



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

A Framework to Enable Interpretation of the Data

J. Steven Hughes

Principal Computer Scientist
Jet Propulsion Laboratory

ESIP GeoSemantics Symposium
January 10th, 2017



Some Definitions

- Semantics is the study of meaning focusing on **relationships**.
- Information Model - In data engineering it is a representation of concepts and the **relationships**, constraints, rules, and operations for a chosen domain.
 - *It provides a sharable, stable, and organized structure of **information requirements** or knowledge for the domain context.¹*

[1] Y. Tina Lee (1999). "Information modeling from design to implementation" National Institute of Standards and Technology.



Use Cases

- What Mars Reconnaissance Orbiter (MRO) High Resolution Imaging Science Experiment (HiRISE) Reduced Data Record (RDR) images have both craters and dunes?
- What coordinate system was used for the HiRISE RDR images?
- An anomalous artifact was found in an Engineering Data Record (EDR) image of Cydonia Mesa collected by HiRISE. For analysis the following are requested:
 - *the calibration files used to calibrate this image*
 - *published instrument design documents.*



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Terms, Definitions, and Relationships

- What Mars Reconnaissance Orbiter (MRO) High Resolution Imaging Science Experiment (HiRISE) Reduced Data Record (RDR) images have both craters and dunes?
- What coordinate system was used for the HiRISE RDR images?
- An anomalous artifact was found in an Engineering Data Record (EDR) image of Cydonia Mesa collected by HiRISE. For analysis the following are requested:
 - *the calibration files used to calibrate this image*
 - *published instrument design documents.*

Digital
Object



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Terms, Definitions, and Relationships

What are these things?

- What Mars Reconnaissance Orbiter (**MRO**) High Resolution Imaging Science Experiment (**HiRISE**) Reduced Data Record (**RDR**) images have both **craters** and **dunes**?
- What **coordinate system** was used for the HiRISE RDR images?
- An anomalous artifact was found in an Engineering Data Record (**EDR**) image of **Cydonia Mesa** collected by HiRISE. For analysis the following are requested:
 - the **calibration files** used to calibrate this image
 - **published instrument design documents**.



Terms, Definitions, and Relationships

How are these things related to the data?

- What Mars Reconnaissance Orbiter (**MRO**) High Resolution Imaging Science Experiment (**HiRISE**) Reduced Data Record (**RDR**) images have both **craters** and **dunes**?
 - *identified in*
 - *collected by*
 - *derived from*
- What **coordinate system** was used for the **HiRISE RDR** images?
- An anomalous artifact was found in an Engineering Data Record (**EDR**) image of **Cydonia Mesa** collected by **HiRISE**. For analysis the following are requested:
 - *the calibration files used to calibrate this image*
 - *published instrument design documents.*



The Information Model (IM)

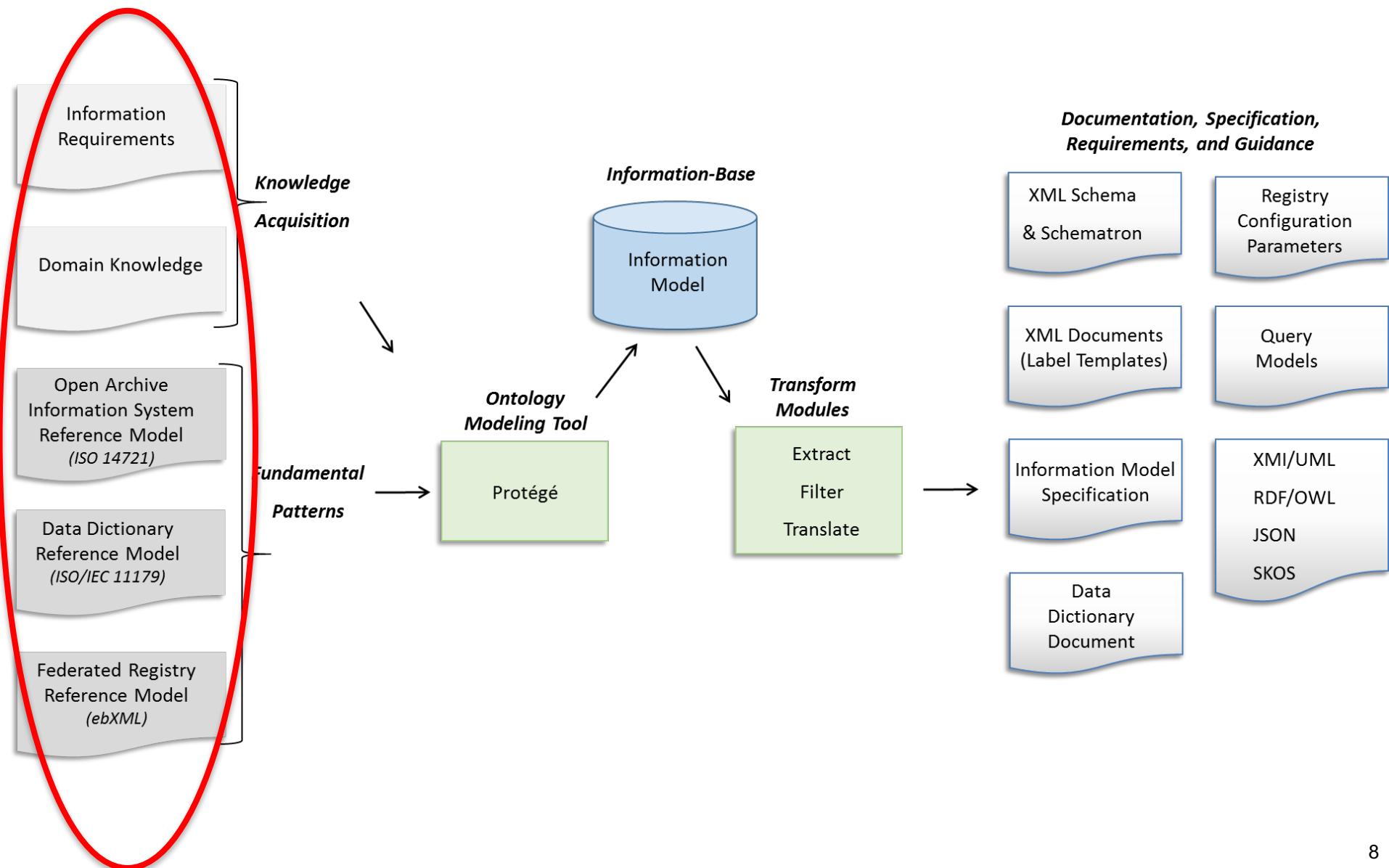
- The Information Model provides the information requirements for the system
 - *Defines the terms in the community and their relationships*
 - *Improves consistency and interoperability and reduces ambiguity*
- Establishes an overarching federated governance model for the metadata
 - *Provides common, discipline, and local governance*
 - *Localizes changes and allows extensions as the community evolves*
 - *Promotes model independence*
- Is effectively the “corner-stone” of the “information model-driven” design paradigm
 - *Allows the system to be configured by and to respond to the information model (information requirements)*
 - *Enables agile development*
 - *Handles diversity*
 - Accommodates new instruments, observation types, data, ...
 - *Reduce the impact of changes on the system*



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Components of the Framework





National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Input

- Foundational Principles
 - *Open Archival Information System Reference Model (OAIS-RM)*
 - *Data Dictionary Reference Model*
 - *Federated Registry Reference Model*
- Community-Specific Input
 - *Information Requirements*
 - *Domain Knowledge*



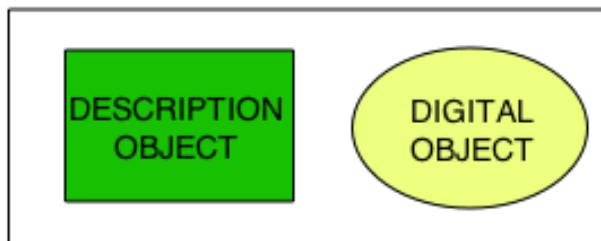
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Open Information Archive System (OAIS) Reference Model

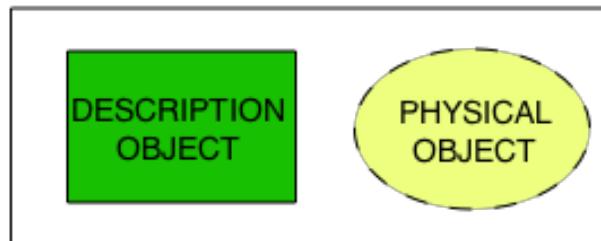
Information Object¹

TAGGED DIGITAL OBJECT



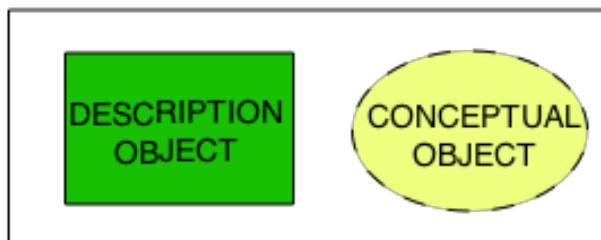
- **digital object:** An object which is real data — for example, a binary image of a redwood tree.

TAGGED NON-DIGITAL OBJECT



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

TAGGED NON-DIGITAL OBJECT

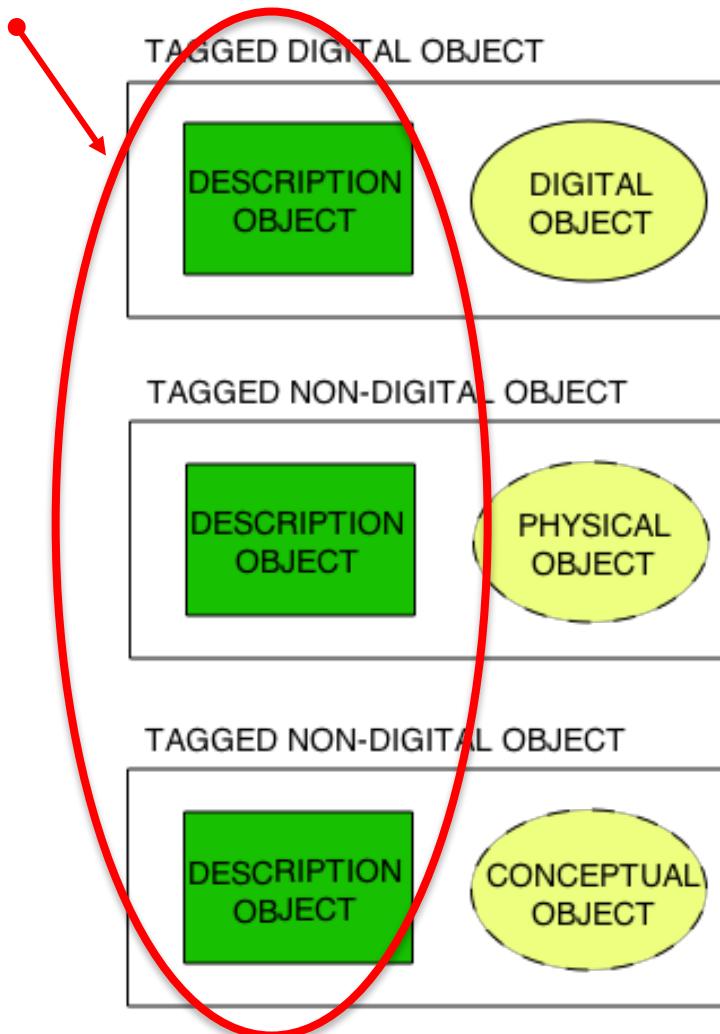


- **conceptual object:** An object which is intangible – for example the Cassini mission and NASA's strategic plan for solar system exploration.



Data Management¹

Description Object Management



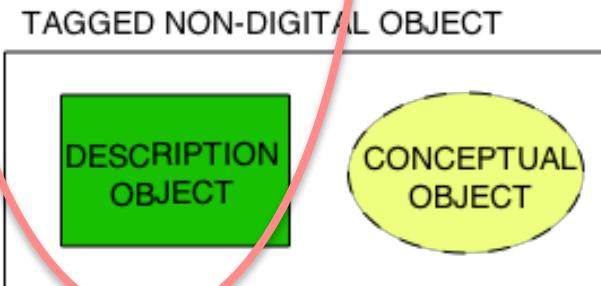
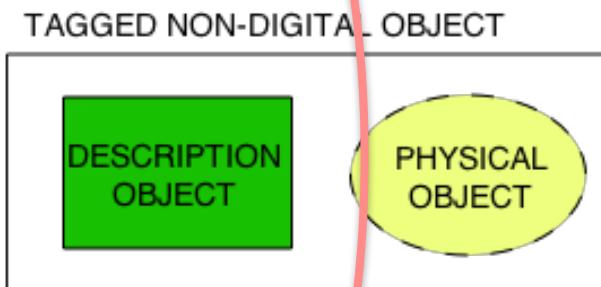
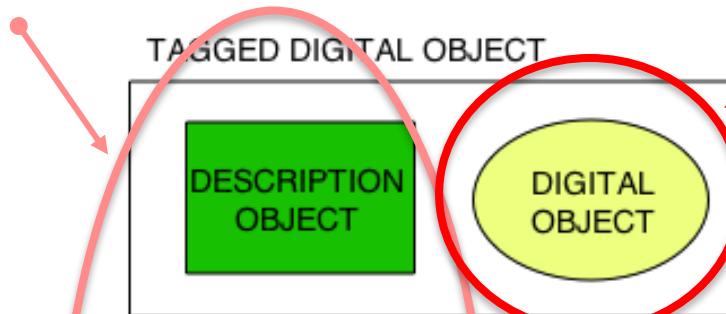
Information Object Model¹

- **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.



Data Object Management

Description Object Management



Information Object Model¹

Data Object Management

- **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.



Information Categories¹

- **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 (Fixity)**
 - *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.*



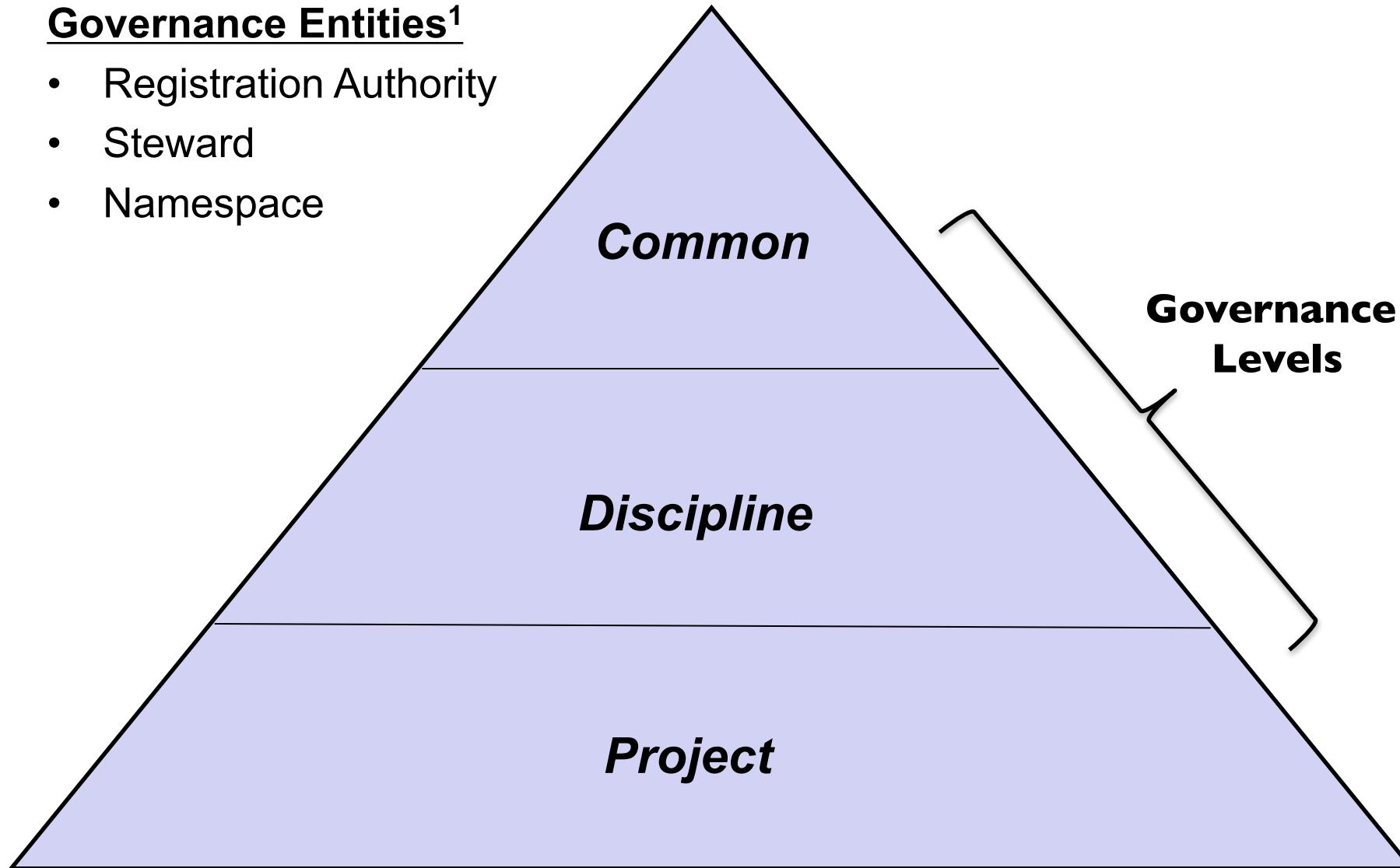
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Data Dictionary Reference Model²

Governance Entities¹

- Registration Authority
- Steward
- Namespace





Registry Reference Model

- ebXML (Electronic Business XML) - Standardizes the secure exchange of data
- Defines key properties of a federated registry
 - *registry database schema*
 - *registry object (generic)*
 - Extensions: products, granules, etc
 - First class objects
 - *digital, physical, and conceptual*
 - *common registry services*
 - tracking/locate/retrieval
 - *core attributes*
 - identification (e.g. <logical_identifier>)
 - versioning (e.g. <version_identifier>)



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Information Requirements

- The community's Requirements and Policies set the foundation for the *information requirements*.

I. The System will provide expertise to guide and assist missions, programs, and individuals to organize and document digital data supporting the institutions goals in science exploration.

1.4 Archiving Standards: The system will have **archiving standards** for science data

1.4.1 The system will **define a standard** for organizing, formatting, and documenting science data

1.4.2 The system will maintain a **dictionary of terms, values, and relationships** for standardized description of science data

1.4.3 The system will define a **standard grammar** for describing science data

1.4.4 The system will **establish minimum content requirements** for a data set (primary and ancillary data)

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

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

1.4.7 The system will keep **abreast of new developments** in archiving standards

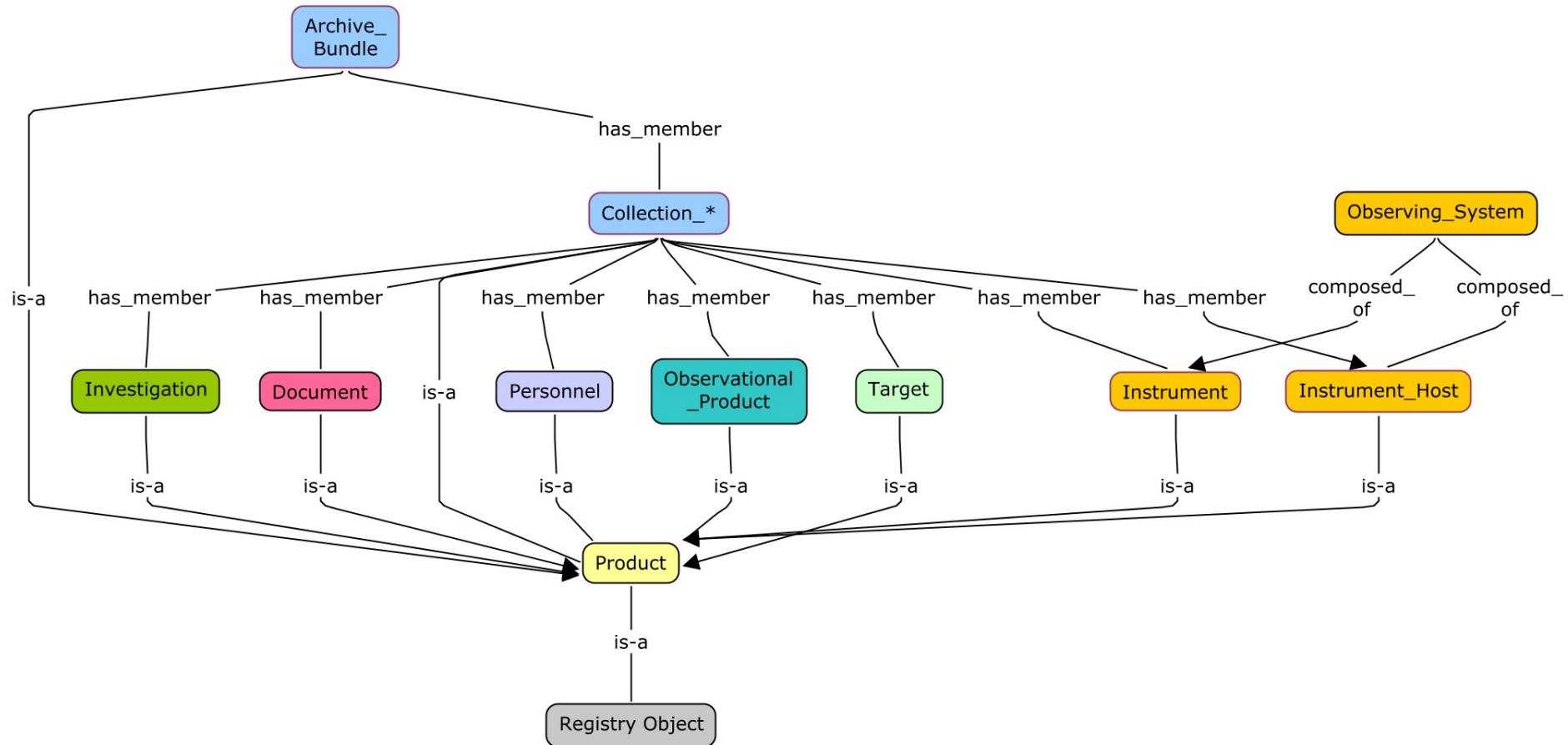


Domain Knowledge

- The experiential (first hand) information about things of interest known by experts in the domain.
- Information about the “things” that should be collected and associated with the data to make and keep it useful.
 - *The data and their structures (representation information)*
 - *The context within which the data was used and collected*
 - Investigations/Missions/Campaigns
 - Observing Systems/Instruments
 - Personnel
 - Data collection targets of interest



Domain Knowledge

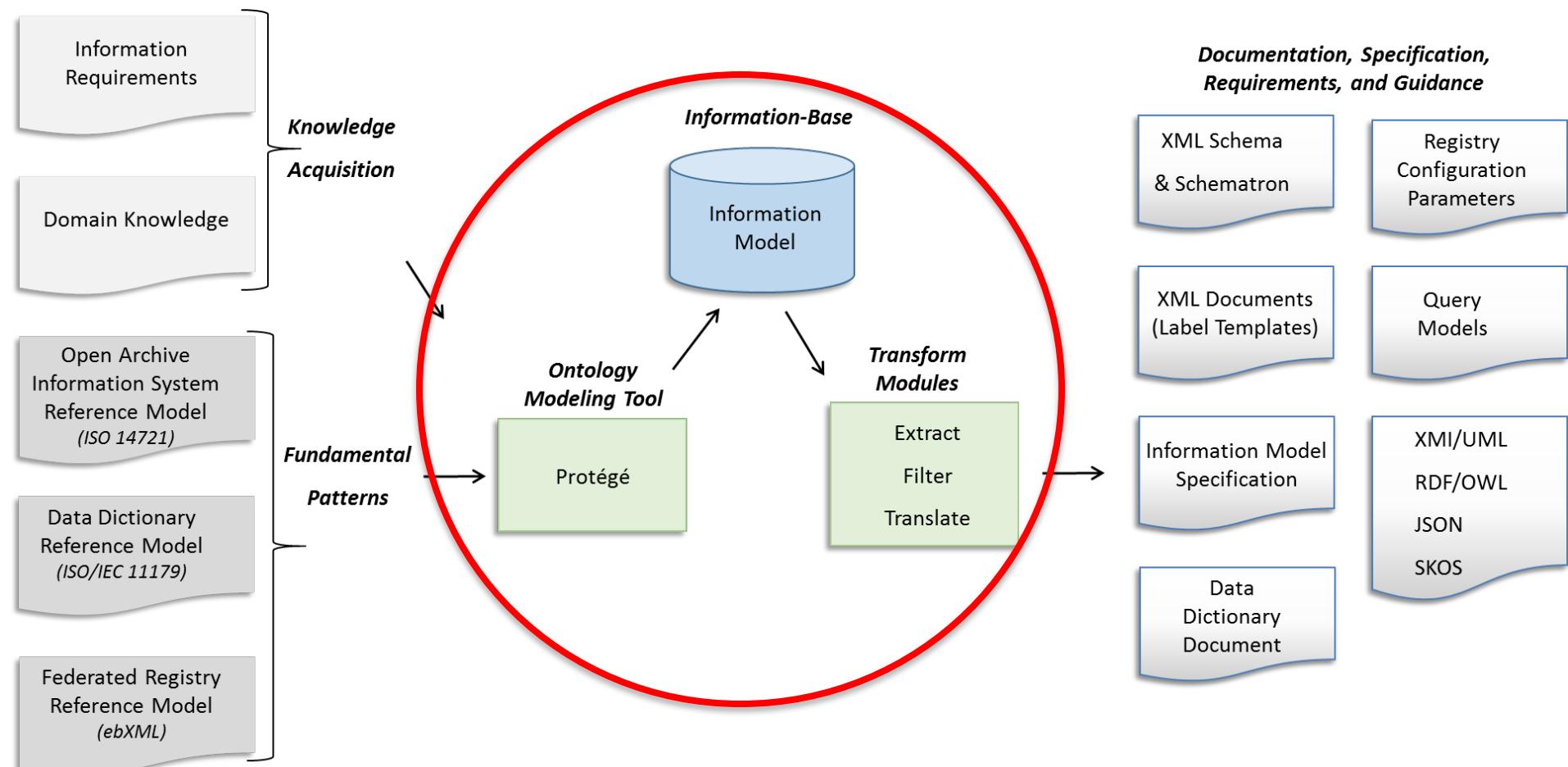




National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Process





The Information Model Database

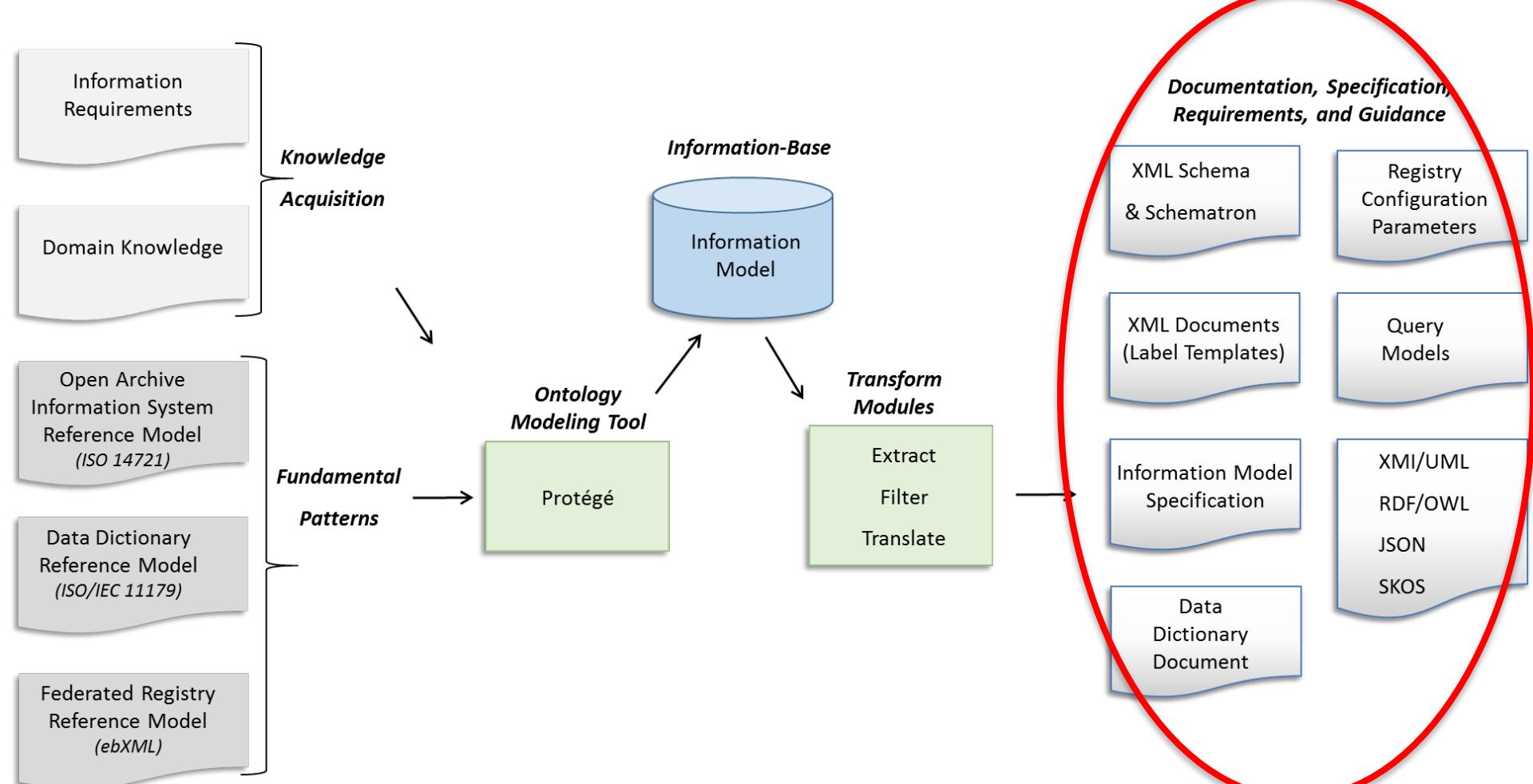
- All “things of interest” are defined in an object modeling tool.
 - *All objects and their attributes and relationships.*
 - *Typically an ontology modeling tool is used*
 - Necessary but not necessarily sufficient
- A master database is created by merging the object model(s) and the data dictionary.
- The contents of the master database is filtered and written to system files in various formats.
 - *Used by the data providers, registry, harvester, search engine, validator and other system tools and services.*



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Output





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" ...>
      ...
    </xs:sequence>
  </xs:extension>
</xs:complexContent>
</xs:complexType>

<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>
    </sch:rule>
  </sch:pattern>
</sch:schema>
```



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Product Label Template

Common Area

Identification_Area

Logical_Identifier

Version_Id

Observation_Area

Time_Coordinates

Primary_Result_Summary

Investigation_Area

Observing_System

Target_Identification

***Discipline_Area
Mission Area***

Reference_List

Internal_Reference

External_Reference

File_Area_Observational

File

Header

Array_2d_Image



The Framework at Work

- What Mars Reconnaissance Orbiter (**MRO**) High Resolution Imaging Science Experiment (**HiRISE**) Reduced Data Record (**RDR**) images have both **craters** and **dunes**?
 - **Spacecraft, instruments, digital images, documentation, and calibration files** are all defined and related in the information model as classes.
 - **Labeled objects** are created, validated and ingested into the registry.
 - All **registered objects** are considered first-class that can be tracked, located and retrieved..
 - The **Imaging Atlas (catalog)** uses semantics from the information model and harvested metadata from **labeled objects** to provide field- and facet-based search.
 - **Image content annotations** were generated using a visual salience **landmark** detector plus a deep learning neural network classifier.
 - Enables **Image Atlas** to provide content-based search for **HiRISE RDR images** of Mars



Semantics at Work

- What coordinate system was used for the HiRISE RDR images?
 - *Coordinate systems are defined as a class (with subclasses) in a discipline level cartography model*
 - *The label for the image indicates that the coordinate system used is planetocentric latitude and east positive longitude direction.*
- An anomalous artifact was found in an Engineering Data Record (EDR) image of Cydonia Mesa collected by HiRISE. For analysis the following are requested:
 - the calibration files used to calibrate this image
 - published instrument design documents.
 - Documents are either referenced as registered objects or via bibliographic citations.
 - Features are classified and defined in a feature catalog.



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Thank You!

PDS4 Documents

<https://pds.jpl.nasa.gov/pds4/doc/index.shtml>



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Backup



Data Dictionary Schema

Data Element

- *Name*
- *Submitter, Steward*
- *Definition*
- *Namespace*
- *Source of definition*
- *Change log*
- *Version*
- *Concept*
- *Alternate Names*
- *Definition in different 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



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Documentation Links

PDS4 Documents: <https://pds.jpl.nasa.gov/pds4/doc/index.shtml>

IM Spec: https://pds.jpl.nasa.gov/pds4/doc/im/v1/index_1700.html

DD (html): https://pds.jpl.nasa.gov/pds4/doc/dd/v1/PDS4_PDS_DD_1700.html

DD (pdf): https://pds.jpl.nasa.gov/pds4/doc/dd/v1/PDS4_PDS_DD_1700.pdf

Release Notes:

https://pds.jpl.nasa.gov/pds4/doc/im/v1/PDS4Build7a_Release_1700_160928.pdf

XSD: https://pds.jpl.nasa.gov/pds4/schema/released/pds/v1/PDS4_PDS_1700.xsd

SCH: https://pds.jpl.nasa.gov/pds4/schema/released/pds/v1/PDS4_PDS_1700.sch

XML: https://pds.jpl.nasa.gov/pds4/schema/released/pds/v1/PDS4_PDS_1700.xml

JSON: https://pds.jpl.nasa.gov/pds4/schema/released/pds/v1/PDS4_PDS_JSON_1700.JSON

OWL/RDF: https://pds.jpl.nasa.gov/pds4/schema/released/pds/v1/PDS4_PDS_OWL_1700.rdf



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 dimensional array.

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	



Data Dictionary Document (pdf and html formats)

Array_2D_Image

Name: Array_2D_Image			Version Id: 1.1.0.0
<i>Description:</i> The Array 2D Image class is an extension of the Array 2D class and defines a two dimensional image.			
<i>Namespace Id:</i> pds	<i>Steward:</i> pds	<i>Role:</i> concrete	<i>Status:</i> Active
<i>Class Hierarchy:</i> Tagged_Digital_Object :: Byte_Stream :: Array :: Array_2D :: Array_2D_Image			
<i>Attribute(s)</i>	<i>Name</i>	<i>Cardinality</i>	<i>Value</i>
	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
<i>Association(s)</i>	<i>Name</i>	<i>Cardinality</i>	<i>Class</i>
	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" ,  
            "attributeld": [  
                "0001_NASA_PDS_1.pds.Array.pds.offset"  
            ]  
        }  
    ]  
}
```



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

Registry Configuration Parameters

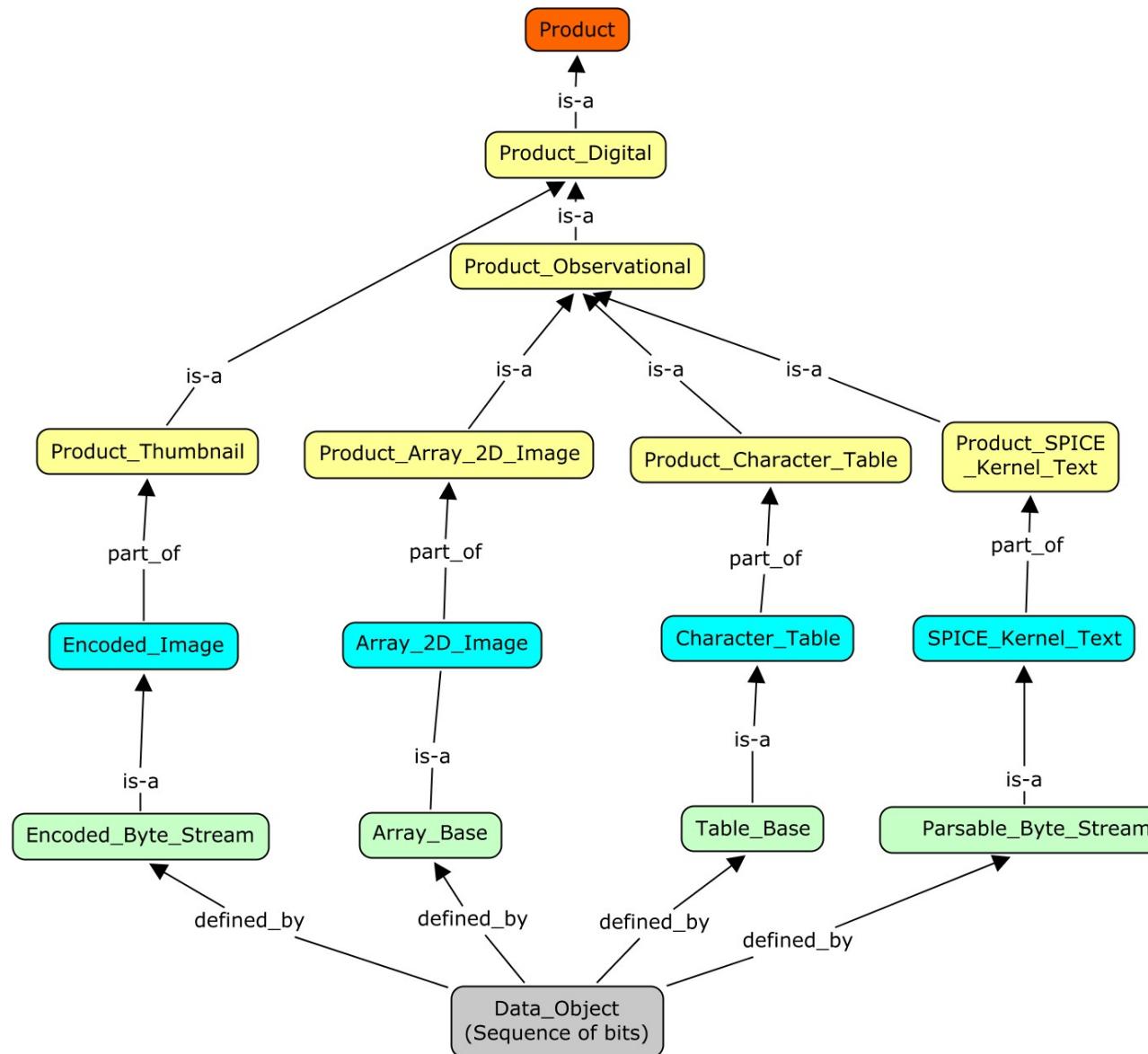
```
reg_object_type:  
Product_Observational  
metadata: {  
    slot1: start_date_time  
    slot2: stop_date_time  
    slot3: version_id  
    slot4: title  
    slot5: product_class  
    slot6: logical_identifier  
    slot7: alternate_title  
    slot8: alternate_id  
    slot9: version_id  
}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



National Aeronautics and
Space Administration

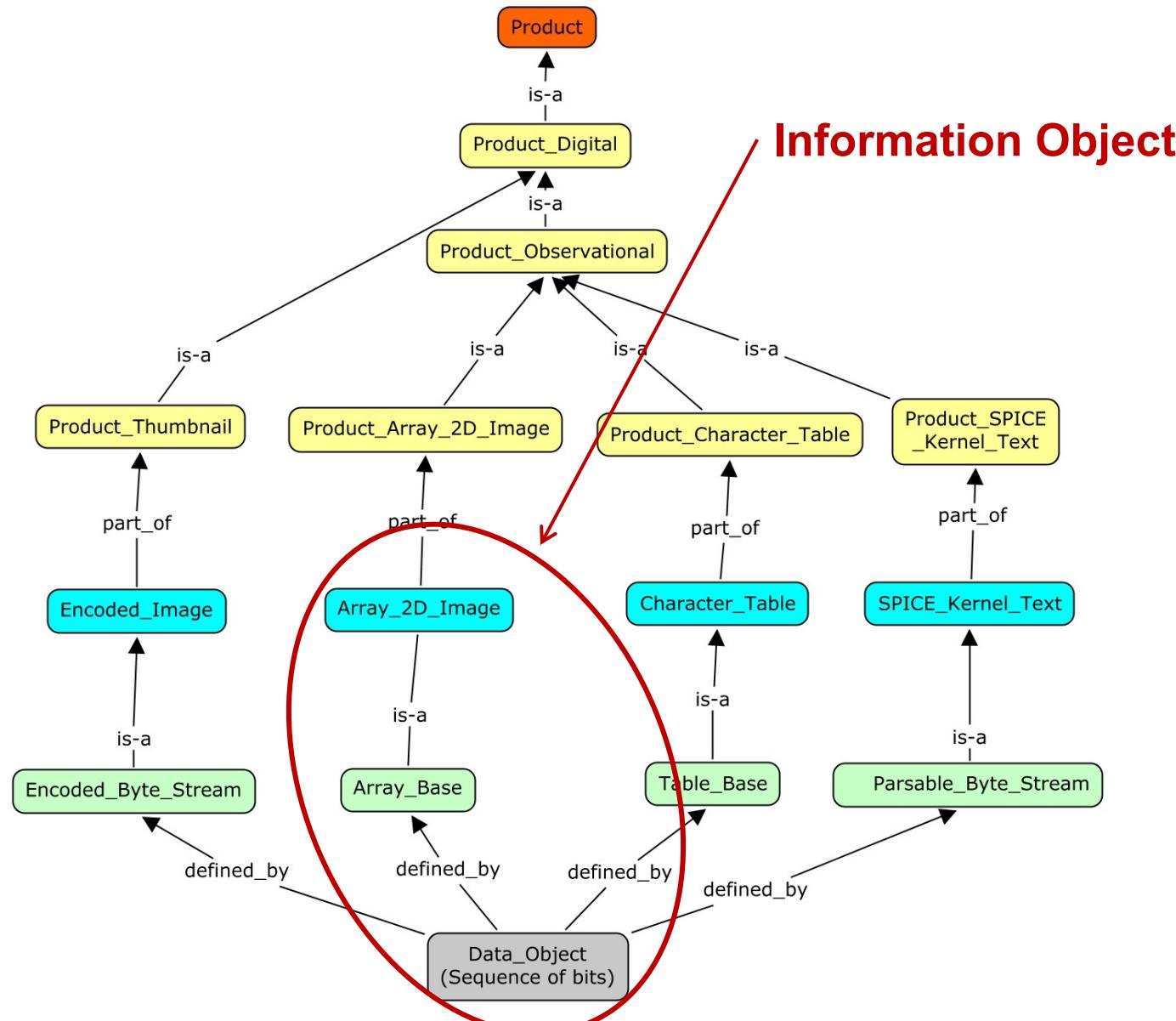
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Product Concept Map





Domain Knowledge and Information Objects



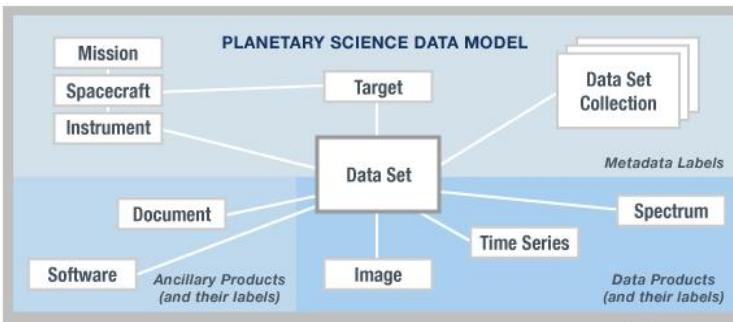


National Aeronautics and
Space Administration

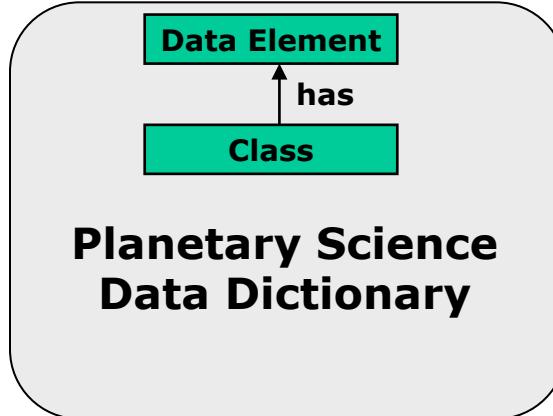
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

From Information Model to Product Label

Information Model



Expressed As



Used to Create

XML Schema

Validates

Extracted/Specialized

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



Reference Models

