PDS 2010 Project Overview: Description, Plans and Status

PDS Design Review II
Greenbelt, MD

Dan Crichton
June 21, 2011

Agenda

- Project Overview and Organization
- Project Status
- Architecture Decisions and Approach
- System Builds and Delivery Plans
- Transition
- Schedule

PDS 2010

- PDS 2010 is 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.

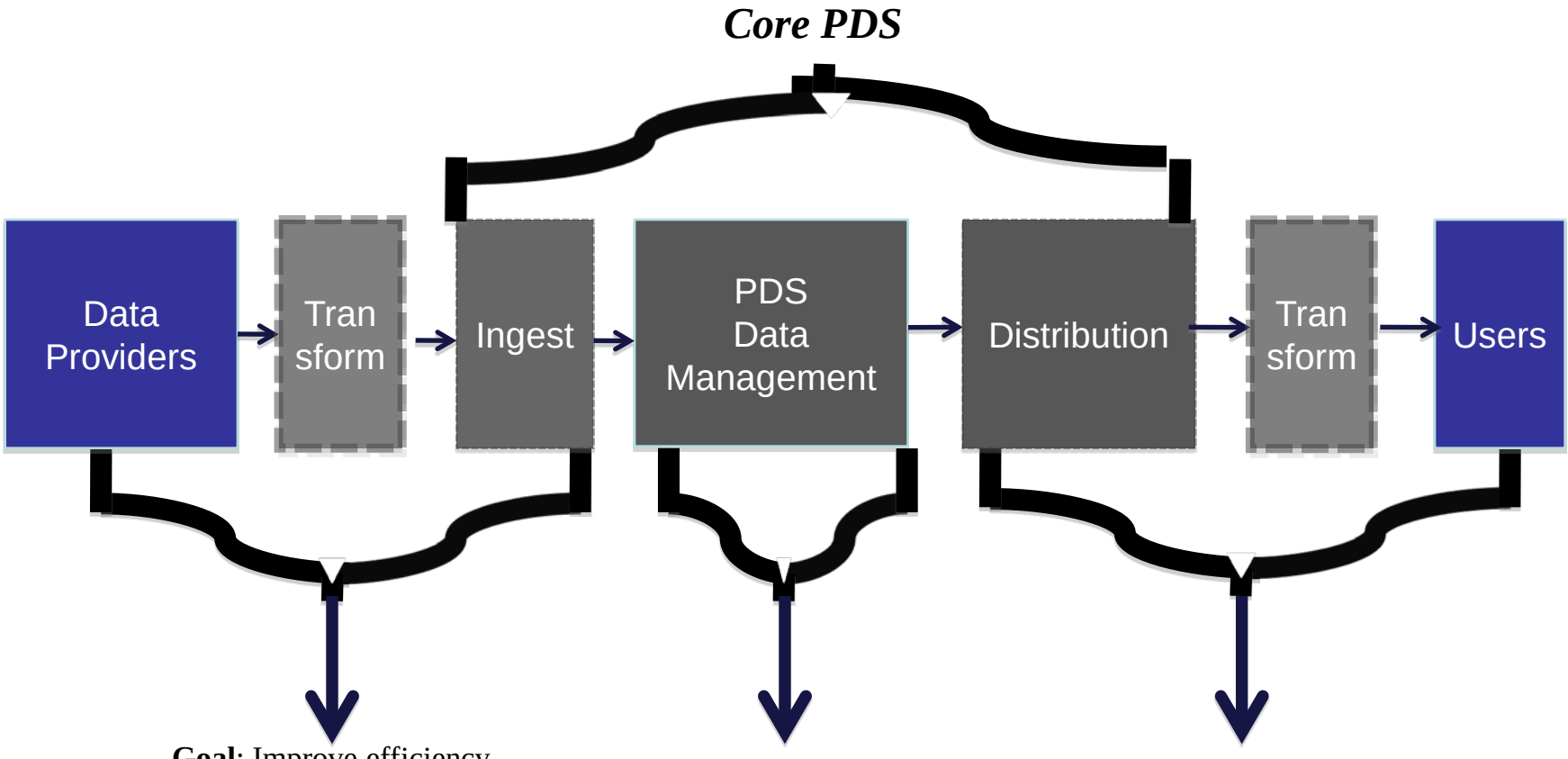
PDS Challenges

- Number and diversity of missions and instruments
  - PDS is currently receiving data from 110 instruments from 15 active missions as well as concurrently working with missions in development
  - New mission data nodes being added to PDS (LROC, for example)
- Requirements for preservation of data and for usability are sometimes in conflict
- Budget pressures which affect archiving/usability across data providers/missions, PDS and the users
- International archiving and standards coordination
- Increasing volume of data
  - In 2001, the PDS archive was 4 TBs
  - In 2010, the PDS online archive is over 100 TBs
  - In 2011, the PDS online archive is over 200 TBs
- **Replacing aging technology, tools, standards and processes**

Motivation for PDS 2010

- 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 very challenging
- Ultimately, new software technologies and standards provides an opportunity to greatly improve the operation and usability of the PDS long-term

Major PDS Functions



* Note: In PDS3, these weren't really split

Structured Project Approach

- Phased approach with builds that introduce increasing functionality
- Two key working groups that include members from across PDS
  - DDWG = Data Design WG
  - SDWG = System Design WG
  - Use of several collaboration mechanisms
- Full lifecycle planned out with deliverables (project plan, requirements, design, CM/build, test, release)
  - Reviews at key points in the process

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

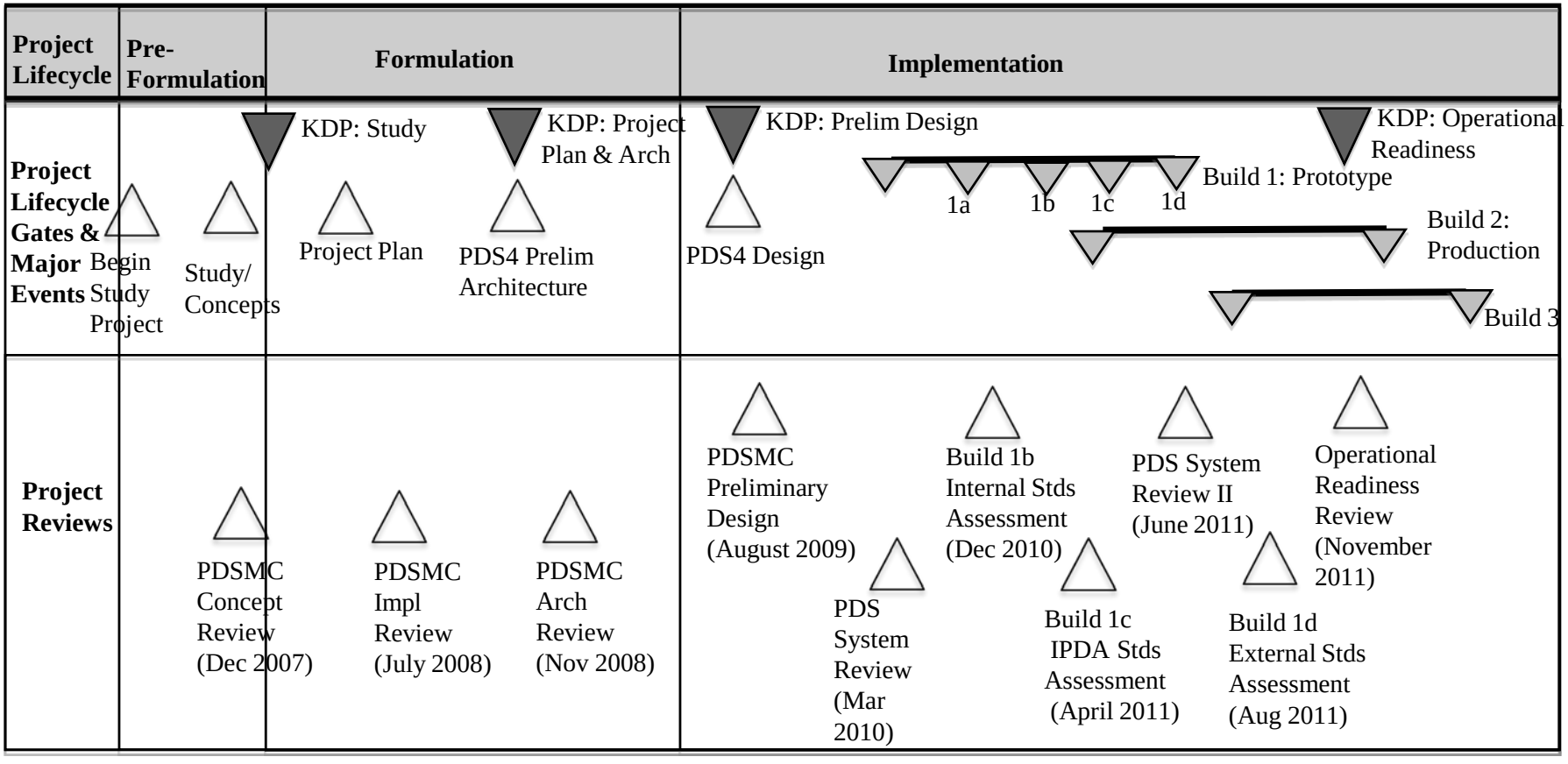
Project Reporting and Management

- 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

Baseline Plan

- Deliver a beta PDS4 Data Standard by October 2010 to begin coordinating PDS4 product definitions and systems development
- Deliver a prototype infrastructure build by October 2010 to test ingestion of PDS4 sample data products
  - Incremental releases will occur (e.g., build 1b)
- Deliver an initial operational capability for PDS4 by October 2011 to support ingestion and distribution of PDS4 data products
- Deliver additional tools and a framework to support PDS4 transformation and science services/distribution capabilities -- Summer 2012

Project Lifecycle



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

Working Groups

- DDWG Members: Steve Hughes (chair), Richard Chen (EN), Amy Culver (IMG), Patty Garcia (IMG), Ed Guinness (GEO), Mitch Gordon (Rings), Sean Hardman (EN), Lyle Huber (ATMOS), Chris Isbell (IMG), Steve Joy (PPI), Ronald Joyner (EN), Debra Kazden (PPI), Joe Mafi (PPI), Thomas Morgan (MGMT), Lynn Neakrase (ATMOS), Anne Raugh (SBN), Elizabeth Rye (EN/IMG), Dick Simpson (RS)

- SDWG Members: Sean Hardman (chair), Tom Stein (Geo), Todd King (PPI), Jordan Padams (IMG/EN), Mike Martin (MGMT), Paul Ramirez (EN)

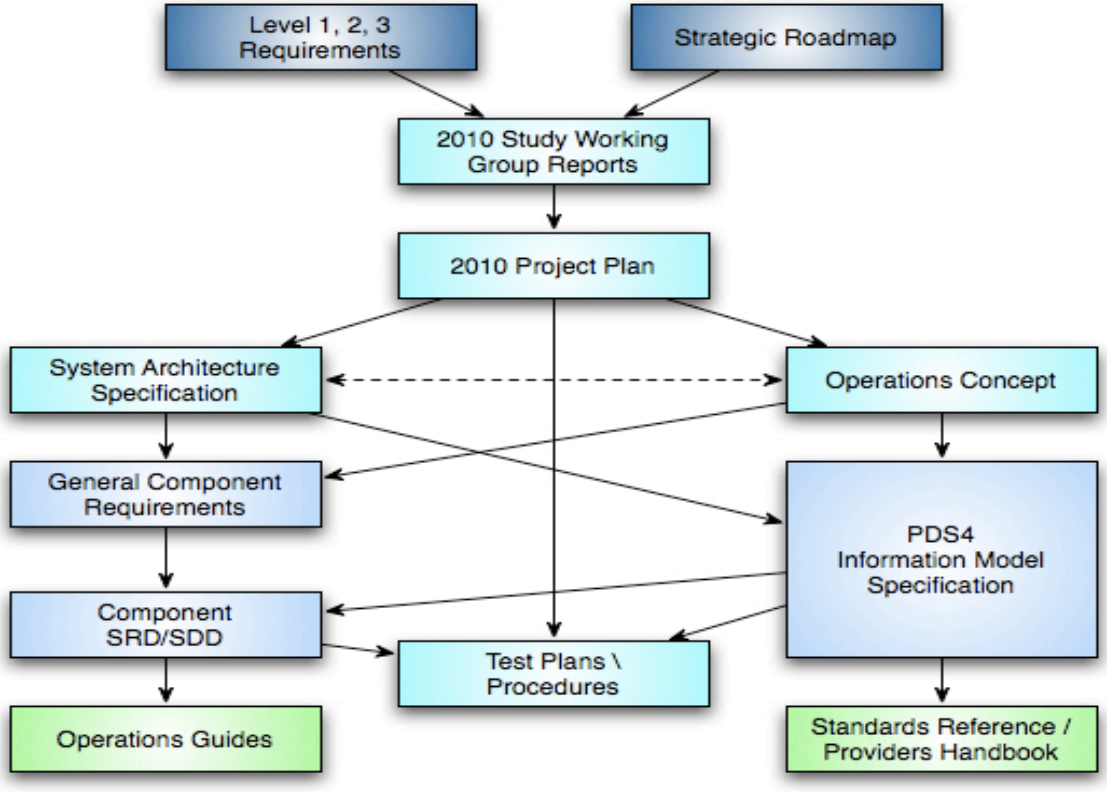
Summary of Progress to Date

- Project plan defined and being executed
- Funding in place for the project
- Requirements in place (approved by MC 3/2010)
- PDS-wide Architecture defined
- Prototype PDS products defined using maturing PDS4 specification
- 3 system builds completed: build 1a, build 1b, 1c
- Multiple reviews of project, system and data standards
- Specifications for key infrastructure services
- Project documentation in place
- System and standards documents developed
- Prototypes conducted to support technical trade decisions
- Transition plan developed for transitioning both software and data over time

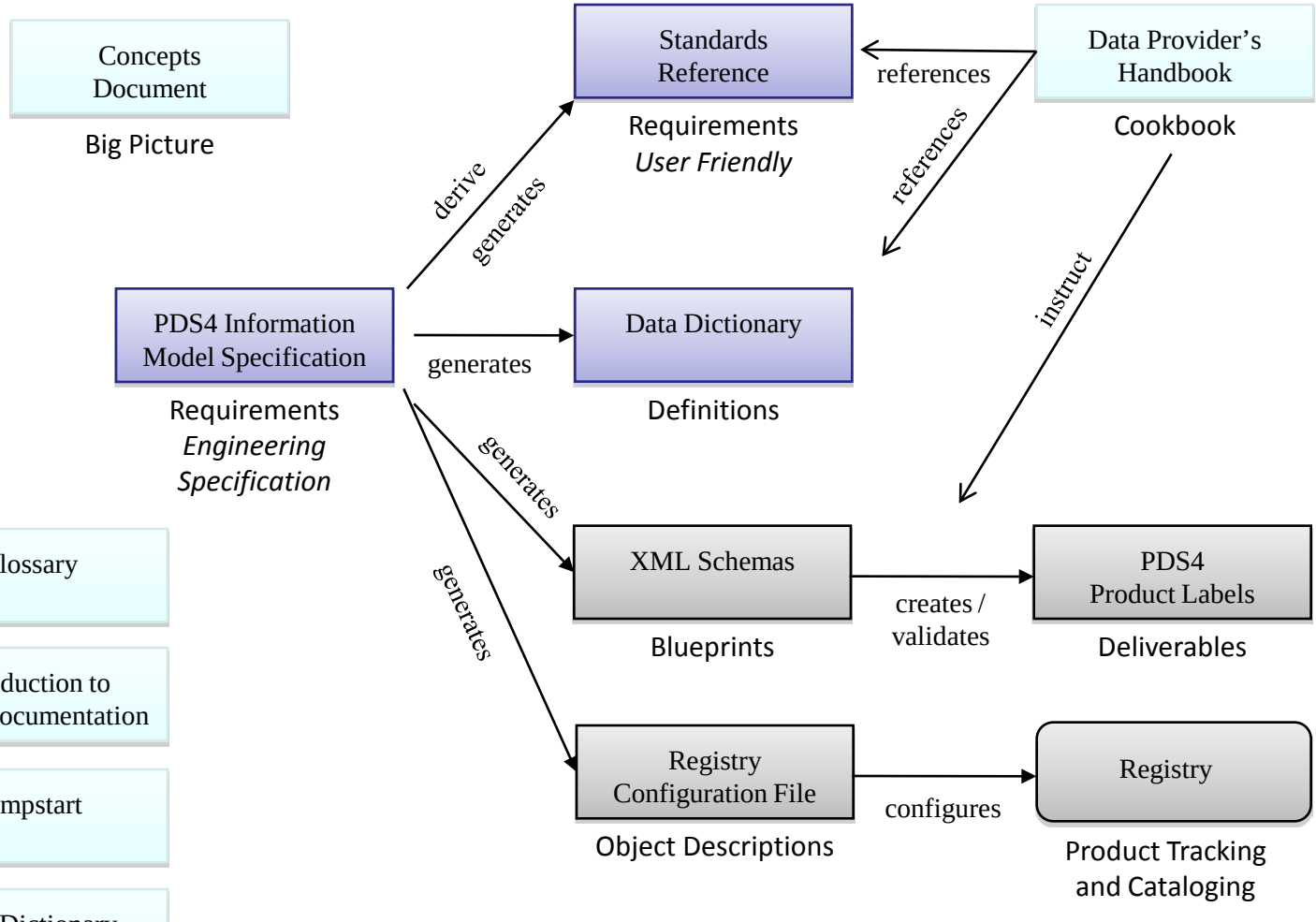
Key Project Documentation

- 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

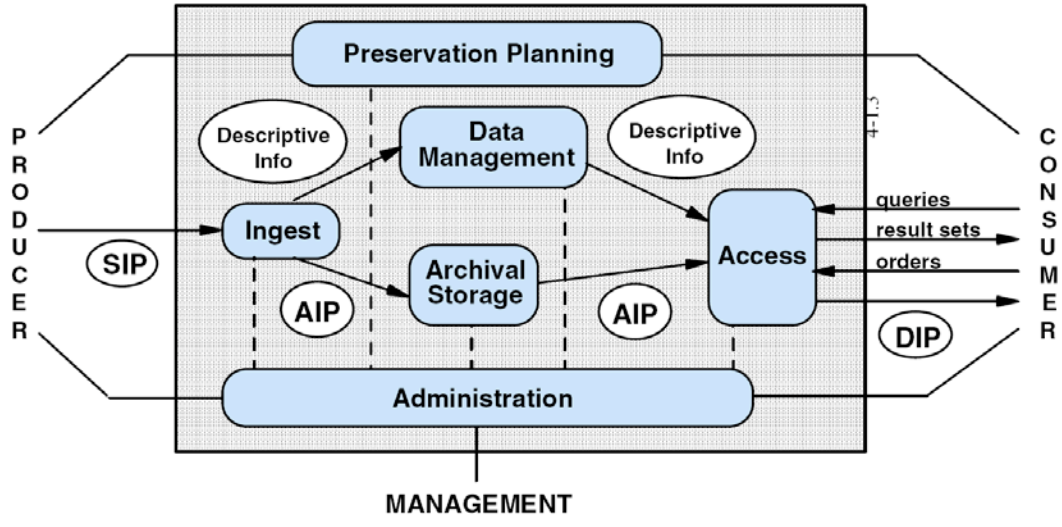


PDS4 Documents & Relationships



System Functions of PDS

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



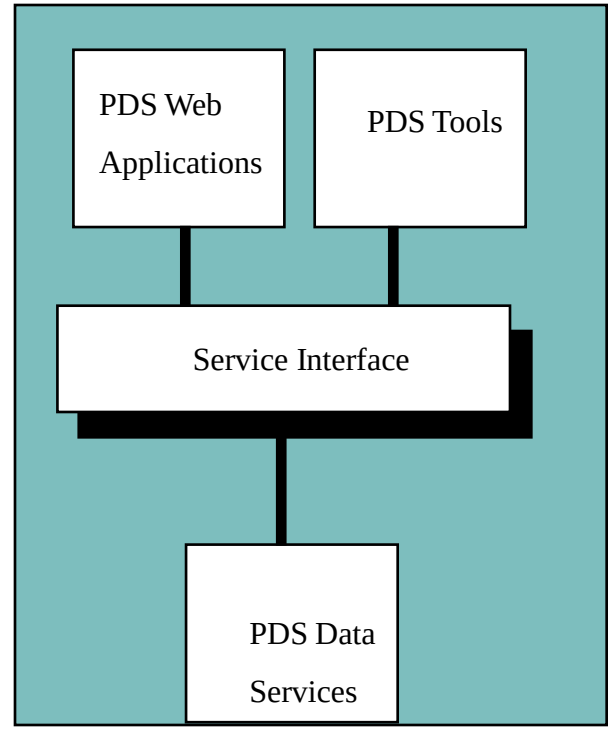
Reference Model for Open Archive Information System, CCSDS 650.0-B-1, January 2002

Key Design Decisions & Recommendation from Aug 2009

- Replace PDS3 ad hoc information model with a PDS4 information model that is now managed in modern tools (DDWG) (in Build 1 prototype)
- Replace ad hoc PDS3 product definitions with PDS4 products that are defined in the model (DDWG) (in Build 1 prototype)
- Require data product formats to be derivations from a core set; Support transformation from the core set (DDWG) (in Build 1 prototype)
- Replace “homegrown” PDS data dictionary structure with an international standard (ISO 11179 RIM) (DDWG) (in Build 1 prototype)
- Adopt a modern data language/grammar (XML) where possible for all tool implementations (SDWG) (in Build 1 prototype)
- Adopt system of registries to support improved tracking and access (SDWG) (in Build 1 prototype)
- Support remote access to data and services to bring the federation together both for ingestion and distribution (SDWG) (planned for Build 2+3)

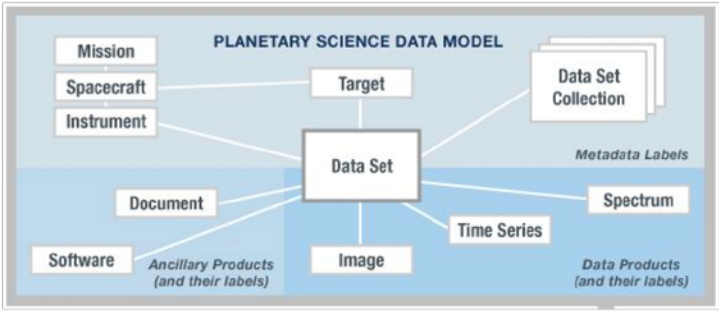
System Design Approach

- Based on a distributed information services architecture (aka SOA-style)
  - Allow for common and node specific network-based services
- System includes services, tools and applications
- Use of online registries across the PDS to track, search and share information about PDS holdings across the federation
- Implement distributed services that bring PDS forward into the online era of running a national data system
  - With good data standards, they become critical to ultimately improving the usability of PDS
  - Support “on-the-fly” transformation from PDS to other data formats



Data Design Approach

Information Model



Used to Create

Validates

Label Schema

Expressed As

Extracted/Specialized

Data Element

has

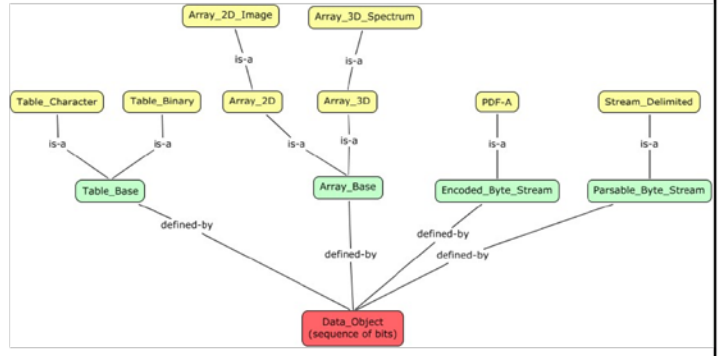
Class

Planetary Science Data Dictionary

Product

Tagged Data Object

(Information Object)



Describes

Data Object

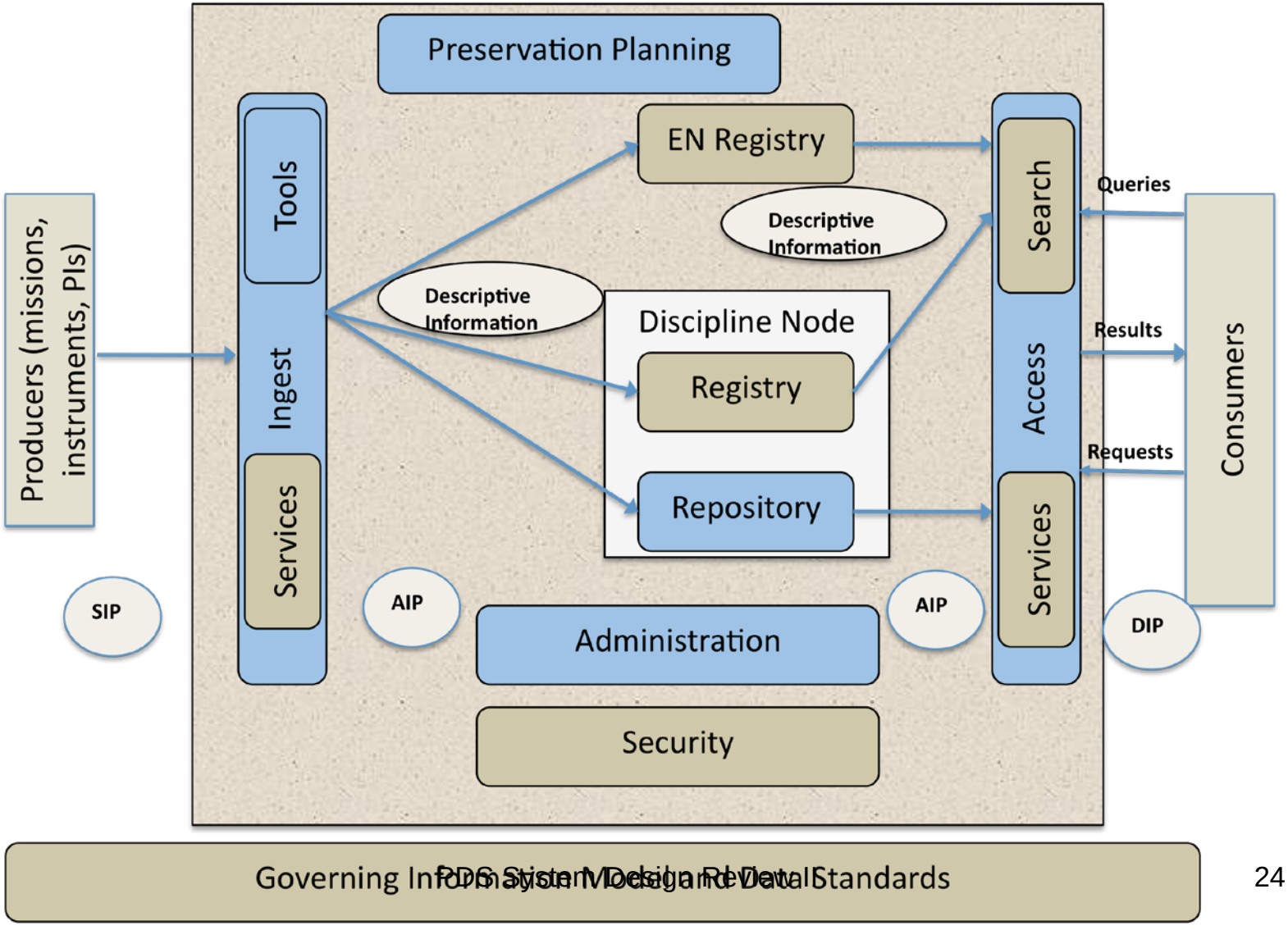
PDS4 Technical Implementation Differences

Function	PDS 3 Implementation	PDS 4 Implementation
Ingestion	Manual process for submission; tools based on PDS internal standards	Automated ingestion; XML-based tools for design, validation and submission
Data Management	Independent data management systems across PDS	Integrated data registries across the PDS to allow for end-to-end tracking and search; interoperability with international partners
Storage Management	All data being migrated online	All data stored online
Preservation Planning	Missions deliver data formatted for the archive	Data maintained in a few simple formats that allow for transformation and long-term use
Distribution/Access	Data distributed in archival format	Data distributed in user formats; user services to better support analysis

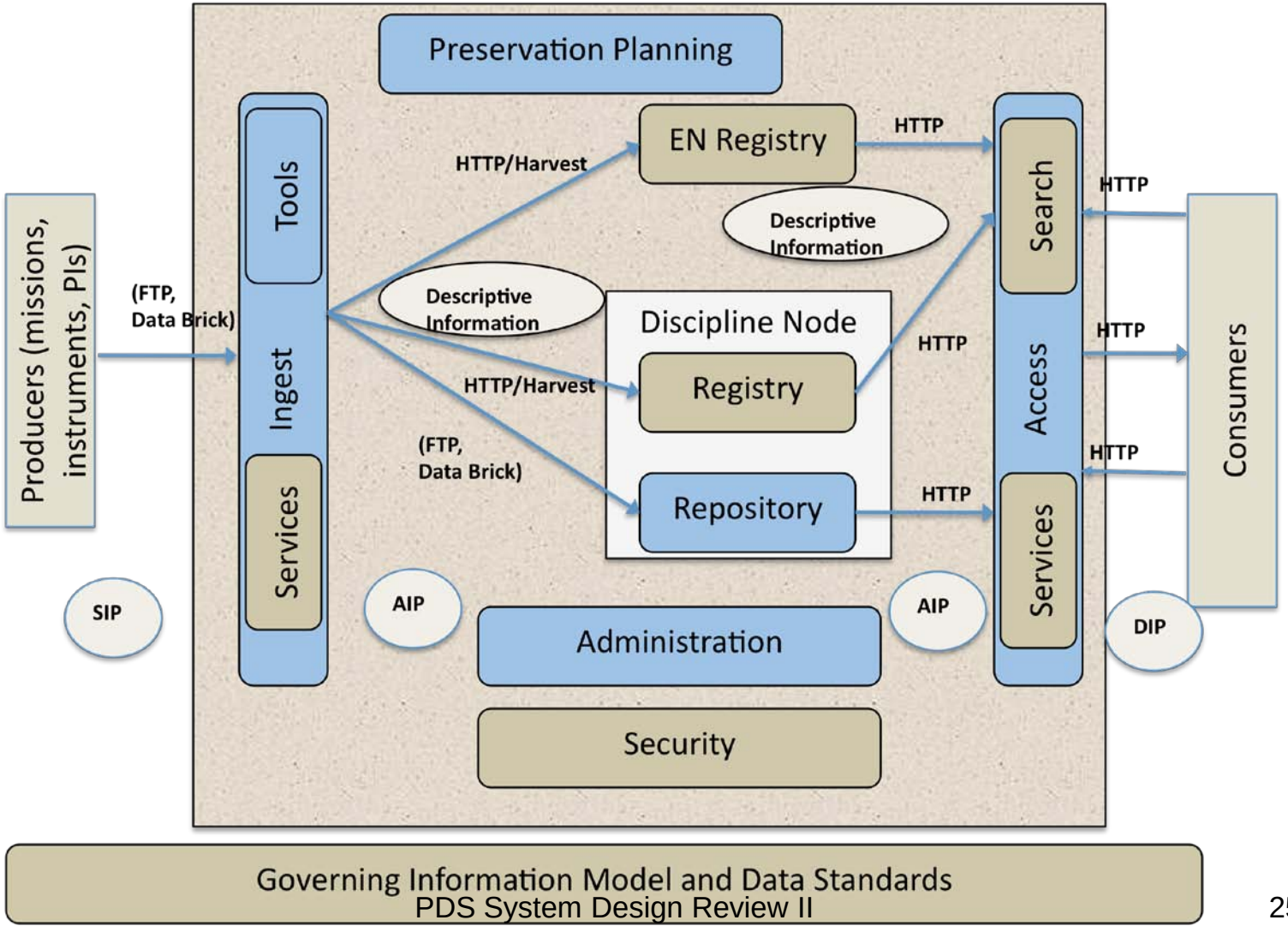
PDS4 Data *Standards* Differences

Function	PDS 3 Implementation	PDS 4 Implementation
Data Model	High level information model; ad hoc model for each data set/product	Entire PDS model captured as an explicit model (ontology) defining all aspects including data, missions, instruments, etc
Data Dictionary	Based on a PDS internal structure	Captured using a rigorous, well-defined structure based on the ISO/IEC 11179 standard; elements organized into namespaces to allow for international coordination
Grammar	Object Description Language (ODL) used to capture metadata and annotate data sets, products, and catalog files	Extensible Markup Language (XML) used to capture PDS metadata; Standard XML tools used

PDS to OAIS Mapping



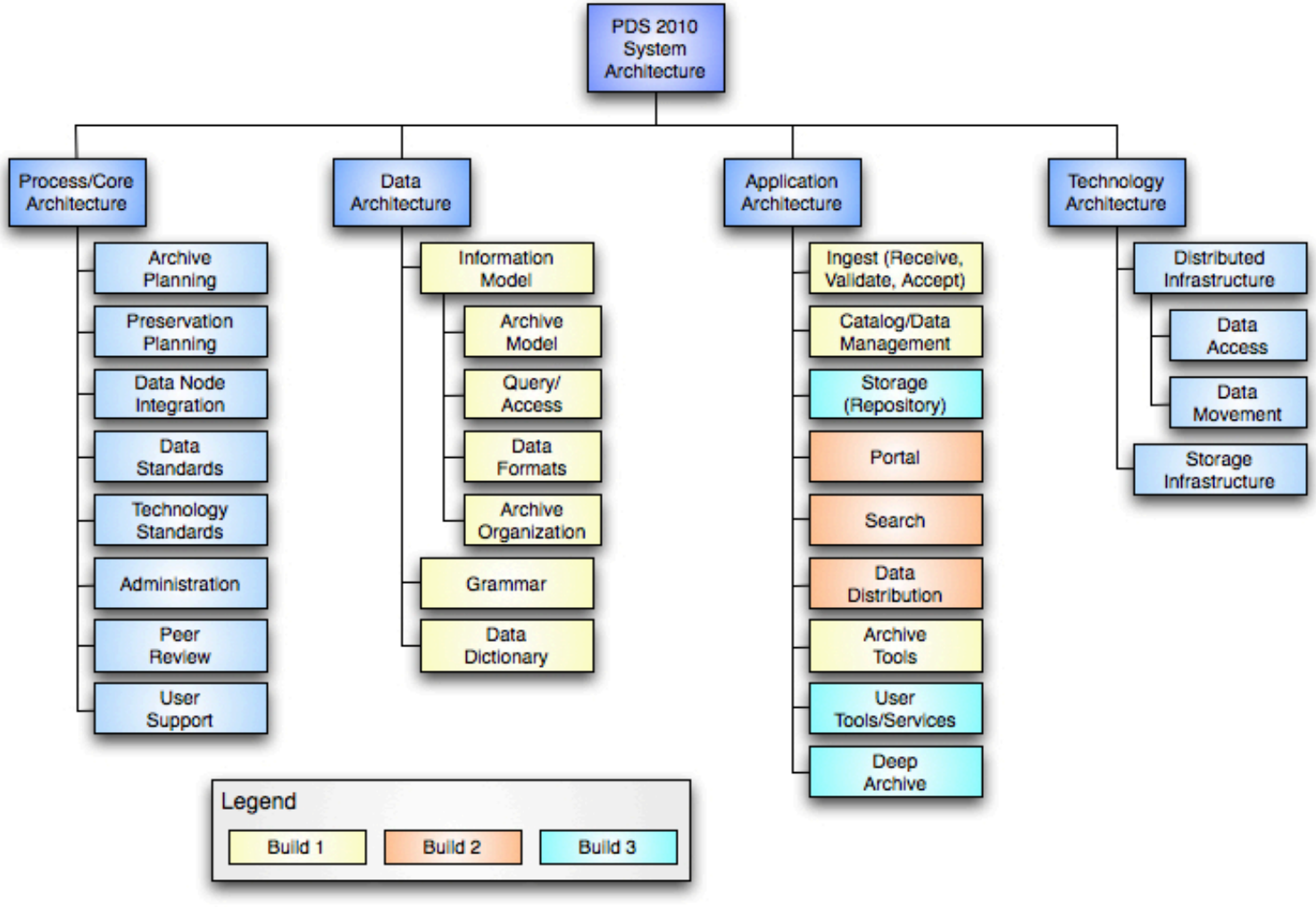
PDS to OAIS Mapping



Planned System Builds

Phase	Purpose	Release	Date
I Prototype Build 1 Ingestion	• Release a prototype Ingest Subsystem • Baseline PDS4 model, standards reference (beta release) • Enable PDS3 to PDS4 catalog migration • Support testing of Node interfaces • Support PDS4 product prototypes	• PDS4 beta 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 • Production 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	Summer 2012

Build-to-Architecture Element Mapping

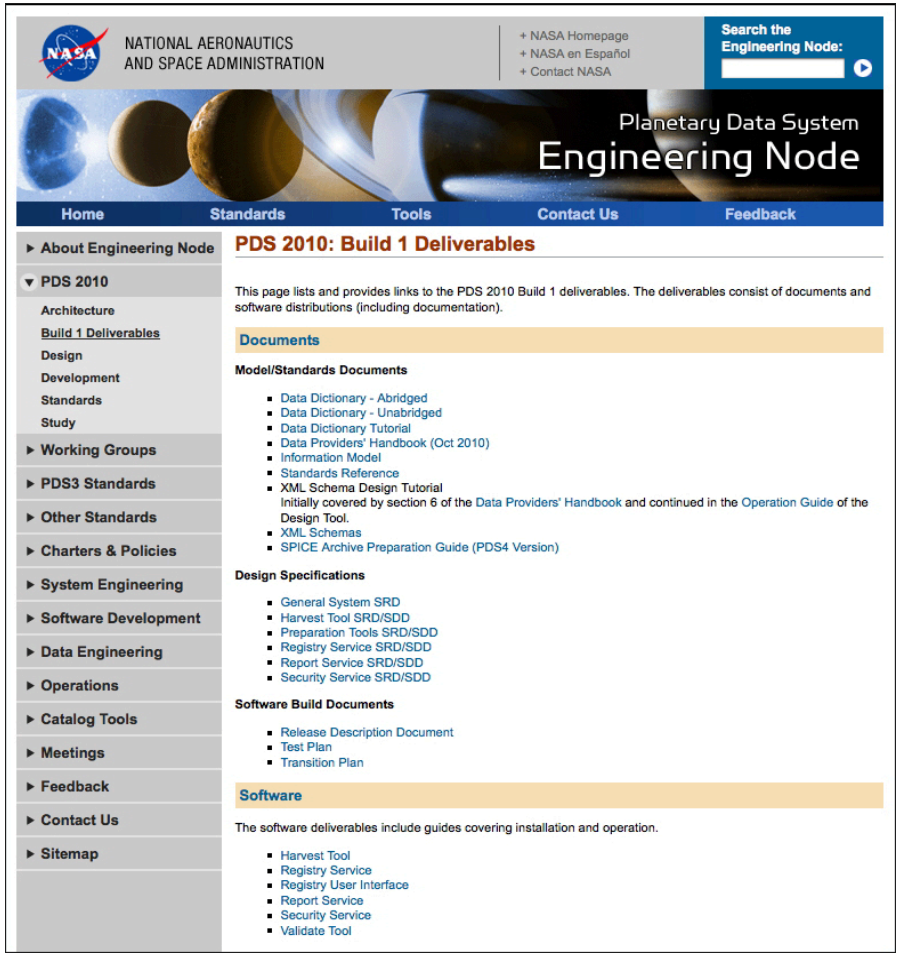


Build 1(a,b,c): Prototype Build

- Delivered Oct 2010 (end of)
  - Build 1b delivered Dec 2010
  - Build 1c delivered Mar 2011

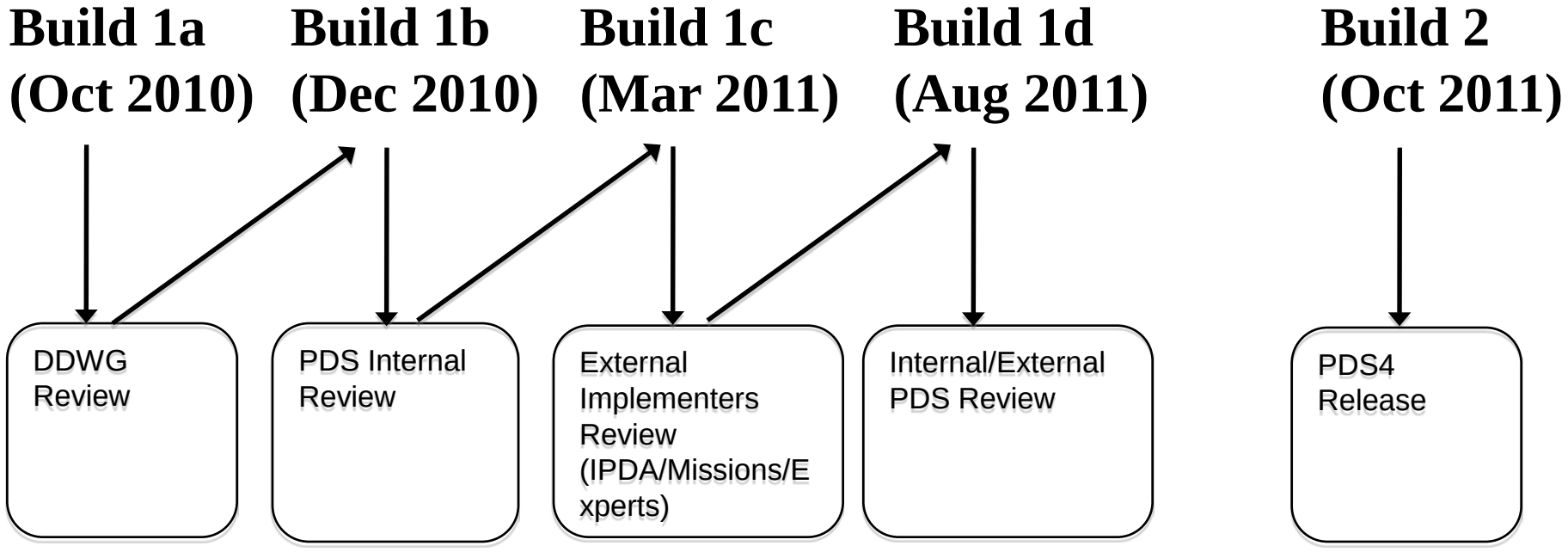
- Data Design (beta PDS4 standard)
  - Data Model
  - Standards Reference (draft)
  - Data Providers Handbook (draft)
  - Data Dictionary (draft)
  - Core Product XML Schemas

- System Design
  - Initial services to support ingestion and replace central catalog (Harvest, Registry, Security)
  - Initial validation tool for PDS4
  - Initial Report Service
  - Documentation



http://pds-engineering.jpl.nasa.gov

Build 1: PDS4 Standards Build / Assessment Alignment



"1B" Assessment

- Emerging PDS4 data standard reviewed as part of build 1b assessment

- PDS4 Build 1b was completed on January 15, 2011 by an internal PDS assessment team
  - The purpose is to allow for improvements in the design and implementation of PDS4 ensuring that PDS has a process for involving and including comments from its stakeholders early
  - Each PDS node participated
  - More detail to come

- The review was performed on the PDS4 emerging standard, specifically
  - Concepts Document
  - Standards Reference
  - Data Dictionary
  - Data Providers Handbook
  - Glossary

“1C” Assessment

- Emerging PDS4 data standard under review as part of build 1c assessment by the IPDA PDS4 Project Team
- Two phased approach to assessment covering documents (Phase I); prototype products (Phase II)
- Phase I completed
- Phase II underway

“1d” Assessment

- Currently planned for August to get input from external stakeholders
- Review of key standards documents
- Final review prior to build 2 “ORR”

System Testing

- Distribution packages available for download and test:
  - Validate Tool, Harvest Tool and Registry Service/UI.
  - Design Tool consists of documentation and links to the third-party software.
  - Report and Security services are EN hosted applications.

- System testing occurs after each build
  - Test plan posted under project page at EN site

- Much of the testing to date has been at EN covering tools and services (e.g., harvest/registry)
  - But, nodes are beginning to test as well
  - This occurring as we approach build 2

Build 2 Release Plan

- Complete PDS4 build 2 (October 2011)
- Conduct Operational Readiness Review (November 2011)
  - Assemble review to assess any final issues that need to be addressed prior to production release of PDS4
  - Finalize steps for release
  - More on next page
- Release PDS4 (January 2012)
  - Switch over software system and supporting catalog
  - Post PDS4 standards documents
- Continue future build/release process
- Begin Migration

Transition Planning

- With Build 1 in place, we are actively working on system transition
  - Processes to support data delivery/ingestion
  - Support PDS3 pipelines and data to ingest into PDS4 system
  - Migration of central catalog into registry system
  - Integration of PDS3 web interfaces to support PDS4
    - E.g., Homepage will be driven by the registry and central catalog will be shutdown.

- Key driver is helping missions plan for and design PDS4 products at first based on PDS4 standards
  - Beebe led analysis for planning PDS4 mission needs, dates, product types, etc to link the schedule (in backup of this presentation)
  - Actual receipt from missions won't be until PDS4

- More to come…

Mission Support Plans / Alignment

- PDS has been working to align its build 2 release plans with the early needs of upcoming missions
- LADEE, MAVEN and Phobos Grunt have been used as drivers
  - They are PDS3 by default, but working with PDS to plan for PDS4
- An analysis of their data standards requirements was performed by PDS
  - They require a subset which is what is planned for build 2 (more to come…)

MC Policy on PDS4 Compliance

- NASA/PDS hereby specifies 1 November 2011 as the transition date from PDS3 to PDS4. NASA missions confirmed for flight after that date will be required to archive their data according to PDS4 standards, while missions confirmed for flight prior to that date are allowed to continue using PDS3. Missions allowed to use PDS3 are encouraged to consider the advantages of switching to PDS4 and NASA will consider proposals for funding augmentations to accomplish the change. Data providers in R&A programs will be required to comply with PDS4 standards for any programs with proposal deadlines after 1 November 2011. Data providers whose proposals were submitted prior to that date are encouraged to consider the advantages of complying with PDS4 standards but will be allowed to complete their deliveries according to PDS3 standards. – November 2010

Migration

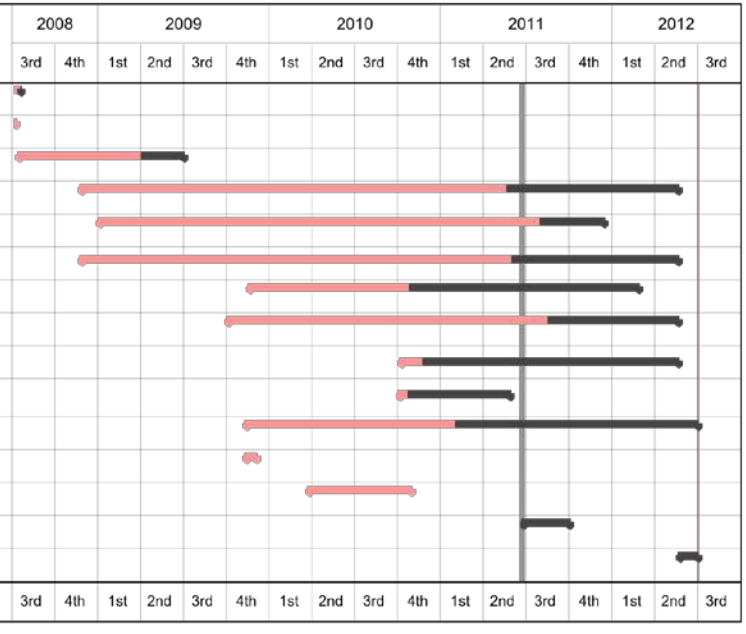
- The migration plan assumes that each node would develop a plan based on demand, available resources, etc
- Some nodes have already begun working on this
  - Atmos will report on their efforts tomorrow
- At the last tech session, it was discussed that this should be one of the major discussions at a Fall Tech session
  - The timing will be right with the build 2 release planned

Project Resources

- 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
  - For the most part, the project doesn't have direct funding, but rather is level of effort across PDS
- In 2010, 2011 and 2012, the PDS requested additional overguide for PDS nodes to cover PDS 2010 which was provided.

High Level Schedule

	Activity Name	Duration (Work Days)	Start Date	Finish Date
1	Concept/Study Phase...	240.00	8/20/07	7/18/08
10	Project Planning...	133.00	1/7/08	7/9/08
20	Architecture/System Engineering...	254.00	7/10/08	6/30/09
34	Projects	911.00	11/21/08	5/18/12
35	P1. Data Standards Project...	767.00	1/2/09	12/12/11
83	P2. Distributed Infrastructure Project...	911.00	11/21/08	5/18/12
106	P3. Tools Project...	595.00	11/16/09	2/24/12
113	P4. Distributed Catalog System Project...	688.00	9/30/09	5/18/12
126	P5. Portals, Search and Distribution Project...	425.00	10/4/10	5/18/12
137	P6. Data Movement and Delivery Project...	171.00	10/1/10	5/27/11
143	Builds	690.00	11/9/09	6/29/12
144	B0. Infrastructure Build...	20.00	11/9/09	12/4/09
147	B1. Ingestion Build...	160.00	3/22/10	10/29/10
151	B2. Distribution Build...	74.00	6/21/11	9/30/11
155	B3. User Capabilities Build...	30.00	5/21/12	6/29/12

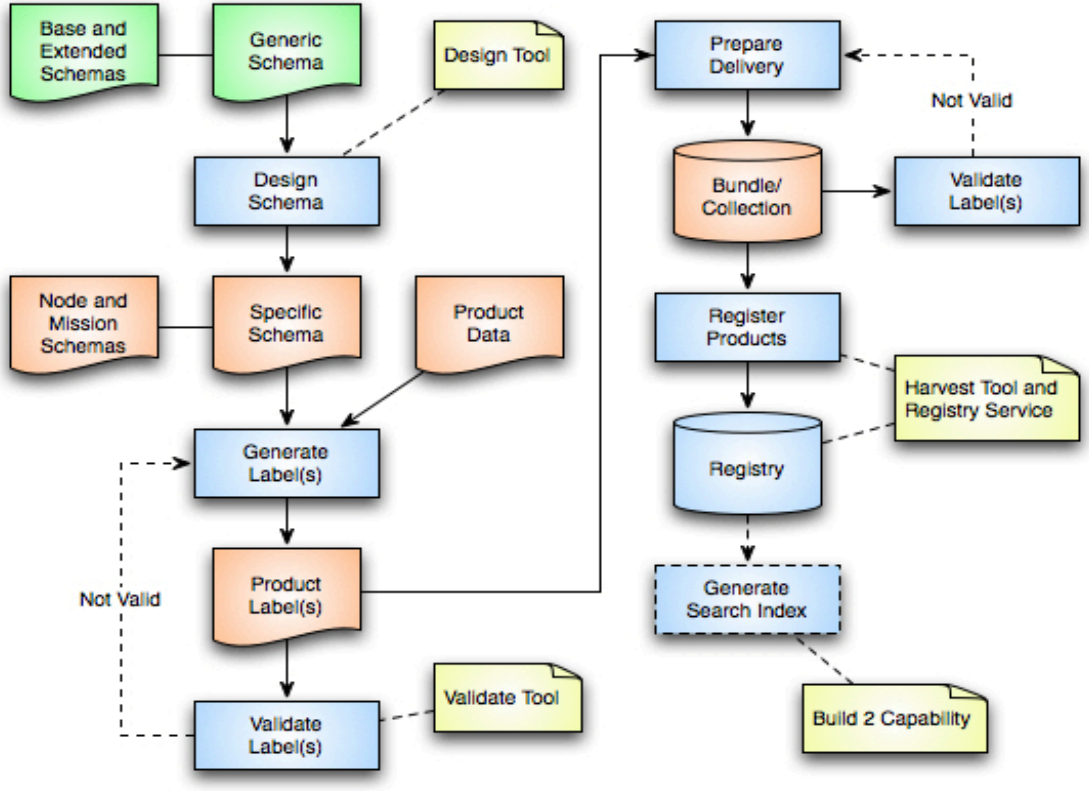


Conclusion

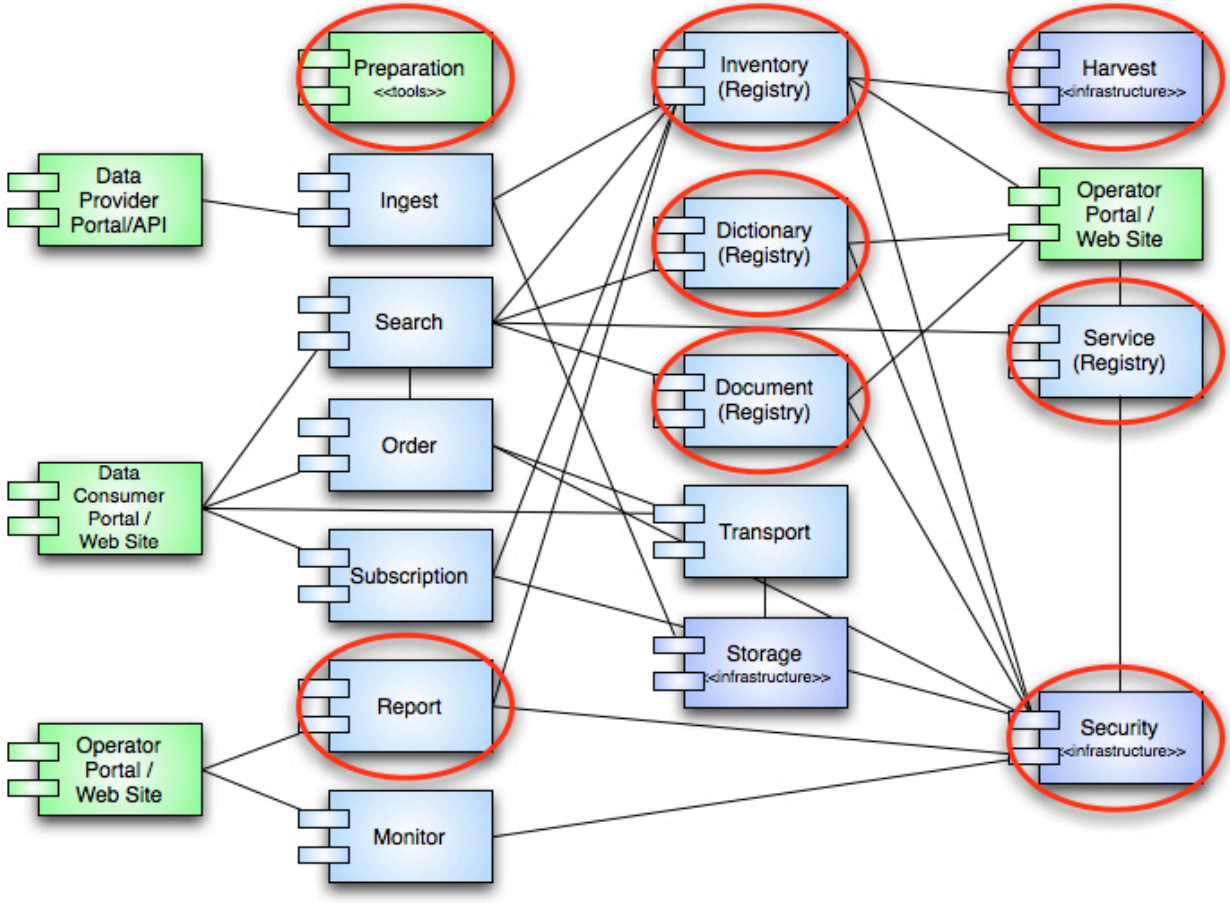
- Continued excellent cross-PDS collaboration
- Good balance in achieving new development on top of running an operational system
- Established a build/delivery structure that is working well
- Community engagement
- A lot of interest in the PDS architecture
  - Many presentations for different venues and disciplines

Backup

System Role in Product Lifecycle



System Design Concept



Results (LADEE)

Mission	Instruments	Product Types	Product Design Start Date (Est)	Launch Date	PDS Product Ingestion Date	PDS4 Project Alignment	Data Supplier Document & Tool Needs
LADEE	Neutral Mass Spectometer (NMS) Dust Dector (LDEX) Ultraviolet Spectrometer (UVS) Ancillary Data	ASCII Tables ASCII Tables Tables/Spectra Documents SPICE? Browse Products	Early 2012	May 2013	Orbit insertion + 6 months	2011 Operational Deployment for PDS4 will sync with Product Design start date.	Documentation: -Concepts document -User guide/tutorial /DPH -Standards Ref – -Data Dictionary **Tools:** -Create local data dictionary -Label design (modify generic schema into a specific one and generate label template) -Label validation -Validation tool to test that the label correctly describes the data

Results (MAVEN)

Mission	Instruments	Product Types	Product Design Start Date (Est)	Launch Date	PDS Product Ingestion Date	PDS4 Project Alignment	Data Supplier Document & Tool Needs
MAVEN	Mass Spectrometer (NGIMS) Ultraviolet Spectrometer (IUVS) 6 field & particles instruments (package) Ancillary Data	ASCII Tables FITS files Combination of CDF, ASCII Tables Documents SPICE? Browse Products	SIS in Progress; design in late 2011/early 2012	November 2013	Six months after orbit insertion	2011 Operational Deployment for PDS4 will sync with Product Design start date.	Documentation: -Concepts document -User guide/tutorial /DPH -Standards Ref – -Data Dictionary **Tools:** -Create local data dictionary -Label design (modify generic schema into a specific one and generate label template) -Label validation -Validation tool to test that the label correctly describes the data

Results (Phobos-SRM)

Mission	Instruments	Product Types	Product Design Start Date (Est)	Launch Date	PDS Product Ingestion Date	PDS4 Project Alignment	Data Supplier Document & Tool Needs
Phobos-Grunt	TV System (TVs)	FITS files ASCII tables	Working on a pipeline and sample dataset	Dec 2011	Plan to prepare the pipeline in advance to be ready to release the data almost immediately after the acquisition to make the data available for the landing site selection	Prepare the pipeline for review by the middle of 2011	1. Concepts Manual 2. Observation Data format descriptions (a subset of the full standards reference) 3. Mission dictionary creation interface definition 4. Quick & Dirty local dictionary creation tool 5. A working set of example schemas (dictionary schemas, generic schemas, derived schemas) to use in learning XML and XML Schema.
	Gamma ray Spectrometer (GSP) Neutron Spectrometer (NSP) Laser TOF Mass Spectrometer (MSS) Seismometer (SSM) Long-Wave Penetrating Radar (LWR)				Archived not in PDS but in RPDA (Russian Planetary Data Archive).		

TGO – A Problem Mission

The ExoMars Trace Gas Orbiter (TGO)

•Mark Allen is US Program Scientist, Stephen Spohn is heading up data planning

•Olivier Witasse is ESA Program Scientist & may be PS for the lander –Good News!

•NASA Dominates the Orbiter and ESA the Lander

•Launch Jan or Feb 2016 & Data Delivery ~ May 2017

•No NASA money for data planning until 2013

Problems

- Data management plan not settled between NASA & ESA
- Small NASA Investment in a demanding mission with 5 instruments
- 4 Instrument PIs set up for PDS3 deliveries – other ESA/unknown

•Mars Atmospheric Trace Molecule Occultation Spectrometer (MATMOS) Wennberg/Caltech

•High-resolution solar occultation and nadir spectrometer (SOIR/NOMAD) Vandaele/Belgium

Results (EXOMARS)

Mission	Instruments	Product Types	Product Design Start Date (Est)	Launch Date	PDS Product Ingestion Date	PDS4 Project Alignment	Data Supplier Document & Tool Needs
EXOMARS TRACE GAS ORBITER	Mars Atmospheric Trace Molecule Occultation Spectrometer (MATMOS) Wennberg/Caltech	Table	2013	Jan or Feb 2016	May 2017	PDS4 development will not be impacted	Documentation and tool support will be required
	High-resolution solar occultation and nadir spectrometer (SOIR/NOMAD) Vandaele/Belgium	?? Table, images, cubes??					
	ExoMars Climate Sounder (EMCS) Schofield/JPL	Table/like MRO					
	High-resolution Stereo Color Imager (HiSCI) McEwen/U of Ariz.	Images/cubes HiRise					
	Mars Atmospheric Global Imaging Experiment (MAGIE) Cantor/Malin SSS	Images Malin					
	Ancillary Data						