PDS/IPDA presentation to the IVOA

Steven Joy, Baptiste Cecconi

Outline

- Planetary Data System (PDS) activities
  - PDS Overview
  - PDS Archive Process
  - PDS4 Implementation and differences from PDS3
  - PDS4 Implementation of DOIs
  - PDS4 registry and search
- International Planetary Data Alliance (IPDA) activities
  - IPDA Overview
  - IPDA Structure
  - IPDA Goals and Challenges
  - IPDA Progress and Future Directions
- Solar System Interest Group (SSIG)
  - Comments
- Summary

Planetary Data System (PDS) Overview

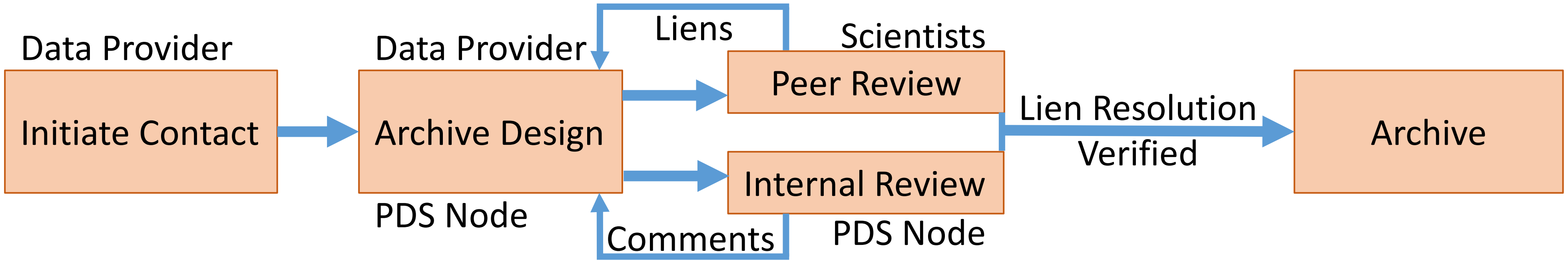
- PDS is, and has always been foremost a data archive and secondarily a data system
  - As an archive, PDS requires that data and documentation are stored in standard, non-proprietary formats
    - This often precludes the use of common user formats such as MS Word for documents or IDL savesets
  - PDS is an open archive, available to anyone, anywhere
- PDS is organized on the principal that planetary data are diverse, and that the best curators of such a diverse archive are domain experts (scientist and technologists) in the various fields of planetary science. As such, PDS is a federation of Science Discipline and Support Nodes:
  - **Atmospheric Sciences (ATM)** – *New Mexico State Univ.* – **Nancy Chanover**
  - **Geosciences (GEO)** – *Washington Univ, St. Louis* – **Ray Arvidson**
  - **Imaging and Cartography (IMG)** – *USGS Flagstaff* – **Lisa Gaddis**
  - **Planetary Plasma Interactions (PPI)** – *UCLA* – **Ray Walker**
  - **Ring Moon System (RMS)** – *SETI* – **Mark Showalter**
  - **Small Bodies (SBN)** – *Univ of Md., College Park* – **James (Gerbs) Bauer**

  - **Navigation and Ancillary Information (NAIF)** – *JPL* – **Chuck Acton**
  - **Engineering (ENG)** – *JPL* – **Dan Crichton**
  - **Management (MNG)** – *GSFC* – **Tom Morgan**
- Most of the Science Discipline Nodes have multiple sub-nodes that provide expertise across many aspects of the discipline that are not available at the Node location

Planetary Data System (PDS) – Some History

- PDS 1 went operational in 1989
  - High level catalog for finding datasets by mission, instrument, spacecraft and target
  - Object Description Language (ODL) is invented for product labeling and catalog information
  - Data archived and distributed on 9-track tape
- PDS 2 (1990)
  - Catalog information streamlined, longer text descriptions, fewer keyword/values
  - CD-ROM (CD-R) becomes the primary archive and distribution media
- PDS 3 (1992)
  - PDS sets up and maintains a web presence
  - Movement to online distribution (~2002) of electronic volumes (PDS-D)
  - Online mass storage and data bricks replace CD/DVD as archive distribution media
- PDS 4 (2014)
  - Prototype build 1 in 2010, operational in 2014 with LADEE as first mission to use new standard
  - Model driven architecture with XML labels
  - Distributed registries and search services replace the catalog and support direct file level access
  - New data standards and reduced set of archive data formats and structures
  - International collaboration

PDS Archive Process



- The PDS archive process should be an iterative process between the Node and the data provider(s) during which the content and structure of the archive are negotiated, delivered, reviewed, and revised

  1. **Archive Design** – Providers propose a collection of data and documentation for archive and then work with the PDS Node to refine the design. Documentation is written, sample products with labels are delivered for internal review.
  2. **Peer Review** – Once the Node and data provider agree that the data are ready to archive, a formal peer review is set up with the review panel consisting of at least two independent science reviewers, PDS standards reviewers, and the data providers. The review results in one of three outcomes:
     - **Archived**: Data and documentation are accepted for archive – no issues found, no liens.
     - **Certified with Liens**: Data are certified as scientifically useful as they are but there are some issues that need to be fixed before they can be archived. Data can be released to the public in the interim.
     - **Not certified**: Data or documentation have deficiencies that prevent them from being scientifically useful until the issues are corrected. Archive will need another review before being released to the public.
  3. **Lien Resolution** – The data provider corrects issues reported during review and resubmits the data for archive.
  4. **Archive** – All issues reported during the review have been addressed. Online version has usage warnings removed and the archive is sent to the NSSDCA for deep archive.

PDS4: An International Planetary Archiving Effort

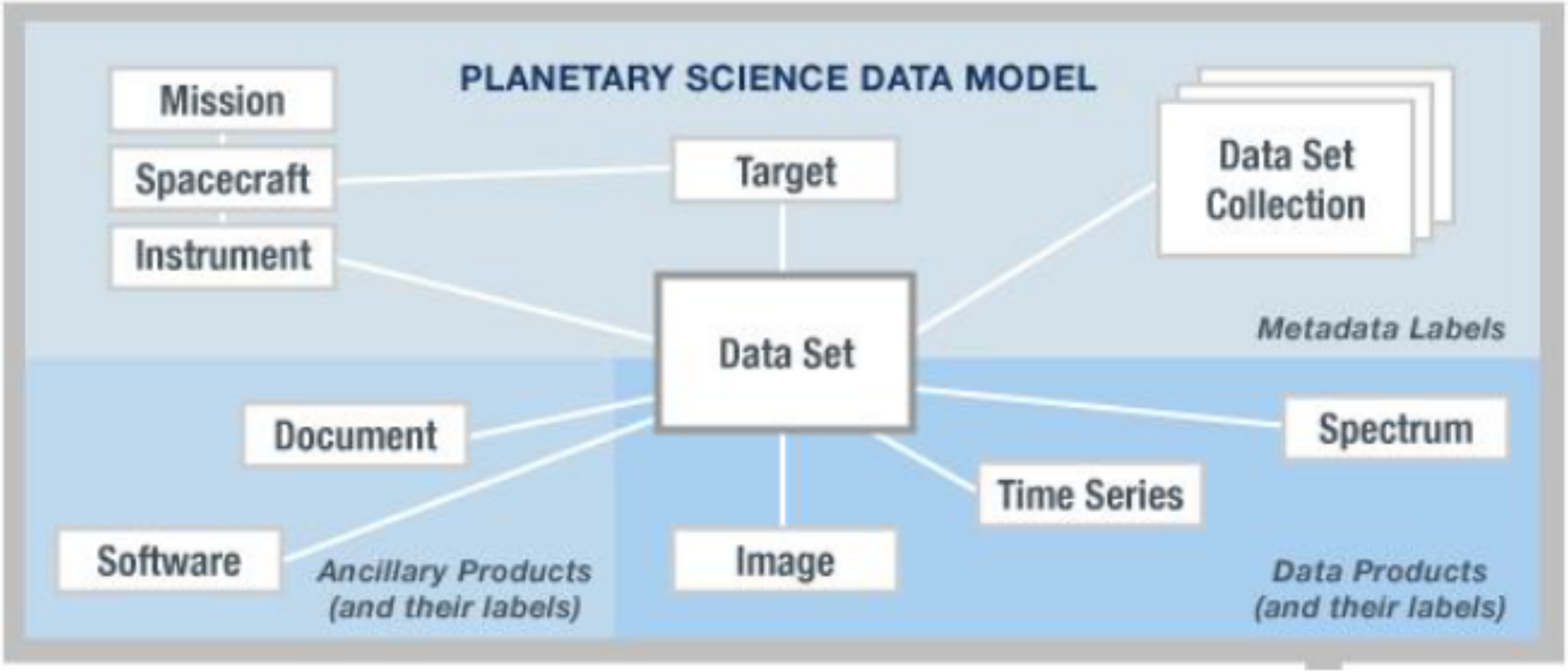
- PDS4 is an international planetary archive data model and standard that is being used to build compatible archives
- An **explicit information architecture**
  - All planetary data tied to a common model to improve <u>validation</u> and <u>discovery</u>
  - Use of XML, a well-supported international 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.
- An **explicit software/technical architecture**
  - Distributed services both within PDS and at international partners
  - Consistent protocols for access to the data and services
  - Deployment of an <u>open source</u> registry infrastructure to track and manage every product in PDS
  - A distributed search infrastructure

PDS4 Differences from PDS3

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; *separation of the storage and display formats*.

Slide taken from D. Crichton presentation on Operational Readiness – 2013-09-13

PDS4 Data Model



The PDS4 data model grows from the concept of a data set, its contents, and how is can be described and documented in a searchable data system

A dataset consist of collection of digital objects (data, documents, calibration, etc.) that when described by meta-data (labels) become digital products

Additional metadata are required the describe the conceptual objects (missions, targets, etc.) in order to support searches

Slide taken from D. Crichton presentation on Operational Readiness – 2013-09-13

PDS4 Archive Organization

- There are 3 primary types of products in PDS4:
- **Basic Products** are the smallest unit of a PDS4 archive. They consist of an individual label and the associated file or files.
- Related basic products of the same type may be grouped together into a **Collection**.
- Related collections may be grouped together into a **Bundle**.

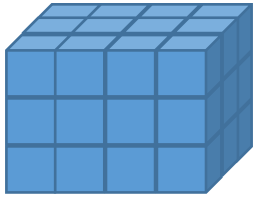
Slide taken from J. Mafi PDS4 training presentation, April 2018

PDS4 Logical and Version Identifiers

- A **Logical Identifier** (**LID**) is a unique ID that may be used to identify and reference any PDS4 product.
  - LIDs must be unique across the entire data system
  - LIDs take the form of a **Uniform Resource Name** (URN)
  - **LID: urn:nasa:pds:bundle:collection:product**
    - urn, agency, and organization are static, but may vary by archiving organization (e.g. “urn:esa:psa”, “urn:jaxa:darts”, etc.)
    - **bundle** is a bundle identifier (e.g. “maven-swea-calibrated”)
    - **collection** is a collection identifier (e.g. “data-svy-pad”)
    - **product** is an identifier for the individual product
- The product **Version Identifier** (**VID**) may be appended to the LID to form a **LIDVID**
  - A double colon (::) is the delimiter to separate the VID from the LID
  - **LIDVID: urn:nasa:pds:bundle:collection:product::vid**
- When a product is requested using only a LID, the most recent version of the product is returned. Specific versions of products can be requested by specifying the full LIDVID.

Fundamental Data Structures

PDS4 archive products must be describable using one of the following fundamental structures:



Array – homogenous binary structures of 1 to 16 dimensions in which all of the elements have the same data type.

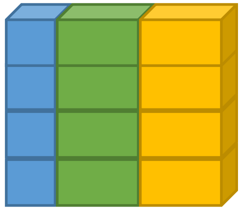
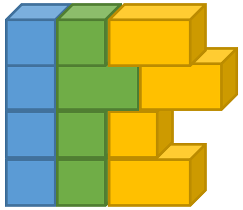
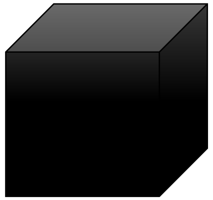


Table – ASCII or binary data with a repeating record structure made up of fixed-width fields.



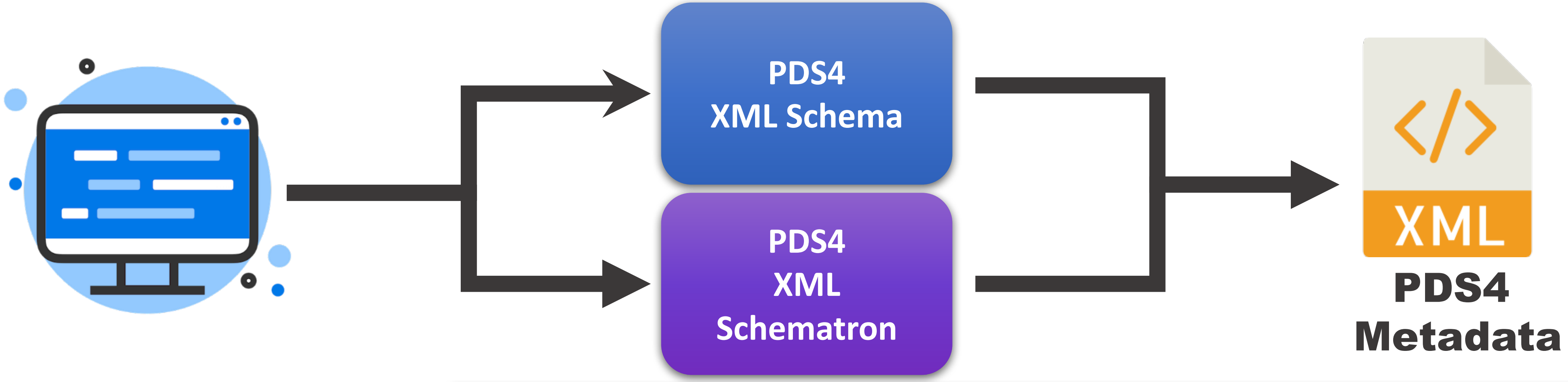
Parsable Byte Stream – ASCII data with a repeating record structure made up of variable width fields separated by a field delimiter (e.g. CSV).



Encoded Byte Stream – Files formatted according some established standard (e.g. PDF).

Slide taken from J. Mafi PDS4 training presentation, April 2018

PDS4 Implementation



The structure and content of PDS4 metadata is defined by a formal **Information Model**.

PDS4 is implemented in XML and expressed in terms of XML **Schema** and **Schematron** files.

- Schema define the metadata structure
- Schematron provide rule-based constraints on elements and content

Slide taken from J. Mafi PDS4 training presentation, April 2018

Anatomy of a PDS4 Label

XML Declaration

- XML identification tag; Schematron identification (optional)

Product Tag

- Root tag; Namespace declarations; Schema identification

Identification Area

- Product identifying information

Observation/Context Area

- Product provenance/background

Reference List

- Links to relevant products and publications

File Area

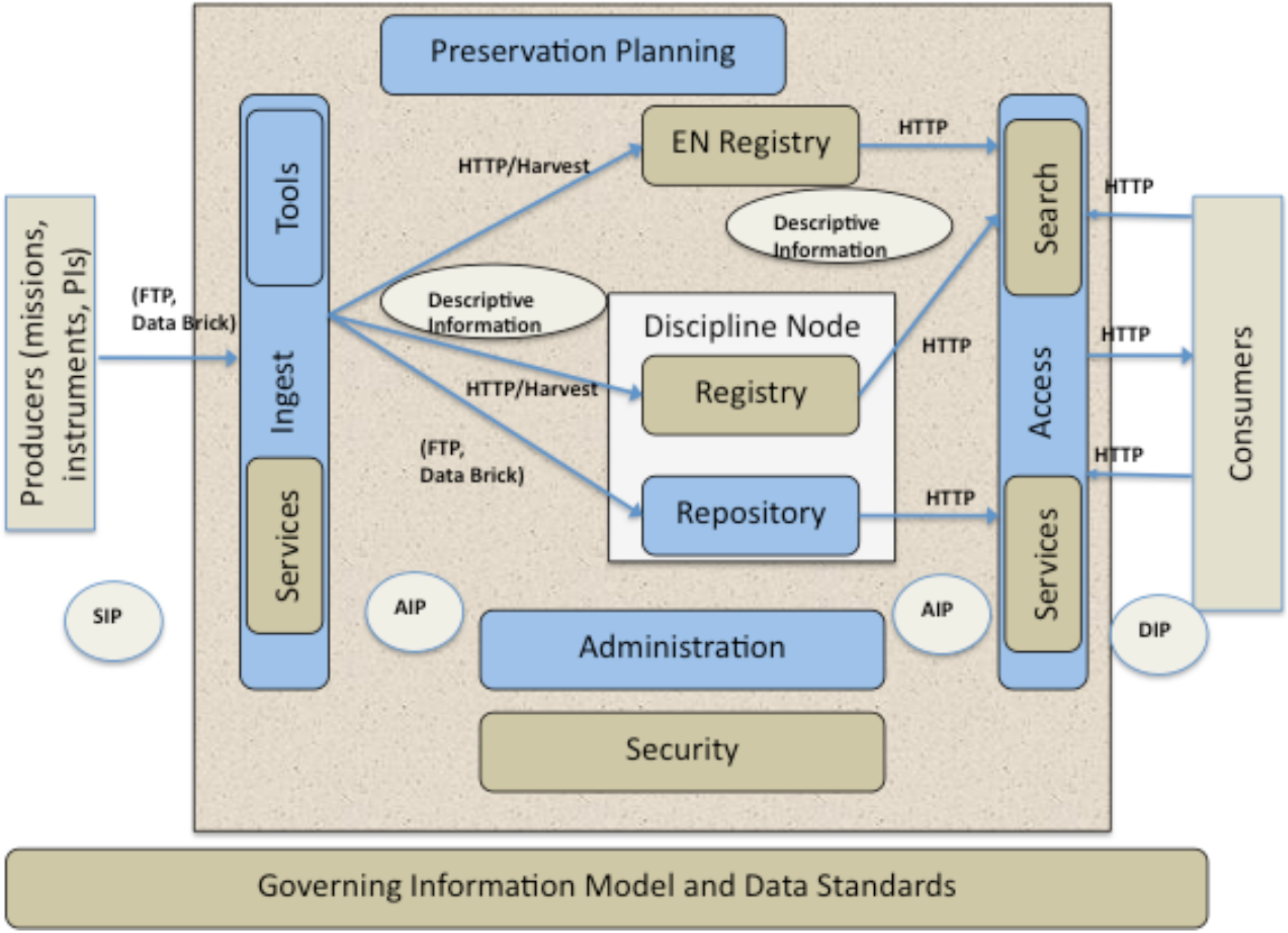
- File format and/or structural information

Slide taken from J. Mafi PDS4 training presentation, April 2018

PDS4 Use of DOIs

- The most recent version of the PDS4 data model does not yet support the use of DOI’s
    - The change request for including this attribute was still under review at the freeze date of the last build.
    - DOI’s will be included in the next build that will become public in October 2018
- PDS is currently generating DOIs for all PDS4 collections, and for selected documents and data products
    - Data archive SIS (software interface specification) and calibration procedure documents will be assigned DOIs – lists of project personnel will not be
    - Only special data products that are expected to be sited in the literature frequently are likely to be assigned DOIs – no data products to date to my knowledge
- PDS is registering DOIs with the U.S. Dept. of Energy Office of Scientific and Technical Information (OSTI)
    - DOIs can only be assigned by LIDVID since they refer to a specific version of a product or document that must always be valid
    - A DOI may be reserved for a collection or document so that it can be included in a label prior to release (publication)

Metadata Harvesting and Registration

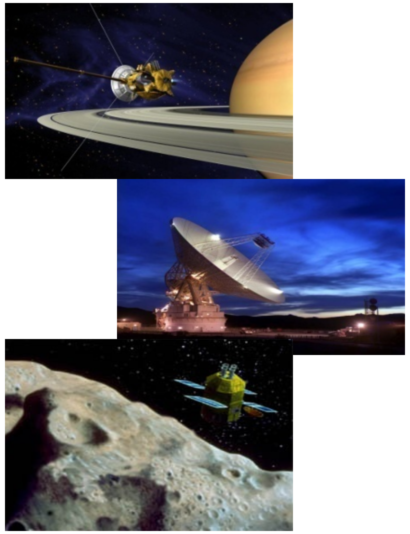


JPL
Jet Propulsion Laboratory
California Institute of Technology

Slide taken from D. Crichton presentation on Operational Readiness – 2013-09-13

International Planetary Data Alliance

http://planetarydata.org



Mission of IPDA*

- Facilitate global access to, and exchange of, high quality scientific data products managed across international boundaries
- Support construction of compatible archives
- Support sharing of tools and software services
- Define data standards within the IPDA, including the data models and derived dictionaries, based on the NASA Planetary Data System (PDS) that is the de-facto standard for all planetary data at the time of the IPDA founding

* Extracted from IPDA Charter, July 2007

Structure

.IPDA

Steering Committee Members

28 Members
12 Countries / International Institutions
Chair: Tom Stein, Washington University
Deputy Chair: Christophe Arviest, ESA

Technical experts group

~20 Members

General activities

Annual meeting, usually in July or August, 20-30 participants
Regular teleconferences every 3 months 10-20 participants
Participation in related meetings : **COSPAR**, EGU, AGU, EPSC, etc...
IPDA Website https://planetarydata.org

IPDA Goals and Progress

1. Support construction of compatible planetary science data archives
   - In 2012, the IPDA endorsed PDS4 as the archiving standard for planetary data
   - Implementation, or a plan to do so, of the PDS4 standard across agencies for mission archive including ESA, IKI, ISRO, JAXA,KARI, NASA, UAE.
2. Support the sharing of tools and software services
   - The PDS4 validation tool us used by several IPDA members which improves interoperability between agencies.
   - The Tool Registry is another example of a virtual clearinghouse of planetary data related tools created by a variety of producers across the globe.
3. Facilitate global access to, and exchange of, high quality scientific data products managed across international boundaries
   - At present, there are REST based access services are in place between ESA, ISRO, and NASA.
   - High-level search between ESA and NASA has been demonstrated and is in place for both PDS3 and PDS4 archives.
   - Projects are underway to develop citation linkage with publications and to create bridges between the IPDA and the International Virtual Observatory Alliance (IVOA).

IPDA Future Challenges

1. Support construction of compatible archives
   - Guidelines on how to build high quality archives (content and review)
   - Improved support and documentation including tutorials for implementation
2. Support sharing of tools and software services
   - Service registry for GIS technologies and tools as suggested at the Planetary GIS meeting
   - Coordinated tool development across international boundaries
3. Facilitate global access to, and exchange of, high quality scientific data products managed across international boundaries
   - API access to archives
   - Continue to develop, evolve, and promote the Open Planetary Data Access Protocol (PDAP) services at JAXA including web mapping services
   - Integrate ISRO data into international search and access
   - Product level search and access across agencies
   - Integrate access with computation
   - Expand interoperability across space sciences

Solar System Interest Group (SSIG) of the IVOA

Charter/Goals

- I reviewed the SSIG charter and the only real suggestion that I have is that the statement be amended, probably in the "Goals" area, to add a statement about self-promotion
- Something like:

  Promote awareness in the scientific community of the ongoing efforts of the astrophysics, heliophysics, and planetary sciences discipline data systems/archives to support system interoperability and data exchange.

- Without public awareness and support, it will be difficult to have our efforts funded
  - Value to the community needs to be demonstrated by our actions and communicated to our peers through the coordinated efforts of our members

Solar System Interest Group (SSIG) of the IVOA

Reference Frames

- I’d like to strongly endorse adding support for common planetary references frames
- Key issues are naming conventions and the communication of definitions
  - Currently within the planetary community there are multiple names for the same frame
  - There are also frames that appear to be the same but have subtle differences that are not well communicated in the naming conventions and/or descriptions
- NAIF/SPICE is currently in the process of formally defining and naming a wide range of dynamic planetary/satellite reference frames that have been previously used in archiving or defined in the literature.
  - This effort imposes a uniformity in naming conventions
  - Formal mathematical frame definitions
  - Documentation of the original frame references and names (i.e. alias lists for many frames)
- If the NAIF/SPICE frame names/IDs become recognized as the proper way to uniquely specify planetary frames, and they are adopted by the archives and recorded in the metadata, then we may have a path forward for the exchange of frame information between data systems
  - I’m not suggesting that the IVOA community adopt or use SPICE, only that we considering leveraging the SPICE effort to formalize frame naming conventions

Summary

- PDS is a long standing NASA organization responsible for data archive
- PDS4 is a new information model driven approach to modernizing the archive standards used within the planetary science community
- PDS4 is being developed with international participation
  - Through the IPDA, member organizations can make standards change requests to the PDS4 data model and have a voting member on the change control review board
  - Designed to support interoperability between distributed planetary data archive systems
- IPDA is an international organization of planetary data archives with members from at least twelve countries
  - IPDA promotes the adoption of common standards, cross-system search and retrieval, and the sharing of software, tools, and methods of data exchange
- The SSIG is a fairly new organization within the IVOA the goal of working with other IVOA groups to review, assess, and propose low-impact adjustments to the IVOA standards to support solar system sciences
  - Standardizing lists of coordinate systems and reference frames is one of several areas where this group can have a positive impact