Planetary Data System

The PDS4 Information Model
An Implementation-Agnostic Model for Interoperability

J. Steven Hughes
Anne C. Raugh

DM II - Thursday, May 11
11:00-12:30, Plenary Room

2023 IVOA Interop Meeting

Monday 8 - Friday 12, May 2023
Congress Center, CNR Research
Area, Bologna, Italy

NASA **Jet Propulsion Laboratory**
California Institute of Technology

Topics

- ISO Standards for Digital Repositories

- Ontologies for Semantic Interoperabilty

- Information Model
    - *Model Governance*
    - *Architectural Principles for Interoperability*
        - between Science Disciplines
        - across Tools, APIs, etc
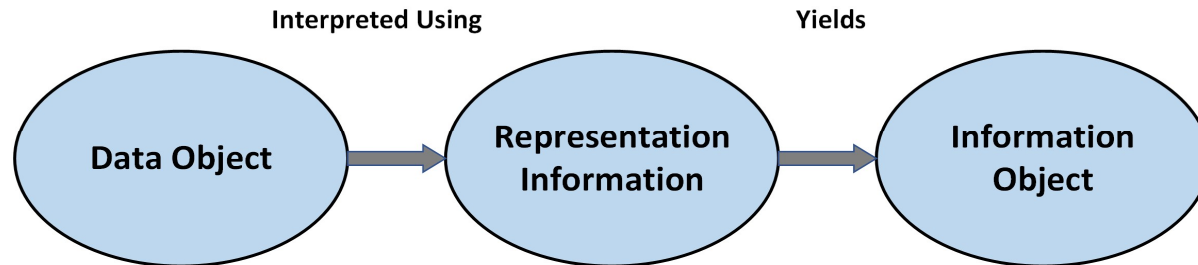
Interoperability

- Two mature ISO level standards exist that provide guidance for the long-term preservation of digital data.

  - *Several important architectural principles have been identified that help enable interoperability.*

1. **Open Archival Information System (OAIS) Reference Model (ISO 14721)**

2. **Metadata Registry Specification (ISO/IEC 11179)**

The Information Object[1]

Interpreted Using   Yields

```
  Data Object  →  Representation  →  Information
                  Information         Object
```

Data Object + Representation Information

= Information Object

- *Data Object = Physical Object or a Digital Object*
- *Representation Information is the information that maps a Data Object into more meaningful concepts so that the Data Object may be understood.*
- *The Information Object is a fundamental building block in the development of a common understanding.*

[1] Open Archival Information System (OAIS) Reference Model (ISO 14721)
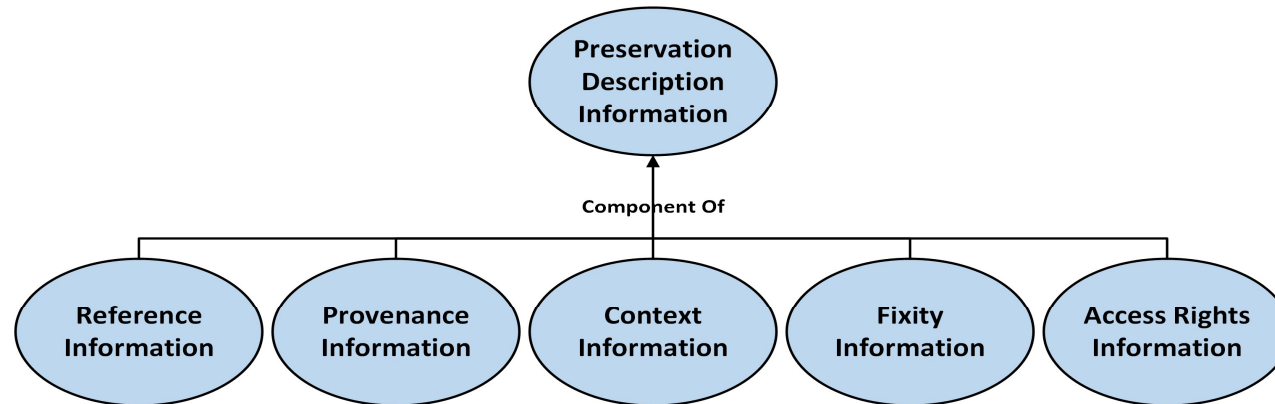
Preservation Description Information (PDI)

- The information, which along with **Representation Information** which is necessary for adequate preservation of the **Data Object**

- PDI is an **Information Object.**

Categories of PDI

Preservation Description Information

Component Of

Reference Information | Provenance Information | Context Information | Fixity Information | Access Rights Information

- **Reference Information** is necessary for referencing this data as well as referencing data that is in a meaningful relationship with this data.

- **Provenance Information** provides the history of the data and is essential for authenticity.

- **Context Information** is the information that helps orient the data within an environment.

- **Fixity Information** is required to ensure that data in general has not been unintentionally altered

- **Access Rights Information** identifies the access restrictions pertaining to the data, including the legal framework, licensing terms, and access control.

PDI (cont)

- Each category of PDI is itself an **Information Object**
  - *This ensures that each has its own **Representation Information** to ensure that that it can be interpreted.*
  - *For example, each instance of **Provenance Information** has its own **Representation Information** so that the consumer can understand it.*

- **Context Information** in particular has an important role in enabling data reusability
  - *Defines the relationships of the **Data Object** to the other **Things** within its environment*
  - *Relationships add **semantic** information.*

Metadata Registry[1]

Data Element Concept	Conceptual Domain
Data Element	Value Domain
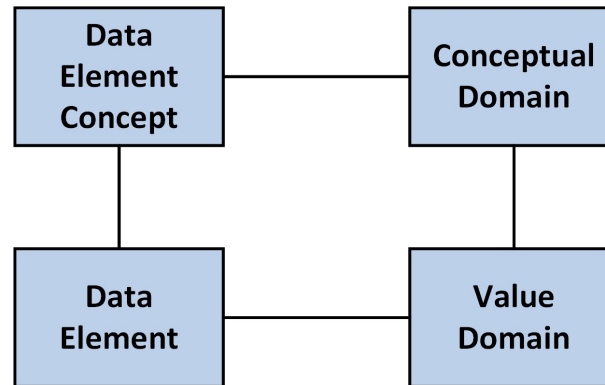
- A data dictionary schema for **Data Element** information
  - *Example the "start_time" of an event.*
  - *alternate names, definition sources, definitions in other languages, effective dates, submitting organization, and stewardship.*
  - *data representation, units of measurement, effective dates, submitter, and steward.*
  - *permissible values and value meanings*

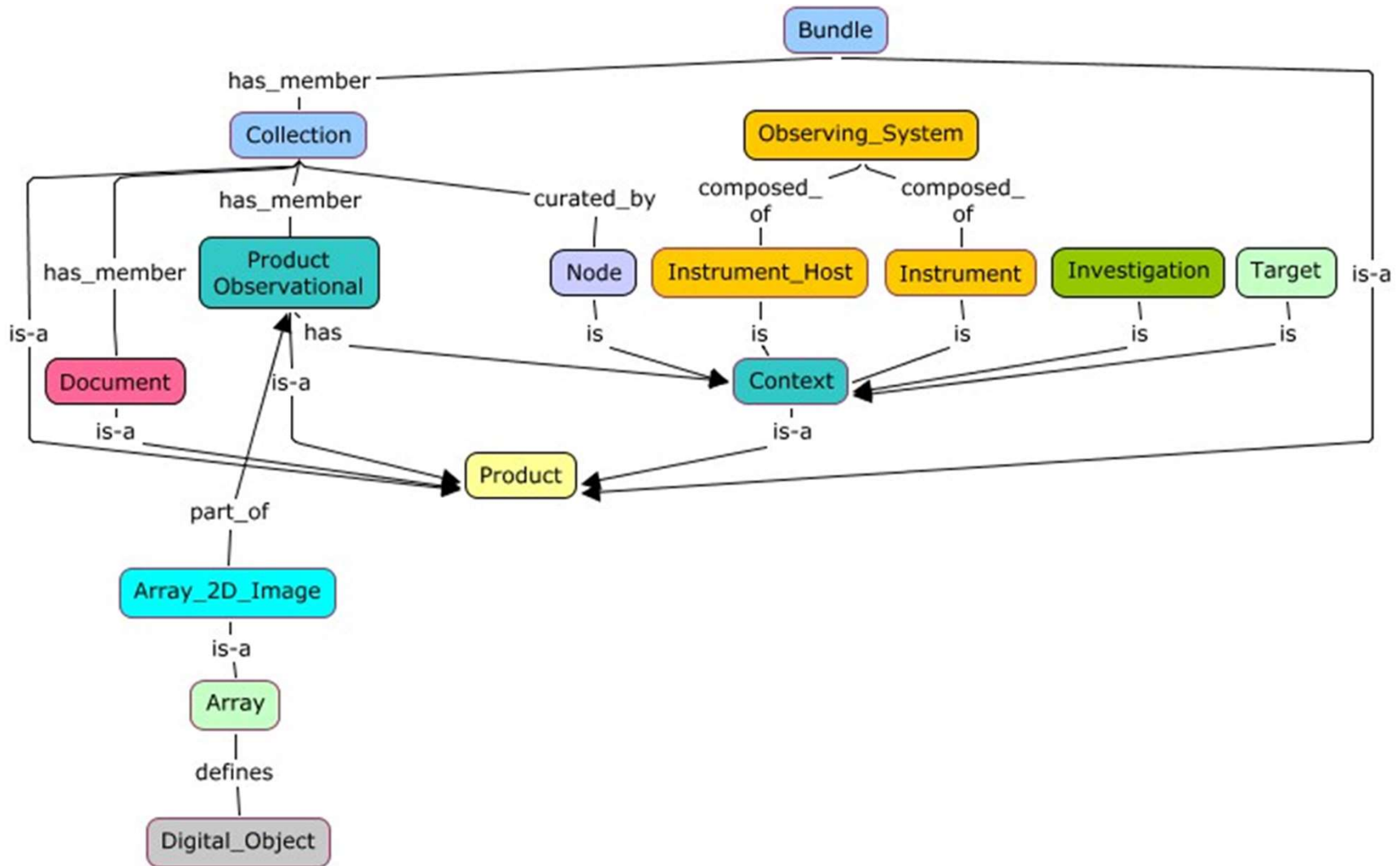[1] Metadata Registry Specification (ISO/IEC 11179)

Interoperability

- In 2001, Uschold [1] argued that a "single shared ontology" is critical for developing a digital library that enables semantic interoperability across disciplines.

[1] M. Uschold and Gruninger. M., "Ontologies and Semantics for Seamless Connectivity," SIGMOD Record, vol. 33, 2004.

PDS4 Concept Map

PDS4 Ontology (1)

| Classes | Slots | Forms | Instances | Queries |

CLASS BROWSER

CLASS EDITOR

For Project: ● UpperModel_190822_Build10a_...

For Class: ● Array_2D_Image   (instance of :STANDARD-CLASS)

Class Hierarchy

Name

Documentation

Constraints

- :THING
  - :SYSTEM-CLASS
  - ChangeLog
  - Data_Object
  - Data_Type
  - Product
  - Product_Components
  - Tagged_Digital_Child
  - Tagged_Digital_Object
    - Byte_Stream
      - Array
        - Array_1D
        - Array_2D
          - Array_2D_Image
          - Array_2D_Map
          - Array_2D_Spectrum
        - Array_3D
      - Encoded_Byte_Stream
      - Parsable_Byte_Stream
      - Table_Base
    - Composite_Structure

Array_2D_Image

The Array 2D Image class is an extension of the Array 2D class and defines a two dimensional image.

Role

Concrete ●

Template Slots

Name	Cardinality	Type	
associated_Special_Constants	single	Instance of Special_Constants	
associated_Statistics	single	Instance of Object_Statistics	
axes	required single	Integer	value=2
axis_index_order	required single	String	value=Last Index Fastest
data_object	required single	Instance of Digital_Object	
description	single	String	
has_Axis_Array	required multiple (:...	Instance of Axis_Array	
has_Display_2d_Image	single	Instance of Display_2D_Image	
has_Element_Array	required single	Instance of Element_Array	
local_identifier	single	String	
local_internal_reference	none	Instance of Local_Internal_Reference	
md5_checksum	single	String	
name	single	String	
offset	required single	Integer	

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

PDS4 Ontology - 2
ISO/IEC 11179 – Metadata Registry

Information Model (IM)

- "An information model is a representation of concepts, relationships, constraints, rules, and operations to specify data semantics for a chosen domain of discourse." [1]

- The **Information Model (IM)** defines a knowledge base for the community.

- The **IM** remains independent of the implementation

- It provides a sharable, stable, and organized structure of **information requirements** for information systems development

[1] Lee, Y. T. 1999. Information Modeling: From Design To Implementation. In Proceedings of the Second World Manufacturing Congress, ed. S. Nahavandi and M. Saadat, 315-321. Canada/Switzerland: International Computer Science Conventions.

From Information Model to Product Label

Information Model

PLANETARY SCIENCE DATA MODEL

Mission
Spacecraft
Instrument
Target
Data Set Collection
Data Set
Document
Software
Image
Time Series
Spectrum
Metadata Labels
Ancillary Products (and their labels)
Data Products (and their labels)
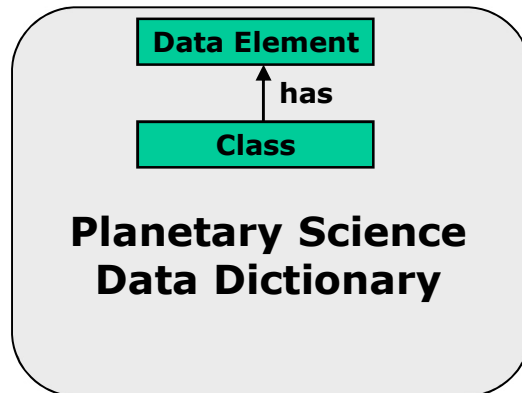
Used to Create

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

14

Model Driven Architecture
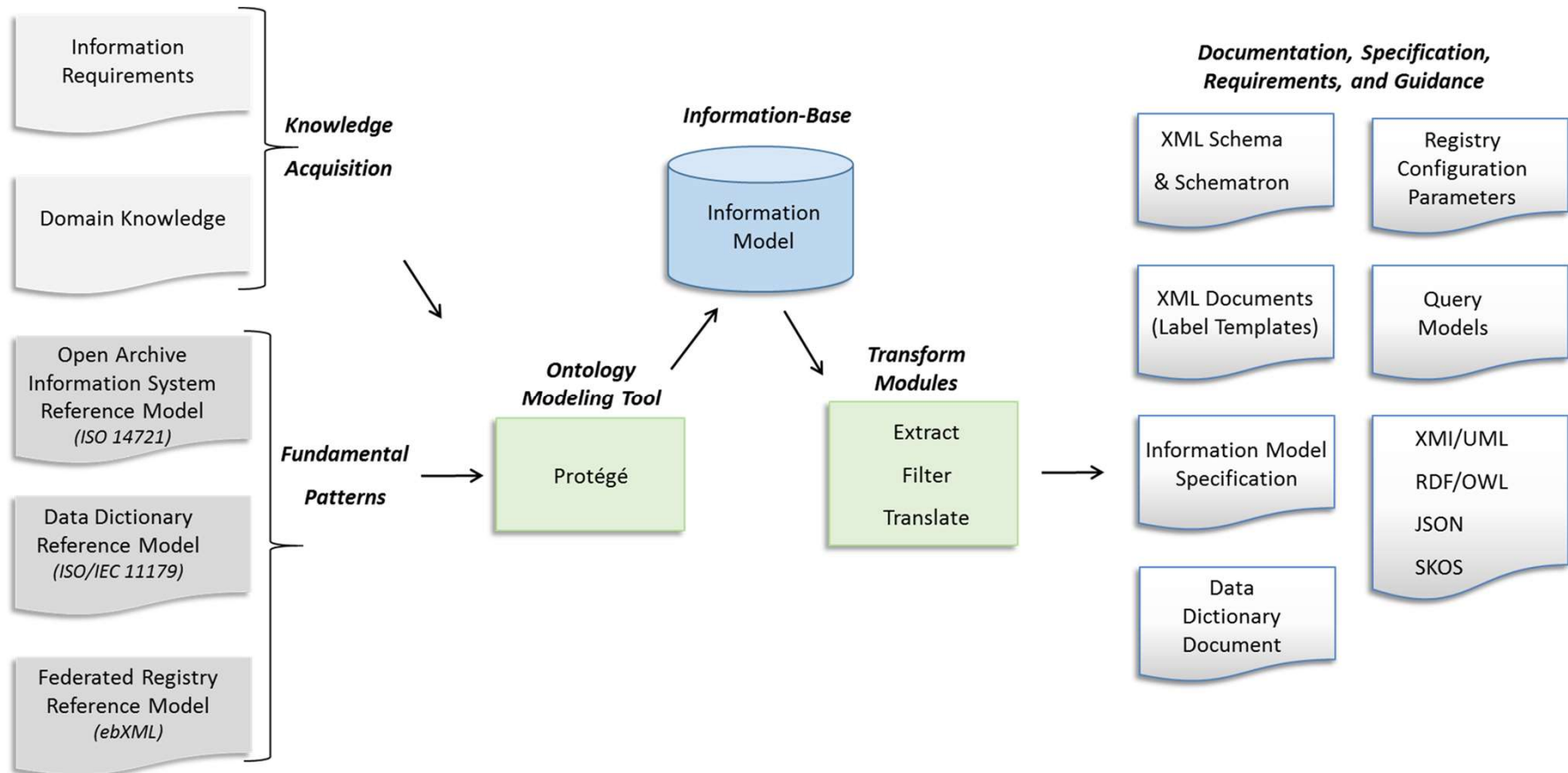
Information Requirements

Domain Knowledge

Knowledge Acquisition

Open Archive Information System Reference Model *(ISO 14721)*

Data Dictionary Reference Model *(ISO/IEC 11179)*

Federated Registry Reference Model *(ebXML)*

Fundamental Patterns

Ontology Modeling Tool

Protégé

Information-Base

Information Model

Transform Modules

Extract

Filter

Translate

Documentation, Specification, Requirements, and Guidance

XML Schema & Schematron

Registry Configuration Parameters

XML Documents (Label Templates)

Query Models

Information Model Specification

XMI/UML

RDF/OWL

JSON

SKOS

Data Dictionary Document

Artifacts Generated from the Model
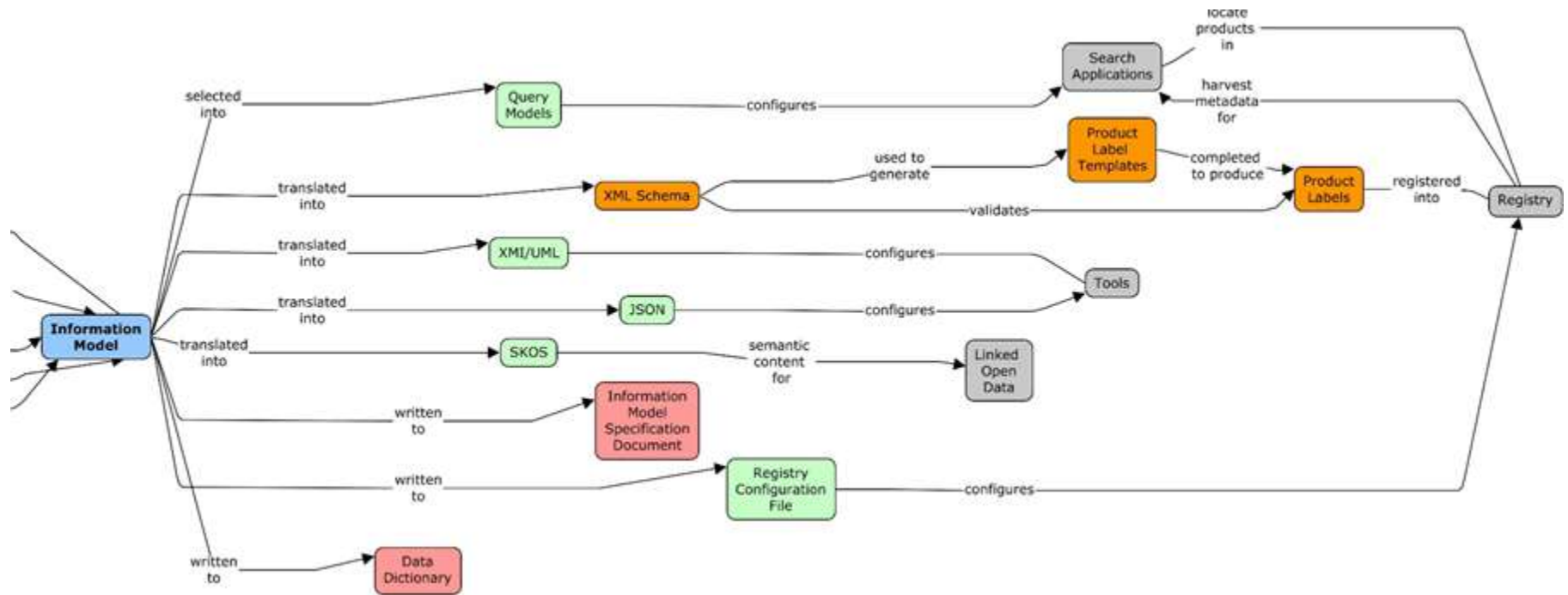
XML Schema and Schematron Files

```xml
<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>
```

17

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>
```

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"
              ]
           }
        }
```

RDF File (Semantic Processing)

```
<kb:NonEnumeratedValueDomain
rdf:about="&pdsns;NEVD.0001_NASA_PDS_1.pds.Array.pds.offset"
  kb:dataIdentifier="0001_NASA_PDS_1.pds.Array.pds.offset"
  kb:versionIdentifier="1.13"
  kb:maximumCharacterQuantity="TBD_maximum_characters"
  kb:minimumCharacterQuantity="TBD_minimum_characters"
  kb:maximumValue="TBD_maximum_value"
  kb:minimumValue="0"
  kb:pattern="TBD_pattern"
  kb:valueDomainFormat="TBD_format"
  kb:defaultUnitId="byte"
  rdfs:label="0001_NASA_PDS_1.pds.Array.pds.offset">
  <kb:representedBy1 rdf:resource="&pdsns;DE.0001_NASA_PDS_1.pds.Array.pds.offset"/>
  <kb:representedBy2 rdf:resource="&pdsns;CD_Integer"/>
  <kb:datatype rdf:resource="&pdsns;ASCII_NonNegative_Integer"/>
  <kb:unitOfMeasure rdf:resource="&pdsns;Units_of_Storage"/>
  <kb:registeredBy rdf:resource="&pdsns;RA_0001_NASA_PDS_1"/>
  <kb:steward rdf:resource="&pdsns;pds"/>
  <kb:submitter rdf:resource="&pdsns;Submitter_PDS"/>
  <kb:terminologicalEntry
rdf:resource="&pdsns;TE.0001_NASA_PDS_1.pds.Array.pds.offset"/>
  <kb:administrationRecord rdf:resource="&pdsns;DD_1.13.0.0"/>
  <kb:registrationAuthorityIdentifier rdf:resource="&pdsns;0001_NASA_PDS_1"/>
</kb:NonEnumeratedValueDomain>
```

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	

21

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

22

Governance

- The Information Model (IM) consists of a hierarchy of models.

  - *Common model*

  - *Discipline models*

  - *Mission models*

- The partitioning of the IM into individual models reduces the impact of change.

- Each model

  - *has a steward that manages the development and maintenance of that model.*
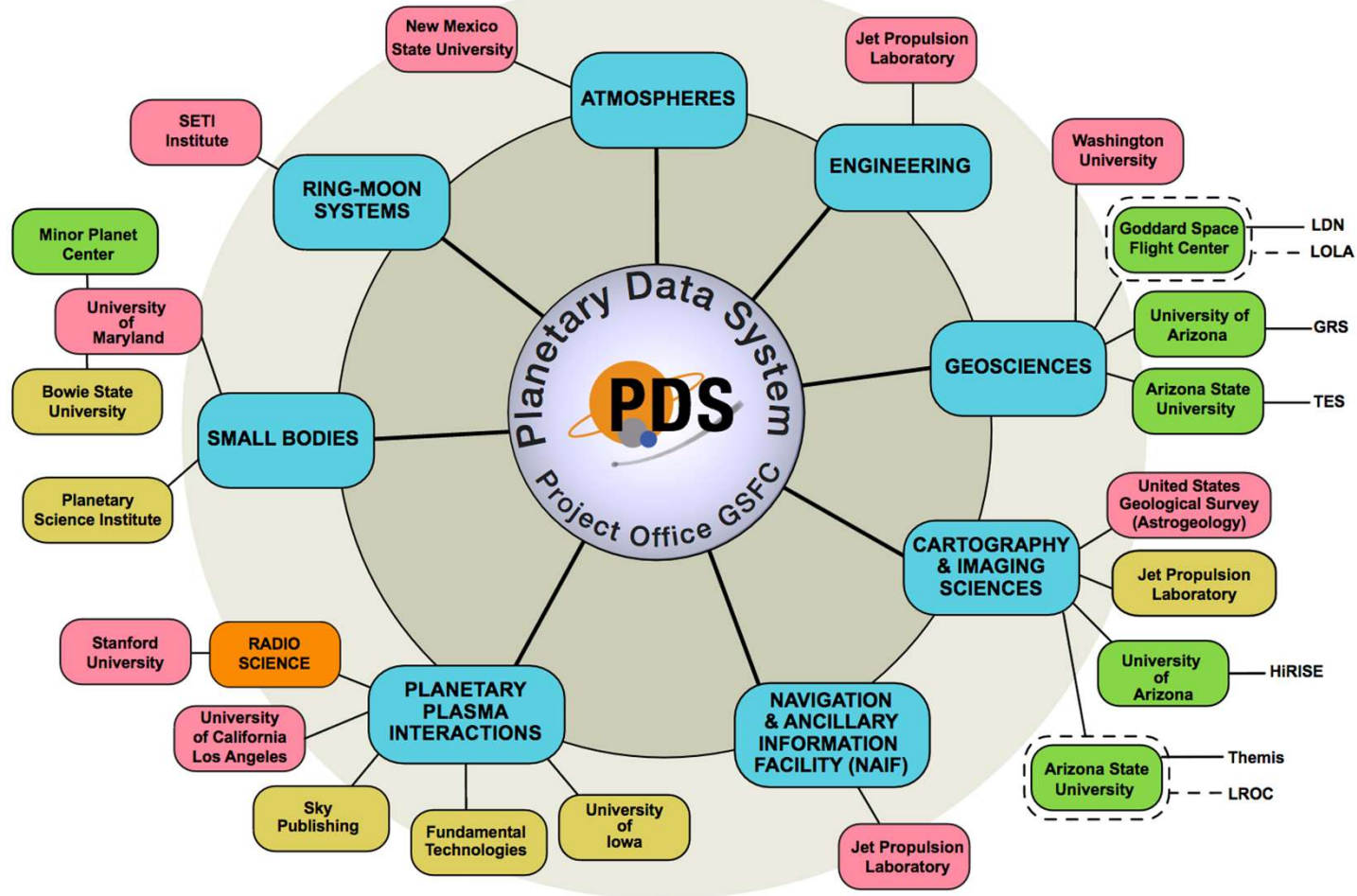
  - *has a unique namespace*

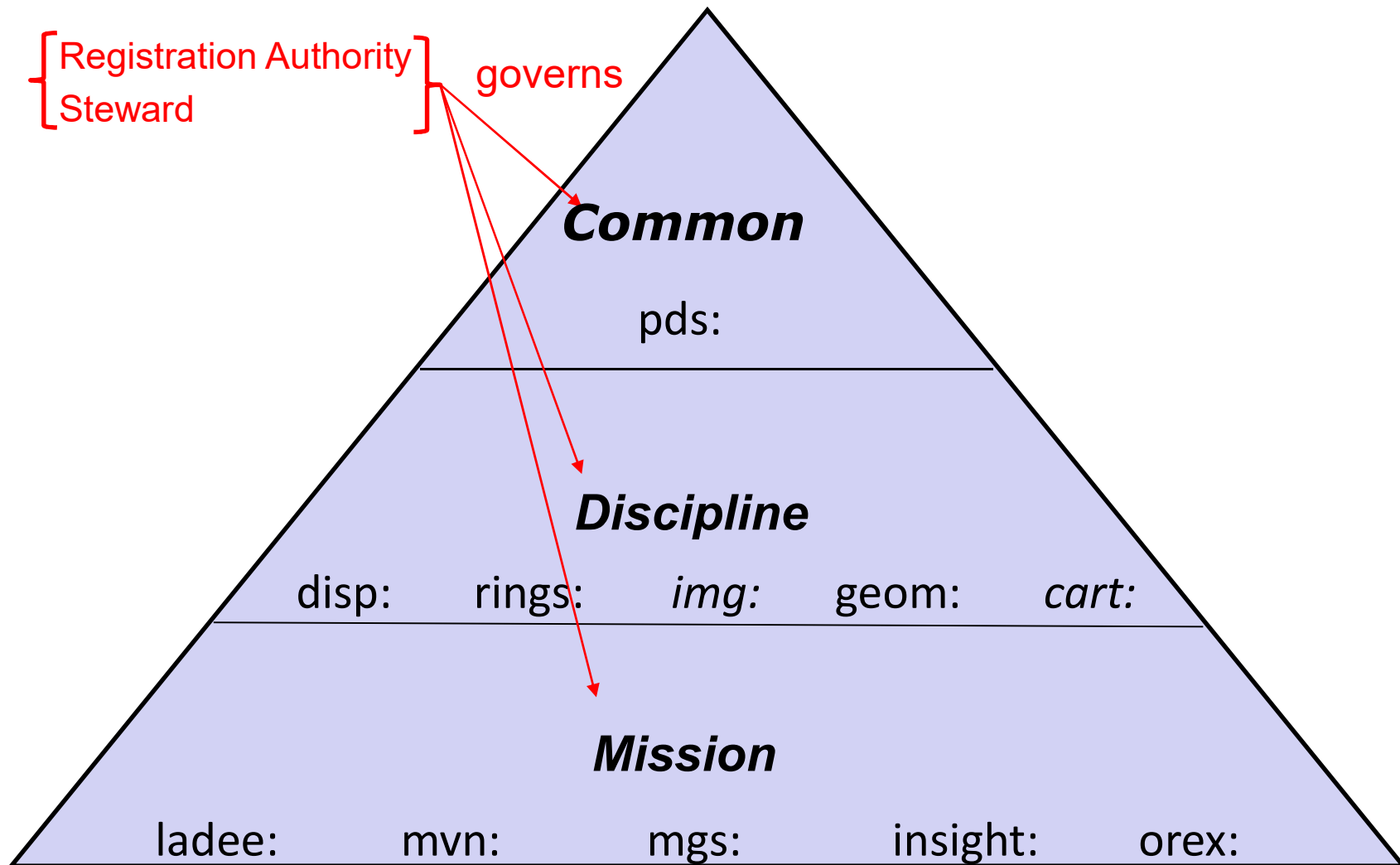# Diverse Disciplines

**Information Model
Multi-level Governance**

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Registration Authority
Steward

governs

Common

pds:

Discipline

disp: rings: *img:*   geom: *cart:*

Mission

ladee: mvn: mgs: insight: orex:

25

Extensions to the "Common"

Local Data Dictionaries (LDDs)

| Common		
	PDS4_PDS_1B10	PDS4 Common
Discipline		
	PDS4_ALT_1000	PPI Node's Alternate
	PDS4_CART_1B10_1930	Imaging Node's Cartography
	PDS4_DISP_1B10	Imaging Node's Display
	PDS4_GEOM_1B10_1700	Geometry
	PDS4_IMG_1B00_1600	Imaging Node
	PDS4_IMG_SURFACE_1B10_1100	Surface Imaging
	PDS4_MSN_SURFACE_1B00_1100	Surface Mission Information
	PDS4_MSN_1B00_1100	Generic Mission
	PDS4_MULTI_1900_1000	multidimensional data
	PDS4_PARTICLE_1900_1100	PPI Node's Particle
	PDS4_PROC_1900	Processing History
	PDS4_RINGS_1800_1500	Rings Node
	PDS4_SP_1C00_1100	Spectral
	PDS4_SPECLIB_1000	Spectral Library
	PDS4_WAVE_1000	PPI Node's Wave
Mission		
	BOPPS_1100	BOPPS
	PDS4_CASSINI_1B00_1200	Cassini
	PDS4_DAWN_1B00_1000	Dawn
	PDS4_INSIGHT_1B00_1850	Insight
	PDS4_JUNO_1900	JUNO
	LADEE_1100	LADEE
	PDS4_MESS_1B00_1020	Messenger
	MGS_1700	*Mars Global Surveyor*
	MPF_1700	*Mars Pathfinder*
	PDS4_MVN_1021	MAVEN
	OSIRIS-Rex _1700	*OSIRIS-Rex*
	PDS4_VG1_1900_1000	VOYAGER 1 (vg1)
	PDS4_VG2_1900_1000	VOYAGER 2 (vg2)

Trustworthy Digital Repository (TDR)

- A Trustworthy Digital Repository (TDR) is an organization or system responsible for the long-term preservation and access to digital materials, such as data, records, documents, and other digital objects.

  - *A TDR must be able to guarantee the authenticity, integrity, and usability of its digital content over time.*

- The aim of a TDR is to ensure that digital content remains accessible, usable, and reliable for as long as it is needed.

  - *This is particularly important for materials that have long-term value, such as cultural heritage objects, scientific research data, and government records.*

Architectural Principles as a Graph
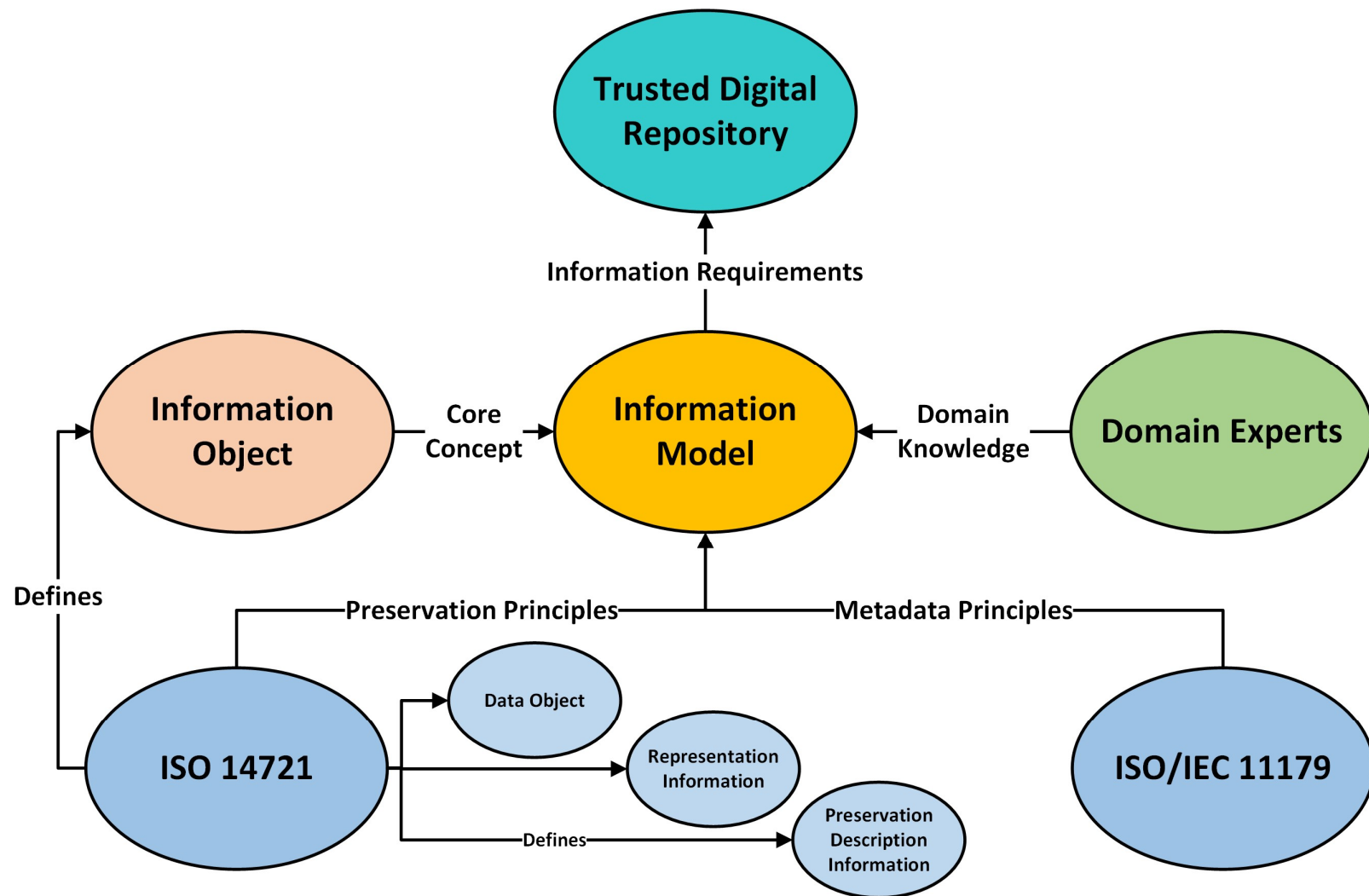
National Aeronautics and Space Administration

Jet Propulsion Laboratory
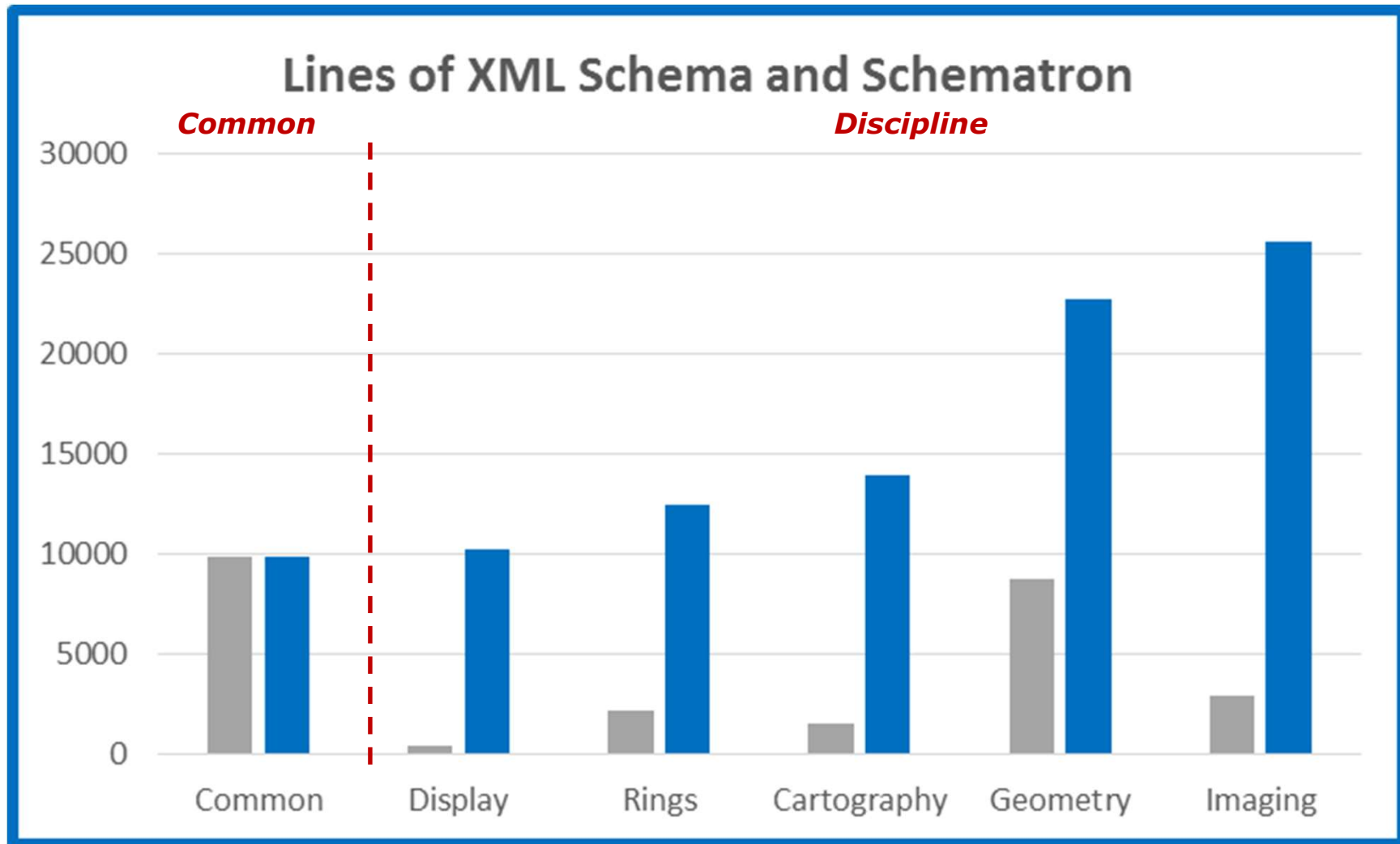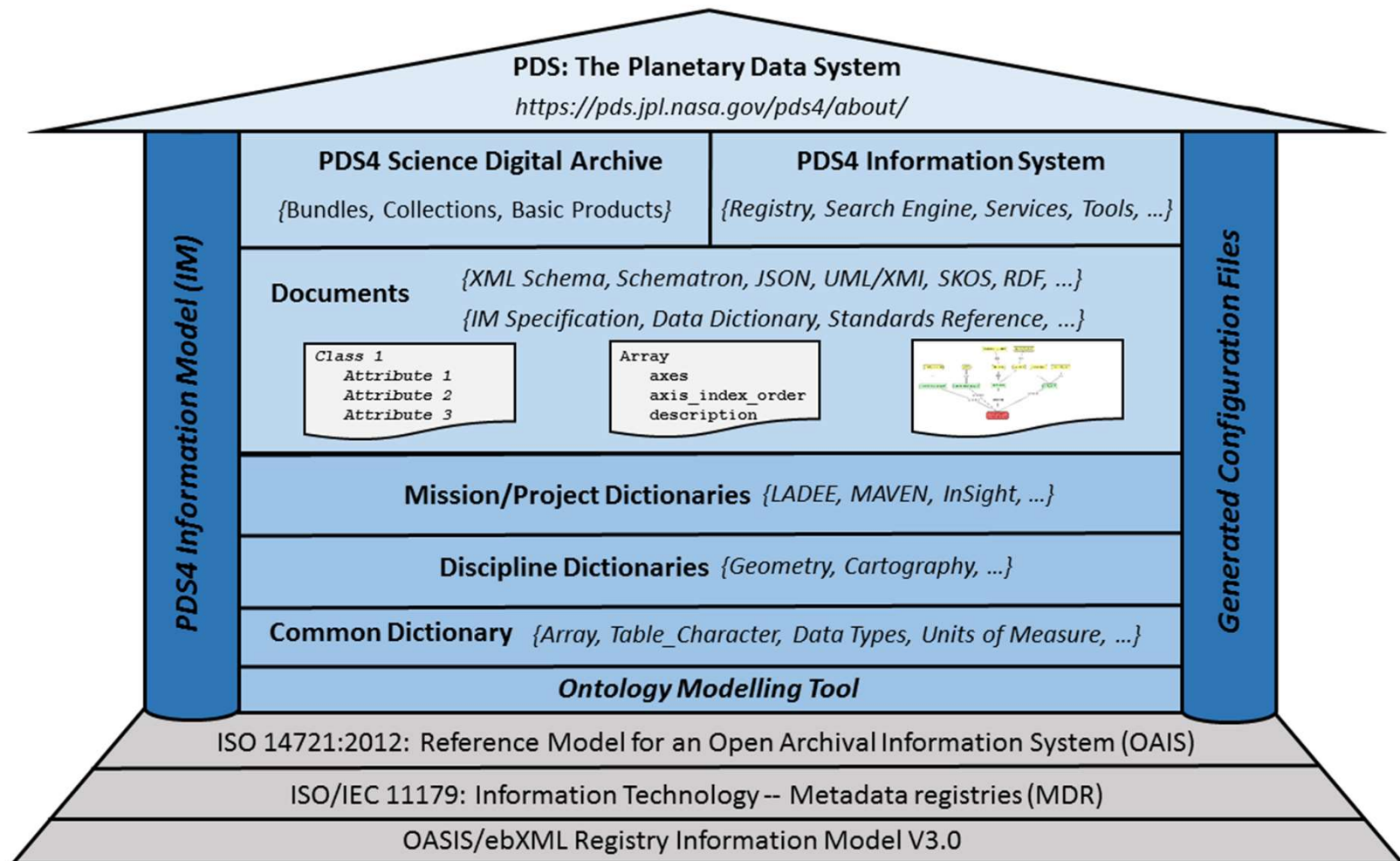California Institute of Technology
Pasadena, California

28

Conclusion

- A framework for capturing the information required to support data interoperability has been developed using principles adopted from two ISO information systems standards.

- Key principles have been identified that support interopeability.

- However this is simply a framework.

  - *The really hard work involves acquiring the appropriate knowledge from domain experts and populating an domain information model using the framework as a guide.*

Backup

Auto-generated Validation Code

Lines of XML Schema and Schematron

Common · Display · Rings · Cartography · Geometry · Imaging

PDS4

Information Model
and Generated Artifacts