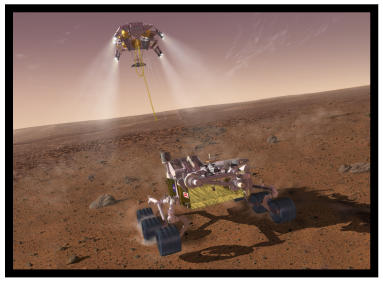
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California



A Reference Model for Science Data Archives

J. Steven Hughes
steve.hughes@jpl.nasa.gov
NASA Jet Propulsion Laboratory (JPL)
California Institute of Technology

Earth Science Information Partners (ESIP)
Winter Meeting 2018
January 9-11th, 2018 – Bethesda, MD

2018 Geosemantics Symposium
Monday January 8, 2018

Overview

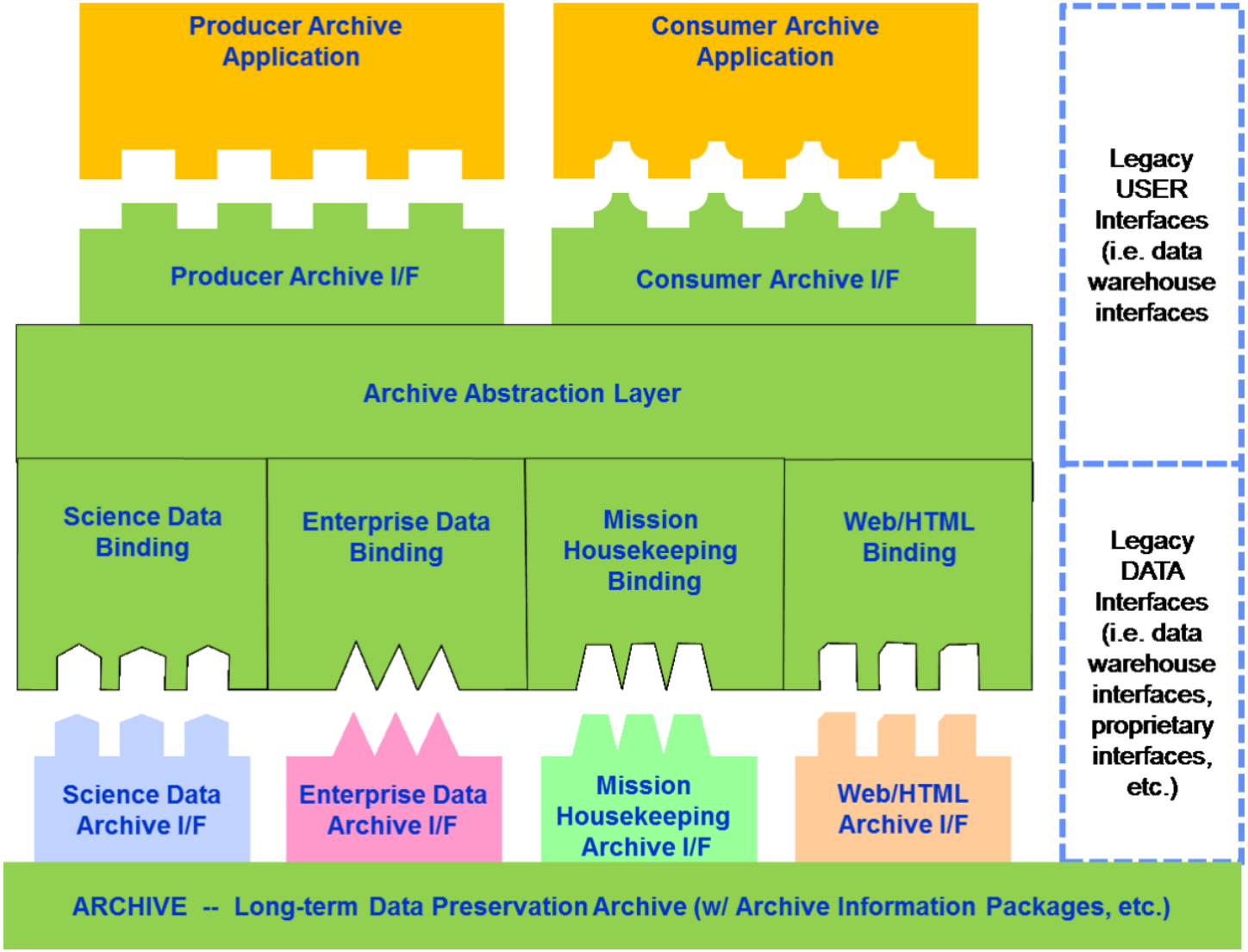
- The Consultative Committee for Space Data Systems (CCSDS) Data Archive Interoperability (DAI) working group has developed the core elements for the underlying processes for digital preservation.
  - *Reference Model for an Open Archival Information System (OAIS) – ISO 14721, CCSDS 650.0-M-2*
  - *Information Preparation to Enable Long Term Use (IPELTU)*

- The DAI working group is now addressing interoperable protocols and interface specifications.
  - *Enable the access, merging and interoperable re-use of the data*
  - *Support for the fundamental scientific technique of checking reproducibility of results.*

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Abstract Protocol Architecture[2]



[2]Data Archive Ingest (DAI) WG Report to the CCSDS Management Council (CMC), Figure 2: Notional Data Archive Architecture, March 2017

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Data Archive Architecture Reference Model

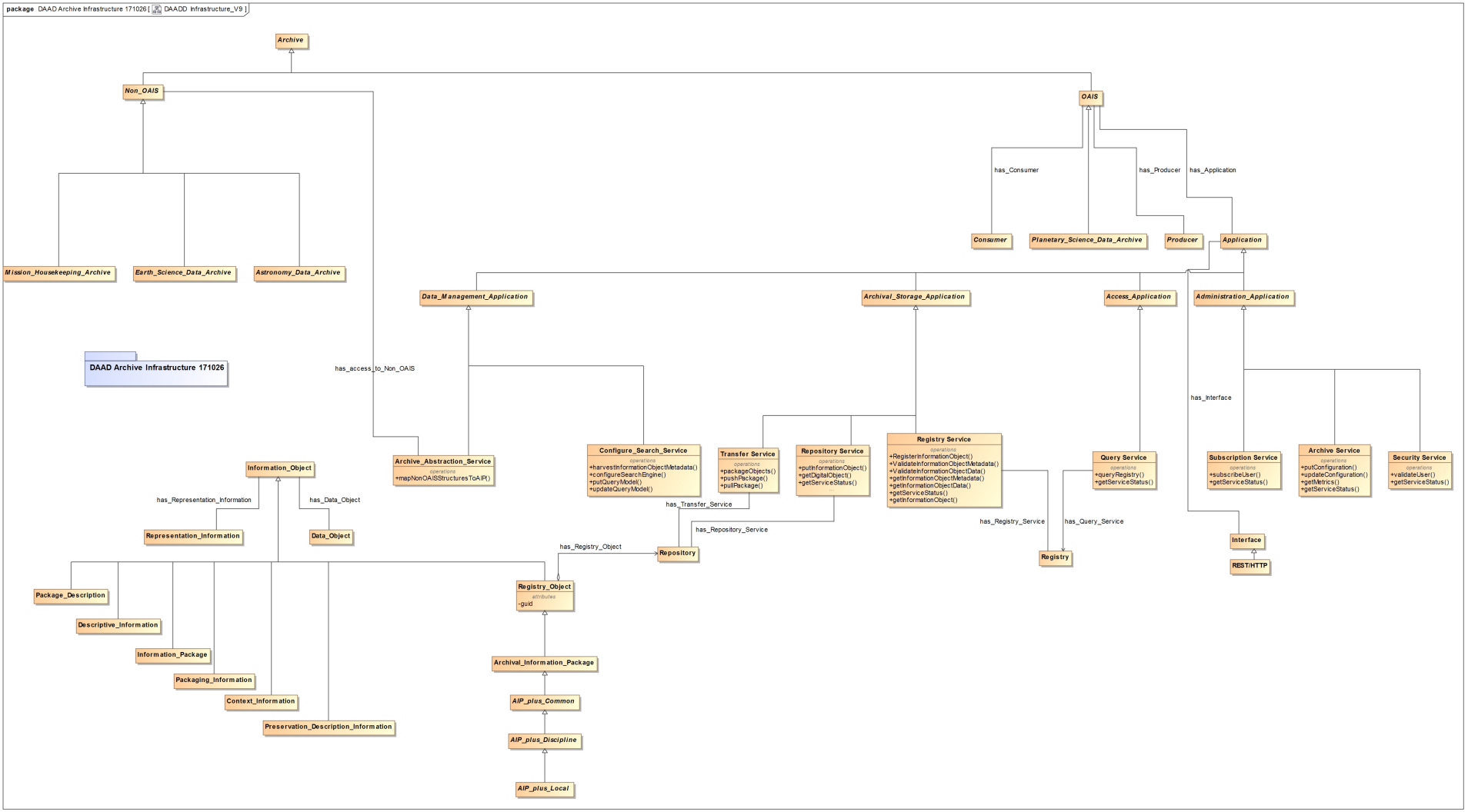
- The Data Archive Architecture Reference Model (DAARM) is an implementable model for trusted digital repositories[1].
  - *Trusted Digital Repository - A repository whose mission is to provide usable, long-term access to digital resources for a designated community.*
  - *The model is an integration of concepts and standards from:*
    - Open Archival Information System (OAIS) Reference Model[1]
    - ISO/IEC 11179 Metadata Registry (MDR) standard
    - Three decades of digital repository development for science research.
  - *The intended scope of the reference model is for digital archives in general.*

[1] ISO 14721:2012 (CCSDSS 650.0-P-1.1) Open archival information system (OAIS) -- Reference model

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Draft UML Model



National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Information Architecture[1]

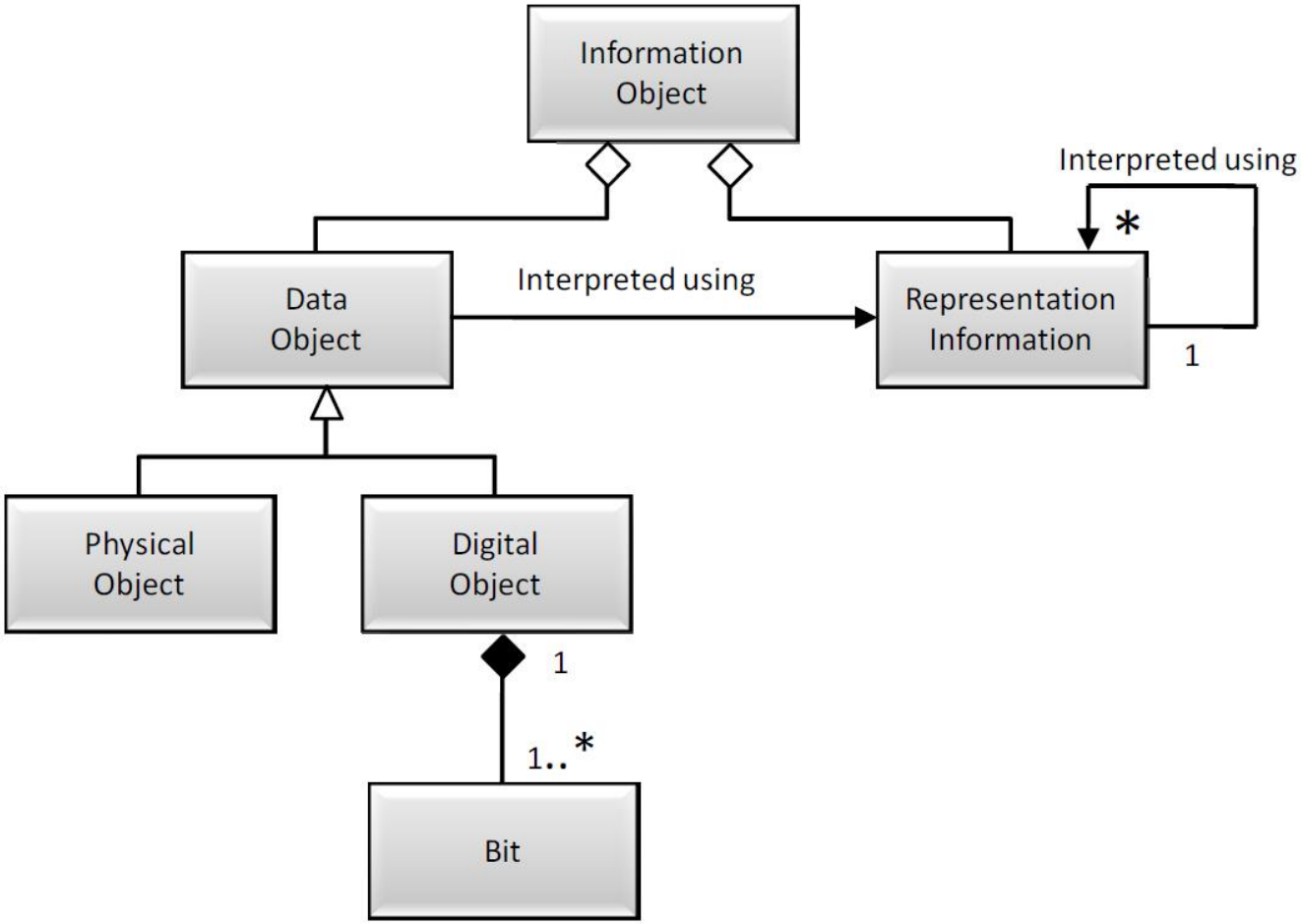


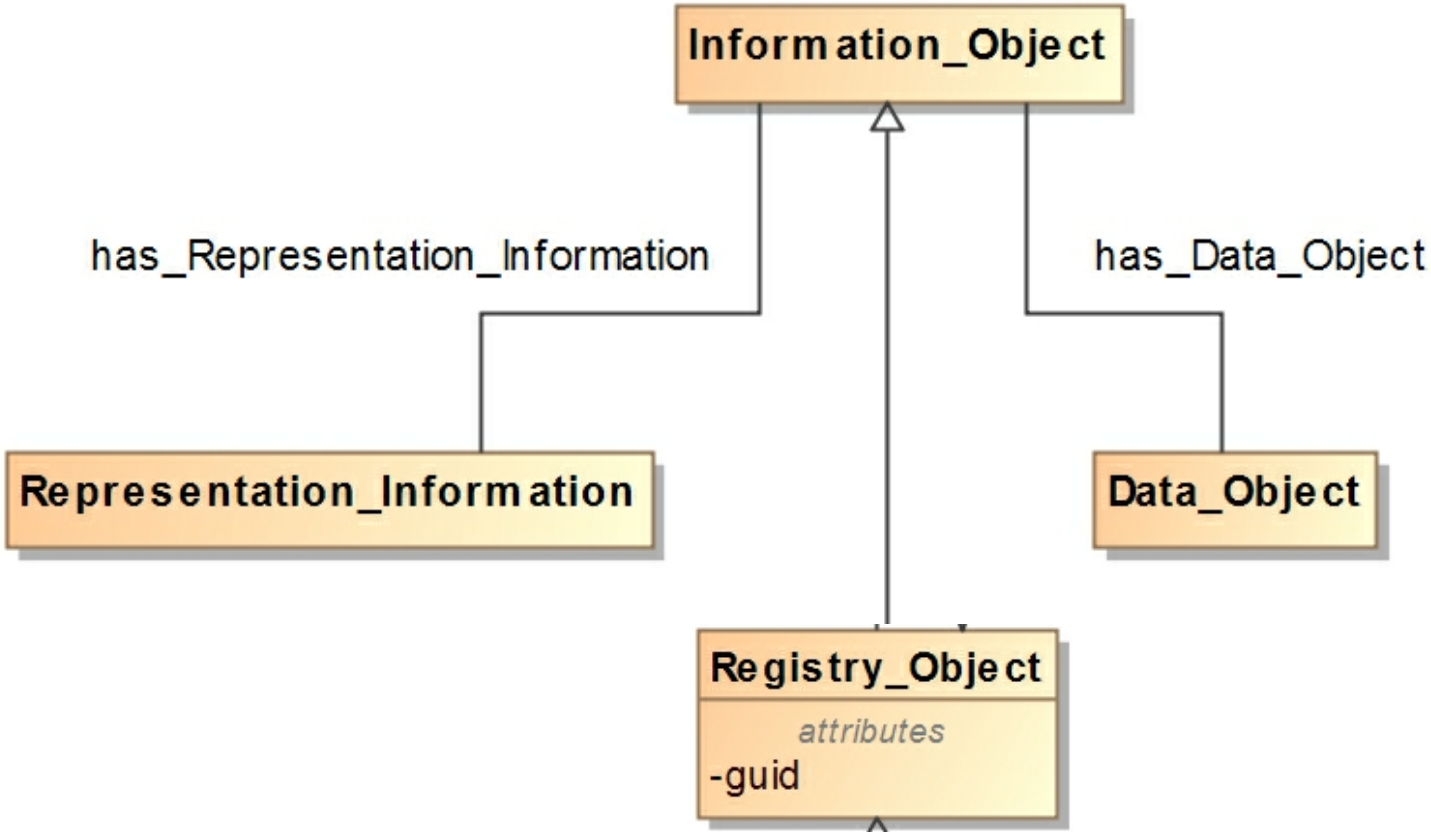
Figure 4-10: Information Object

[1] ISO 14721:2012 (CCSDSS 650.0-P-1.1) Open archival information system (OAIS) -- Reference model

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Registry Object



Information Categories[1]

1. Identification - Provides a unique and immutable identifier for each data object.

2. Representation/Format - Provides meaning for a data object and allows it to be interpreted.

3. Integrity - Ensures the data object has not been unintentionally altered.

4. Provenance - Provides the history of the data object and is essential for authenticity and reproducibility.

5. Context - Describes the environment in which the data object was created.

6. Reference - Allows the data object to be referenced.

7. Access Rights - Defines the access restrictions pertaining to the data object , including the legal framework, licensing terms, and access control

8. Quality* - Provides a scheme for assessing and assigning a quality measure to the data object .

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Functional Entities[1]

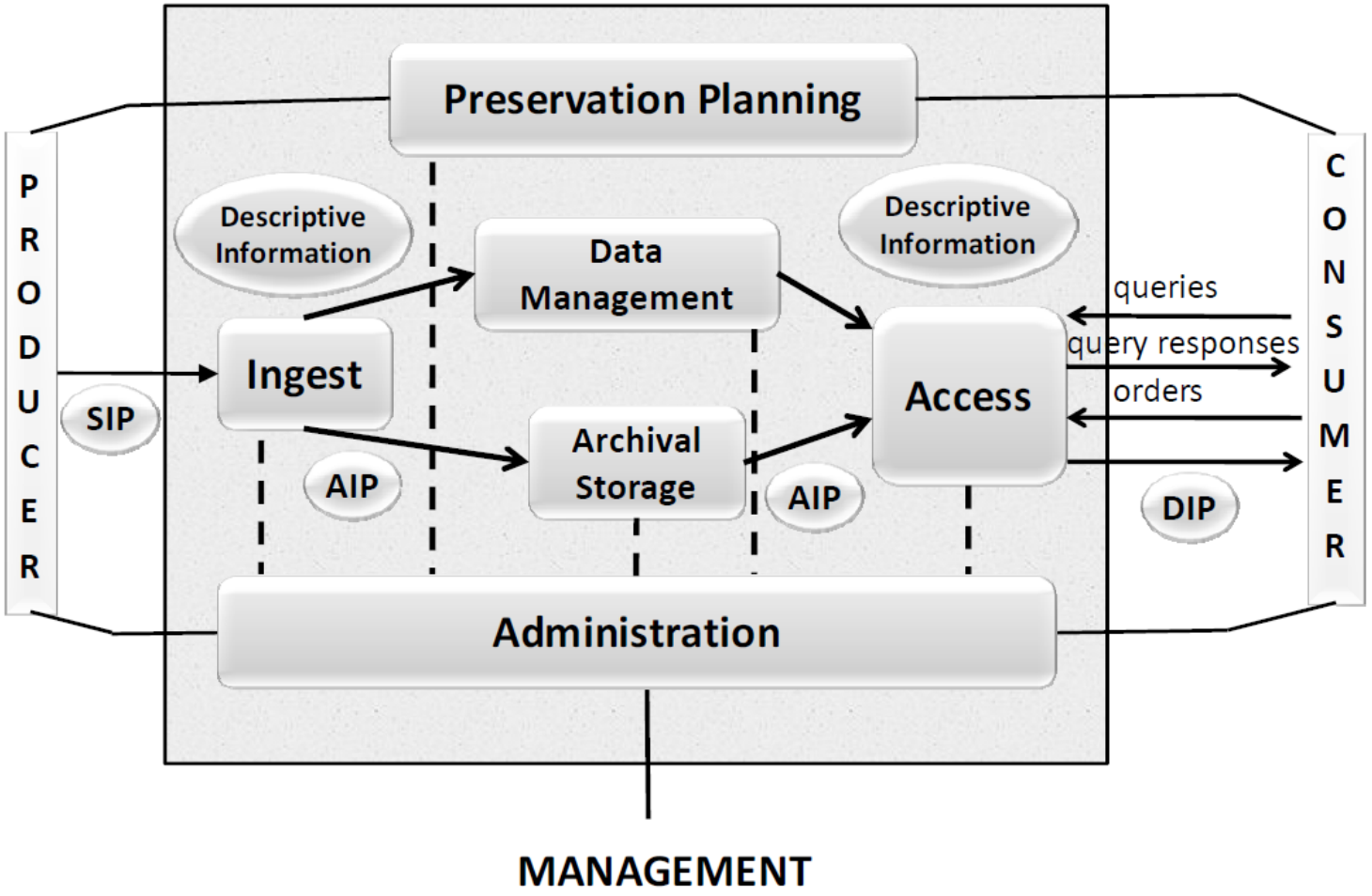


Figure 4-1: OAIS Functional Entities

[1] ISO 14721:2012 (CCSDSS 650.0-P-1.1) Open archival information system (OAIS) -- Reference model

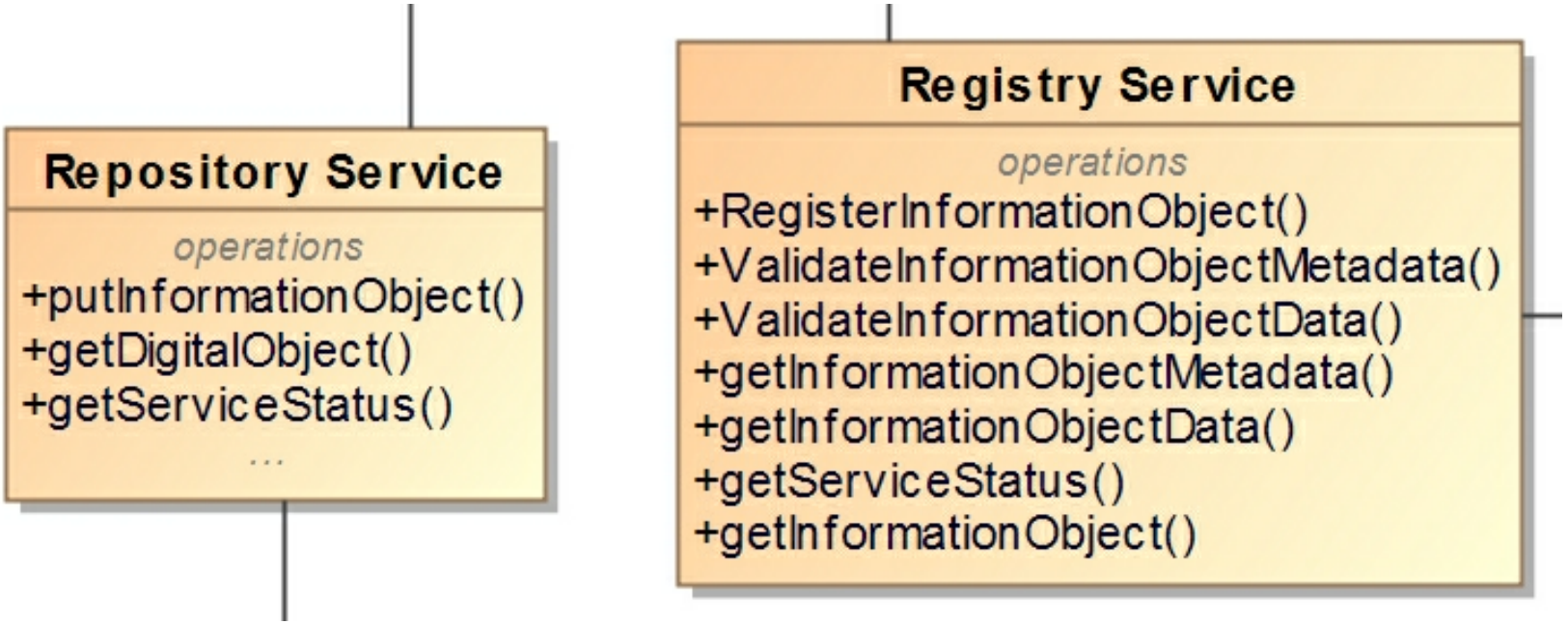
Functional Entities[1]

- Ingest - Accept information objects from producers, prepares them for storage, and ensures that they become established.
- Archival – Store and retrieve Information Objects.
- Data Management – Maintaining administrative information, for example consumer access statistics.
- Access - Make the archival information holdings and related services visible to Consumers.
- Administration - Control the operation of the other functional entities.
- Preservation Planning - Monitoring the environment to ensure that the information stored remains usable by the Designated Community.

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Archival
Store and Retrieve Information Objects



National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Information Model Definitions

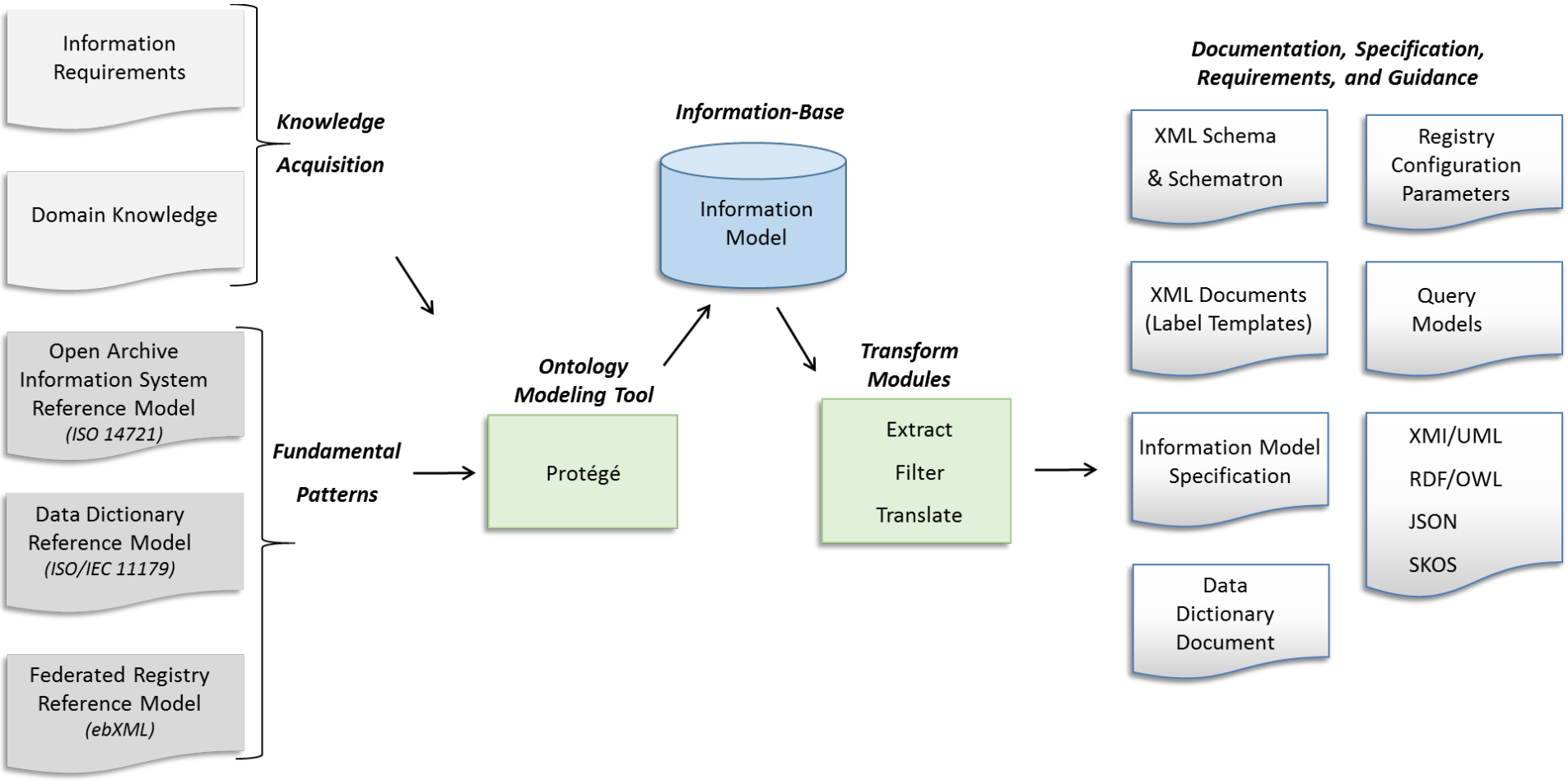
- "An information model is a representation of concepts, relationships, constraints, rules, and operations to specify data semantics for a chosen domain of discourse." [1]
- It provides a sharable, stable, and organized structure of information requirements or knowledge for the domain context.
- Information Modeling is an essential discipline within Data Science

[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.

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Information Model (IM)



National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Information Model Roles

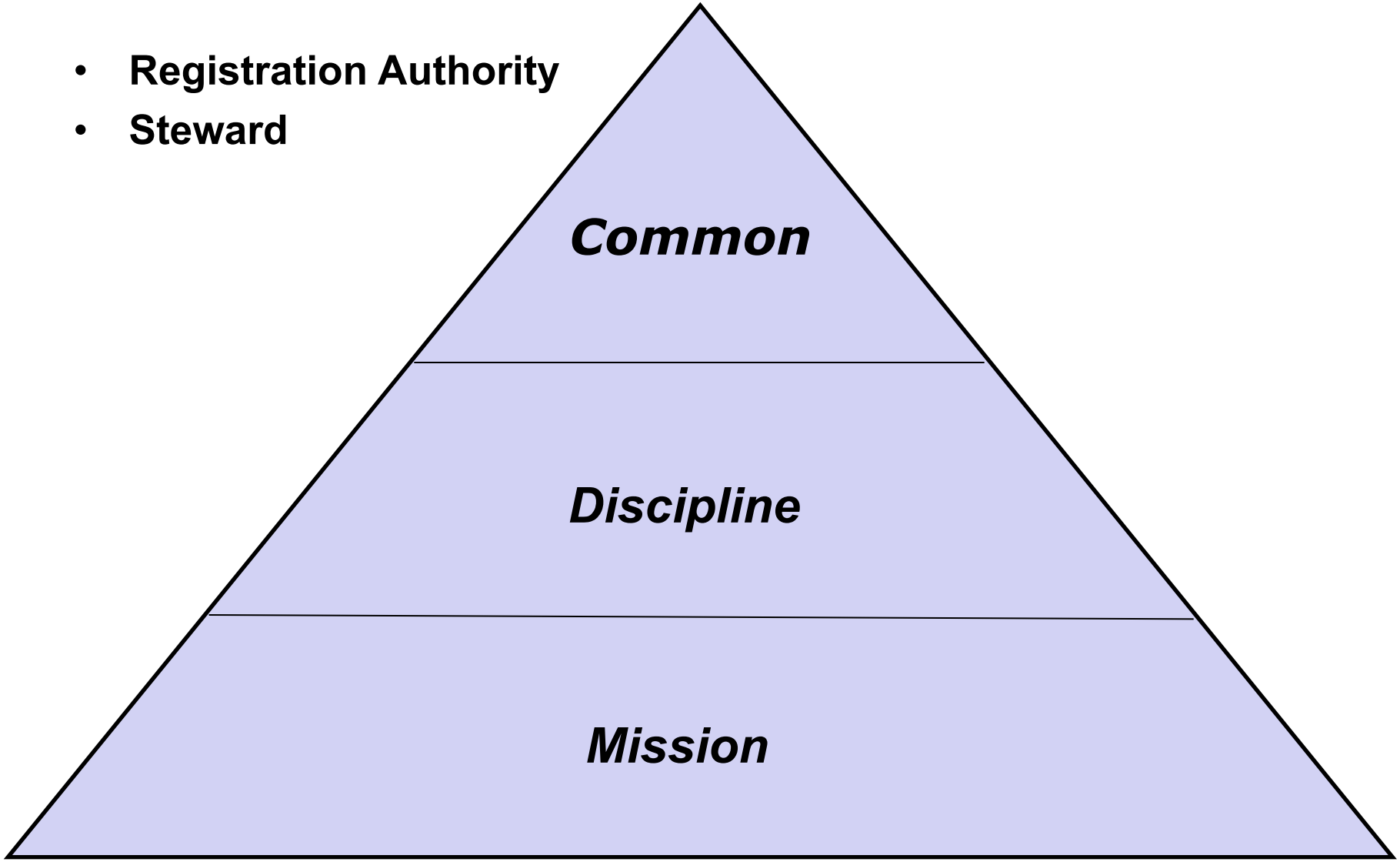
- Requirements: The IM is the primary source for information requirements.
- Governance: A multi-level governance scheme reduces the impact of change as the science community grows and evolves.
- Semantics: The IM provides named relationships to support semantic technologies
- Usability: The IM provides the metadata needed to interpret and use the data.
- Interoperability: The IM is designed by discipline experts to provide interoperability, at multiple levels.
- Configuration: Extracts from the IM are used to configure tools and services

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Multi-level Governance

- Registration Authority
- Steward



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

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

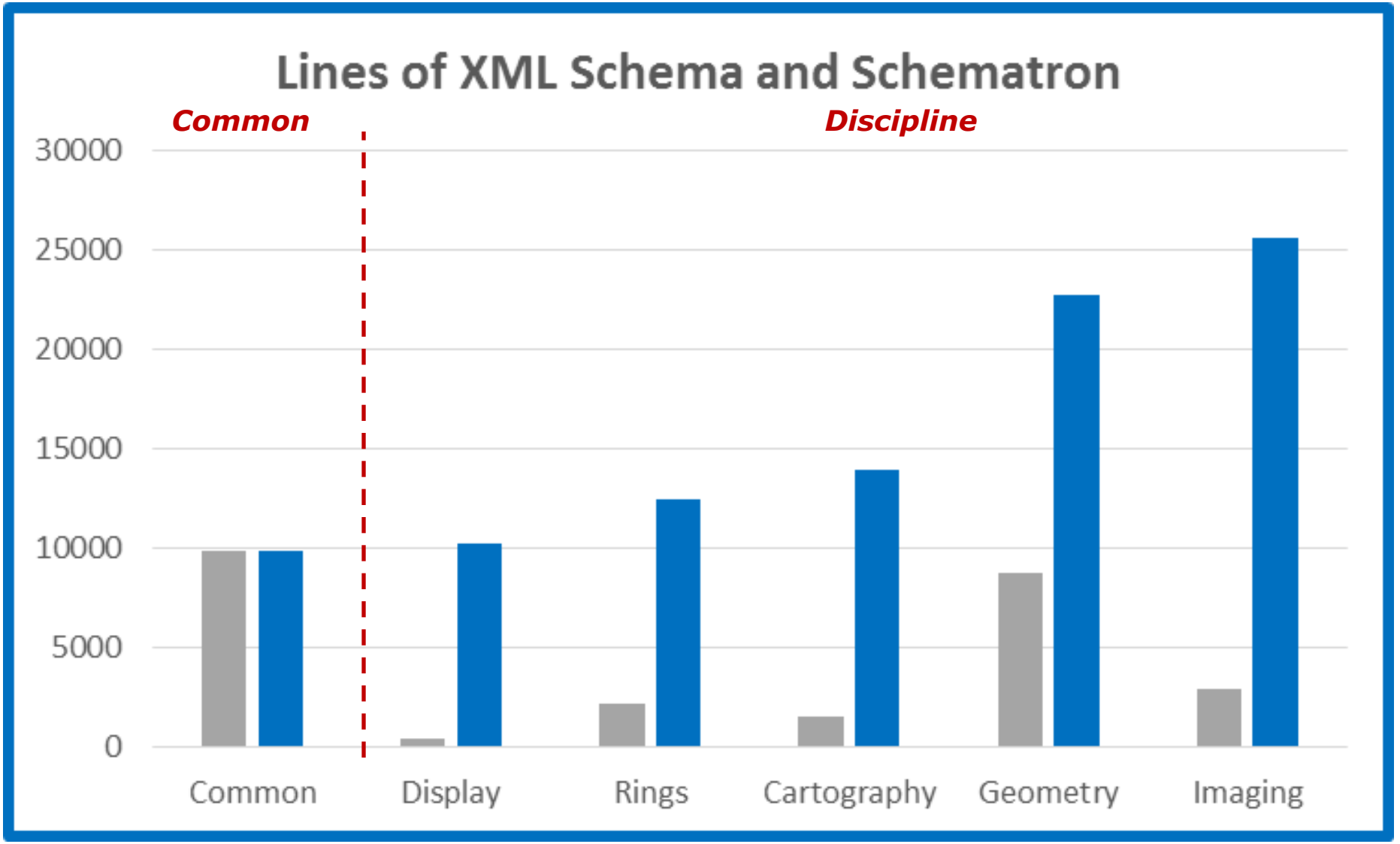
Model Components
Common, Discipline and Mission Dictionaries

Registration Authority	Steward Id	Namespace Id*	XML Schema Namespace	Logical Identifier Prefix	Governance Level	Steward	Oversight
0001_NASA_PDS_1	pds	pds	http://pds.nasa.gov/pds4/pds/v1	urn:nasa:pds:	Common	PDS EN Node*****	CCB
0001_NASA_PDS_1	atm	atm	http://pds.nasa.gov/pds4/atm/v1	urn:nasa:pds:	Discipline	PDS ATM Node	
0001_JAXA_DARTS_1	darts	darts	http://pds.nasa.gov/pds4/darts/v1	urn:jaxa:darts:	Discipline	DARTS (JAXA)	
0001_NASA_PDS_1	en	dph	http://pds.nasa.gov/pds4/dph/v1	urn:nasa:pds:	Discipline	PDS EN Node	
0001_NASA_PDS_1	geo	geo	http://pds.nasa.gov/pds4/geo/v1	urn:nasa:pds:	Discipline	PDS GEO Node	
0001_NASA_PDS_1	geo	geom	http://pds.nasa.gov/pds4/geom/v1	urn:nasa:pds:	Discipline	PDS GEO Node	
0001_NASA_PDS_1	img	cart	http://pds.nasa.gov/pds4/cart/v1	urn:nasa:pds:	Discipline	PDS IMG Node	
0001_NASA_PDS_1	img	disp	http://pds.nasa.gov/pds4/disp/v1	urn:nasa:pds:	Discipline	PDS IMG Node	
0001_NASA_PDS_1	img	img	http://pds.nasa.gov/pds4/img/v1	urn:nasa:pds:	Discipline	PDS IMG Node	
0001_NASA_PDS_1	naif	naif	http://pds.nasa.gov/pds4/naif/v1	urn:nasa:pds:	Discipline	PDS NAIF Node	
0001_NASA_PDS_1	ops	pds	http://pds.nasa.gov/pds4/pds/v1	urn:nasa:pds:	Discipline	PDS EN Node	
0001_NASA_PDS_1	ppi	alt	http://pds.nasa.gov/pds4/alt/v1	urn:nasa:pds:	Discipline	PDS PPI Node	
0001_NASA_PDS_1	ppi	particle	http://pds.nasa.gov/pds4/particle/v1	urn:nasa:pds:	Discipline	PDS PPI Node	
0001_NASA_PDS_1	ppi	ppi	http://pds.nasa.gov/pds4/ppi/v1	urn:nasa:pds:	Discipline	PDS PPI Node	
0001_NASA_PDS_1	ppi	wave	http://pds.nasa.gov/pds4/wave/v1	urn:nasa:pds:	Discipline	PDS PPI Node	
0001_ESA_PSA_1	psa	psa	http://psa.esa.int/psa/v1	urn:psa:esa:	Discipline	ESA PSA	
0001_NASA_PDS_1	rings	rings	http://pds.nasa.gov/pds4/rings/v1	urn:nasa:pds:	Discipline	PDS Rings Node	
0001_NASA_PDS_1	rs	rs	http://pds.nasa.gov/pds4/rs/v1	urn:nasa:pds:	Discipline	PDS RS Node	
0001_ROS_RSSA_1	rssa	rssa	http://pds.nasa.gov/pds4/rssa/v1	urn:ros:rssa:	Discipline	RSSA (IKI)	
0001_NASA_PDS_1	sbn	sbn	http://pds.nasa.gov/pds4/sbn/v1	urn:nasa:pds:	Discipline	PDS SBN	
0001_NASA_PDS_1	sbn	sp	http://pds.nasa.gov/pds4/sp/v1	urn:nasa:pds:	Discipline	PDS SBN	
0001_NASA_PDS_1	atm	ladee	http://pds.nasa.gov/pds4/mission/ladee/v1	urn:nasa:pds:	Mission	PDS ATM Node	
0001_NASA_PDS_1	atm	ladee	http://pds.nasa.gov/pds4/ladee/v1	urn:nasa:pds:	Mission	PDS ATM Node	
0001_NASA_PDS_1	geo	insight	http://pds.nasa.gov/pds4/mission/insight/v1	urn:nasa:pds:	Mission	PDS GEO Node	
0001_NASA_PDS_1	img	mgs	http://pds.nasa.gov/pds4/mission/mgs/v1	urn:nasa:pds:	Mission	PDS IMG Node	
0001_NASA_PDS_1	img	mpf	http://pds.nasa.gov/pds4/mission/mpf/v1	urn:nasa:pds:	Mission	PDS IMG Node	
0001_NASA_PDS_1	sbn	orex	http://pds.nasa.gov/pds4/mission/orex/v1	urn:nasa:pds:	Mission	PDS SBN	
0001_NASA_PDS_1	ppi	mvn	http://pds.nasa.gov/pds4/mission/mvn/v1	urn:nasa:pds:	Mission	PDS PPI Node	
0001_NASA_PDS_1	ppi	mvn	http://pds.nasa.gov/pds4/mvn/v1	urn:nasa:pds:	Mission	PDS PPI Node	
0001_NASA_PDS_1	sbn	bopps	http://pds.nasa.gov/pds4/mission/bopps/v1	urn:nasa:pds:	Mission	PDS SBN	

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Information Requirements Generated from the Dictionaries



National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Usability

- A desk assessment of PDS4 against ISO 16363[1], the instrument for assessing a repository against the OAIS Reference Model[2], found that 92% of the metrics of the ISO 16363 standard were satisfied
    - Governance and Organizational Viability
    - Digital Object Management
    - Infrastructure and Security Risk Management.
- *Maintain the value of the data over time*

[1] ISO 16363:2012 (CCSDS 652.0-R-1) Audit and certification of trustworthy digital repositories
[2] ISO 14721:2012 (CCSDSS 650.0-P-1.1) Open archival information system (OAIS) -- Reference model

Next Steps

- **CCSDS Data Archive Interoperability (DAI) Working Group**
  - *Write and review CCSDS Blue Book*
    - CCSDS Fall 2018 Technical Meetings (Spring and Fall)
    - Develop two working prototypes
    - Reference Model Review
      - *JPL's Center for Data Science and Technology – D. Crichton*
      - *NASA Planetary Data System (PDS) System Development – S. Hardman*
      - *CCSDS Systems Architecture (SAWG) Chair – P. Shames*
      - *JPL's Multimission Ground System and Services (MGSS) Project – C. Radulescu*
      - *Life Storage of Mission Data (LSMD) task – M. McAuley*
      - *FernUniversität in Hagen - M. Hemmje*
      - *Engineering Data Management (EDM) task – L. Jewell*
      - *Information Retrieval and Data Science Group – C. Mattmann*

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Acknowledgements

- CCSDS Data Archive Interoperability (DAI) Working Group
  - Bruce Ambacher
  - Robert Downs
  - John Garrett
  - David Giaretta
  - Matthias Hemmje
  - Mike Kearney
  - Terry Longstreth
  - Don Sawyer

- JPL's Center for Data Science and Technology – Dan Crichton
- NASA Planetary Data System (PDS) – Sean Hardman, Ronald Joyner
- JPL Principal Data Scientist, USC Adjunct Associate Professor – Chris Mattmann
- JPL's Multimission Ground System and Services (*MGSS*) Project – Costin Radulescu
- CCSDS Systems Architecture (SAWG) Chair – Peter Shames
- Life Storage of Mission Data (LSMD) task – Mike McAuley
- Engineering Data Management (EDM) task – Laura Jewell

This research was carried out by the Jet Propulsion Laboratory, managed by the California Institute of Technology under a contract with the National Aeronautics and Space Administration.

References

- *Reference Model for an Open Archival Information System (OAIS), ISO 14721:2012.*
- *Data Archive Ingest (DAI) WG Report to the CCSDS Management Council (CMC), Figure 2: Notional Data Archive Architecture, March 2017*
- *DAI Architecture Analysis, SEA System Architecture WG, Slide 16, Alternative Standardized Archive System Architecture Deployment Option (3), May 2017*
- *Planetary Data System PDS4 Information Model Specification, Version 1.8.0.0, March 2017.*
- *Planetary Data System - PDS4 System Architecture Specification September 1, 2013, Version 1.3.*
- *CCSDS Reference Architecture for Space Information Management (RASIM) CCSDS 311.0-M-1*
- *The Semantic Planetary Data System, PV2005, Edinburgh, November 2005.*
- *PDS-D – The Planetary Data System Distribution Subsystem. Lunar and Planetary Science XXXIV (2003)*
- *The Planetary Data System - Distributed Inventory System, IEEE Forum on Research and Technology Advances in Digital Libraries, 1999. Proceedings.*

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Thank You

Questions and Answers

PDS homepage: https://pds.nasa.gov/

Acknowledgements - This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Status Continued

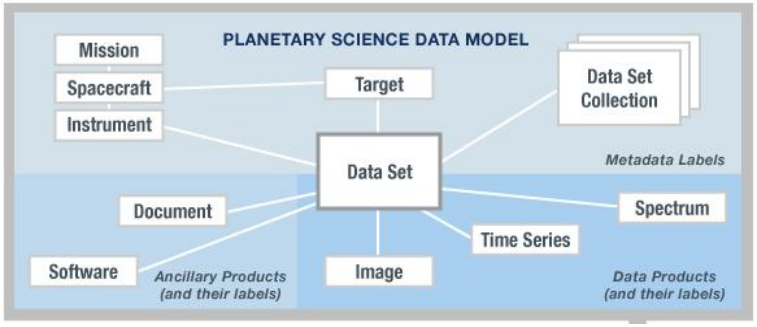
- **Utilize the Cornerstone Framework (NPO-49832) for model capture and management.**
  - *Cornerstone is the framework used to capture and manage the PDS4 Information Model.*
  - *Provides a framework for model-driven information system development*
  - *Maintains Information Model independence.*

National Aeronautics and Space Administration

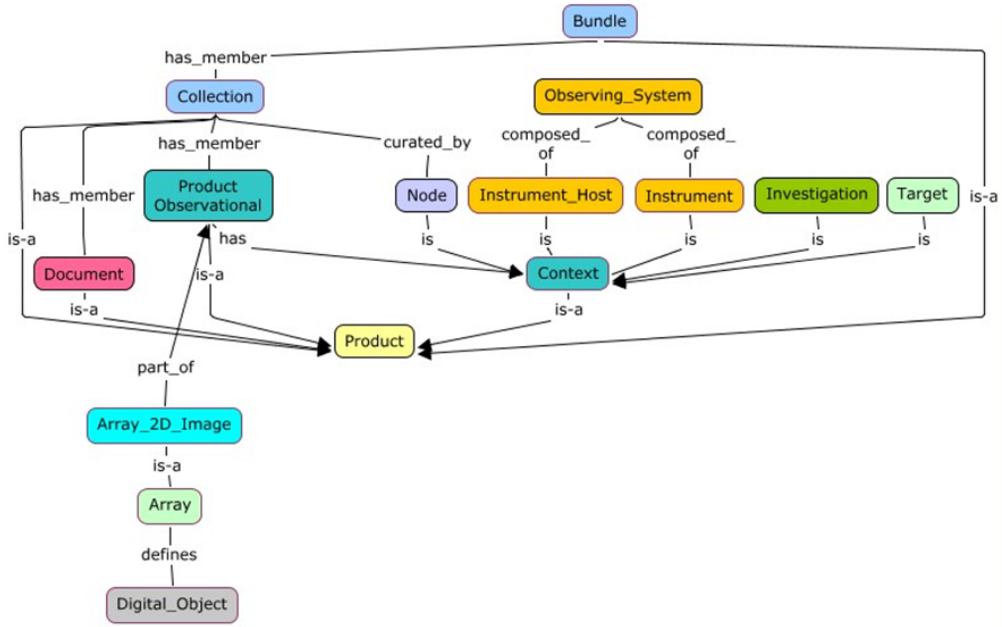
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

View Points

Community's View



Information Modeler's View



Repository View

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>
```

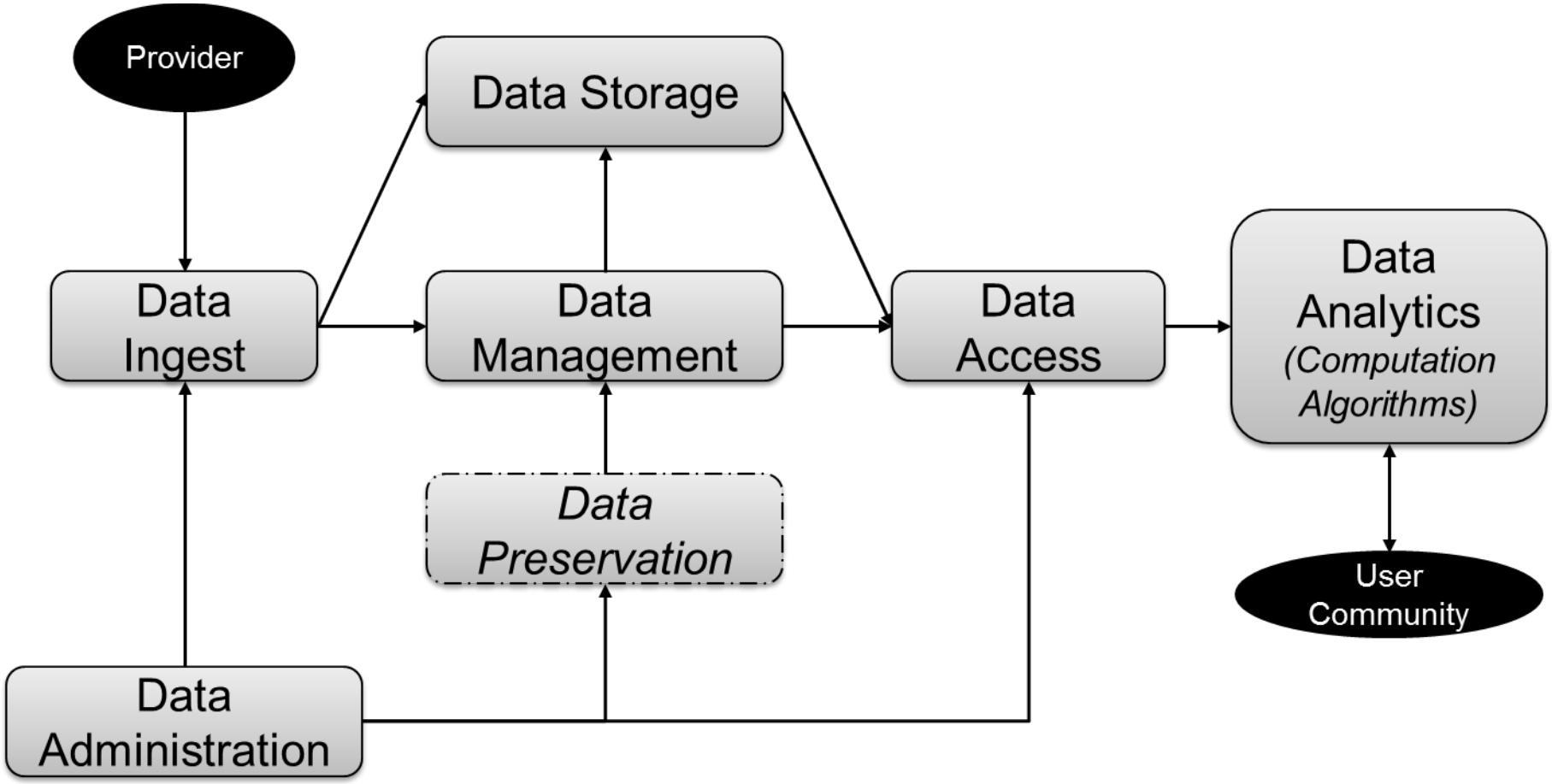
↓ **Describes**

Data Object

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Conceptual Architecture[1]



[1] ISO 14721:2012 (CCSDSS 650.0-P-1.1) Open archival information system (OAIS) -- Reference model

Semantics

- All registry objects are first class products.
  - *All products have a Persistent Identifier (PID)*
  - *Named relationships are used to relate objects (semantic)*
    - data, documents, people, software, and contextual objects
  - *Supports Linked Open Data.*