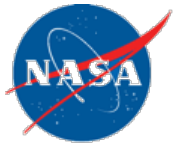


National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Overview of the PDS4 Data Standards

Steve Hughes

PDS Technical Session

September 21-23, 2016

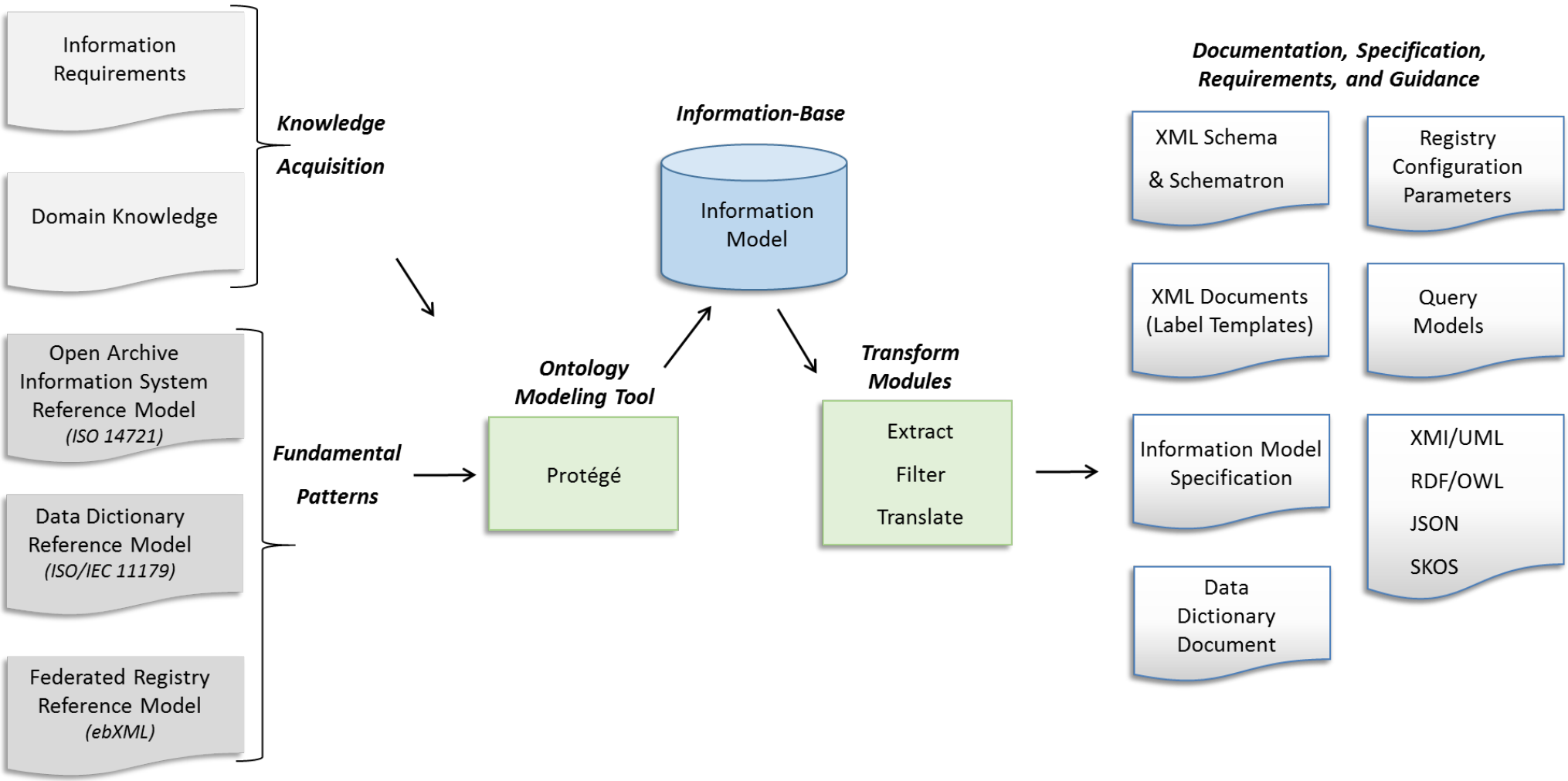
Topics

- **Overview of the PDS4 Standards**
- Source and Production of the PDS4 Standards
- PDS4 Standards and the Product Development Lifecycle

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Overview



The PDS Information Model (IM): Addresses Data Variety

- **PDS4 Information Model plays a key role in defining the PDS information requirements**
  - *Defines entities in the Planetary Science community and their relationships*
  - *Establishes an overarching governance model for PDS data*

- **The PDS4 system is enabled by an "information model-driven" approach where the information model is the corner-stone of the system**
  - *Handles the diversity of different disciplines*
  - *Enables federated governance*
  - *New instruments, observation types and data can be accommodated*
  - *Allows the system to be configured by the information model*
  - *Ensures updates to the model do not break the software*
  - *Provides metadata definitions that are tied to the model to increase consistency*

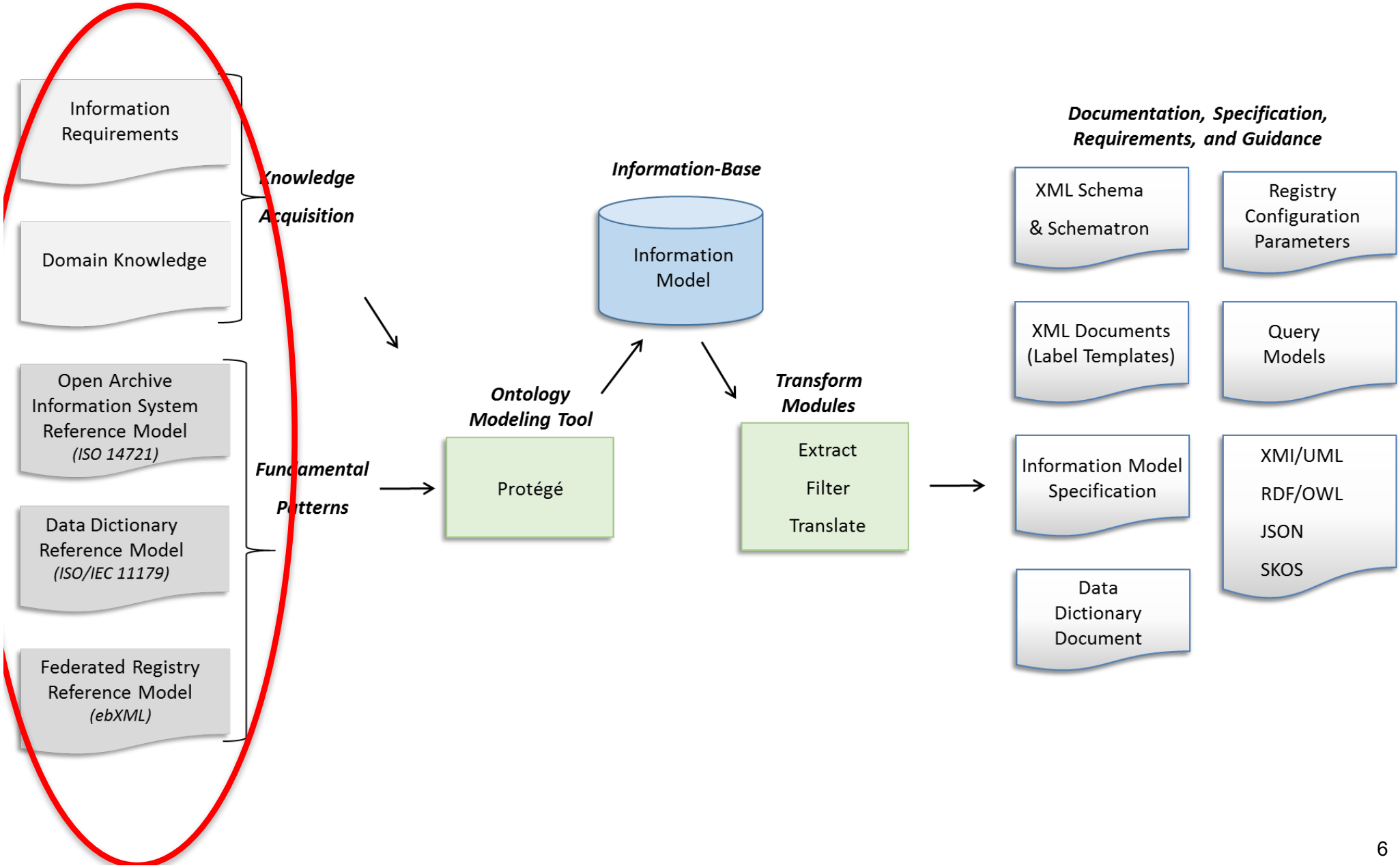
Topics

- Overview of the PDS4 Standards
- **Source and Production of the PDS4 Standards**
- PDS4 Standards and the Product Development Lifecycle

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Input to the Information Model



The Model Consists of Two Parts

- **PDS-specific**
  - *Information Requirements*
  - *Domain Knowledge*

- **Reference Models for Information Systems**
  - *Open Archival Information System Reference Model (OAIS-RM)*
  - *Data Dictionary Reference Model*
  - *Federated Reference Model*

PDS Information Requirements

- The PDS Requirements and Policies set the foundation for the *information requirements*. In particular,
  - *Level 1/2/3 requirements (e.g., 1.4)*
  - *Format policy (e.g., what structures can we support)*

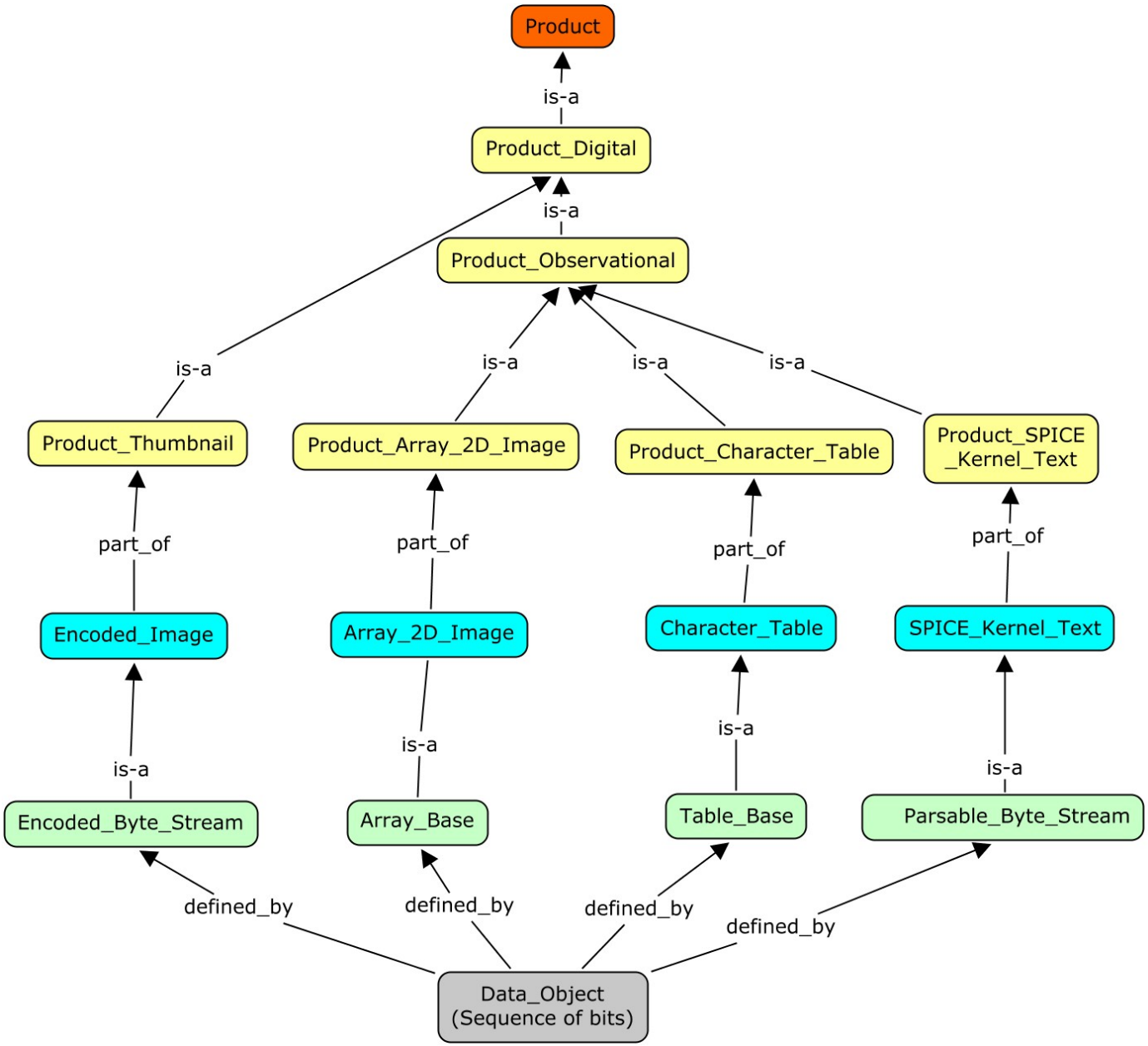
PDS-Specific Domain Knowledge

- Domain knowledge is expert information about the "things" in the domain (Planetary Sciences) that should be described and associated with the data to make and keep it useful.
  - *Fundamental Structures and their extensions*
    - Array, Table_*, etc
  - *Context Information*
    - Investigation, Observing Systems, etc
  - *Integrity (Checksums)*
  - *Reference (Relationships between "things")*
  - *Documents*

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Example Domain Knowledge (1)

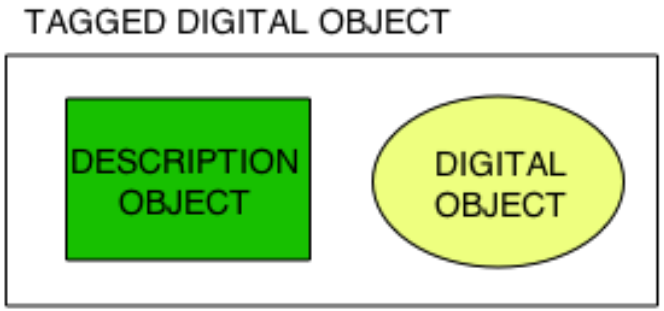


National Aeronautics and Space Administration

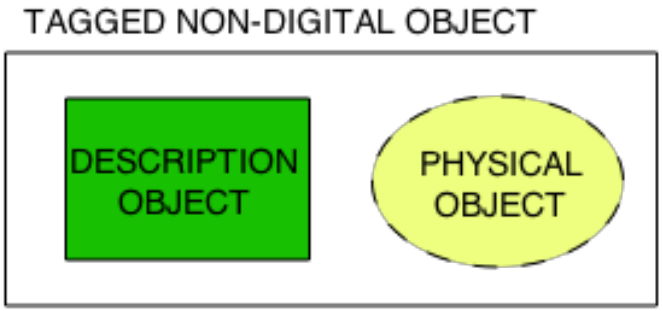
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Archive Reference Models: *Open Information Archive System (OAIS) Reference Model*

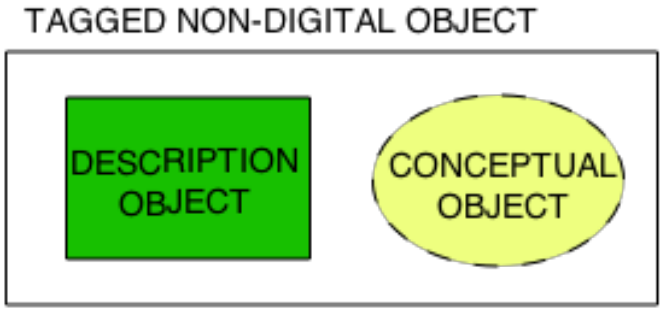
Information Object Model[1]



- **digital object: An object which is real data — for example, a binary image of a redwood tree.**



- **physical object: An object which is physical or tangible – for example the planet Saturn and the Venus Express magnetometer.**



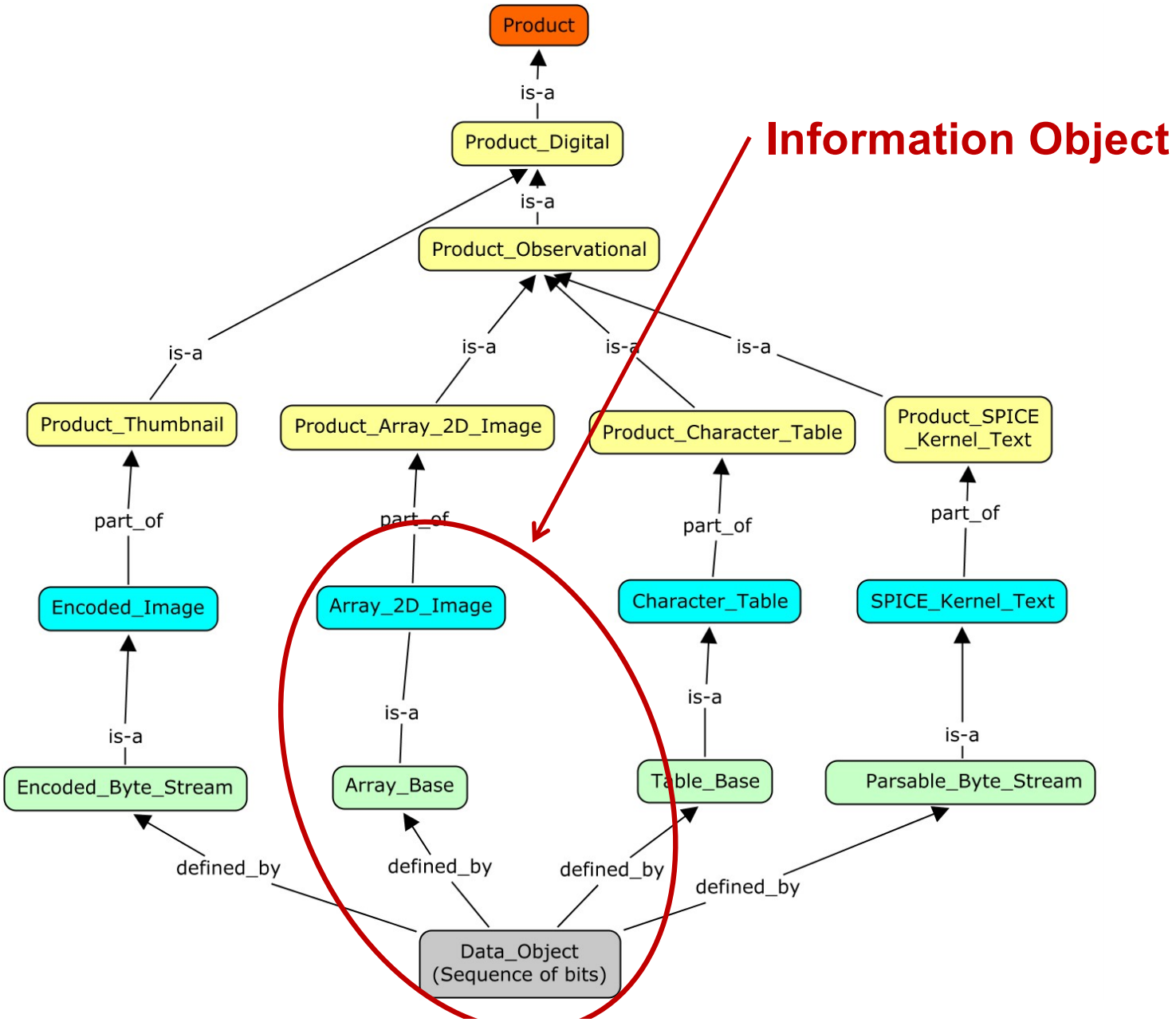
- **conceptual object: An object which is intangible – for example the Cassini mission and NASA's strategic plan for solar system exploration.**

[1]Open Archival Information System (OAIS) Reference Model - ISO 14721:2003

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Domain Knowledge and Information Objects



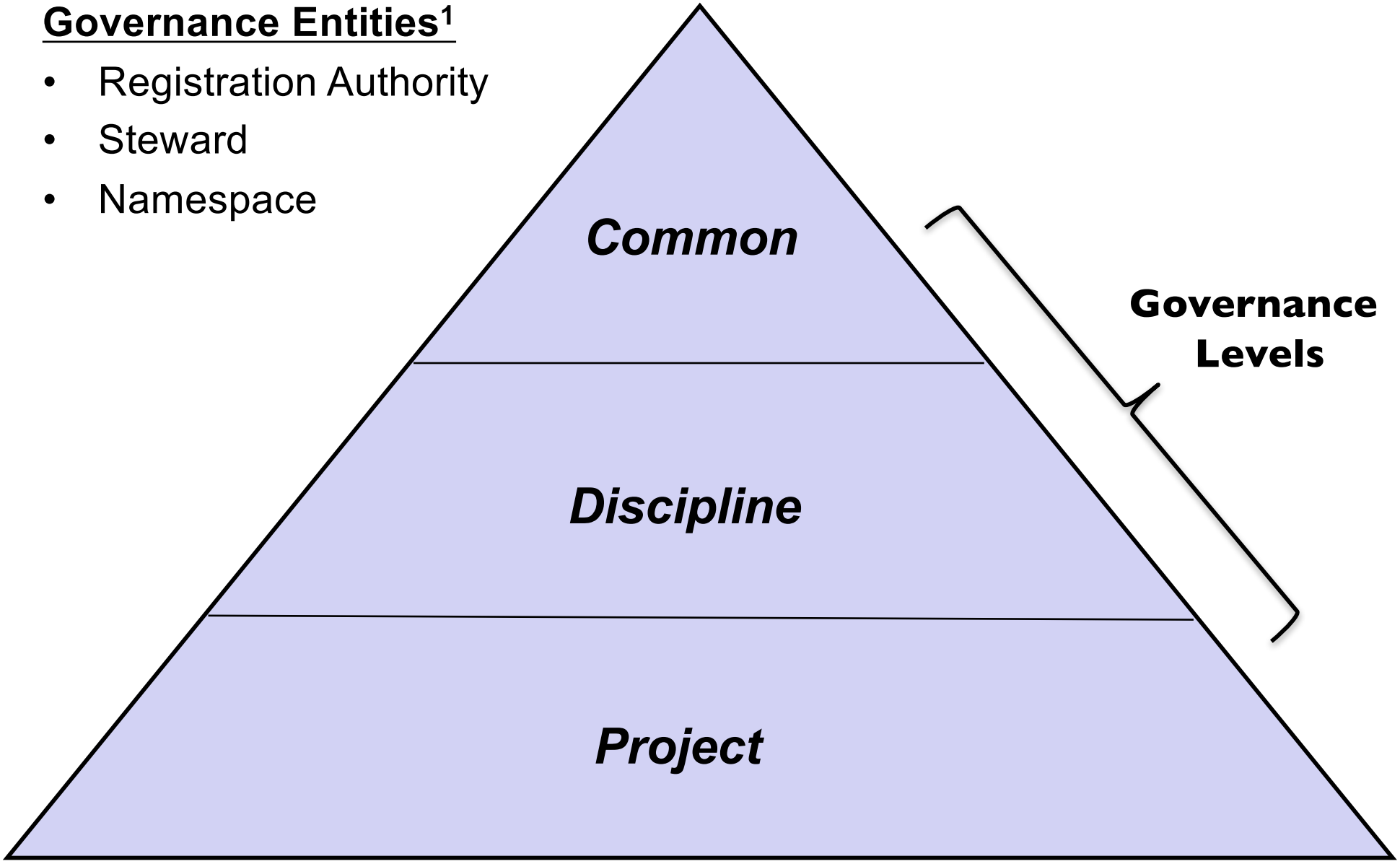
National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Archive Reference Models: *Data Dictionary Reference Model*

Governance Entities[1]

- Registration Authority
- Steward
- Namespace



[1]ISO/IEC 11179:2003 - Volume 3 – Metadata Registry Specification

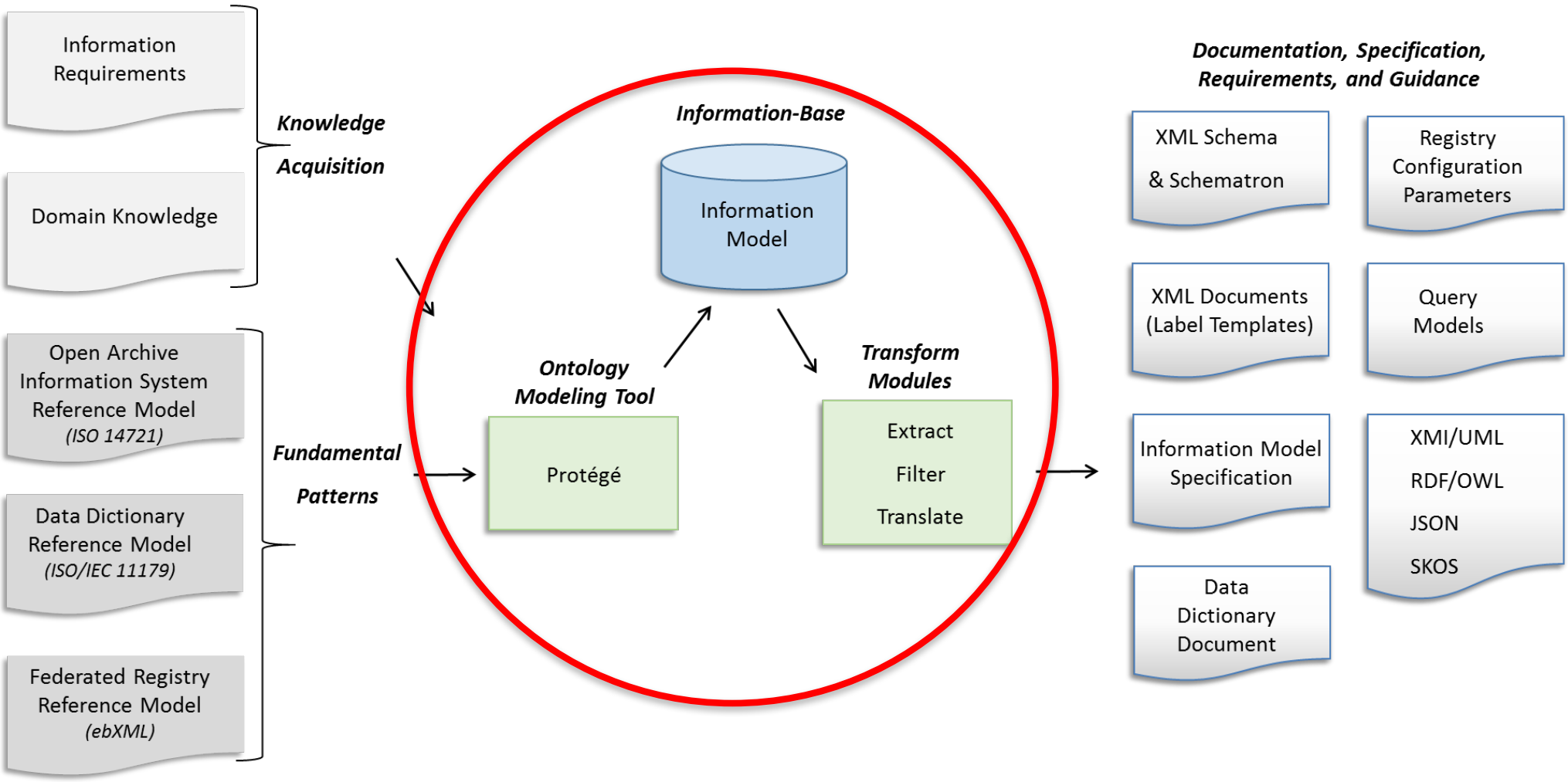
Archive Reference Models: *Registry Reference Model*

- ebXML (Electronic Business XML) - Standardizes the secure exchange of data
- Defines key properties of a federated registry:
  - *registry database schema*
  - *generic registry object*
  - *core attributes*
    - identification (e.g. <logical_identifier>)
    - versioning (e.g. <version_identifier>)
    - tracking
  - *common registry services*

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Information Model Management



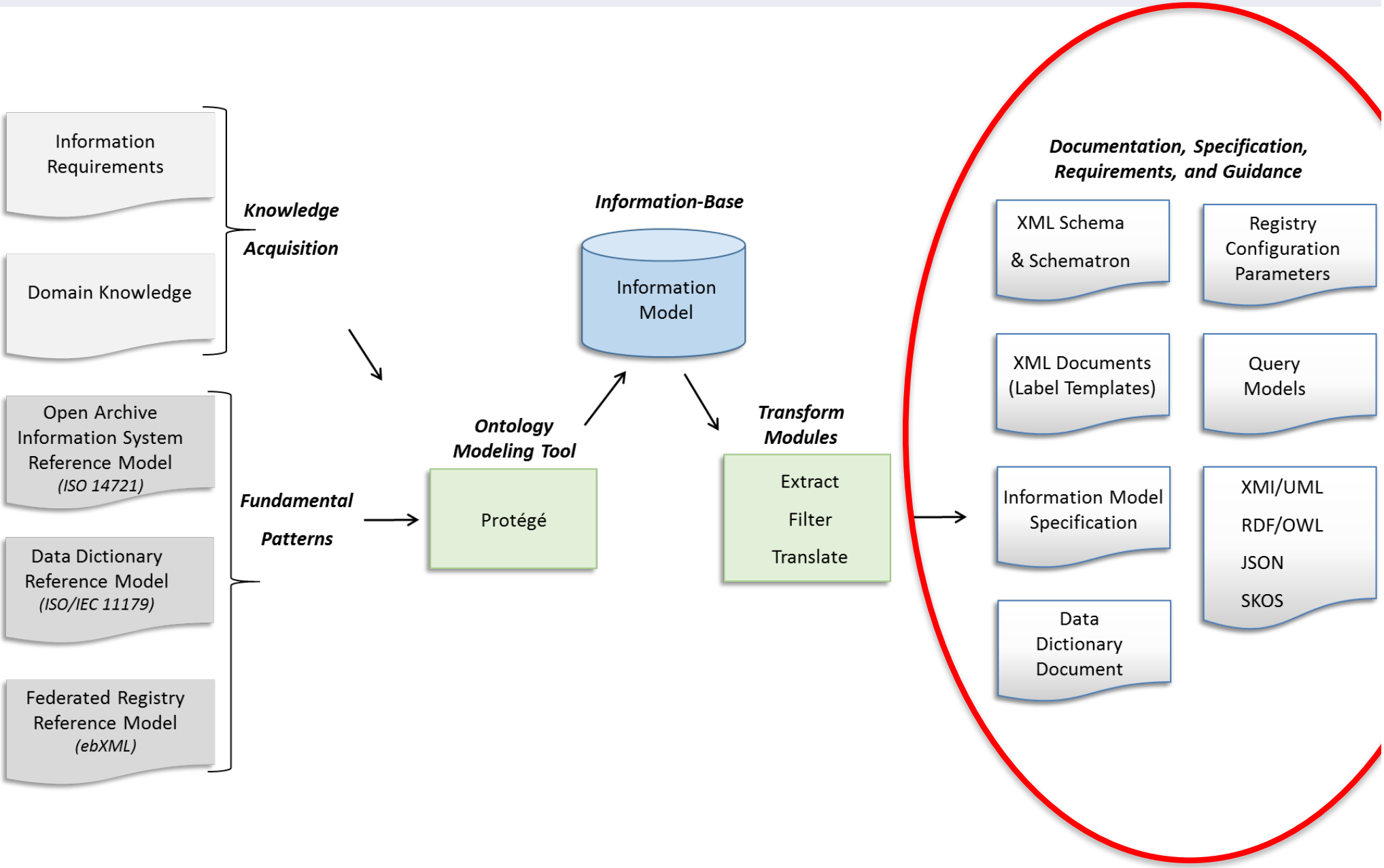
The PDS Information Model Database

- All of the source information for the PDS Information Model database is managed using the Protégé software.
  - *This database describes all objects and their relationships.*
- The database content is a merge of the Domain Model and Data Dictionary
  - *Includes the Common (pds:) and all ingested local data dictionaries (LDDs)*
- A specialized tool has been written to export the database to formatted files used by data providers, registry, harvester, search engine, validator and other tools.

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Information Model Output



Information Model Specification

9.4 Array_2D_Image

Root Class:Tagged_Digital_Object
Role:Concrete
Class Description:The Array 2D Image class is an extension of the Array 2D class and defines a two dimen
Steward:pds
Namespace Id:pds
Version Id:1.1.0.0

	Entity	Card	Value/Class	Ind
Hierarchy	Tagged_Digital_Object			
	. Byte_Stream			
	. . Array			
	. . . Array_2D			
	. . . . Array_2D_Image			
Subclass	none			
Attribute	none			
Inherited Attribute	axis_index_order	1	Last Index Fastest	
	description	0..1		
	offset	1		
	axes	1	2	R
	local_identifier	0..1		
	name	0..1		
Association	has_Display_2d_Image	0..1	Display_2D_Image	
Inherited Association	associated_Special_Constants	0..1	Special_Constants	
	associated_Statistics	0..1	Object_Statistics	
	data_object	1	Digital_Object	
	has_Element_Array	1	Element_Array	

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

XML Product Templates and Labels

```
<Product_Observational
   <Identification_Area>
      <logical_identifier>urn:nasa:pds:example.dph.sampleproducts:exampleproducts:array2d_image …
      <version_id>1.0</version_id>
      <title>MARS PATHFINDER LANDER Experiment</title>

      <Array_2D_Image>
         <local_identifier>MPFL-M-IMP_IMG_GRAYSCALE</local_identifier>
         <offset unit="byte">0</offset>
         <axes>2</axes>
         <axis_index_order>Last Index Fastest</axis_index_order>

         <Element_Array>
            <data_type>UnsignedMSB2</data_type>
            <unit>data number</unit>
            <scaling_factor>1</scaling_factor>
            <value_offset>0</value_offset>
         </Element_Array>
         <Axis_Array>
            <axis_name>Line</axis_name>
            <elements>248</elements>
            <sequence_number>1</sequence_number>
         </Axis_Array>
         <Axis_Array>
            <axis_name>Sample</axis_name>
            <elements>256</elements>
            <sequence_number>2</sequence_number>
```

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

XML Schema and Schematron Files

```
<xs:complexType name="Array">
  <xs:annotation>
    <xs:documentation>The Array class defines a homogeneous N-dimensional array of scalars. …
  </xs:annotation>
  <xs:complexContent>
    <xs:extension base="pds:Byte_Stream">
      <xs:sequence>
        <xs:element name="offset" type="pds:offset" minOccurs="1" maxOccurs="1"> </xs:element>
        <xs:element name="axes" type="pds:axes" minOccurs="1" maxOccurs="1"> </xs:element>
        <xs:element name="axis_index_order" type="pds:axis_index_order" minOccurs="1“ …
        <xs:element name="description" type="pds:description" minOccurs="0" maxOccurs="1"> …
        <xs:element name="Element_Array" type="pds:Element_Array" minOccurs="1“  …
        <xs:element name="Axis_Array" type="pds:Axis_Array" minOccurs="1“ …
          …


  <sch:pattern>
    <sch:rule context="pds:Array/pds:axis_index_order">
      <sch:assert test=". = ('Last Index Fastest')">
       The attribute pds:axis_index_order must be equal to the value 'Last Index Fastest'.</sch:assert>
```

Data Dictionary Document (pdf and html formats)

Array_2D_Image

Name: **Array_2D_Image**			*Version Id:* **1.1.0.0**
Description: The Array 2D Image class is an extension of the Array 2D class and defines a two dimensional image.			
Namespace Id: pds	*Steward:* pds	*Role:* concrete	*Status:* Active
Class Hierarchy: Tagged_Digital_Object :: Byte_Stream :: Array :: Array_2D :: Array_2D_Image			
Attribute(s)	*Name*	*Cardinality*	*Value*
	name	0..1	None
	local_identifier	0..1	None
	offset	1..1	None
	axes	1..1	2
	axis_index_order	1..1	Last Index Fastest
	description	0..1	None
Association(s)	*Name*	*Cardinality*	*Class*
	has_Element_Array	1..1	Element_Array
	has_Axis_Array	2..2	Axis_Array
	associated_Special_ Constants	0..1	Special_Constants
	associated_Statistics	0..1	Object_Statistics
	data_object	1..1	Digital_Object

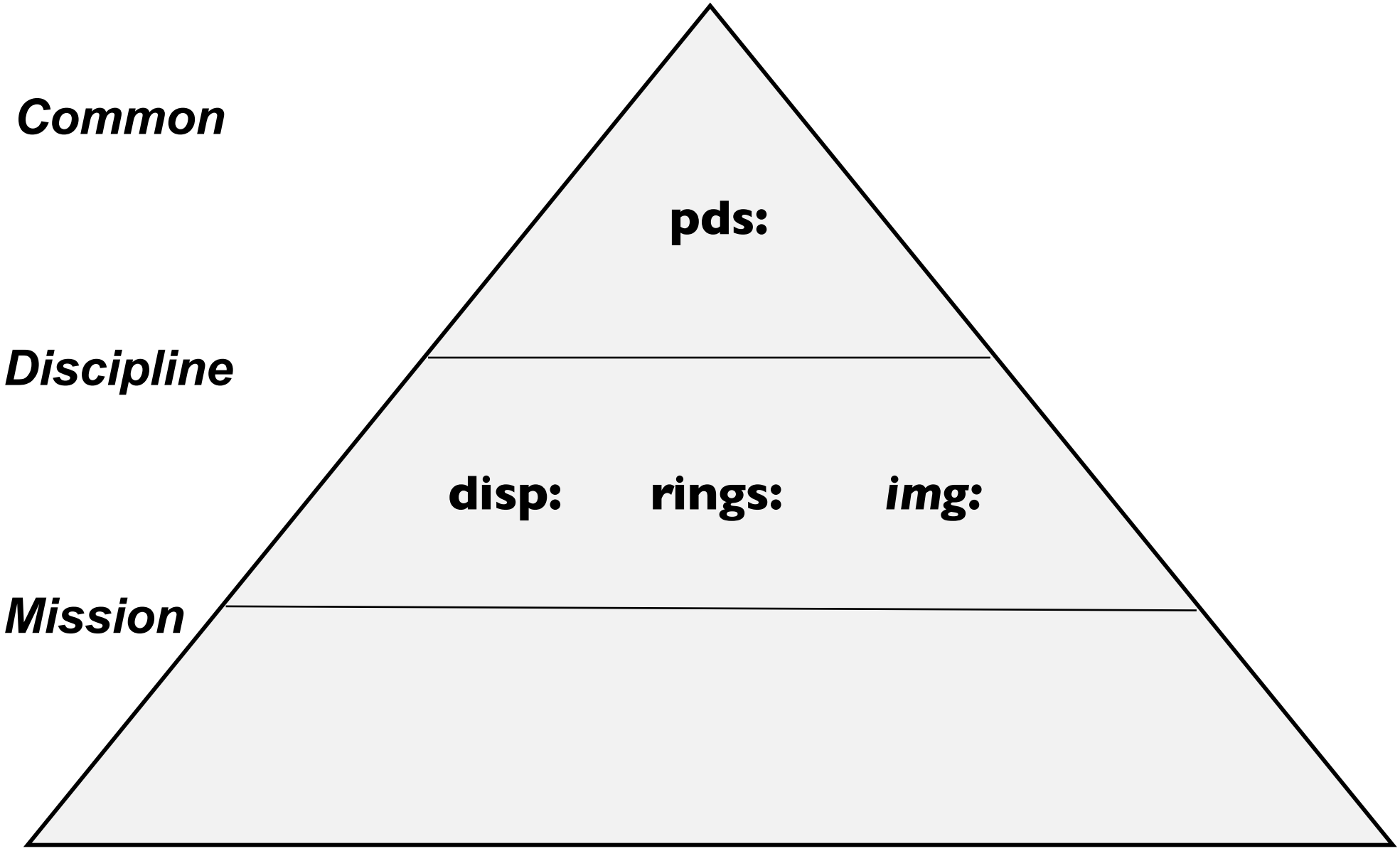
JSON File (Tool Configuration)


```
"class": {
        "identifier": "0001_NASA_PDS_1.pds.Array_2D_Image" ,
        "title": "Array_2D_Image" ,
        "registrationAuthorityId": "0001_NASA_PDS_1" ,
        "nameSpaceId": "pds" ,
        "steward": "pds" ,
        "versionId": "1.1.0.0" ,
        "description": "The Array 2D Image class is an extension of the Array 2D class and …
          , "associationList": [
              {"association": {
                "identifier": "0001_NASA_PDS_1.pds.Array.pds.offset" ,
                "title": "offset" ,
                "isAttribute": "true" ,
                "isChoice": "false" ,
                "isAny": "false" ,
                "minimumCardinality": "1" ,
                "maximumCardinality": "1" ,
                "classOrder": "1010" ,
                "attributeId": [
                  "0001_NASA_PDS_1.pds.Array.pds.offset"
                 ]
              }
        }
```

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Dictionaries in JSON File V1.6.0.0



Registry Configuration Parameters

reg_object_type:
Product_Observational
metadata: {
slot1: start_date_time
slot2: stop_date_time
slot3: version_id
slot4: title
slot5: product_class
slot6: logical_identifier
slot7: alternate_title
slot8: alternate_id
slot9: version_id
}

Query Model

9C.2 Constraints on Collections

- Mission Science Data Collection - Constraints

```
<Property_Map>
   <identifier>0001_NASA_PDS_1.pds.Property_Map.pds.QueryModel.MissionScienceDataCollection.pds….
   <description>This Property Map indicates that the attribute Collection.pds.collection_type …
   <Property_Map_Entry>
      <property_name>field</property_name>
      <property_value>collectionType</property_value>
   </Property_Map_Entry>
   <Property_Map_Entry>
      <property_name>facet</property_name>
      <property_value>Collection Type</property_value>
   </Property_Map_Entry>
   <Property_Map_Entry>
      <property_name>constraint</property_name>
      <property_value>Data</property_value>
```

```
 <sch:pattern>
…
   <sch:let name="collType" value="every $ref in ($collTypeRef) satisfies $ref = ('Data')"/>
   <sch:let name="missionType" value="some $ref in ($collMissionRef) satisfies $ref = ('Mission')"/>
   <sch:let name="purposeType" value="every $ref in ($collPurposeRef) satisfies $ref = ('Science')"/>
   <sch:let name="instrumentType" value="some $ref in ($collObsSysCompTypeRef) satisfies $ref =
                                                                                    ('Instrument')"/>
```

Topics

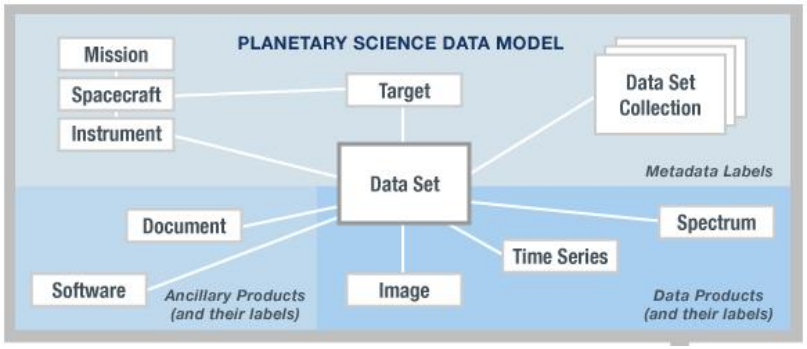
- Overview of the PDS4 Standards
- Source and Production of the PDS4 Standards
- **PDS4 Standards and the Product Development Lifecycle**

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

From Information Model to Product Label - Overview

Information Model



Used to Create

Validates

XML Schema

Expressed As

Extracted/Specialized

Data Element

has

Class

Planetary Science Data Dictionary

Product

Tagged Data Object

(Information Object)

```
<local_identifier>MPFL_M_IMP_IMAGE</local_identifier>
  <offset unit="byte">0</offset>
  <axes>2</axes>
  <axis_index_order>Last_Index_Fastest</axis_index_order>
  <encoding_type>Binary</encoding_type>
  <Element_Array>
    <data_type>SignedMSB4</data_type>
    <unit>pixel</unit>
  </Element_Array>
  <Axis_Array>
    <axis_name>Line</axis_name>
    <elements>248</elements>
    <sequence_number>1</sequence_number>
  </Axis_Array>
  <Axis_Array>
    <axis_name>Sample</axis_name>
    <elements>256</elements>
    <sequence_number>2</sequence_number>
  </Axis_Array>
</Array_2D_Image>
```

Describes

Data Object

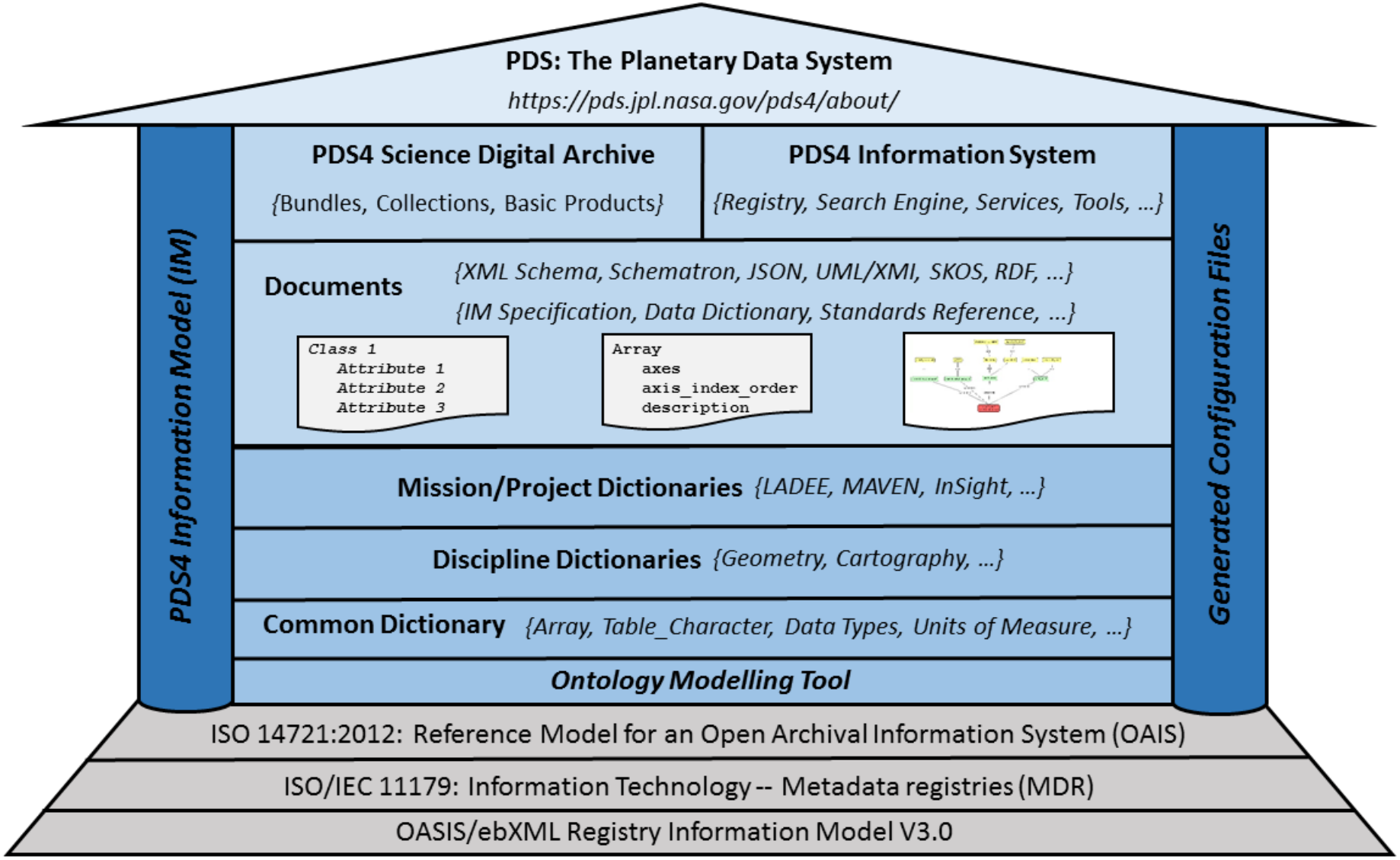
PDS4 Standards and Product Development

	Design	Generate	Validation	Ingest	Distribution	Search	Analysis / Usability
Information Model Specification	X		X				
Data Dictionary Document	X		X				
Standards Reference	X		X				
Data Provider's Handbook	X	X	X	X	X	X	?
Examples	X		X				
XML Schema/Schematron	X	X	X				X
JSON File	X	X	X				X
Query Model	X			X		X	
LDDTool/Ingest_LDD	X		X				
Important Resource	X						
Useful Resource	X						

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The PDS4 Edifice



Thank You!

National Aeronautics and Space Administration

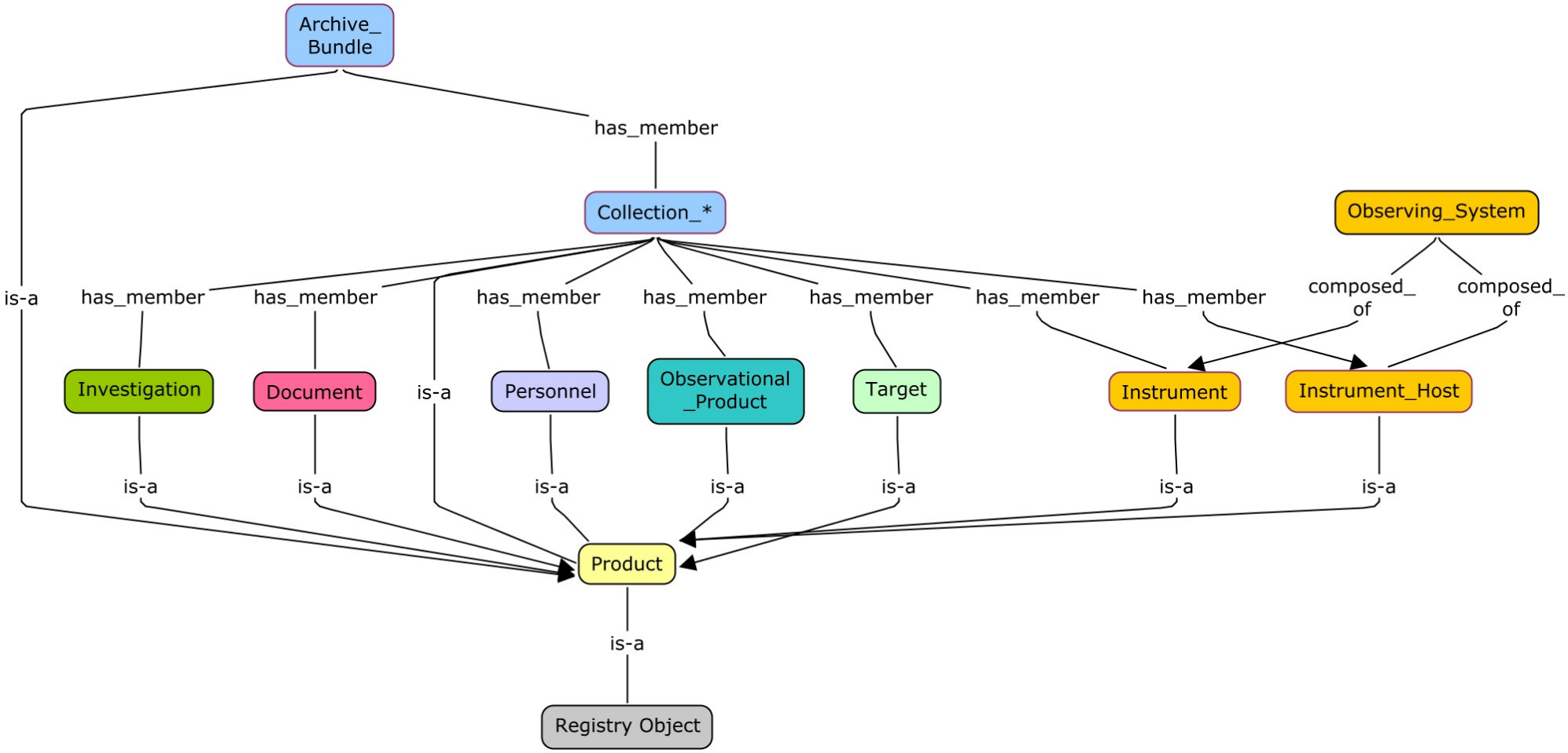
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Backup

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Domain Knowledge



National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Data Dictionary Reference Model

Data Element

- *Name*
- *Submitter, Steward*
- *Definition*
- *Namespace*
- *Source of definition*
- *Change log*
- *Version*
- *Concept*
- *Alternate Names*
- *Definition in multiple natural languages*
- *Classification*
- *Unit of measurement*
- *Effective Dates*

Value Domain

- Permissible Value
- Value Meaning
- Submitter, Steward
- Definition
- Cardinality
- Source of definition
- Change log
- Version
- Concept
- Character Set
- Representation
- Minimum and Maximum Value
- Minimum and Maximum Length
- Alternate encodings
- Effective Dates

PDS Requirements: Standards

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.

1.4 Archiving Standards: PDS will have archiving standards for planetary science data

1.4.1 PDS will define a standard for organizing, formatting, and documenting planetary science data

1.4.2 PDS will maintain a dictionary of terms, values, and relationships for standardized description of planetary science data

1.4.3 PDS will define a standard grammar for describing planetary science data

1.4.4 PDS will establish minimum content requirements for a data set (primary and ancillary data)

1.4.5 PDS will, for each mission or other major data provider, produce a list of the minimum components required for archival data

1.4.6 PDS will develop, publish and implement a process for managing changes to the archive standards

1.4.7 PDS will keep abreast of new developments in archiving standards

PDS Policies: Formats

Required Formats:

1. All data must be provided in one or more of the following formats:

i. Fixed-width binary or ASCII tables that are composed of identically structured records;

ii. N-dimensional homogeneous arrays of binary elements (N<=16);

iii. Variable-width ASCII tables composed of identically defined records, where the individual, variable-width fields in the record are delimited and the variable length records are also delimited;

iv. NAIF/SPICE kernel files.

2. All documentation needed to understand the data must be formatted as flat UTF-8 text, PDF/A-1a (which is preferred), or PDF/A-1b. Figures may be embedded in PDF files. However any figures other than those embedded in a PDF file must be formatted as JPEG, GIF, PNG or TIFF images.*

**"Flat UTF-8" means the file may not contain mark-up language (e.g., formats such as HTML, XML, RTF, and LaTeX do not comply).*

Information Model Roles

- It provides a sharable, stable, and organized structure of **information requirements …**
  - … need to **turn** information requirements (the what) into functional specifications (the how) of an information system.
  - Typically this is done by a systems engineer.
  - Under PDS4, this is done primarily by the information model
    - The contents are extracted and written to various types of formatted files.
    - The files are input to the registry, harvester, search engine, and tools and services for product design, production, validation, and analysis.

Protege

- **An ontology modeling tool**
  - *Use to define "things" and their relationships*
- **PDS4 is captured in two key protégé databases**
  - *Domain Model*
    - defines fundamental structures and context classes
  - *Data Dictionary*
    - Attributes* of an attribute
      - *Identifier and Name* - <offset>
      - *Definition* - The offset attribute provides the …
      - *Data Type* - ASCII_Integer
      - *Extrema* - Minimum Value: 0
      - *Permissible Values* - nill

Information Model Tool - IMTool

- *Workflow*
  - *Input the various protégé databases and configuration files*
  - *Merge the input into an in-memory database*
  - *Output the contents of the database to various file formats.*

Information Categories

- **Identification**
  - *Identification information provides a unique and immutable identifier for any information object that is to be discovered and accessed.*
- **Representation/Format**
  - *Representation information allows a data object to be interpreted. This includes describing the data format.*
- **Integrity**
  - *Integrity information ensures the information object has not been unintentionally altered.*
- **Provenance**
  - *Provence Information provides the history of the data and is essential for authenticity. It must include the provider.*
- **Context**
  - *Context information provides additional information that describes the environment in which the data object was created. For example, context information may describe instruments or light sources.*
- **Reference**
  - *Reference information allows the information objects to be referenced. Identification information is a subset of Reference Information.*
- **Access Rights**
  - *Access Rights information identifies the access restrictions pertaining to the data, including the legal framework, licensing terms, and access control; provider provided access and distribution conditions, and specifications for rights enforcement measures.*
- **Quality**
  - *Quality information provides a scheme for assessing and assigning a quality measure for the data object.*