International Planetary Data Alliance

PDS4 Background Information for Assessors of PDS4 Data Standards

April 12, 2011

Outline

- Background
- Document Set and Reference Materials
- Key PDS4 processes
- Assessment steps

- Review Materials

  http://pds.jpl.nasa.gov/build1creview/

  Username: ipda

  password: 1cReview

PDS4 Data Standards

- Redesigned using current best practices for data system development.
- Designed with fewer, simpler, and more rigorously defined data formats for science data products.
- Used XML, a well-supported international standard, for data product labeling, validation, and searching.
- Implemented 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.

PDS4 Progress

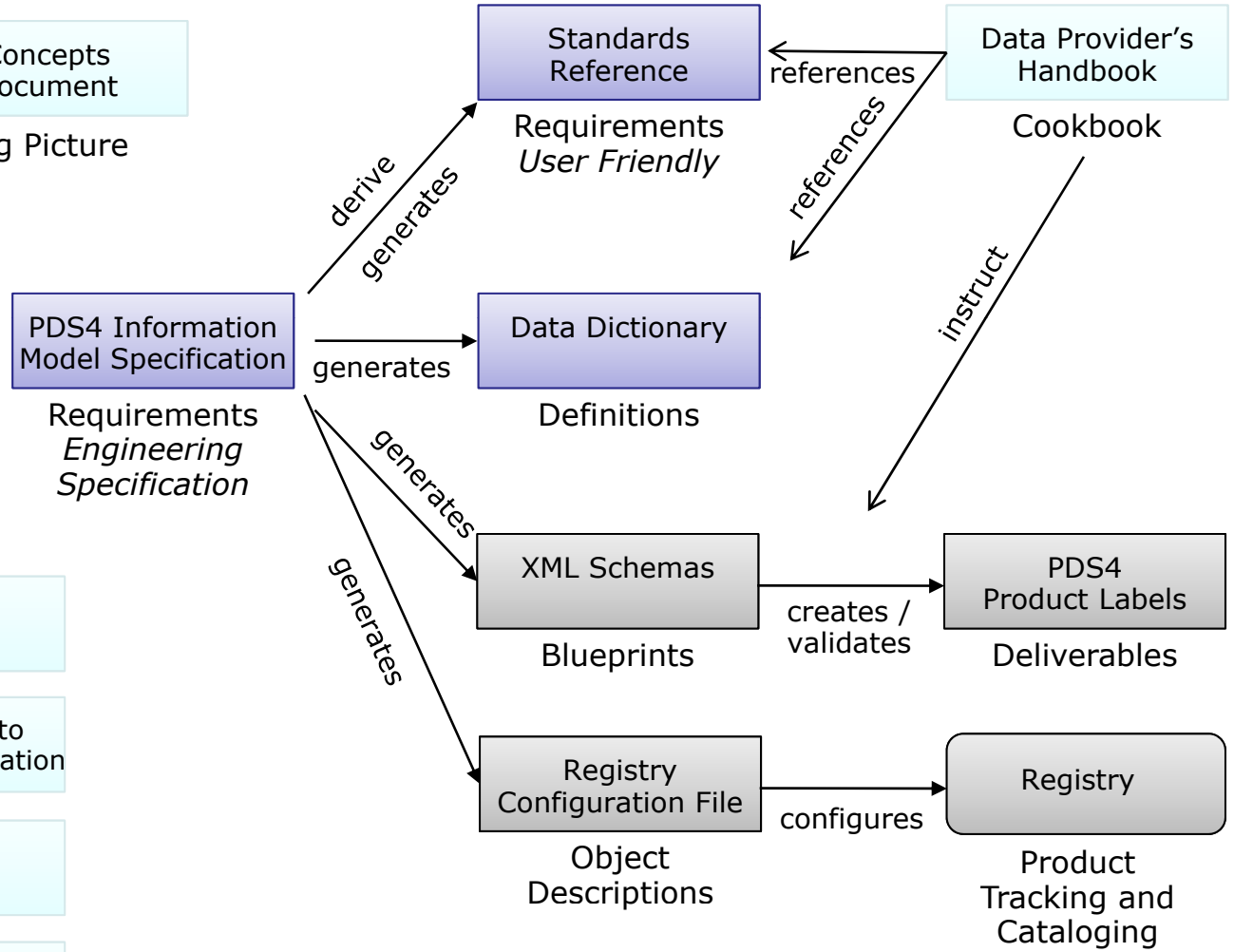
- Data Standards
  - Significant progress in generating data standards
  - Completed internal PDS review
  - Continuing to develop and update
  - IPDA input is critical

- Software Infrastructure
  - Core PDS 2010 infrastructure being tested for ingestion
  - Testing migration of PDS3 databases
  - Working on distribution system

Response to 2010 IPDA Assessment

- Rewrote or redesigned most documents
- Wrote additional informative documents
- Key Issues Addressed:
  - Rewrote many definitions
  - Wrote more supporting material such as tutorials and user guides.
  - Rewrote the PDS4 data provider's handbook
  - Released the standards reference document
  - Currently finalizing local data dictionaries and their management process

PDS4 Documents and their Relationships



PDS4 Data Standards Documents*

The following eight 'documents' on the next two slides describe the Planetary Data System version 4 (PDS4). They are listed roughly from most general to most specific — probably the same order in which you should read them.

1. Introduction – A guide to get you started.*

2. Concepts Document – Introduction to PDS4 key concepts — the view from 10000 feet, avoiding gory details.

3. Glossary – A concise set of definitions for key PDS4 terms. Although primarily intended as a quick reference, the Glossary is organized functionally, presenting terms in the approximate order in which you are likely to encounter them.

4. Jumpstart Guide – A brief introduction to PDS4 in terms of analogous PDS3 vocabulary. Experienced PDS3 users should read it once, noting both the parallels and the differences; then set it aside. People not familiar with PDS3 should skip it; concentrate on the Concepts Document.

5. Data Provider's Handbook – A cookbook to guide data providers step-by-step through the process of developing an archive.

PDS4 Data Standards Documents (cont'd)

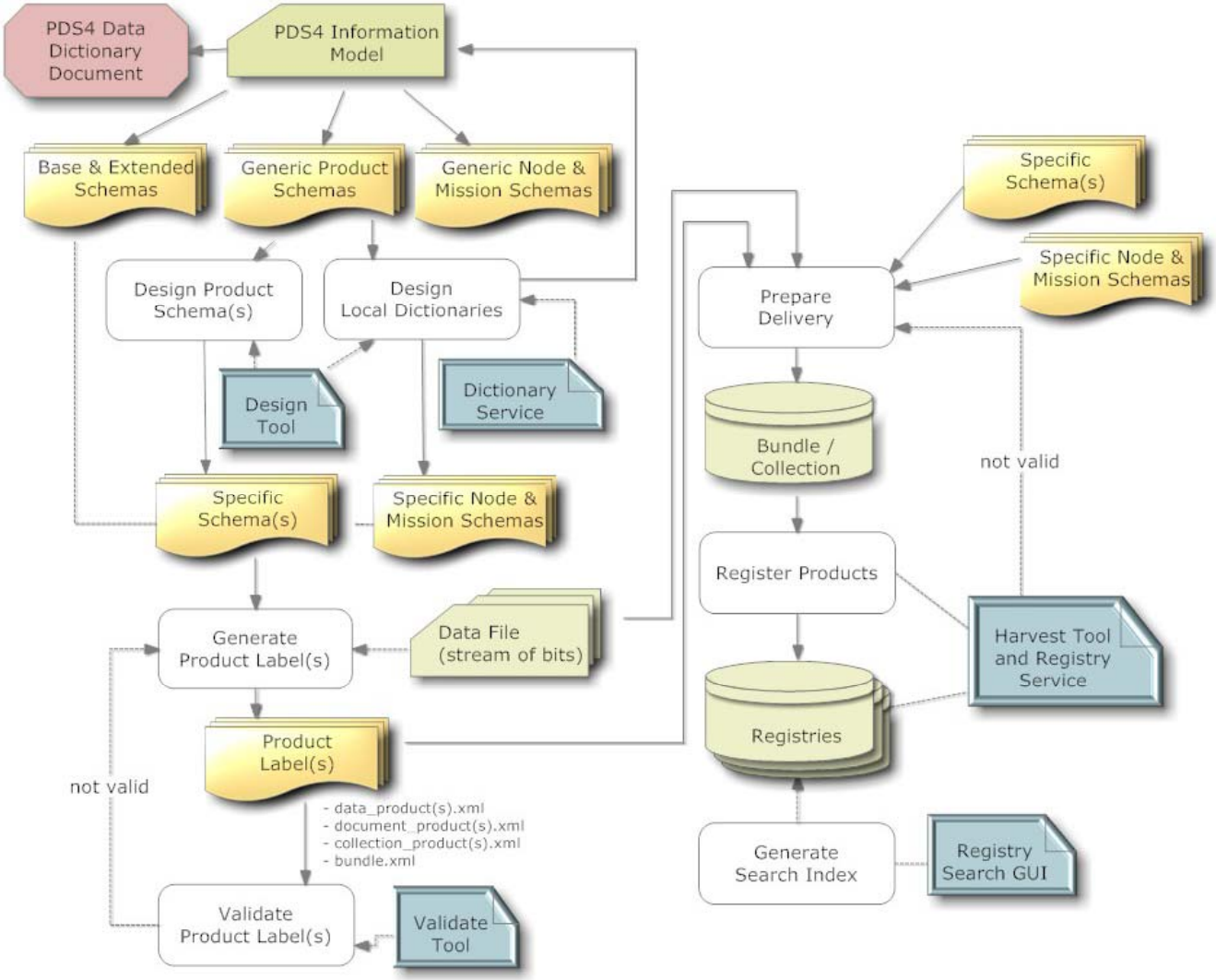
The remaining three are reference documents.

6. Standards Reference – One of the two fundamental reference documents for PDS4. You will need this as you work your way through the Data Provider's Handbook and as you prepare an archive.

7. Data Dictionary – The other fundamental reference for PDS4. It comes in two versions, abridged and unabridged. Use the abridged version unless you encounter a specific instance in which the information in the more detailed unabridged version is required. The abridged version has been abstracted from the unabridged version with the needs of data providers and data end users in mind. It contains full definitions but not all the fine detail or repetition necessary to support the underlying Information Model.

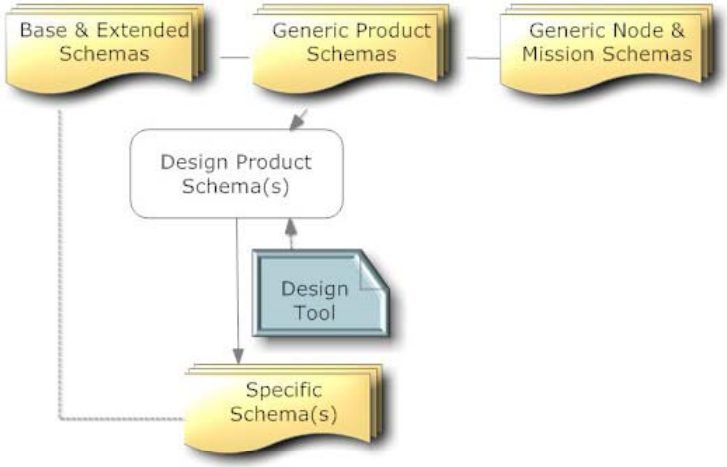
8. Examples – A set of products, collections, bundles, and packages that illustrates design concepts and goals. Frequently referenced by [5] and to be used in conjunction with [5-7] when constructing an archive.

PDS4 Processes Overview



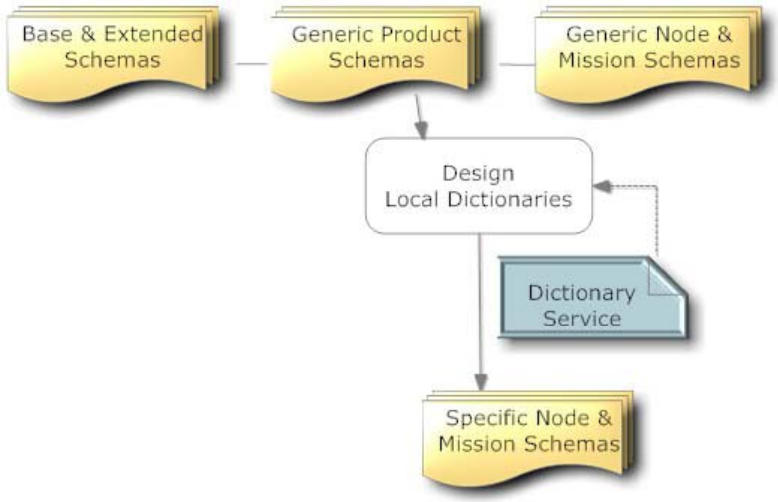
PDS4 Processes
Product Design

- Modifying the "generic" schema to be "specific" to a particular product is a necessary step in setting up the data product pipeline



PDS4 Processes
Local Data Dictionary Design

- Building and using dictionaries that are specific to either a mission or a node is a necessary step in setting up the data product pipeline



Phase 1 - Assessment

- The assessment involves the following steps
  - Read the Introduction document on the Build 1c review page
  - Review the documents listed for Review on the review page
  - Review the example products on the review page
  - Answer a set of questions (on the assessment response sheet)
  - Capture the issues (on the assessment response sheet)

- Materials are posted at the following review site
  http://pds.jpl.nasa.gov/build1creview/
- Please send review materials and input to Steve Hughes by **May 15, 2011.**
- If you have technical questions, please contact Steve Hughes, Ron Joyner, or Mitch Gordon.
- If you have review questions, please contact Steve Hughes or Dan Crichton.

Phase 2 - Prototype

- The prototype involves the following steps:
  - Choose a familiar data set.
    - It is suggested that a simple data set with simple ASCII tables or grayscale images be chosen.
  - Choose the appropriate Generic Schema and create a Specific Schema for your products.
  - Generate a sample XML label and create product labels.
  - If time allows generate labels for a collection, archive bundle, and a document.
  - Report on the process.
- Materials are posted at the following review site
  http://pds.jpl.nasa.gov/build1creview/
- Please send the prototype report to Steve Hughes by **July 15, 2011.**
- If you have technical questions, please contact Ron Joyner, Mitch Gordon, or Steve Hughes.

Contacts

- Steve Hughes – Steve.Hughes@jpl.nasa.gov
- Ron Joyner – Ron.Joyner@jpl.nasa.gov
- Mitch Gordon – mgordon@seti.org
- Dan Crichton – Dan.Crichton@jpl.nasa.gov

THANK YOU FOR YOUR TIME!

Key PDS4 Definitions*

- **digital object**: An object which is real data — for example, a binary image of a redwood tree or an ASCII table of atmospheric composition versus altitude.

- **physical object**: An object which is physical or tangible (and, therefore, does not itself fit into a digital archive). Examples of 'physical objects' include the planet Saturn and the Venus Express magnetometer. Note that an ASCII file describing Saturn is a digital object, not a physical object (nor a component of a physical object).

- **conceptual object**: An object which is intangible (and, because it is intangible, does not fit into a digital archive). Examples of 'conceptual objects' include the Cassini mission and NASA's strategic plan for solar system exploration. Note that a PDF describing the

* from "GLOSSARY OF PDS4 TERMS"

Key PDS4 Definitions (cont'd)

- **description object**: Something that describes an object. As appropriate, it will have structural and descriptive components. Technically speaking, a 'description object' in PDS4 is a 'digital object' — a string of bits; but we assume that we can read it and, on that basis, give it a special name.

- **tagged digital object**: A digital object paired with its companion description object. [Note: In the OAIS RM this is known as an 'information object']

- **tagged non-digital object**: A physical object or a conceptual object paired with its companion description object. [Note: In the OAIS RM this is known as an 'information object']

- **information object**: A data object paired with its description.

Key PDS4 Definitions (cont'd)

- **product**: One or more tagged objects (digital, non-digital, or both) grouped together and having a single PDS-unique identifier. In the PDS4 implementation, the descriptions are combined into a single XML label. Although it may be possible to locate individual objects within PDS (and to find specific bit strings within digital objects), PDS4 defines 'products' to be the smallest granular unit of addressable data within its complete holdings.

- **basic product**: The simplest product in PDS4; one or more data objects (and their description objects), which constitute (typically) a single observation, document, etc. The only PDS4 products that are not basic products are Product_Collection and Product_Bundle. Every basic product must be a primary member of one (and only one) collection. Basic products may be secondary members of any number of collections

Key PDS4 Definitions (cont'd)

collection: A list of basic products, all of which are closely related in some way. A collection is itself a product (because it is simply a list, with its label); but it is not a basic product.

bundle: A list of collections. For example, a bundle could list a collection of raw data obtained by one instrument during a mission lifetime, a collection of the calibration products associated with the instrument, and a collection of all documentation relevant to those collections.