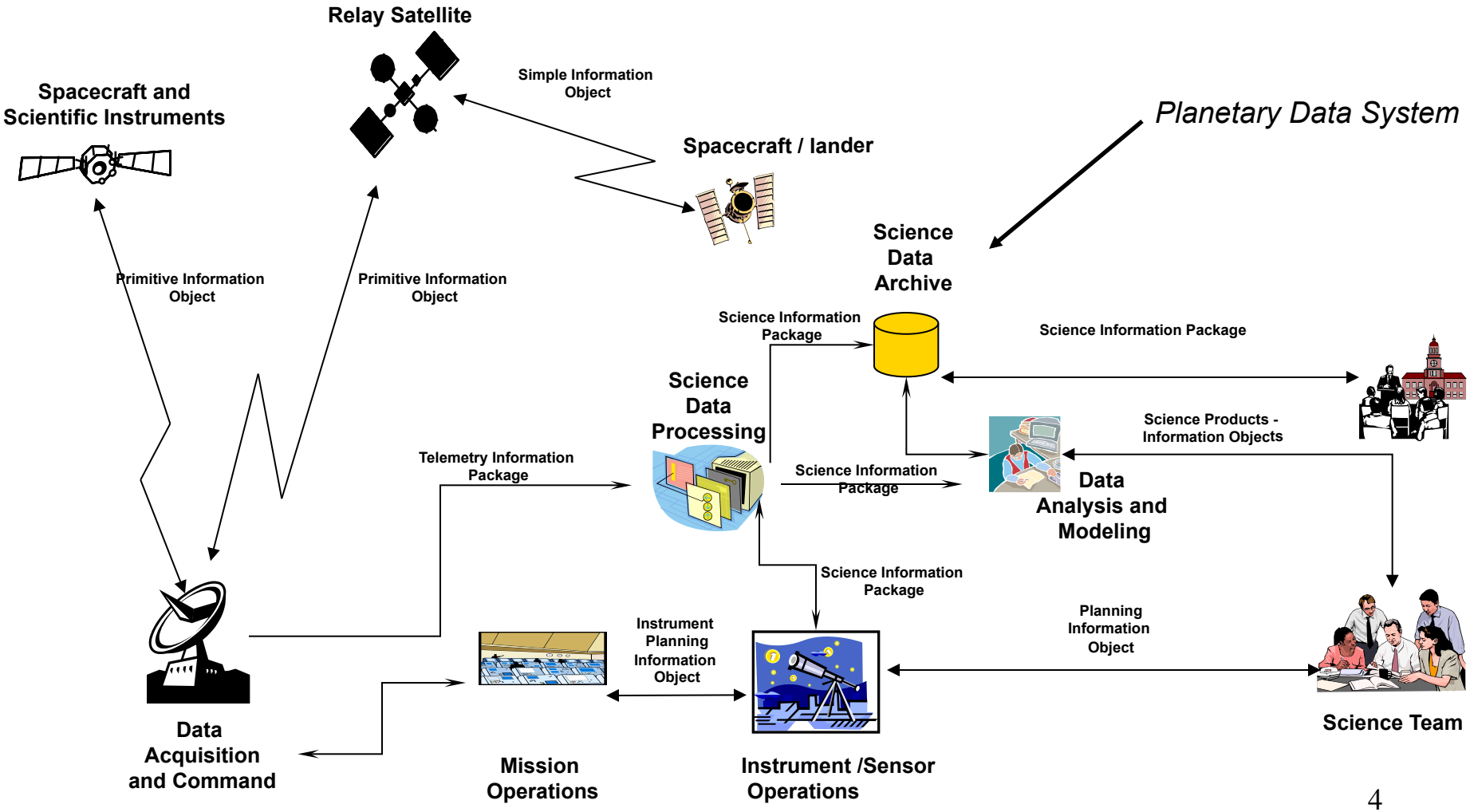
- Engineering Overview/PDS3 Implementation
- Architectural Drivers
- PDS 2010 Project Overview
- High Level Architecture Concept
- Major Design Decisions

Engineering Overview

- PDS operates as a set of distributed nodes that each include
  - Software and Storage Infrastructure
  - Scientific and Data Archiving Tools
  - Databases
  - Expertise to support data archiving
  - Mission-specific data nodes

- Engineering Node provides systems engineering, common software and standards that helps PDS operate as an integrated system
  - A mirror site is in place at UMD

PDS in Context



Credit: CCSDS Reference Architecture for Space Information Management

PDS Requirements

1. PDS will provide expertise to guide and assist missions, programs, and individuals to organize and document digital data supporting NASA's goals in planetary science and solar system exploration

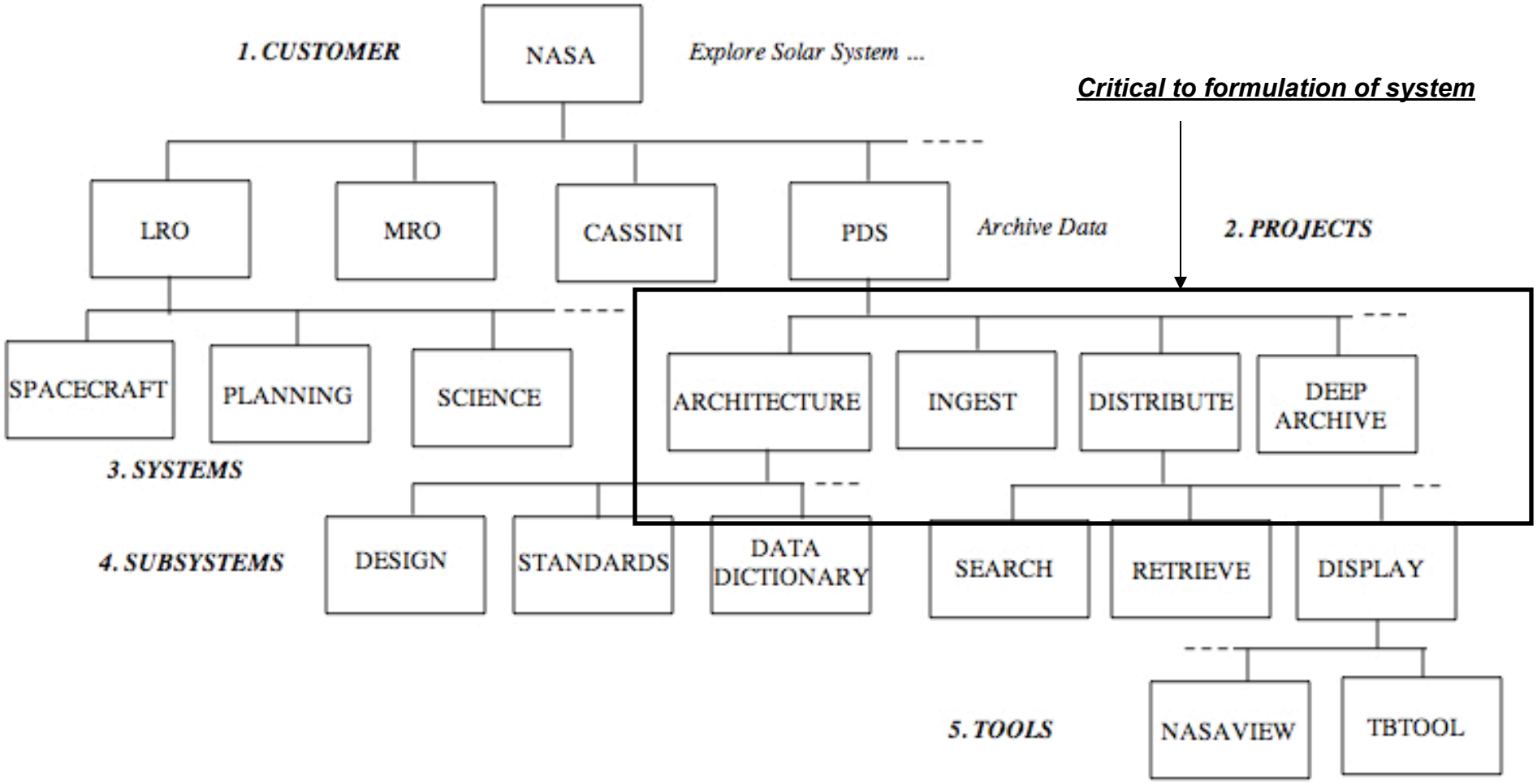
2. PDS will collect suitably and well-documented data into archives that are peer reviewed and maintained by members of the scientific community

3. PDS will make these data accessible to users seeking to achieve NASA's goals for exploration and science

4. PDS will ensure the long-term preservation of the data and their usability

Credit: PDS Level 1,2,3 Requirements. August 2006.

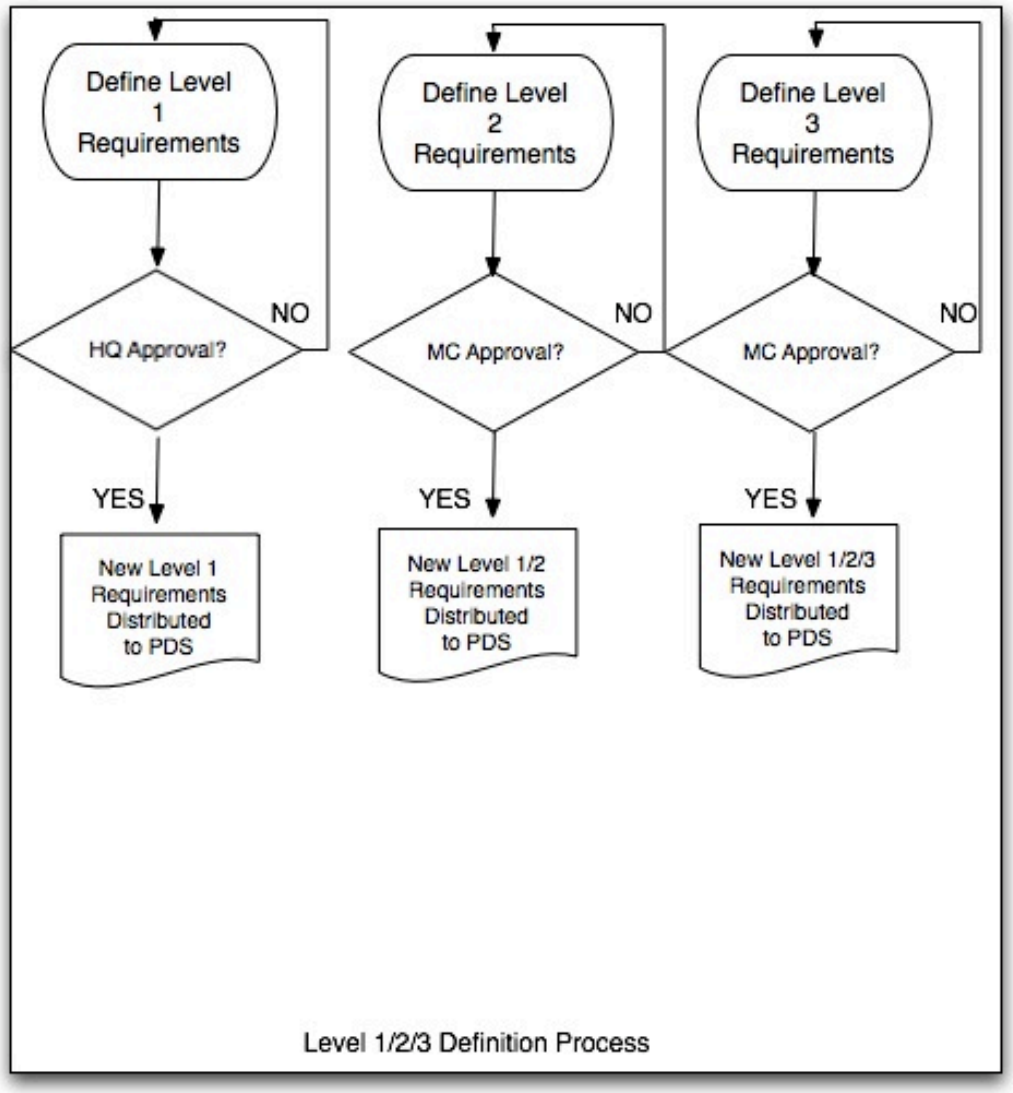
Requirements Organization



* Adapted from Dick Simpson (2007-03-28)

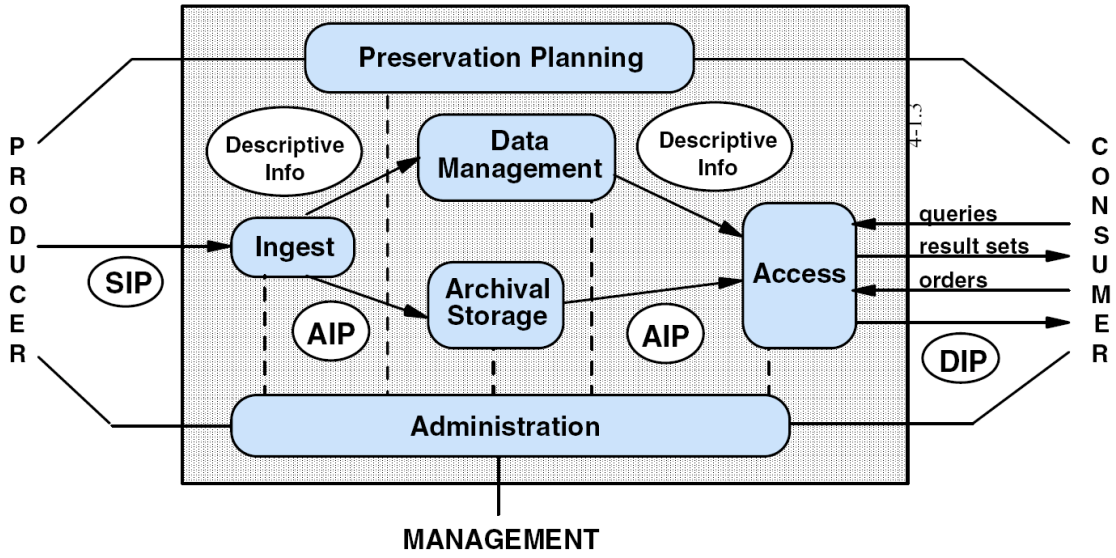
PDS Requirements Approval

- Level 1 - NASA HQs requirements for PDS (Sponsor Level)
- Level 2 - MC derived requirements on PDS from the sponsor (MC Level)
- Level 3 - PDS System Level requirements (top-level)
- Minor editorial and maintenance changes approved by PDS Program Manager and Program Scientist. Additional requirements at levels 1/2/3 approved by MC with Level 1 concurrence at HQ.



System Functions of PDS

- Ingestion
- Data Management
- Storage Management
- Administration
- Preservation Planning
- Distribution/ Access



Reference Model for Open Archive Information System, CCSDS 650.0-B-1, January 2002

Timeline of PDS Technical Implementations

- PDS 1 – 1990
  - High-Level Catalog for finding data sets by mission, instrument, spacecraft and target.
  - Archive volumes stored and distributed on tape.
  - The Object Description Language (ODL) is used for product labeling and capturing catalog information.
- PDS 2 - 1994
  - CD-ROM become the archive and distribution volume of choice.
  - High-Level Catalog simplified by using more text instead of keywords to capture descriptive information.
- PDS 3 - 2000
  - PDS sets up and maintains a WEB presence.
  - Movement to online distribution of products /PDS-D
  - On-line mass storage and storage bricks replace CD/DVD as archive and distribution media.
- PDS4 - 2010
  - Movement to a distributed, service architecture
  - Integrated federation
  - New data standards, data formats and structures

Technical Implementation (PDS3)

Function	Implementation
Ingestion	Based on PDS3 Standards; Tools for design/validation; submission manual process
Data Management	Central data catalog (dataset level); distributed data management at the science data product level
Storage Management	Per 2008 MC, all data now online in federated repositories
Preservation Planning	Missions deliver data formatted for the archive; checksums maintained for archival data; three copies of data
Distribution/Access	Data distributed in archival format; each node provides access to their holdings; central portal points to node web search
Administration	Local and federate management to provide coordination and governance

PDS3 Data Standards

Function	Implementation
Data Model	Governs descriptive aspects of the product; defines standard catalog elements (mission, instrument, etc)
Data Dictionary	Based on a PDS defined structure; changes governed by PDS-wide standards body
Grammar	Object Description Language (ODL) used to capture metadata and annotate data sets, products, and catalog files
Archive Organization	Volume organization with local discipline node optimization

Note: Model changes occur at the data dictionary level; standards defined in a narrative reference

PDS3 Assessment

- The current PDS3 was designed based on an offline system both the standards and software infrastructure have evolved to support online operations

- The growth of PDS, both for NASA and non-NASA missions, has stressed the structure and capabilities of the PDS3 standards

- Software tools, infrastructure, technologies and standards have changed which makes continued maintenance and extension of PDS3 a concern

Agenda

- Engineering Overview/PDS3 Implementation
- Architectural Drivers
- PDS 2010 Project Overview
- High Level Architecture Concept
- Major Design Decisions

PDS Roadmap

- Published in February 2006 by the PDS MC
  - http://pds-engineering.jpl.nasa.gov/projects/PDS4/Exchange/PDS_Roadmap.pdf

- Addresses critical upcoming challenges
  - Mission and Data Provider
  - User and Customer Challenges
  - Challenges Associated with International Collaborations
  - Operational Challenges
  - On-going Challenges given budget constraints

Key Drivers

- Based on the input from the PDS Roadmap, requirements and the Management Council, the following is a categorized list of architectural drivers organized into thematic areas:

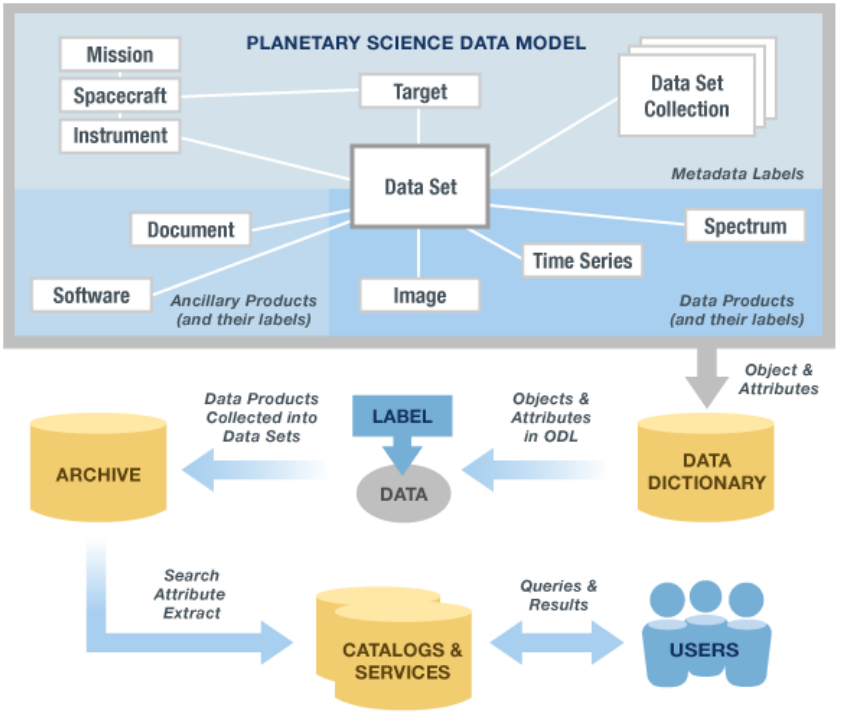
- More Data
  - PDS storage requirements are projected to increase from 50 TBs* to several hundred TBs in the next few years
- More Complexity
  - Missions, instruments, and data are all becoming more complex.
- More Producer Interfaces
  - PDS is facing an increasing number of missions, a greater number and diversity of data providers, and smaller, focused missions. PDS is also working internationally now.
- Greater User Expectations
  - The Internet has led users to expect well-documented data to be readily available via text-based or graphical search systems with data delivery in a variety of formats
- Limited Funding
  - The emphasis on smaller, faster, cheaper missions which often include international partners may limit the ability to provide products suitable for analysis by the broader science community.
- Creating a "System" from the Federation
  - The current PDS nodes operate autonomously and independently with limited distributed access via PDS-D to node repositories.

Major PDS3 Implementation Gaps (relative to PDS requirements)

- Standards for formatting data
- Submission of data to the PDS archive
- End-to-end Tracking (mission deliveries through to deep archive)
- Distributed catalog system (to the product level)
- Uniform protocols for accessing data, metadata and computing resources across the enterprise
- Translation of archival products to selected formats

PDS3 Data Standards Issues

- Lack data formats
- Cumbersome change process
- Lack formal engineering specification
  - Implementation differences across PDS and Internationally
- PDS-specific grammar
- Newer international standards (e.g., data dictionary)
- Long-term preservation vs. Contemporary Formats



PDS Planetary Science Data Model

Technical Implementation Issues

- System components and implementation should be derived from an online, distributed architecture
  - The current PDS evolved from an offline system
- Need to share services between nodes and internationally for ingestion through to distribution
  - Catalog access
  - Data format conversion
  - Search
  - Access
- Tools built on aging PDS3 standards
  - Design and Validation tools
  - Viewing and visualization tools
- Limited automation

Agenda

- Engineering Overview/PDS3 Implementation
- Architectural Drivers
- PDS 2010 Project Overview
- High Level Architecture Concept
- Major Design Decisions

What is PDS 2010?

- A PDS-wide project to upgrade PDS from PDS3 to PDS4
- A transition from a 20-year-old collection of standards and tools to a modern system constructed using best practices for data system development.
- Fewer, simpler, and more rigorously defined formats for science data products.
- Use of XML, a well-supported international standard, for data product labeling, validation, and searching.
- A hierarchy of data dictionaries built to the ISO 11179 standard, designed to increase flexibility, enable complex searches, and make it easier to share data internationally.

Key goals from vision statement*

- Simplified, but rigorous, archiving standards that are consistent, easy to learn, and easy to use
- Adaptable tools for designing archives, preparing data, and delivering the results efficiently to PDS
- On-line services allowing users to access and transform data quickly from anywhere in the system
- A highly reliable, scalable computing infrastructure that protects the integrity of data, links the nodes into an integrated data system, and provides the best service to both data providers and users

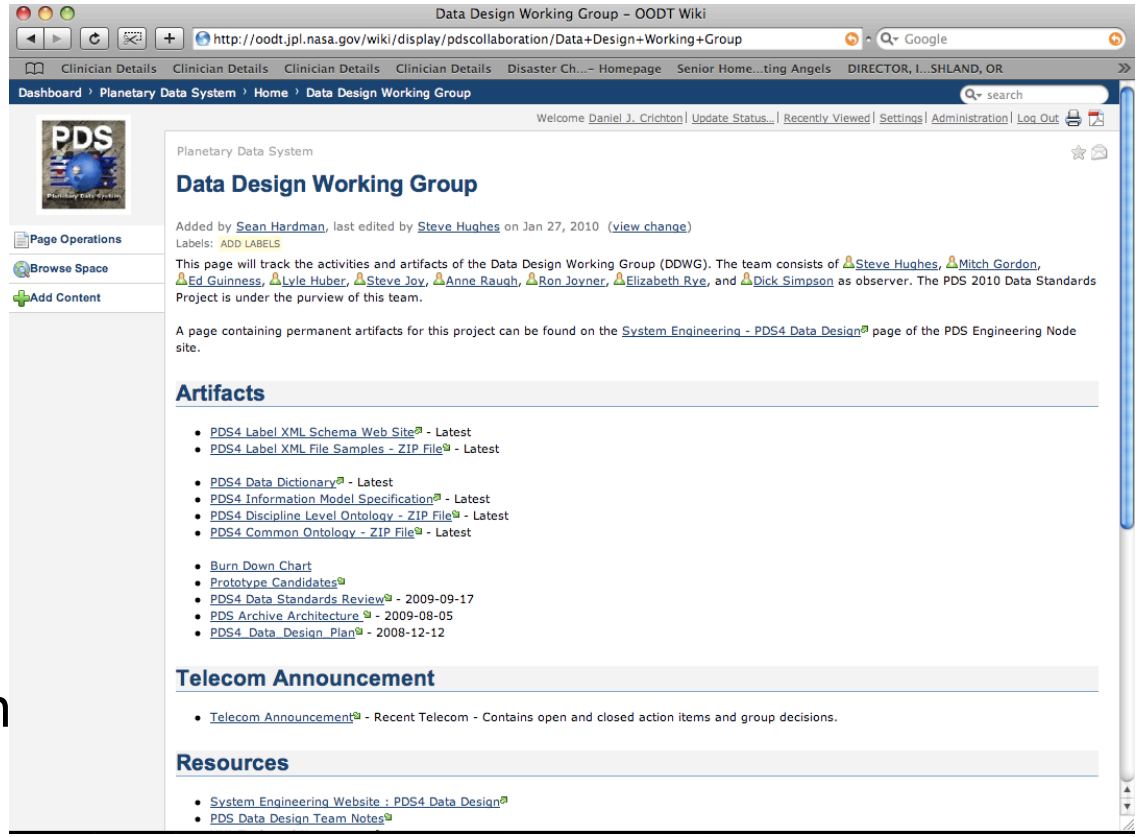
* PDS 2010 Executive Summary, July 2008

Project Concept

- Broken into separate projects
  - Projects have implementation plan, requirements, deliverables, schedule, etc
- Collaborative development across PDS
  - Members from across PDS participate in development lifecycle
- Constrained budget (minimal overguide increment across PDS in FY10, FY11, FY12)
- Phased design and development
  - Data design and standards development is intentionally ahead of the software
  - Address efficiency in ingestion as phase I
  - Address user/access as phase II
- Builds will be used to release increasing capabilities based on PDS4 standards
- FY10 focus is primarily on data standards; FY11 will switch to software

Project Approach

- Study
  - User, Data, Architecture Teams
  - All nodes represented
- Architecture
  - Data (DAWG) and System Architecture (SAWG) Teams
  - All node represented
- Design/Implementation
  - Data Design (DDWG) and System Design (SDWG) Teams
  - All node represented



Example: DDWG Wiki

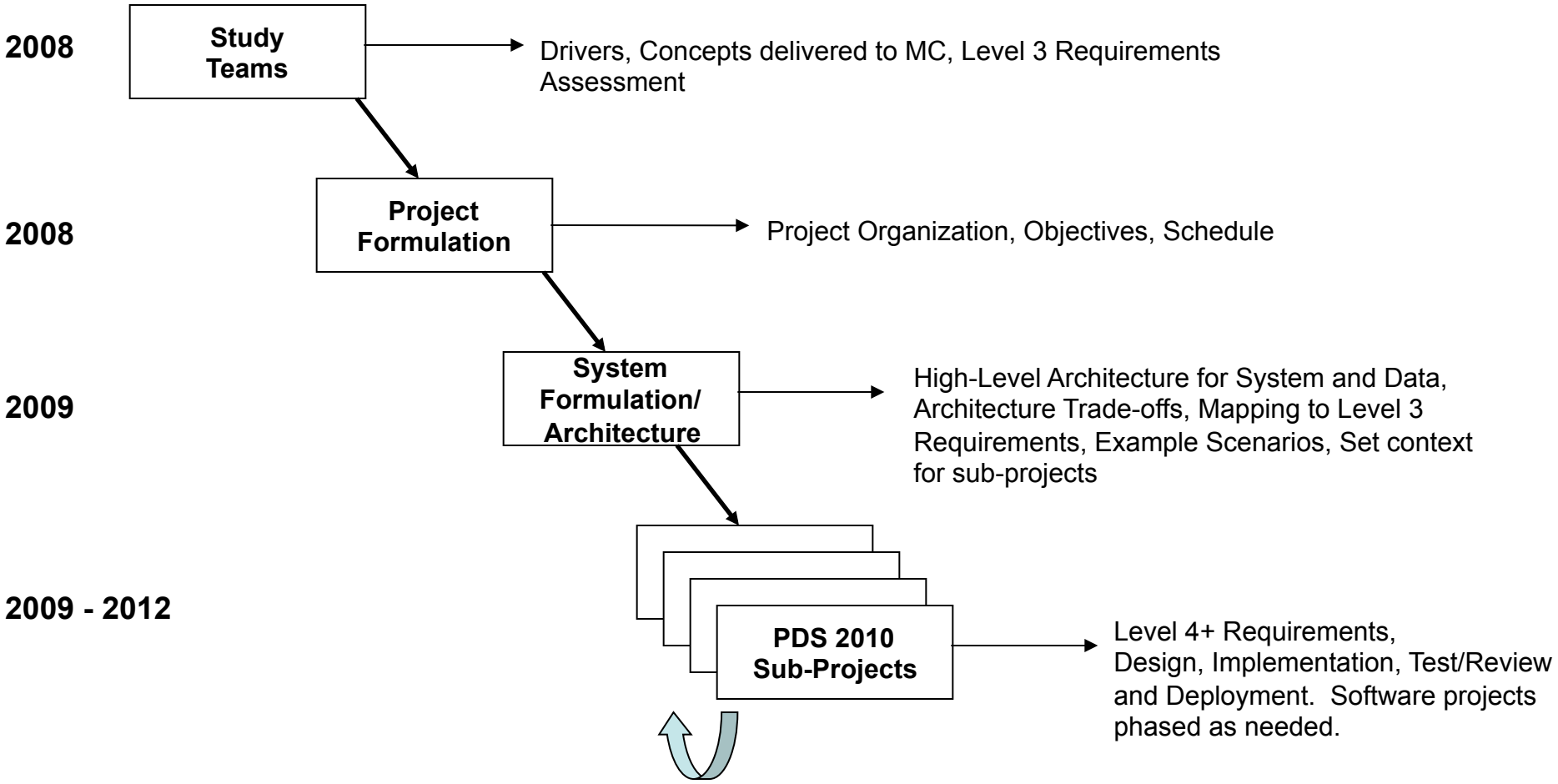
Working Groups

- PDS 2010 has used the "Working Group" concept to develop cross-PDS teams
  - Multi-disciplinary teams of scientists, software and data engineers brought together
- PDS 2010 effectively broken into a data standards team and a software team
  - Individual implementation plans for each will be presented…
- Working Groups hold regular telecons
- Artifacts generated are posted to Wikis for each WG
- WGs report out at Management Council meetings

PDS 2010 Project Leads

- Project Manager: Dan Crichton
- Project Scientist: Reta Beebe
- PDS4 Data Standards Lead: Steve Hughes
- PDS 2010 System Development Lead: Sean Hardman
- Transition/Operations: Emily Law
- NOTE: Involvement from discipline nodes across the PDS is critical; they are intimately involved in each step and support leading these efforts

Engineering Approach to PDS 2010



Key PDS Meetings to Date

- August 2007 – Initial kick off meeting with MC at UCLA to discuss a possible PDS4 effort
- December 2007 – Review concepts with MC at NMSU
- February 2008 – Project formed
- July 2008 – Project plan reviewed with MC at USGS; architecture teams formed
- September 2008 – Technical Session to discuss/review architecture at JPL
- November 2008 – Architecture reviewed with MC at DC meeting
- January 2009 – Design teams formed
- June 2009 – Technical Session to discuss/review design at JPL
- August 2009 – Review of design plans with the Management Council at Caltech
- December 2009 – Discuss prototype PDS4 data products with the MC at NMSU

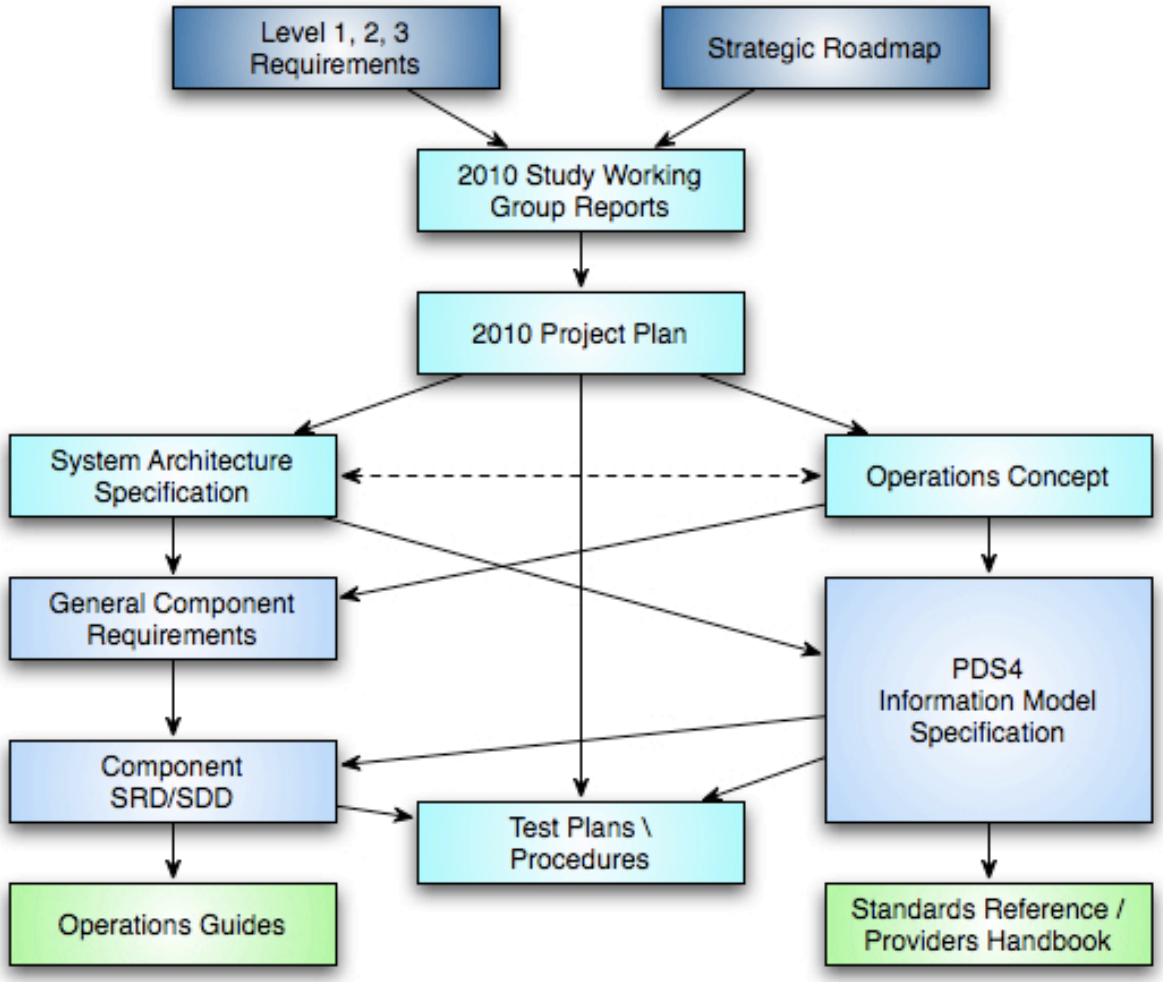
Project Reporting and Management

- Project reported to the Management Council on an regular basis
  - Monthly telecons
  - Management Councils F2F Meetings (3 per year)
- Monthly schedule updates
- Weekly meetings with the leads
- Regular meetings and reporting by the working groups

Key Project Documentation

- Project Plan
  - Contains management approach and implementation plan
- Multiple Concept Papers
  - Architecture, user services and data design
- System Architecture
  - Defines data and software architecture
- Operations Concept
  - Interactions of PDS across the mission phases and from ingestion thru to distribution
- System Design Specifications for services and tools
- In addition, PDS maintains policies, requirements, standards, and schedule information online at the Management Council and Engineering nodes

Document Tree

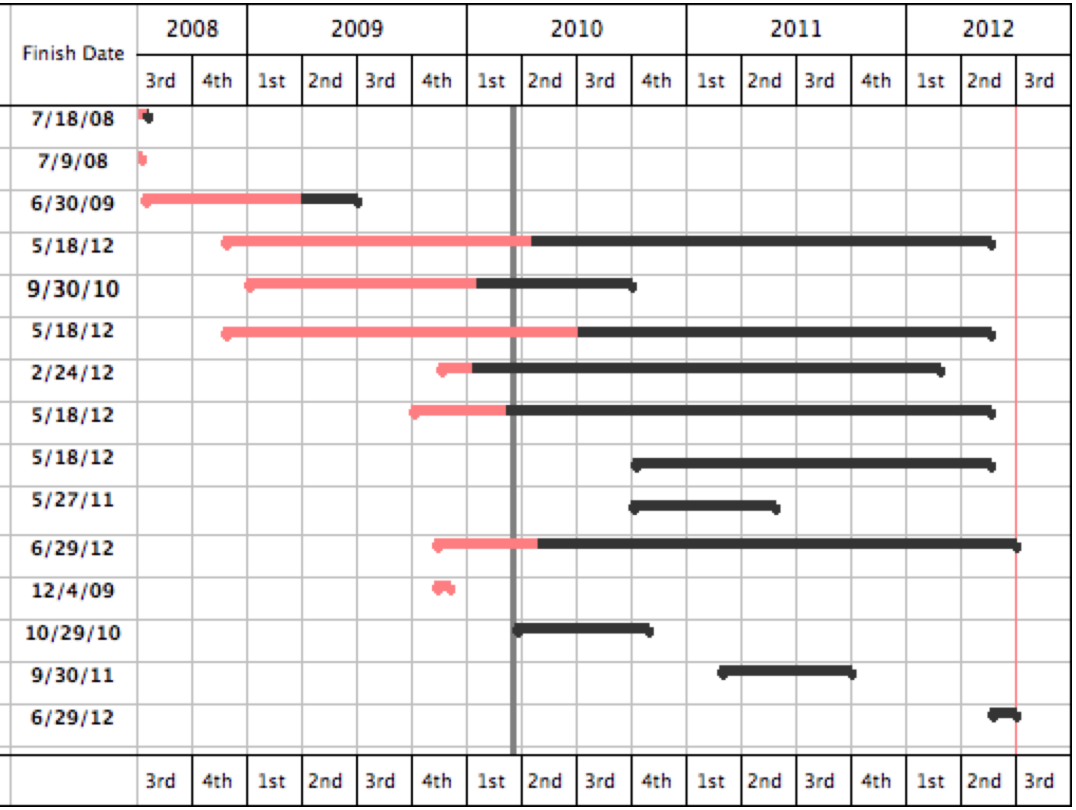


Project Budget

- To be discussed more during the resources section
- Each PDS node operates within a guideline budget that covers node development and operations providing on-going support for PDS
  - Some development, particularly at Engineering, can be done within guidelines by redirecting efforts
- In 2010, 2011 and 2012, the PDS requested additional overguide for PDS nodes to cover PDS 2010

High Level Schedule

Activity Name	Duration (Work Days)	Start Date	Finish Date
Concept/Study Phase...	240.00	8/20/07	7/18/08
Project Planning...	133.00	1/7/08	7/9/08
Architecture/System Engineering...	254.00	7/10/08	6/30/09
Projects	911.00	11/21/08	5/18/12
P1. Data Standards Project...	455.00	1/2/09	9/30/10
P2. Distributed Infrastructure Project...	911.00	11/21/08	5/18/12
P3. Tools Project...	595.00	11/16/09	2/24/12
P4. Distributed Catalog System Project...	688.00	9/30/09	5/18/12
P5. Portals, Search and Distribution Project...	425.00	10/4/10	5/18/12
P6. Data Movement and Delivery Project...	171.00	10/1/10	5/27/11
Builds	690.00	11/9/09	6/29/12
B0. Infrastructure Build...	20.00	11/9/09	12/4/09
B1. Ingestion Build...	160.00	3/22/10	10/29/10
B2. Distribution Build...	157.00	2/24/11	9/30/11
B3. User Capabilities Build...	30.00	5/21/12	6/29/12



Planned System Builds

Phase	Purpose	Release	Date
I Prototype Build 1 Ingestion	• Release a prototype Ingest Subsystem • Baseline PDS4 model, standards reference • Enable PDS3 to PDS4 catalog migration • Support testing of Node interfaces • Support PDS4 product prototypes	• PDS4 info model, standards reference, data dictionary, schemas baseline • PDS 2010 Ingestion subsystem including Harvest, Registry (Inventory, Document, Dictionary, Service), Report and Security services • Initial data provider tool suite • First set of process, documentation and tutorial	October 2010
II Operational Build 2 Distribution	• Initial operational PDS 2010 system and PDS4 Standards • Allow acceptance of PDS4 data into operational archive • Enable data migration from PDS3 to PDS4 • Allow user to search and access both PDS3 and PDS4 data	• E2E PDS 2010 system, Distribution subsystem including Search and Monitor services, revised web site, general portal applications • Complete tool suite • 1st release of PDS4 standards reference, data dictionary • Enhanced process, documentation and tutorial	October 2011
III Operational Build 3 User Capabilities	• Incremental release of operational PDS 2010 system to enhance user capabilities • Support data transformation • Support science services	• Integration of DN applications and science services • Order and Subscription services	June 2012

Key PDS 2010 Review Dates

Review	Supporting Working Group	Date
PDS Technical Group Preliminary Design Meeting	SDWG/DDWG	June 2009
Preliminary PDS MC System and Data Review	SDWG/DDWG	August 2009
PDS4 Product Design and Prototyping	**DDWG**	**November 2009– June 2010**
PDS4 Internal Science Node Exercise and Data Product Development	DDWG	November 2009
PDS4 MC Discussion/Product Review	DDWG	December 2010
PDS4 MC Discussion/Product Review	DDWG	March 2010
System Review, Ingestion Subsystem	SDWG	February/March 2010
PDS Standards Release V1.0	DDWG	End of FY2010
Distribution Subsystem/Delta Review	SDWG	Spring 2011

Transition Planning

- Plan to be discussed Tuesday PM, but details include…
- Analyzed tradeoffs and impacts on stakeholders
- Complete Initial PDS4 Standard (v4.0)
- Allow for phased transition to PDS4 over time
  - Existing pipelines will remain supported during life of mission
  - Support ingestion and distribution of PDS4 data when ready
  - Expect nodes and partners to transition when ready
- Ensure PDS 2010 will serve data from PDS3 and PDS4
- Migrate data sets to PDS4 as needed
  - Migration included in discipline node planning

Summary of Progress to Date

- Project plan defined and being executed
- Funding in place for the project
- Requirements in place
- PDS-wide Architecture defined
- Prototype PDS products defined using maturing PDS4 specification
- Data provider engagement
- Specifications for key infrastructure services
- Prototypes conducted to support technical trade decisions
- Multiple reviews with PDS Management Council and Technical Group
- Assessment projects underway with the International Planetary Data Alliance to coordinate standards
- Transition plan developed for transitioning both software and data over time