Data Preservation and Curation for the Planetary Science Community

AGU 2013
IN11D-04

J.S. Hughes, D.J. Crichton,
R. Joyner, S. Hardman, E. Rye

Topics

- PDS Mission and Vision Statement
- PDS4: The Next Generation PDS
- PDS4 Architecture
  - End-To-End System
  - Information Model
- Conclusion

PDS Mission and Vision

Mission

Facilitate achievement of NASA's planetary science goals by efficiently collecting, archiving, and making accessible digital data and documentation produced by or relevant to NASA's planetary missions, research programs, and data analysis programs.

Vision

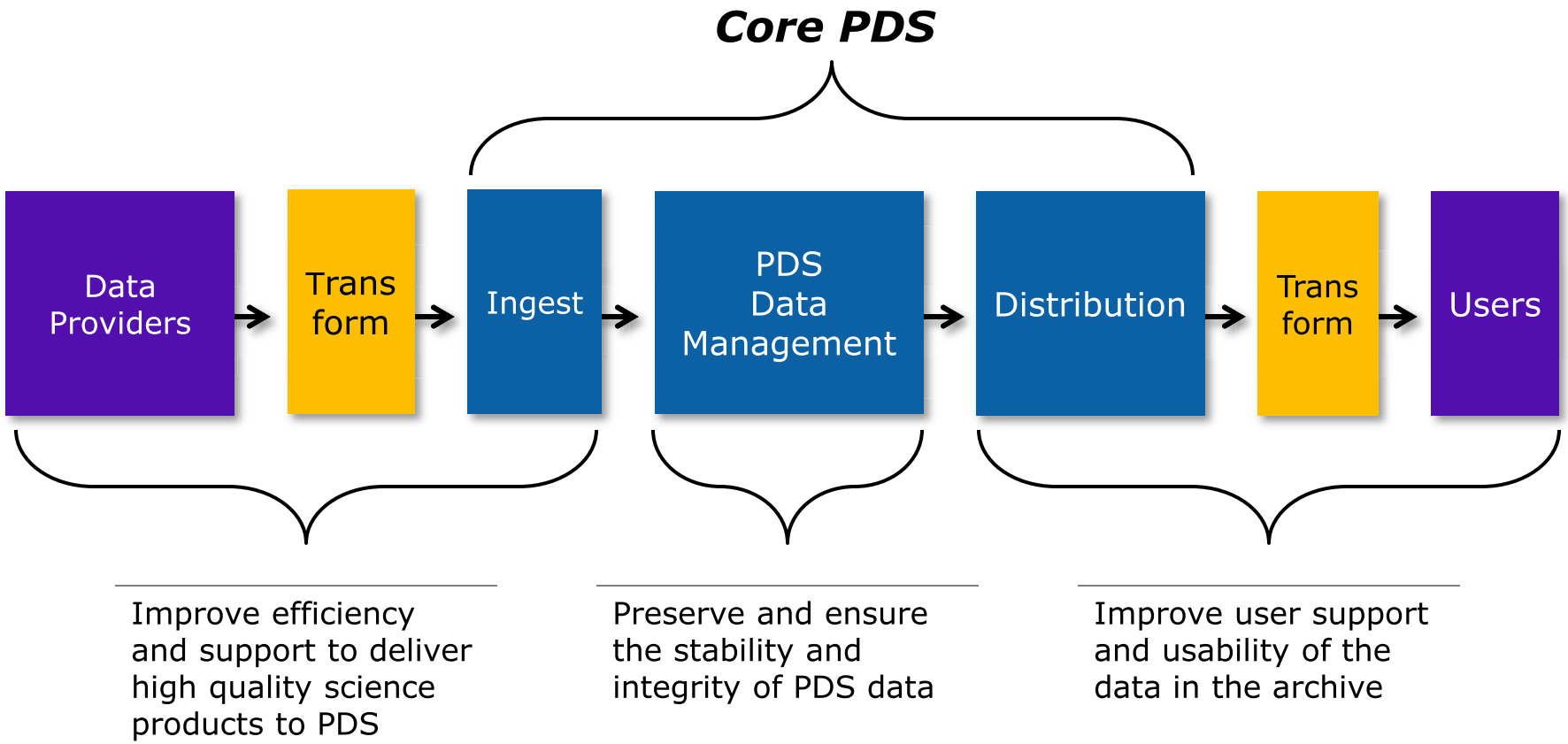
- To gather and preserve the data obtained from exploration of the Solar System by the U.S.
- To facilitate new and exciting discoveries by providing access to and ensuring usability of those data to the worldwide community
- To inspire the public through availability and distribution of the body of knowledge reflected in the PDS data collection

PDS is a federation of distributed discipline and service nodes.

PDS4: The Next Generation PDS

- PDS4 is a PDS-wide project to upgrade from a mature but dated PDS Version 3 to PDS Version 4 a data archive system using modern information technologies.
- An **explicit information architecture**
  - All PDS data tied to a common model to improve validation and discovery
  - Use of XML, a well-supported international standard, for data product labeling, validation, and searching.
  - A hierarchy of data dictionaries built to the ISO 11179 standard, designed to increase flexibility, enable complex searches, and make it easier to share data internationally.
- An **explicit software/technical architecture**
  - Distributed services both within PDS and at international partners
  - Consistent protocols for access to the data and services
  - Deployment of an open source registry infrastructure to track and manage every product in PDS
  - A distributed search infrastructure

An End-to-End System for Data Preservation and Curation



An Information Model for Data Preservation and Curation

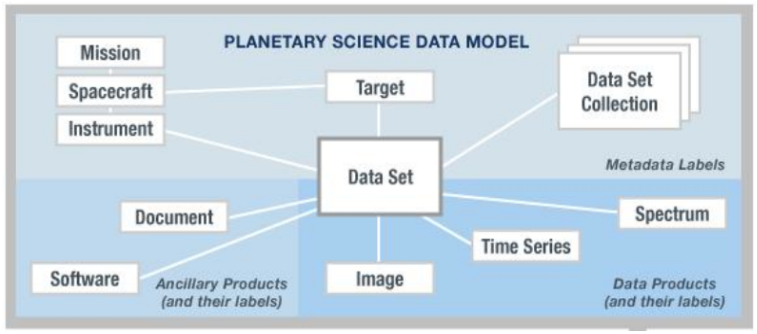
- An information model is a representation of concepts and the relationships, constraints, rules, and operations to specify data semantics for a chosen domain of discourse [1].

- The PDS4 information model consists of:
  - Concepts adopted from several information technology standards.
  - Knowledge acquired from science experts in the planetary sciences.

[1] (2013) The Wikipedia website. [Online]. Available: http://en.m.wikipedia.org/wiki/Information_model.

The Information Model in Context

Information Model



Used to Create

Validates

Label Schema

Expressed As

Extracted/Specialized

Data Element

has

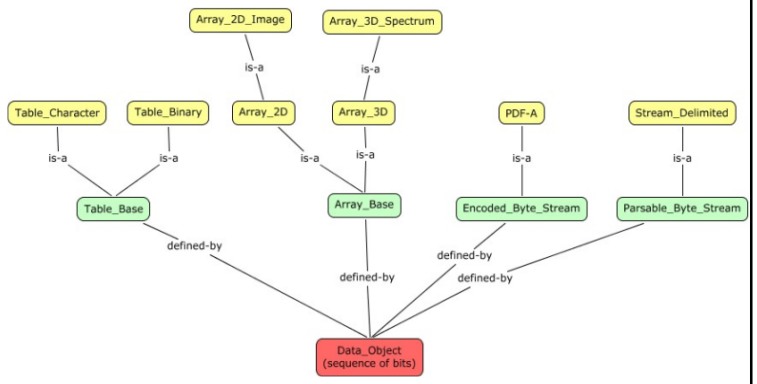
Class

Planetary Science Data Dictionary

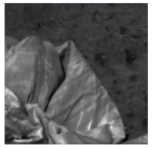
Product

Tagged Data Object

(Information Object)



Describes



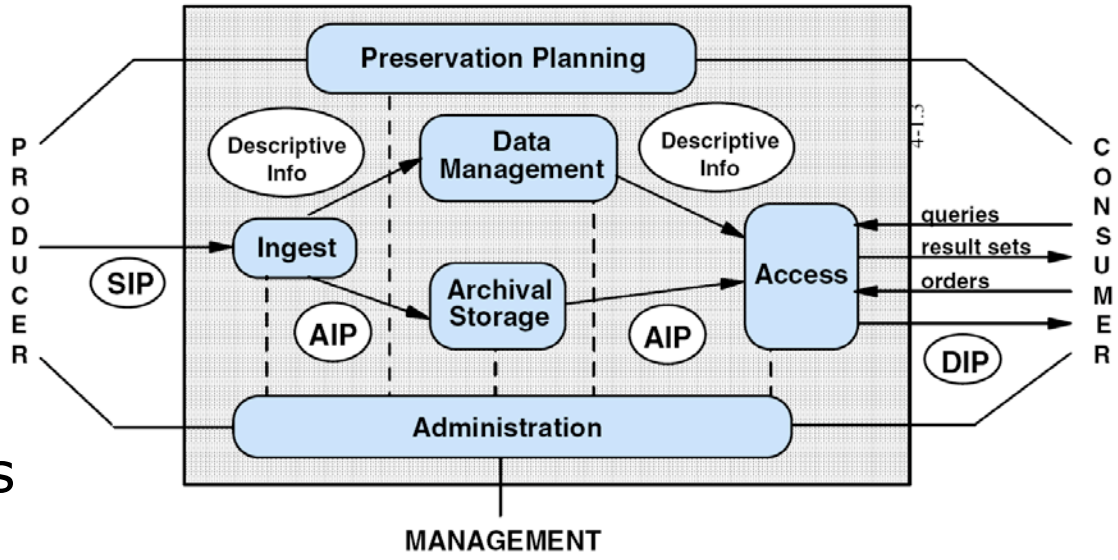
Data Object

Information Technology Standards

- ISO 14721:2003 - Open Archival Information System (OAIS) reference model
  - Defines key metadata components of an archival information system including the information object, representation information, and data object.
- ISO/IEC 11179 Metadata Registry (MDR)
  - Provides a comprehensive schema for the community's data dictionary.
- OASIS/ebXML Registry Information Model V3.0.
  - Provides a federated registry/repository model for registry object's identification, versioning, and tracking.

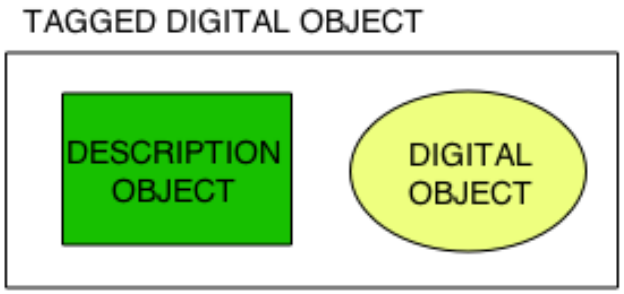
The PDS4 End-to-End System Mapped to OAIS

- Ingestion
- Data Management
- Storage Management
- Administration
- Preservation Planning
- Distribution/Access

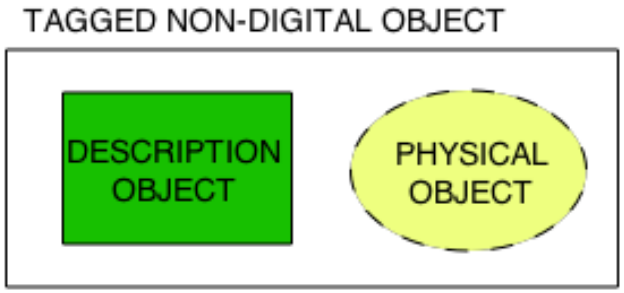


Reference Model for Open Archive Information System, CCSDS 650.0-B-1, January 2002

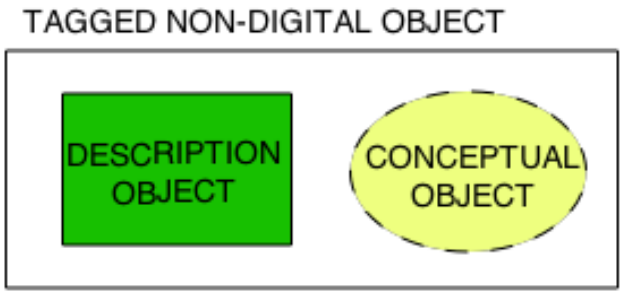
Information Object Model[1]



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



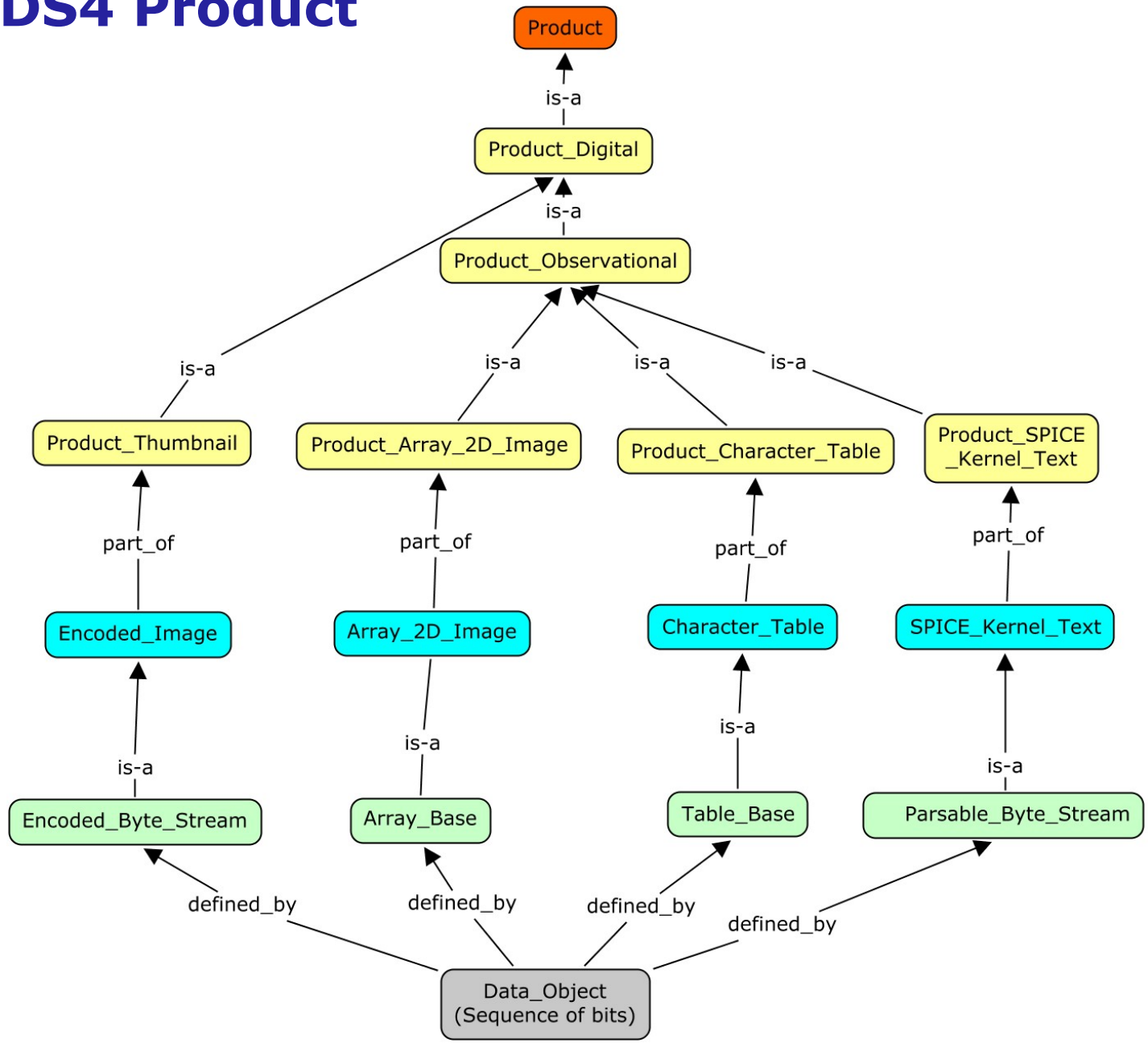
- **physical object**: An object which is physical or tangible – for example the planet Saturn and the Venus Express magnetometer.

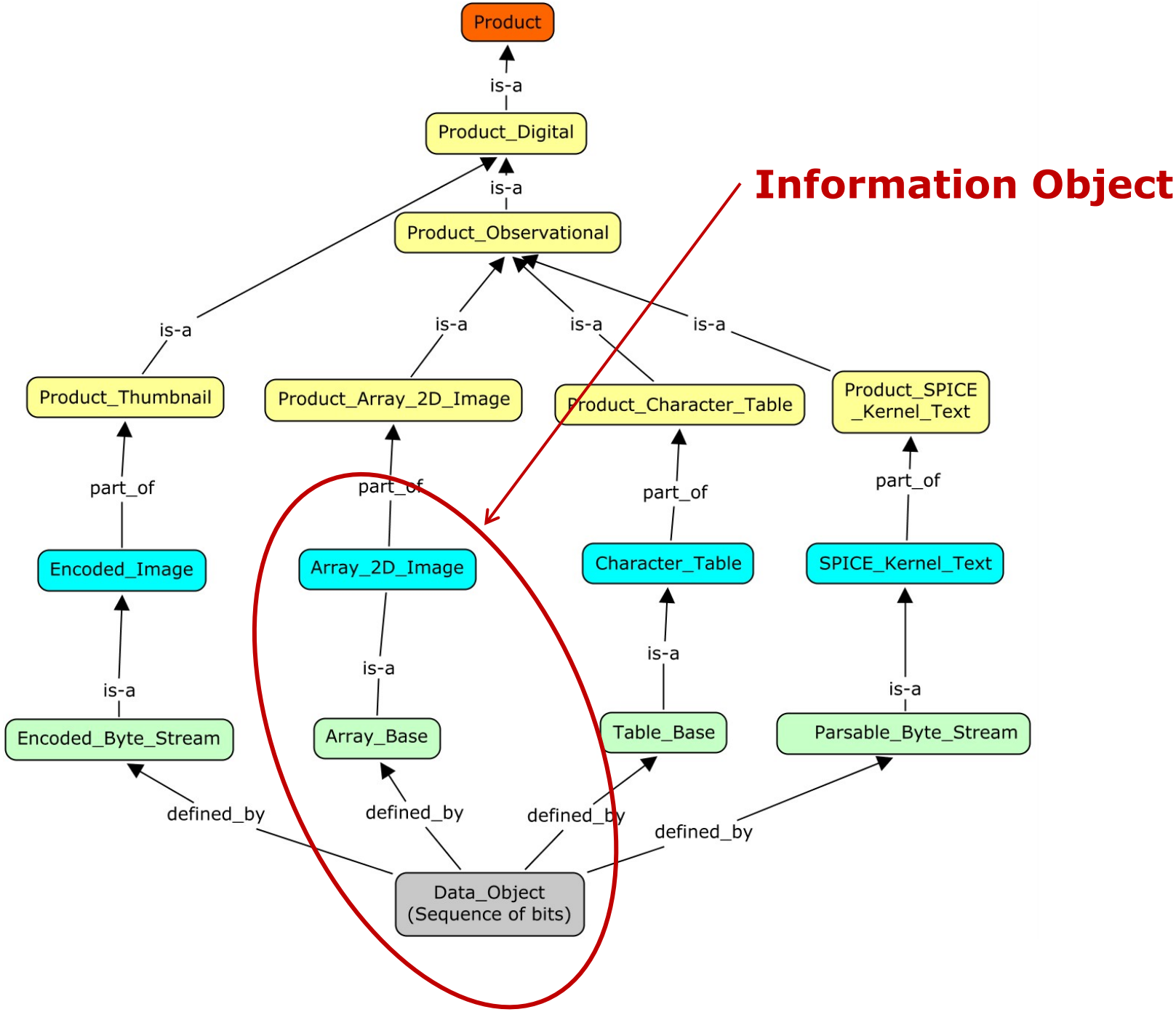


- **conceptual object**: An object which is intangible – for example the Cassini mission and NASA's strategic plan for solar system exploration.

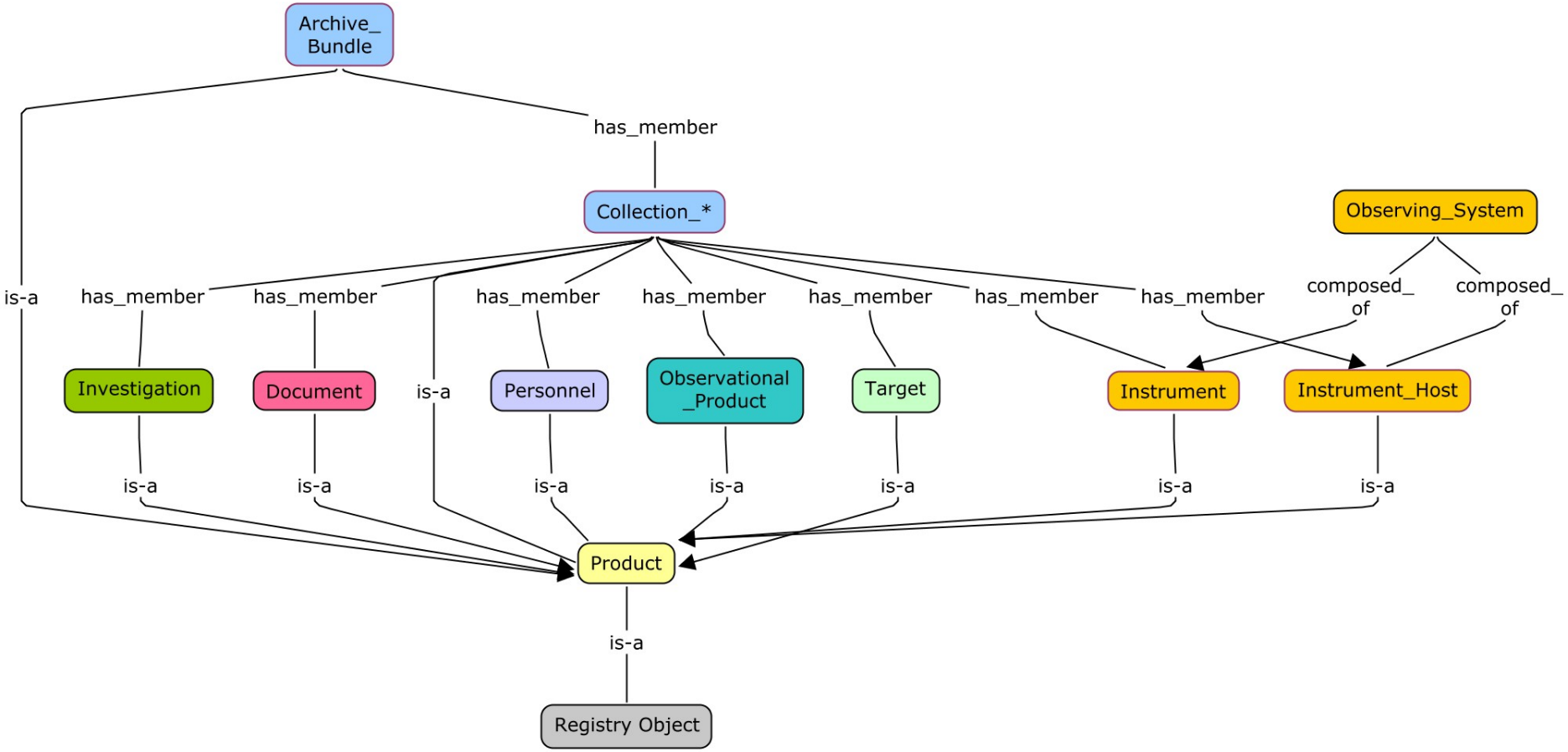
[1]Open Archival Information System (OAIS) Reference Model - ISO 14721:2003

PDS4 Product




Information Object
Product
is-a
Product_Digital
is-a
Product_Observational
is-a
is-a
is-a
is-a
Product_Thumbnail
Product_Array_2D_Image
Product_Character_Table
Product_SPICE
_Kernel_Text
part_of
part_of
part_of
part_of
Encoded_Image
Array_2D_Image
Character_Table
SPICE_Kernel_Text
is-a
is-a
is-a
is-a
Encoded_Byte_Stream
Array_Base
Table_Base
Parsable_Byte_Stream
defined_by
defined_by
defined_by
defined_by
Data_Object
(Sequence of bits)

PDS4 Information Model



Data Dictionary Attributes
ISO/IEC 11179

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

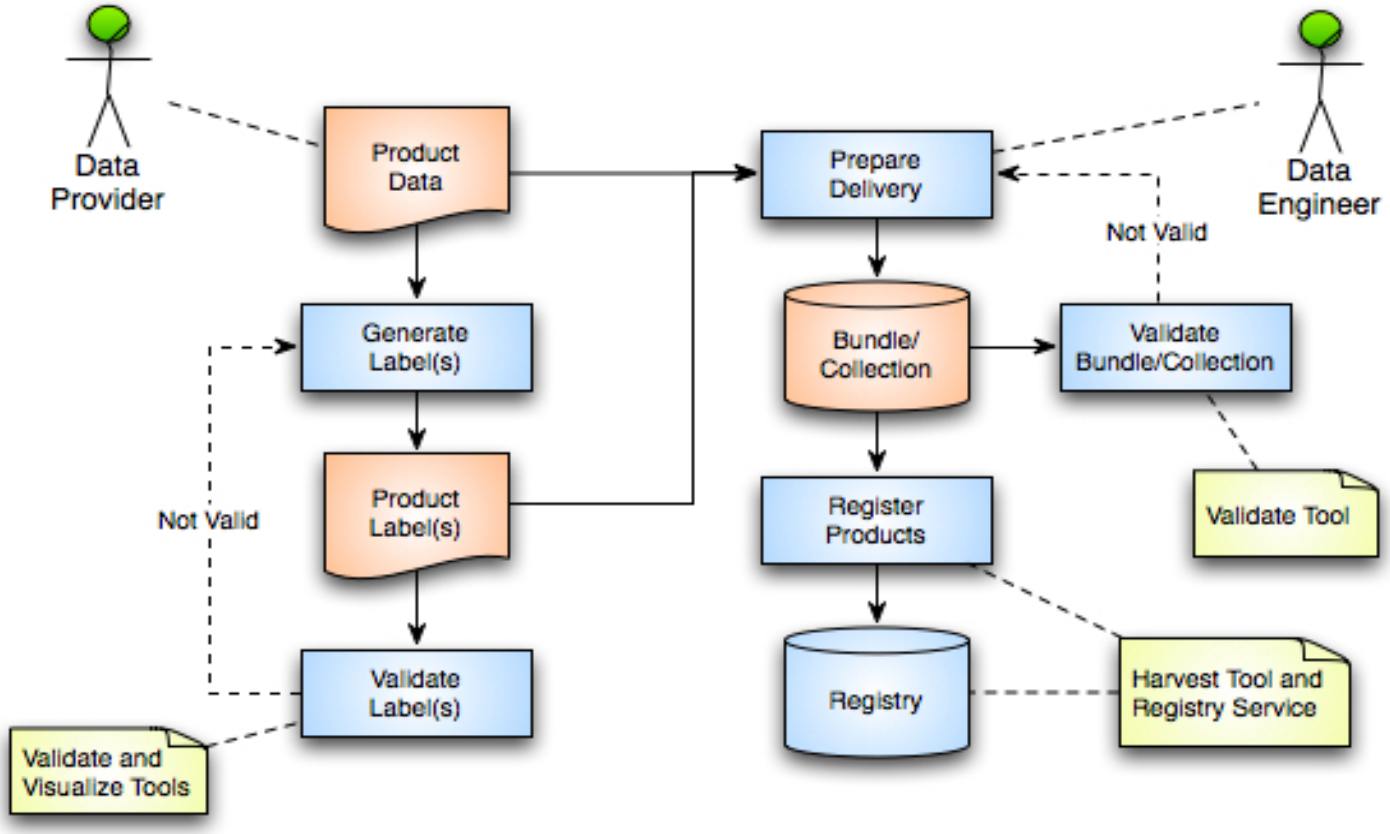
Federated Registry Model

- An ebXML Registry provides services for sharing content and metadata between cooperating registries in a federated environment.
  - Each cooperating registry appears and acts as a single virtual registry/repository within the federated model.

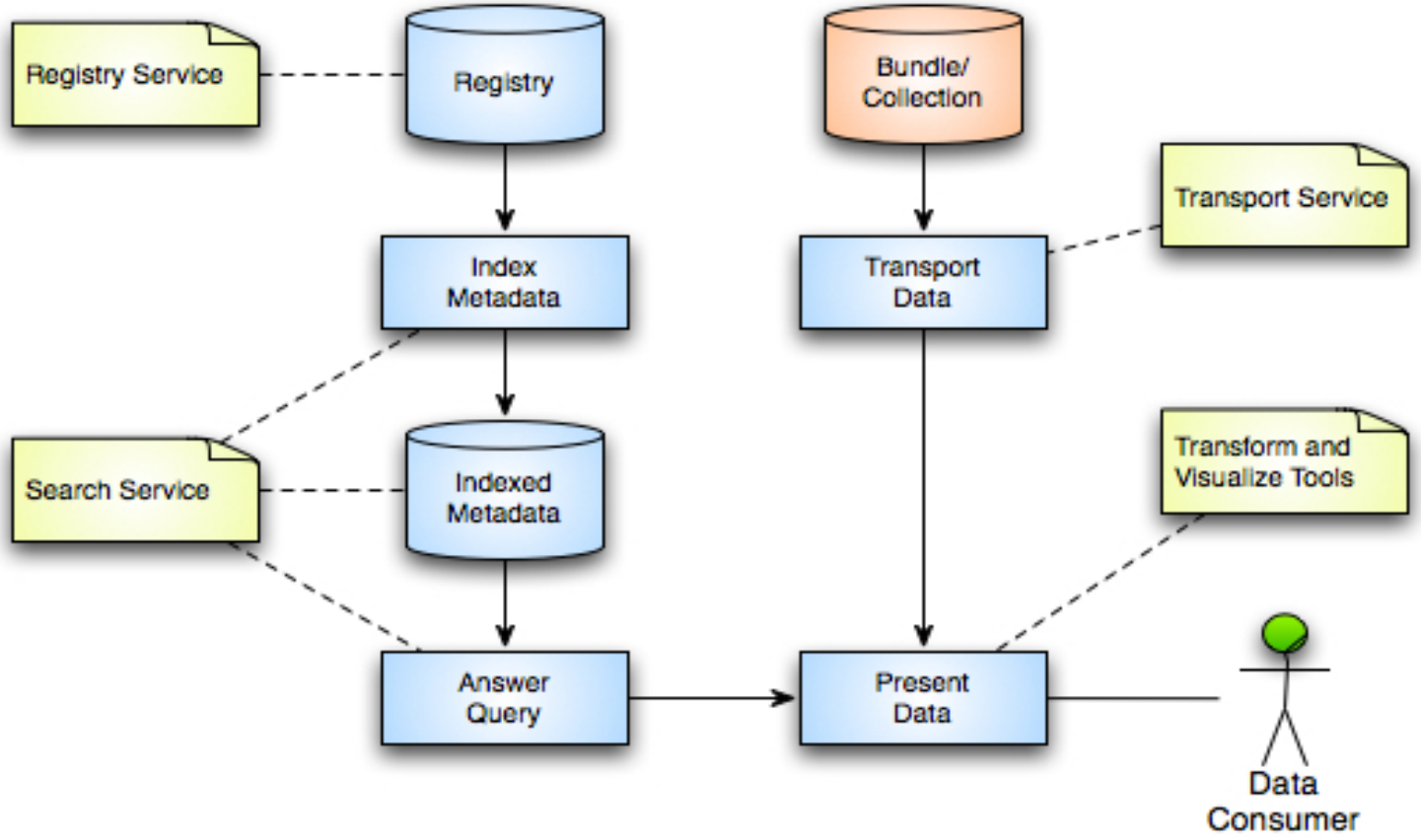
- The benefits of the federated registry approach are evident in seamless information integration and sharing while preserving local autonomy over data (e.g., federated search seamlessly returns results from multiple stores).

- Manages unique and immutable identification, versioning and tracking of products.

Operations Concept
Product Lifecycle (Ingestion)



Operations Concept
Product Lifecycle (Distribution)



Conclusion

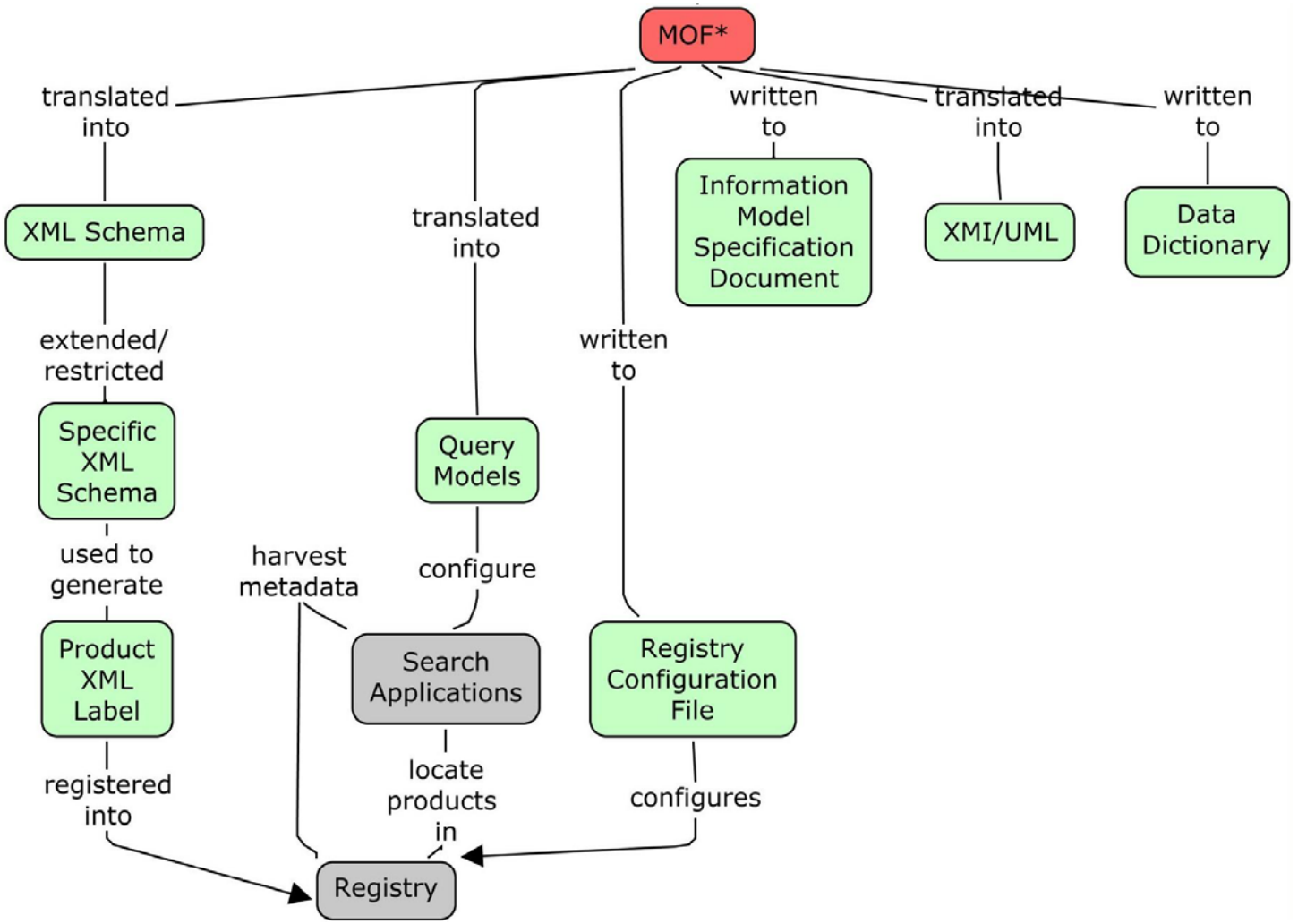
- The PDS has just released PDS4, its next generation data archive system.
  - The PDS4 data standards are the result of a multi-year effort to develop an information model based on accepted standards for data preservation and curation, metadata management, and model development.
  - The information model is used to drive system development
    - Generation of data standards documentation
    - Configuration of federated registries and search engines.

- After two decades of successful support to the planetary science community the PDS is well-positioned to meet the community's requirements for long-term preservation and curation for the next decade.

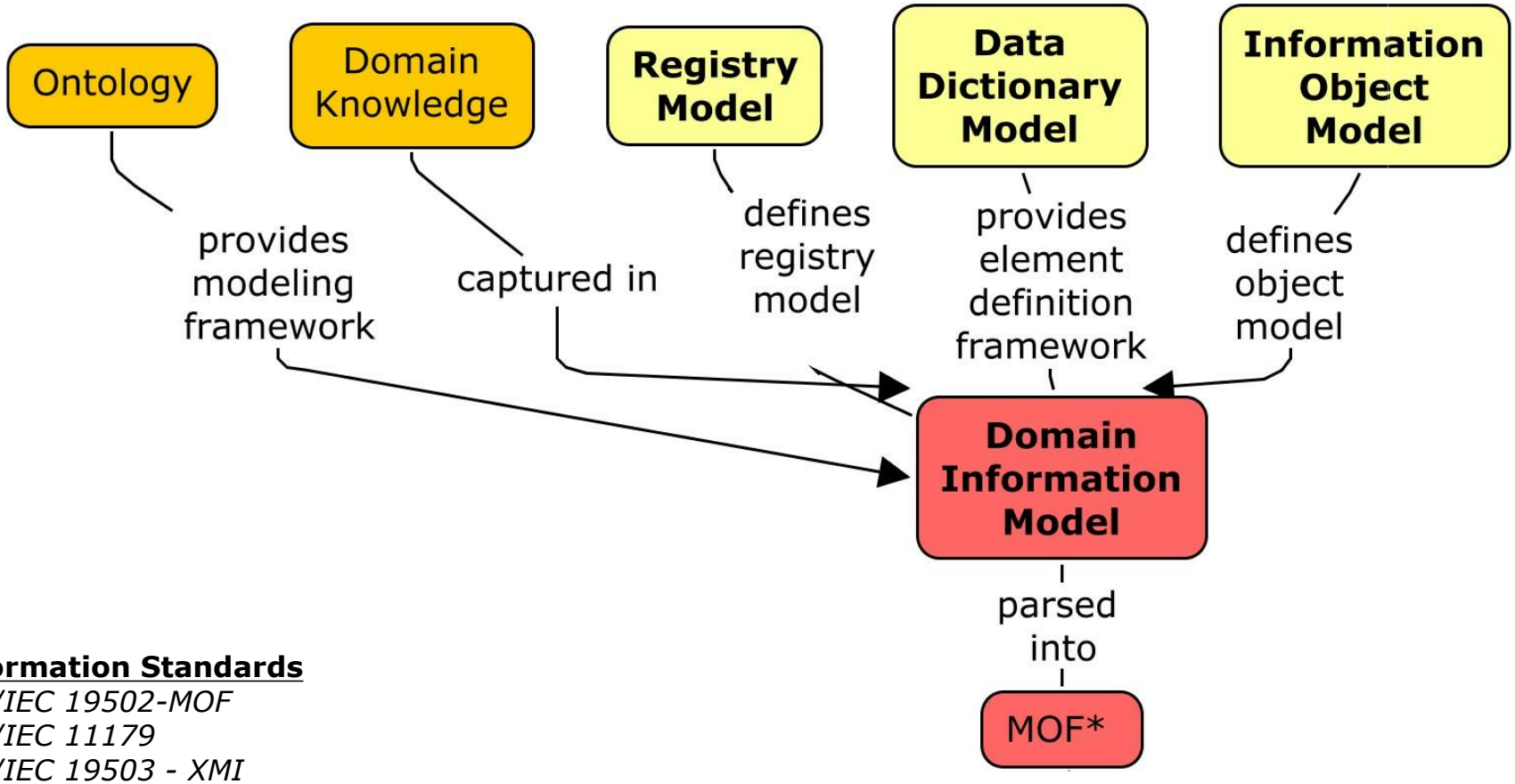
THANK YOU!

Backup

Generated Artifacts



Models in Context



Information Standards
ISO/IEC 19502-MOF
ISO/IEC 11179
ISO/IEC 19503 - XMI
ISO 639-RDF
ISO/IEC 19501 – UML
OWL-DL

Registry in Context