The Next Generation PDS Archive Data Standards

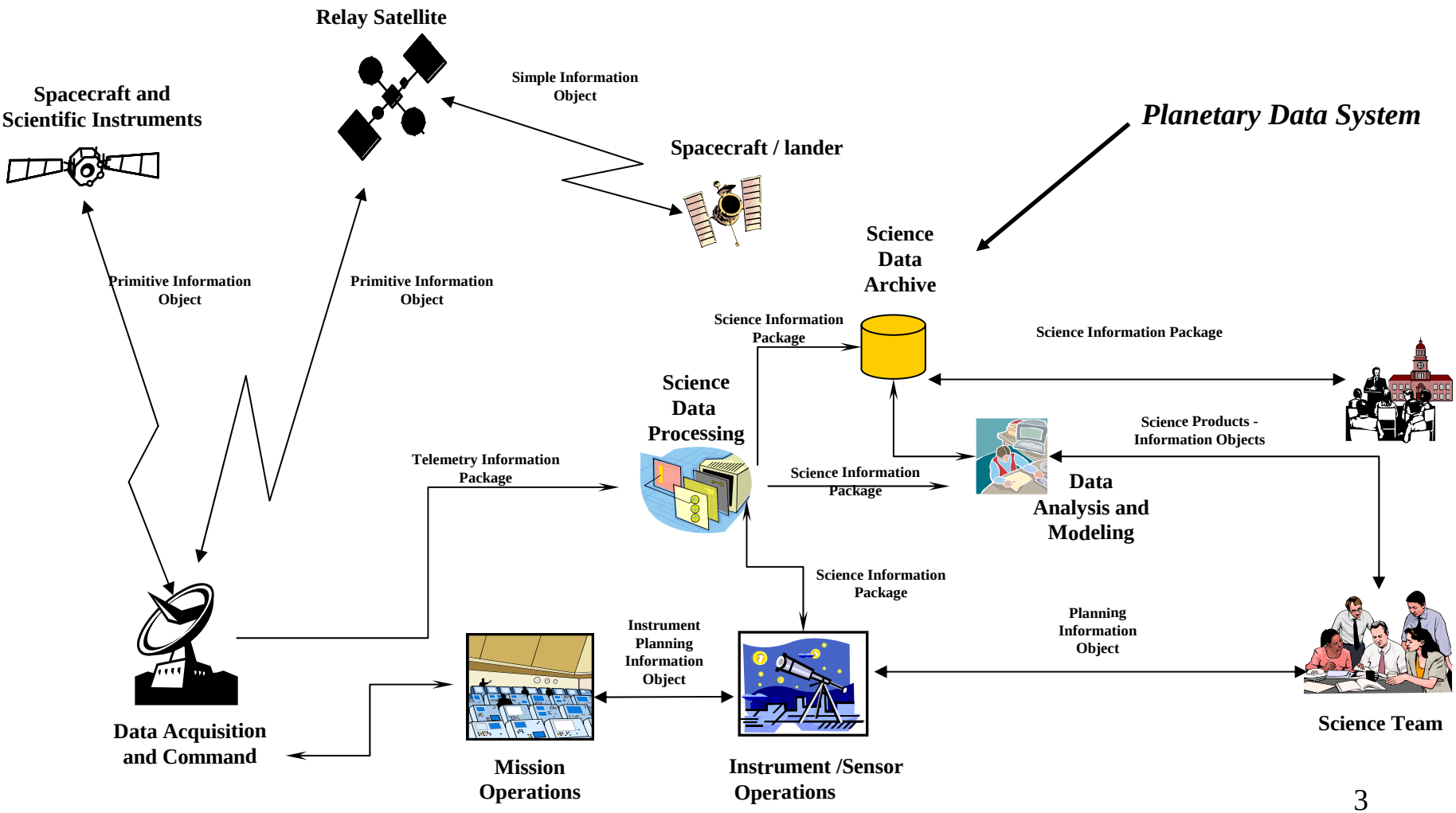
December 3, 2009

J. Steven Hughes, Anne Raugh, Mitch Gordon, Edward Guinness, Ron Joyner, Lyle Huber, Elizabeth Rye, Dan Crichton, Steve Joy, and Dick Simpson

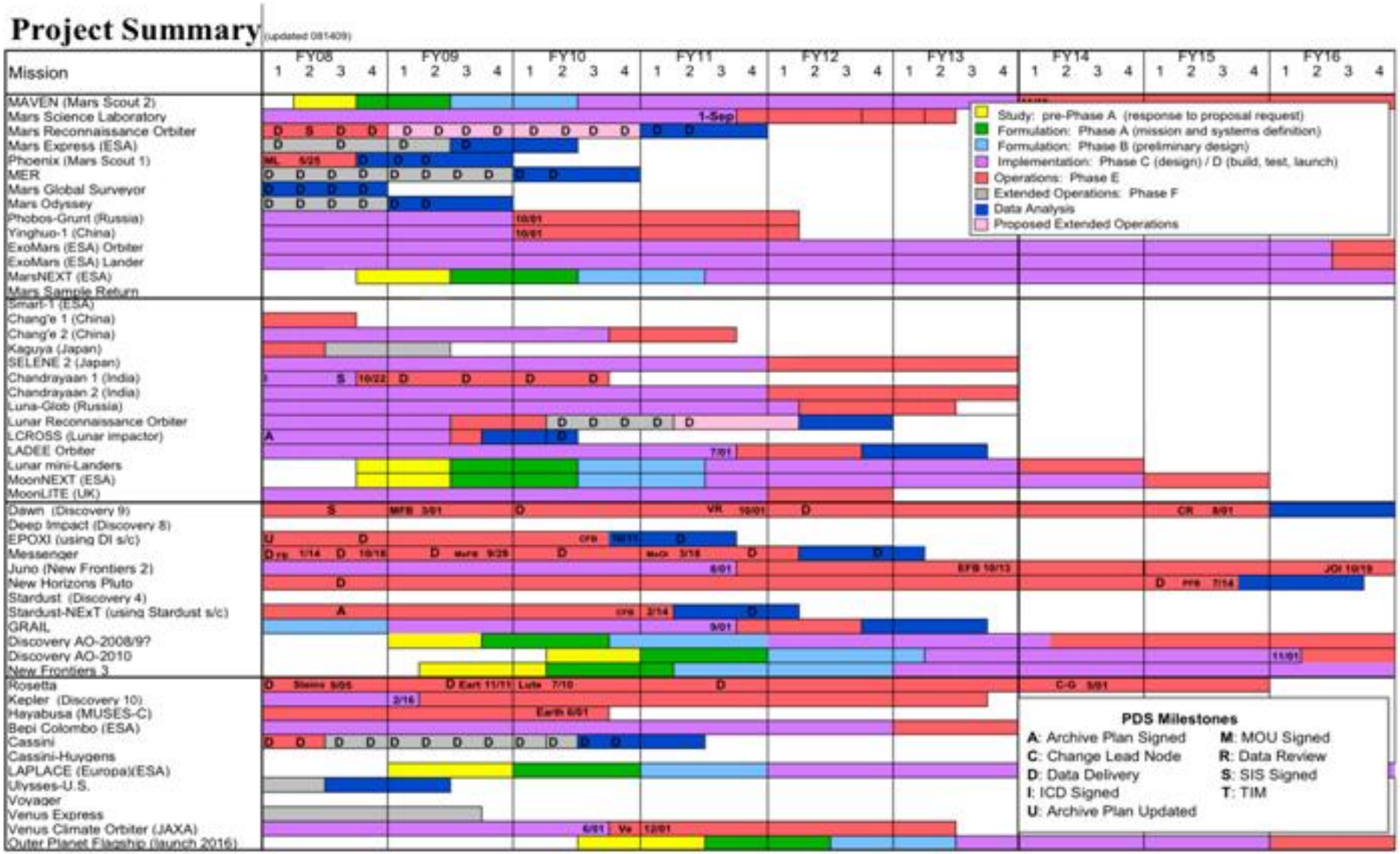
Topics

- Background
- Modernization of the PDS
- Data Design Goals
- Data Driven Development Methodology
- Overview of the PDS4 Data Standards
- Industry Standards
- International Collaboration
- Conclusion

PDS's Role in a Distributed Space Systems Architecture



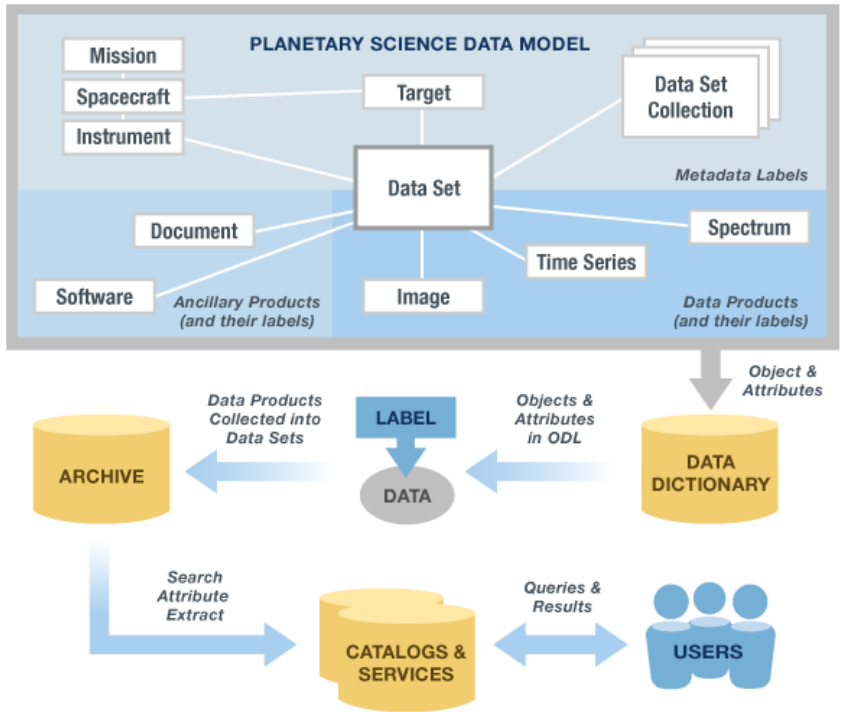
PDS participates in several concurrent missions



~20 concurrent missions will need to be supported by PDS over the next decade

PDS Functions

- PDS serves the scientific community by assuring the availability of high quality and scientifically useful data products
- To accomplish these goals, PDS
  - works with Data Providers to Prepare Archival Quality Data Products
  - provides Access to Data from NASA and International Missions
  - delivers Data to the Scientific Community
  - **establishes a Common Data Model and Data Dictionary for Planetary Data**
  - **sets Archival Standards**
  - preserves the Data
  - assists Scientists in Accessing and Using Planetary Data
  - is responsive to a diverse community of users
  - facilitates Education and Public Outreach



PDS Planetary Science Data Model

PDS3 Data Standards

- The PDS data standards were developed in the late 1980's to define the concepts and terms needed for archiving science data in the planetary science domain.
  - Data standards were innovative for their time however after almost two decades of use:
  - Ambiguity had crept in
  - Data formats had become obsolete
  - Usability software had become difficult to maintain
- These issues have caused significant problems for PDS operations, data providers, and end-users.

Modernization of the PDS

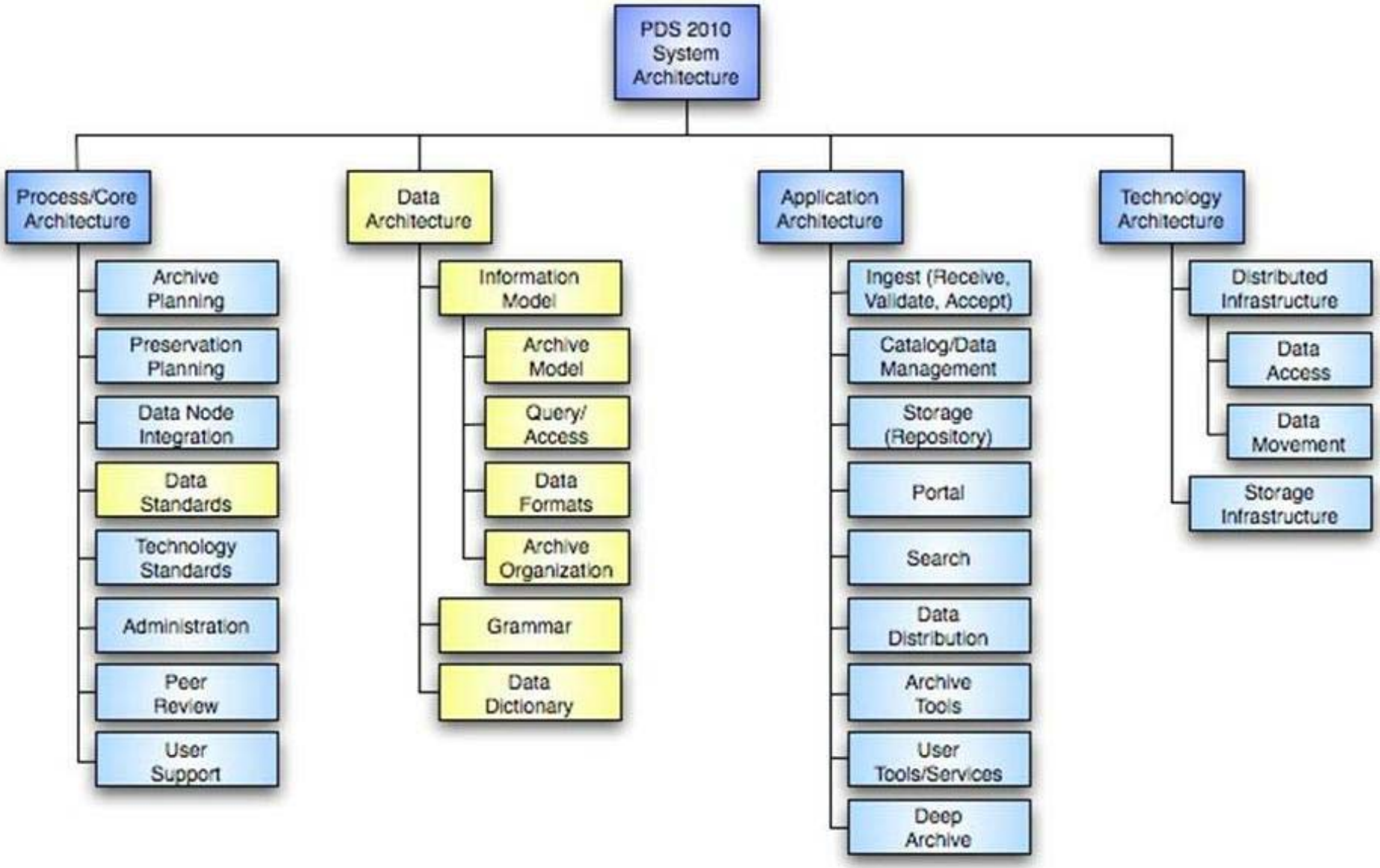
- "PDS 2010" is a plan to move PDS to a fully online, federated system coupled with an upgrade of the PDS Data Standards
  - The architectural approach allows for better leveraging of modern IT technologies
  - Major effort is in 2010 and 2011
- Addresses several drivers which require a modernization to continue to meet demand and users expectations (volume, number of missions, complexity of missions, international missions, better users support, etc)
- Improves "efficiency in the mission interface" and "usability" in the distribution to data to PDS users
- Better addresses the tension between "preservation" and "usability" of data
- Replaces aging technology, tools and processes

Modernization of the PDS (cont…)

Specific goals for PDS 2010

- **Simplified, but rigorous, archiving standards (PDS4) 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
  - Services that allow for operation on the archive, but distribution for usability
- 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

System Architecture

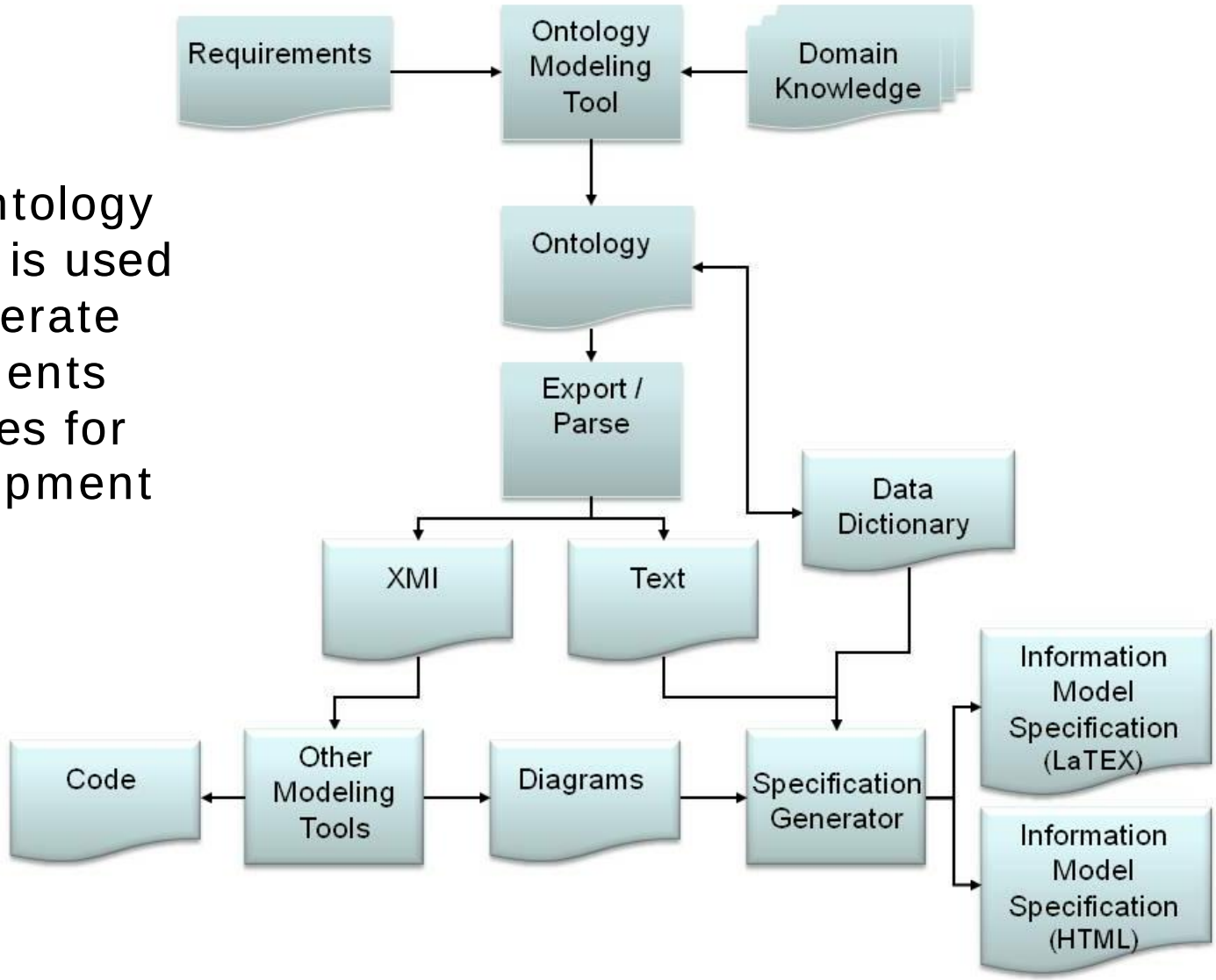


PDS4 Data Design Goals

- Define a few simple data formats
- Make the archive more stable over the long-term
- Make archive preparation more efficient for data providers
- Make the data more accessible and useful to the end-users
- Separate the data architecture from the technology and application architecture

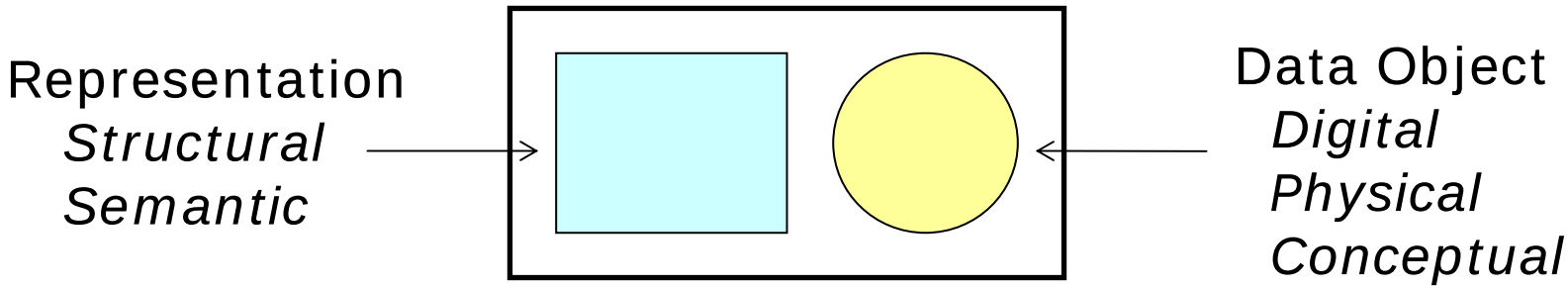
Data-Driven Development Methodology

The ontology model is used to generate documents and files for development



Overview of the PDS4 Data Standards

- The OAIS* Information Object unifies digital, conceptual and physical objects and their descriptions



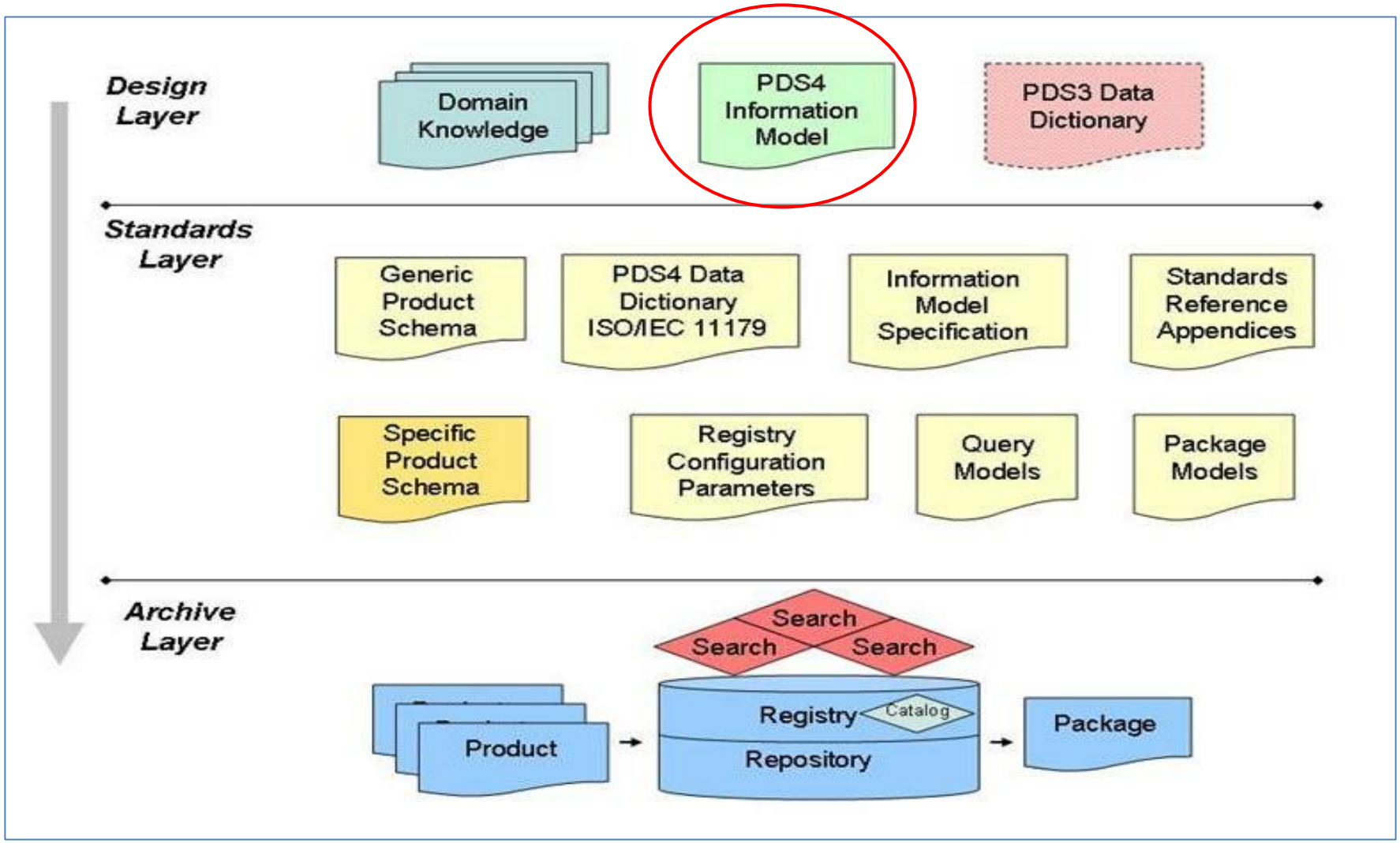
- A product is a uniquely defined package of related information objects
  - Data Product, Software, Document
- A data set is a collection of products

* Open Archive Information System

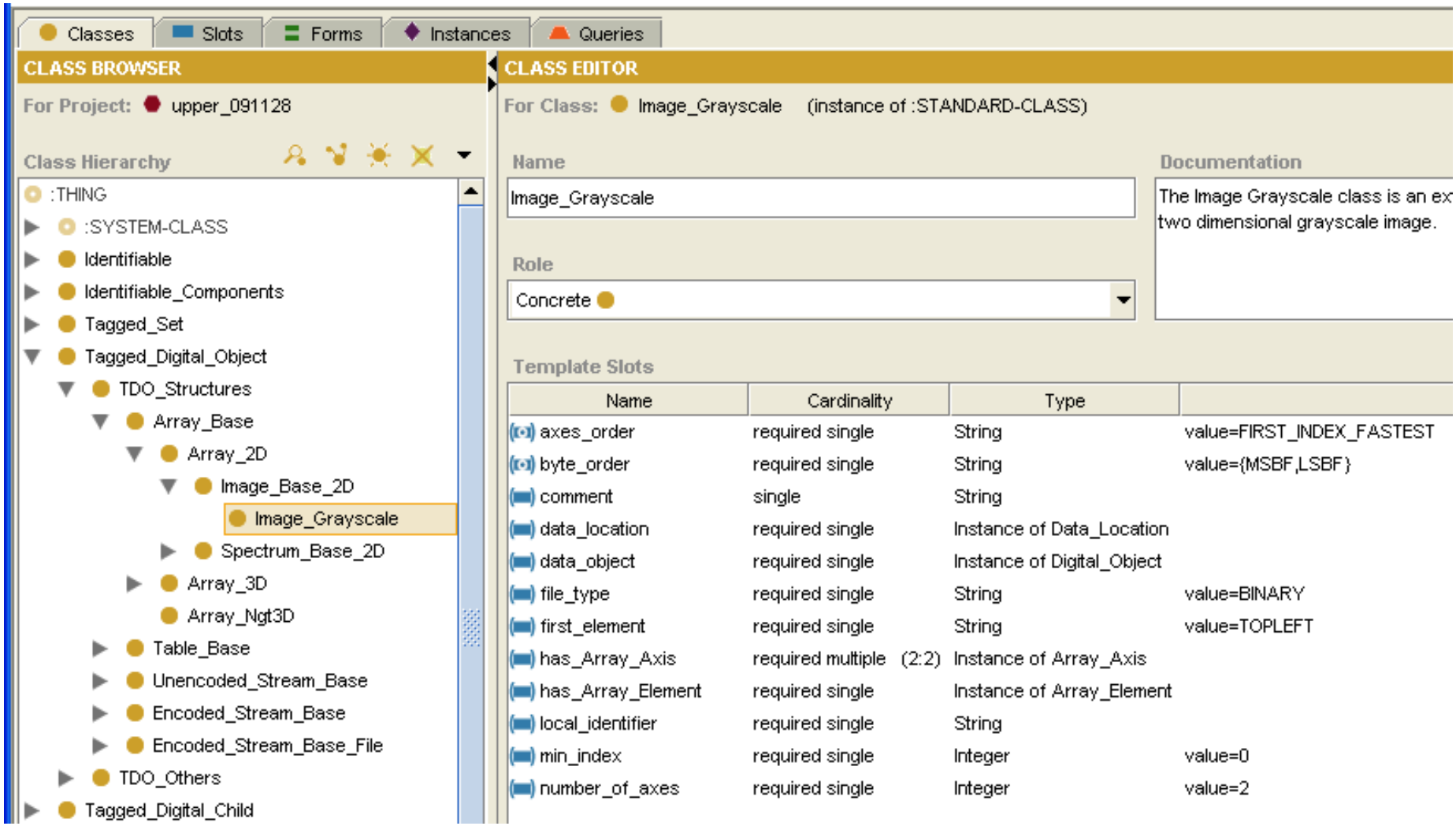
Data Structure

- Four simple data structures form the foundation for all digital objects
  - Array, Table, Parseable Byte Stream, Encoded File
- The definitions of these data structures should not change significantly over time.
- Extensions to and combinations of the data structures are used to describe data to be archived.
  - Examples: 2- and 3- dimensional spectra are defined as extensions to Array_Base
  - Table_Binary and Table_Character are defined as extensions to Table_Base
  - A Spectral cube is described using a combination of the simple data structures.

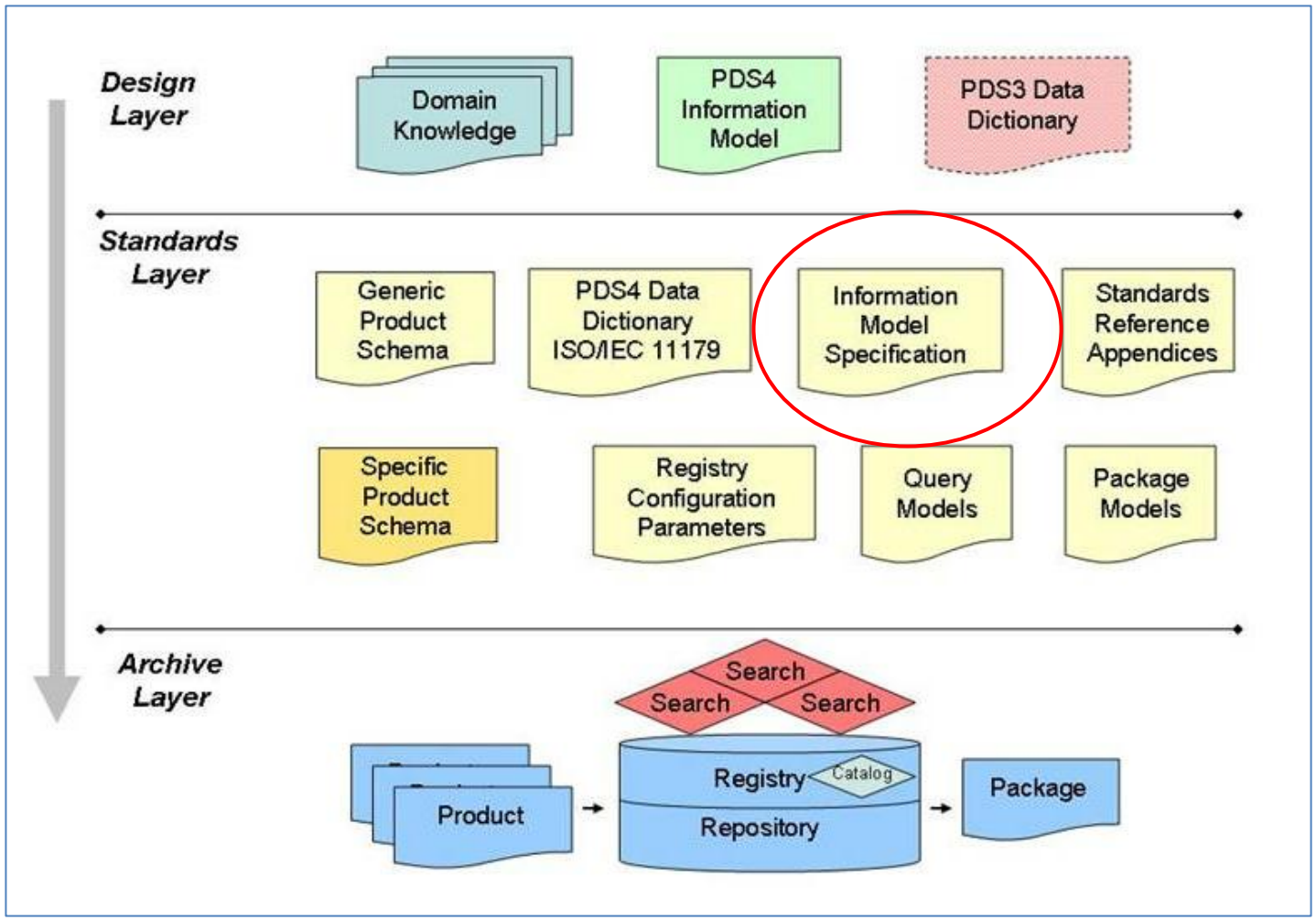
Generated Artifacts



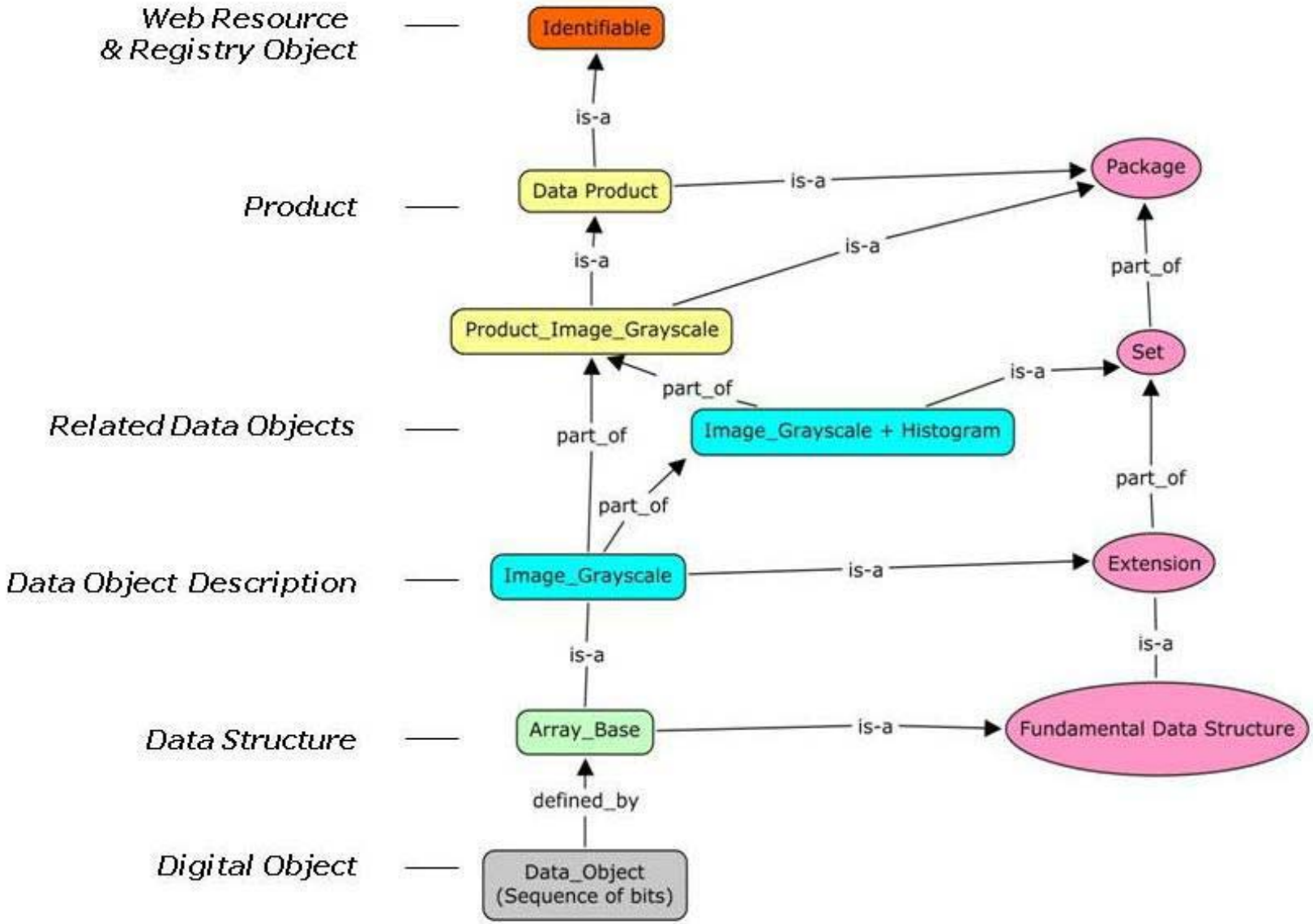
The PDS4 Information Model



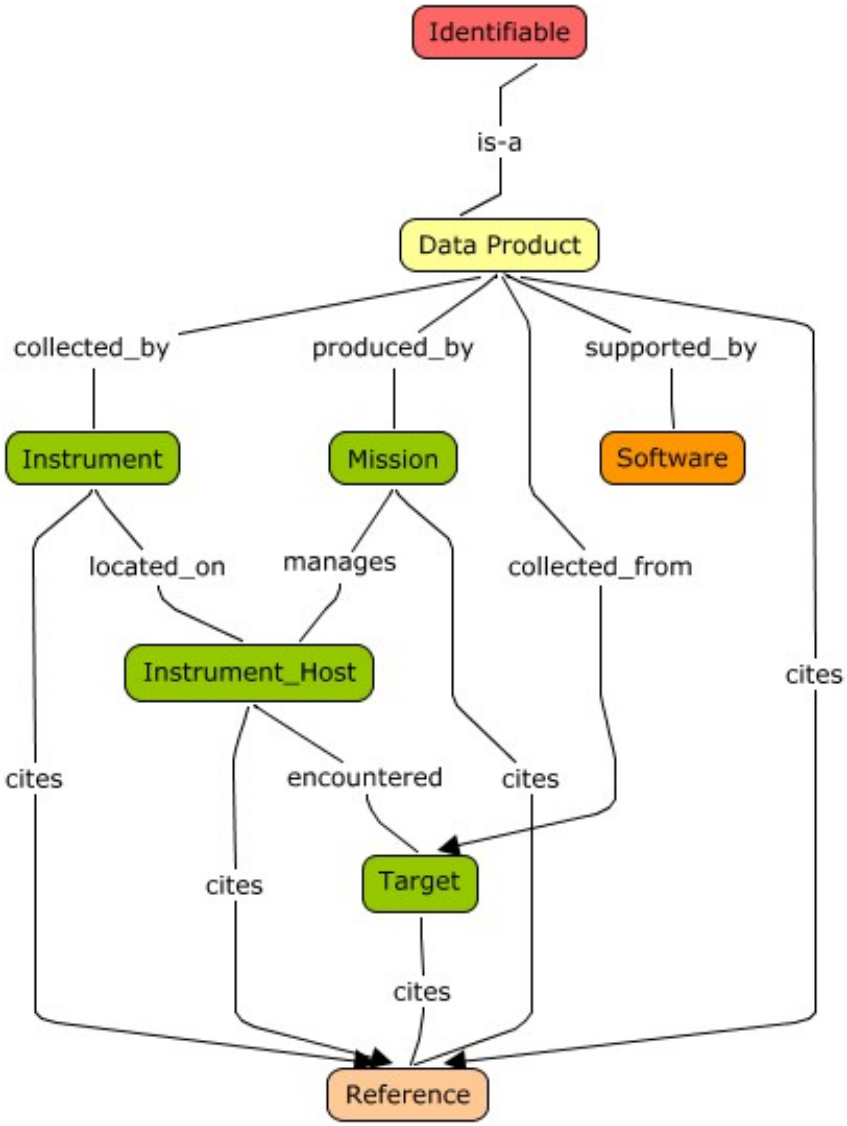
PDS4 Information Model



Basic Components of a Data Product



PDS4 High-Level Concept Map



PDS4 Information Model

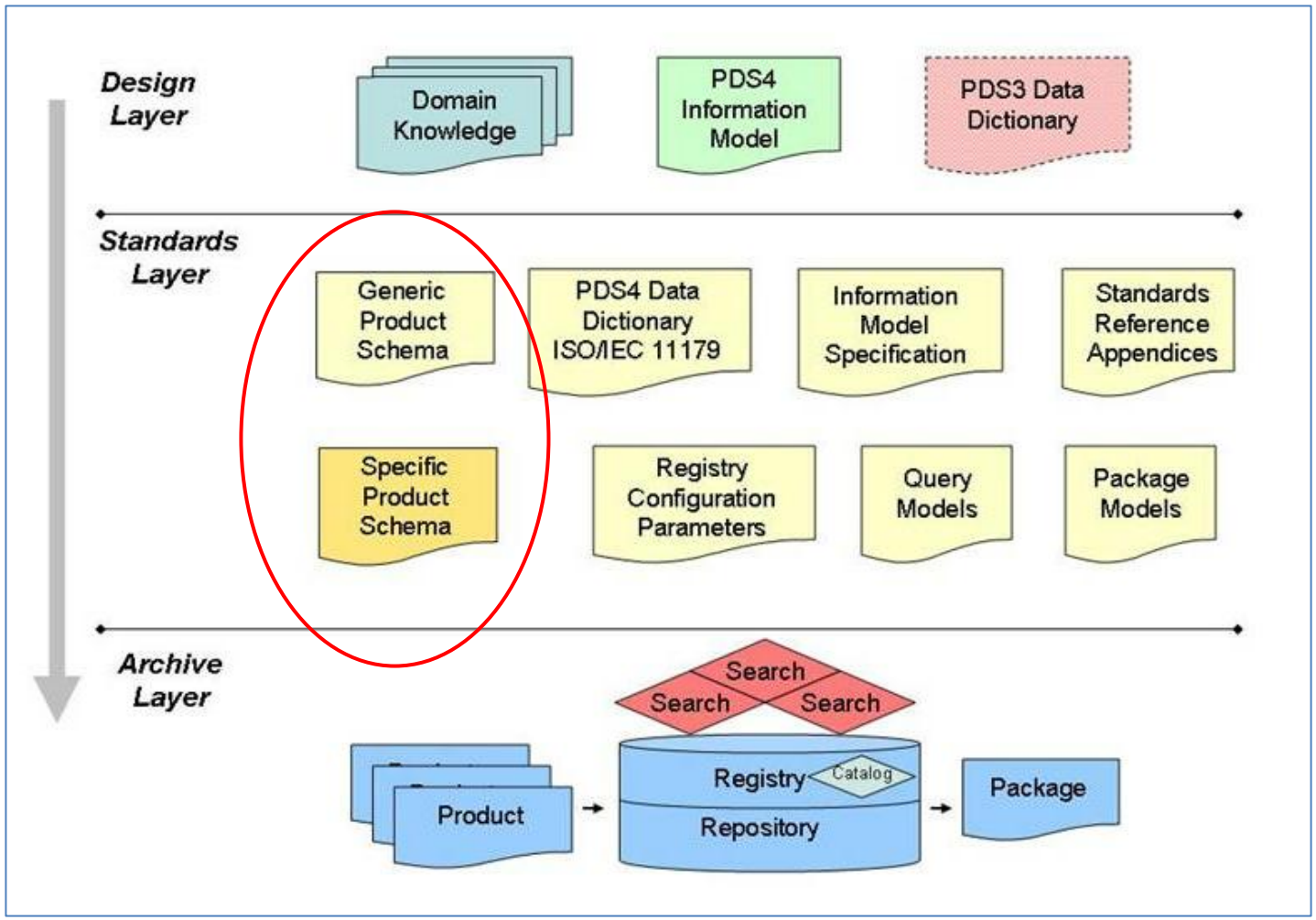
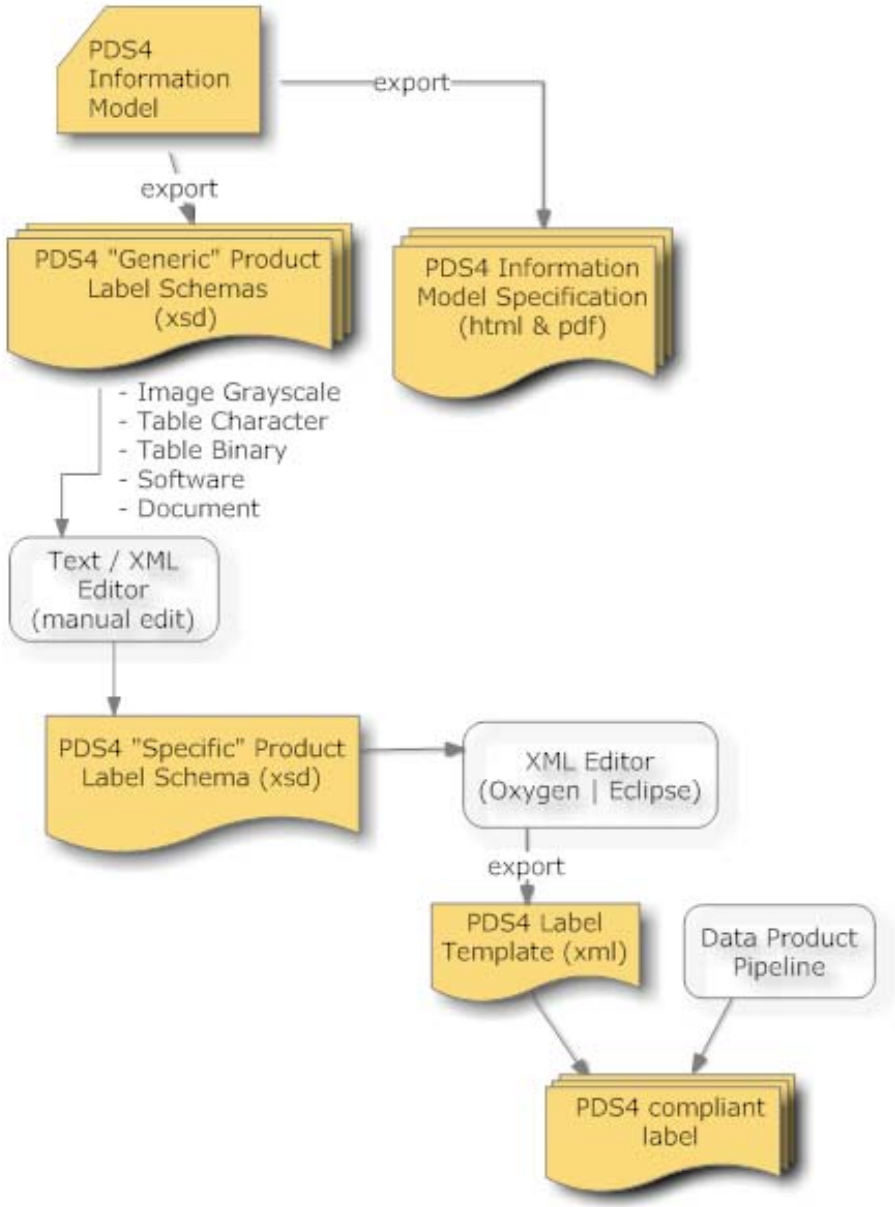
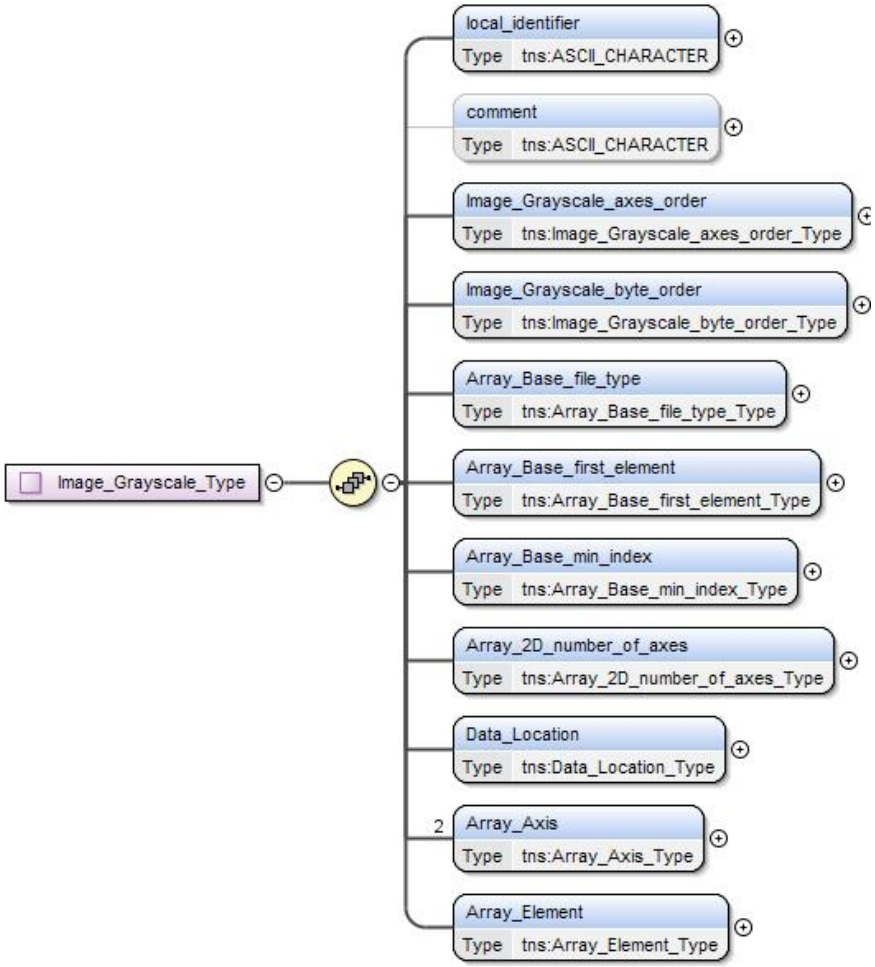


Diagram of the Lifecycle of a Product Label Schema

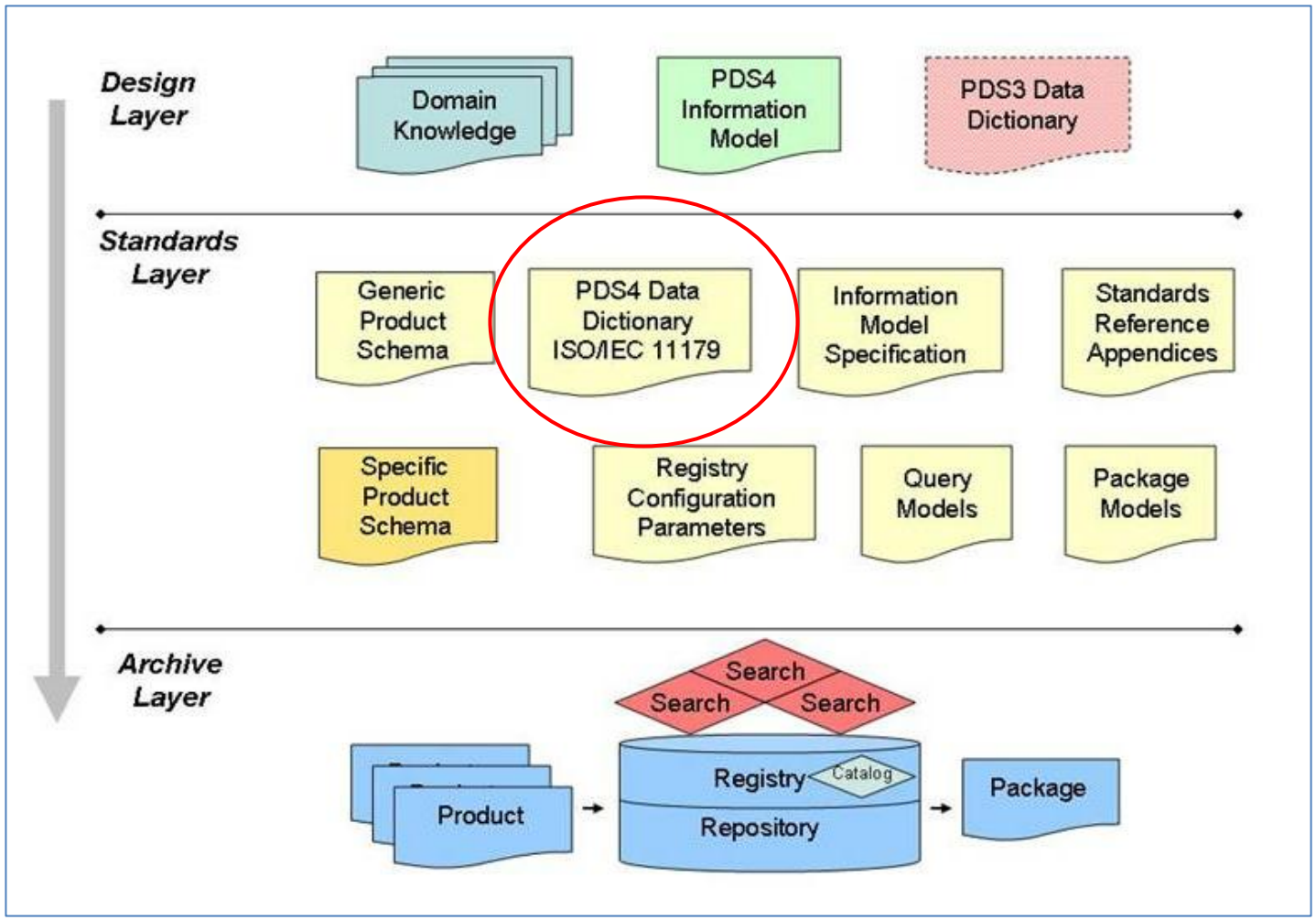


Generic Label Schema and Resulting XML Label



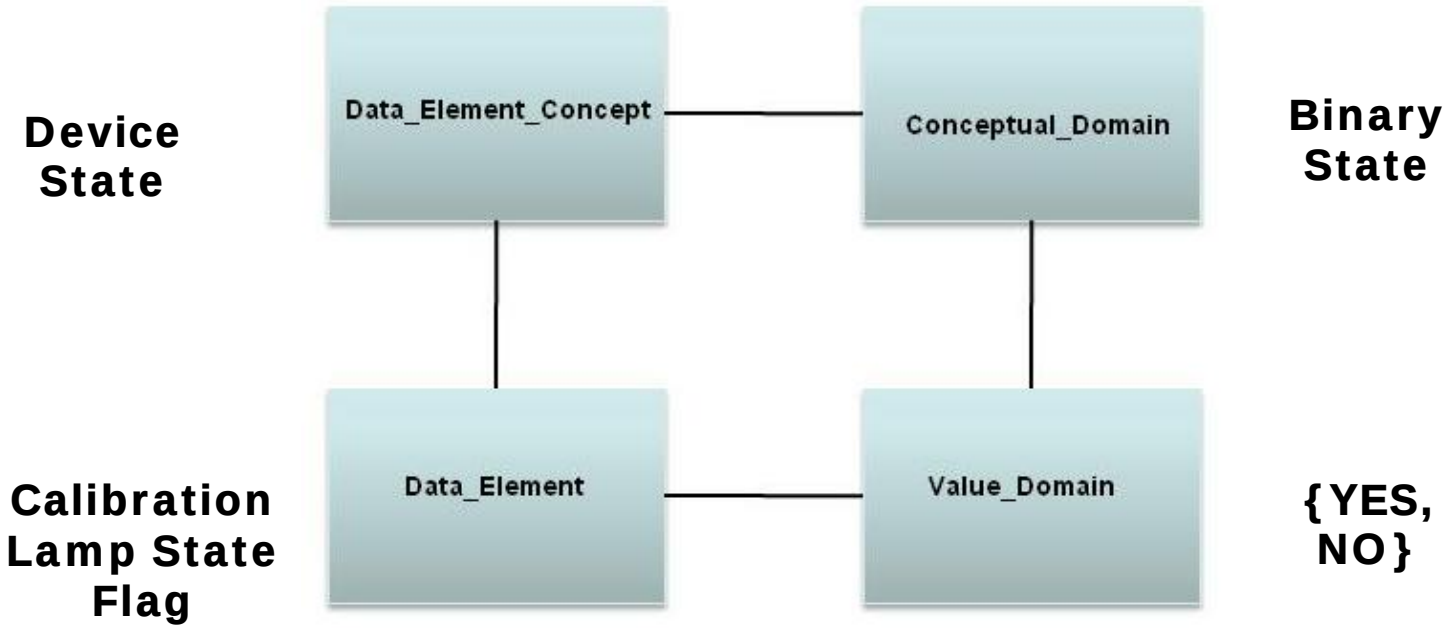

```
<Image_Grayscale>
…
 <Image_Grayscale_axes_order>FIRST_INDEX_FASTEST
 <Image_Grayscale_byte_order>MSBF
 <Array_Base_file_type>BINARY
 <Array_Base_first_element>TOPLEFT
 <Array_Base_min_index>0
 <Array_2D_number_of_axes>2
…
 <Array_Axis>
  <axis_length>248
  <axis_name>LINE
…
 </Array_Axis>
…
 <Array_Element>
  <element_bytes>2
  <element_scaling_factor>N/A
  <element_type>MSB_UNSIGNED_INTEGER
  <element_unit>DATA NUMBER
  <element_value_offset>N/A
 </Array_Element>
</Image_Grayscale>
```

PDS4 Information Model



PDS4 Data Dictionary

The data dictionary is ISO/IEC 11179:2003 compliant.



ISO/IEC 11179 High-level Metamodel

Attribute Definitions

title: **shutter_mode_id** in img:Camera_Parameters
name space id: **img:**
version: **090609n**

- data element - administration_record: Proto_DataDict_090609n
  - description: **This element identifies the state of an imaging instrument's shutter during image acquisition. Note: the instrument shutter mode affects the radiometric properties of the camera.**
  - data_type: **ASCII_CHARACTER**
- value domain - administration_record: Proto_DataDict_090609n
  - minimum_value: **N/A**
  - maximum_value: **N/A**
  - minimum_characters: **N/A**
  - maximum_characters: **20**
  - unit_of_measure_name: **none**
  - permissible value: **BODARK** - begin date: 2009-06-09 - end date: 2019-12-31
  - permissible value: **BOTSIM** - begin date: 2009-06-09 - end date: 2019-12-31
  - permissible value: **BSIMAN** - begin date: 2009-06-09 - end date: 2019-12-31
  - permissible value: **NADARK** - begin date: 2009-06-09 - end date: 2019-12-31
  - permissible value: **NAONLY** - begin date: 2009-06-09 - end date: 2019-12-31
  - permissible value: **WADARK** - begin date: 2009-06-09 - end date: 2019-12-31
  - permissible value: **WAONLY** - begin date: 2009-06-09 - end date: 2019-12-31
- conceptual domain - **Enumerated** - administration_record: Proto_DataDict_090609n
- data element concept - administration_record: Proto_DataDict_090609n

Data Type Definitions

data Type:ASCII_DATE_TIME

description: **ASCII_DATE_TIME indicates a date time in ASCII format.**
data_type_unit: **BYTE**
data_type_is_enumeration: **FALSE**
data_type_length: **null**
conceptual_domain: CD_Time
data_type_concept: DTC_Time
value_domain: VD_Date_Time

conceptual domain: CD_Time

description: **The value-space of a date-and-time datatype is the denumerably infinite set of all possible points in time with the resolution (time-unit, radix, factor). The time-literal denotes the date-and-time value specified by the characterstring as interpreted under ISO 8601. - ISO/IEC 11404**

data_type_concept: DTC_Time

description: **Time is a family of datatypes whose values are points in time to various common resolutions: year, month, day, hour, minute, second, and fractions thereof.**

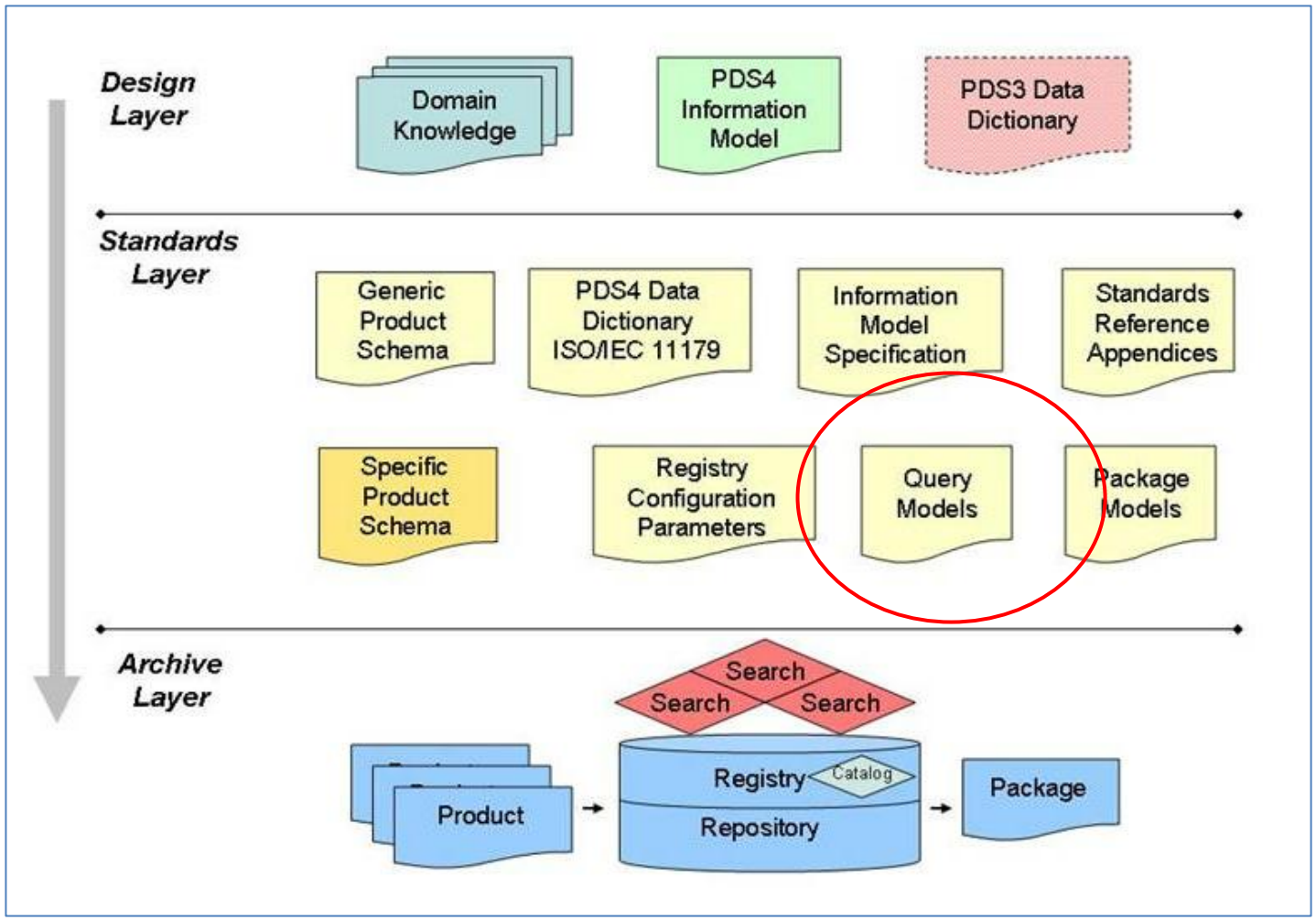
value domain: VD_Date_Time

description: **ISO 8601 - yyyy-mm-ddThh:mm:ss.sss**

Data Dictionary
Class Definition

- **class: Camera_Parameters - Occurs 0 to 1 Times**
  - attribute: **local_identifier** value: ***value***
  - attribute: **comment** value: ***value*** Optional
  - attribute: **edit_mode_id** value: ***value*** Optional
  - attribute: **exposure_duration** value: ***value*** Optional
  - attribute: **filter_id** value: **0, 1, 2, 3, 4, 5, 6, 7, 8, A, B, C1, C2, C3, D, HFM1, LFM1** Optional
  - attribute: **filter_name** value: **A, B, BLUE, BLUE-GREEN, C, CLEAR, D, E, F, GREEN, IR-7270, IR-7560, IR-8890, IR-9680, L1000_R480, L440_R440, L450_R670, L670_R670, L800_R750, L860_R-DIOPTER, L885_R947, L900_R600, L925_R935, L930_R530, L935_R990, L965_R965, LONGWAVE, METHANE-JST, METHANE-U, MINUS BLUE, MI_CLOSED, MI_OPEN, NEAR-INFRARED, NONE, ORANGE, PANCAM_L2_753NM, PANCAM_L8_440NM, PANCAM_LV_602NM, PANCAM_R8_880NM, RED, SHORTWAVE, SODIUM-D, SOLAR UV-22, T11, T15, T20, T7, T9, ULTRAVIOLET, VIOLET** Optional
  - attribute: **gain_mode_id** value: **100K, 10K, 400K, 40K, HIGH, LOW** Optional
  - attribute: **scan_mode_id** value: **.055, 4.0, epf, long, short** Optional
  - attribute: **shutter_mode_id** value: **BODARK, BOTSIM, BSIMAN, NADARK, NAONLY, WADARK, WAONLY** Optional
- **End_Class**

PDS4 Information Model



Data Set Query Model – RDF

RDFS/XML

```
        ...
<rdfs:Class rdf:about="&rdf_;Data_set"
       rdfs:label="Data_set">
      <rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
</rdfs:Class>
<rdf:Property rdf:about="&rdf_;archive_status"
       rdfs:label="archive_status">
      <rdfs:domain rdf:resource="&rdf_;Data_set"/>
      <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
<rdf:Property rdf:about="&rdf_;data_set_id"
       rdfs:label="data_set_id">
      <rdfs:domain rdf:resource="&rdf_;Data_set"/>
      <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
<rdf:Property rdf:about="&rdf_;data_set_name"
       rdfs:label="data_set_name">
      <rdfs:domain rdf:resource="&rdf_;Data_set"/>
      <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
<rdf:Property rdf:about="&rdf_;data_set_release_date"
       rdfs:label="data_set_release_date">
      <rdfs:domain rdf:resource="&rdf_;Data_set"/>
      <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
```

RDF/XML

```
rdf_:stop_time="1987-05-23"
dc:title="ARECIBO MOON RADIO TELESCOPE CALIBRATED ...
    rdf_:reslocation="http://pdsquery.jpl.nasa.gov/query?
       Identifier=ARCB-L-RTLS-3-70CM-V1.0&amp...
rdf_:resclass="data.metadata.dataset"
dc:publisher="NASA.PDS"
rdf_:data_set_release_date="1990-06-15"
rdf_:data_set_name="ARECIBO MOON RADIO TELESCOPE ...
rdf_:rescontext="NASA.PDS"
rdf_:data_set_id="ARCB-L-RTLS-3-70CM-V1.0"
dc:language="en"
rdf_:data_set_terse_desc="ARECIBO MOON RADIO ...
dc:description="ARECIBO MOON RADIO TELESCOPE ...
dc:identifier="ARCB-L-RTLS-3-70CM-V1.0"
rdf_:start_time="1986-07-03"
dc:format="XML/RDF">
 <rdf_:target_name>
   <rdf:Description rdf:about="&terms;moon">
     <rdfs:label>MOON</rdfs:label>
   </rdf:Description>
 </rdf_:target_name>
     ...
</rdf_:Data_set>
```

Industry Standards Referenced and Controlling

- ISO/IEC 11179:3 Registry Metamodel and Basic Attributes specification - Adopted for the data dictionary schema.
- ISO/IEC 11404:2007(E) - Provides the specification for language-independent data types.
- Open Archival Information System (OAIS) Reference Model - Provides a standard for the unification of digital, conceptual, and physical data objects
- XML (Extensible Markup Language) - Rules for encoding documents electronically.
- XML schema - Type description language for XML documents.
- Electronic Business XML (ebXML) federated registry/repository information model – Provides a standard to support federated registry/repository functions.

International Collaboration

- PDS Standards are currently used as the de facto standard for archiving planetary science data
  - ESA has adopted PDS Standards
  - JAXA, ISRO, CNSA and others are working to adopt them

- In 2006, ESA and NASA proposed and started the *International Planetary Data Alliance* to improve efficiency in coordinating data archiving for international missions and improving access to international archives

- In 2008, COSPAR passed a resolution recognizing IPDA and supporting its efforts to establish standards for archiving and sharing planetary science data

- Representatives include: ESA, NASA, ISRO, JAXA, DLR, BNSC, CNES, ASI, CNSA, RSA/IKI with Japan as the current chair

Next Steps

- IPDA Engagement – January 2010
- Community Engagement – February 2010
- PDS 2010 System Review – March 2010
  - Data Standards Management
- Readiness Review – October 2010
  - Preliminary Products
  - First mission

Conclusion

- Starting from fundamental principles and leveraging their combined experience the PDS Data Design Working Group (DDWG) has made significant progress in the development of the next generation PDS data standards.

- The use of shared ontologies and existing industry standards will provide an archive information system that will support more information interconnectedness, correlative science, and system interoperability.

THANK YOU!

Backup Slides