A Reference Architecture for Archival Systems

Use Case With Product Data

Raphael Barbau
barbau@nist.gov

Joshua Lubell
lubell@nist.gov

NASA GSFC Information Science & Technology Colloquium Series
October 15, 2014

National Institute of Standards and Technology
U.S. Department of Commerce

engineering laboratory

Engineering Laboratory
Systems Integration Division

Disclaimer

Certain commercial products are identified to help explain the research. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the products identified are necessarily the best available for the purpose.

Outline

- Preservation and product data
- Standards for preservation
  - Open Archival Information System Reference Model
  - Audit and Certification of Trustworthy Digital Repository
- Our approach
  - Enterprise Architecture and EA frameworks
  - Reference architecture (RA) for archival systems
  - Application to product data
- Potential usage for aerospace industry long-term archival standards
- Conclusion

The Preservation Problem

- Too much digital data!
  - It takes about 15 minutes for the world to churn out new digital information equivalent to the entire collection in US Library of Congress
- Proprietary file formats
  - Expected lifetime of typical manufacturing software application only 3 years
- Short-lived Computing hardware and software
  - Expected lifetime of today's storage/retrieval technologies only 10 years
- Products often outlive computer software/hardware by an order of magnitude
  - Aircraft can last 50 years or more
  - Healthcare records should be preserved through the patient's lifetime, and perhaps beyond

Product Model Data Preservation is *Really* Challenging

- Product lifecycle may be an order of magnitude longer than the modeling s/w lifecycle
  - Examples: aircraft, buildings, power plants
- Semantic richness needed for long-term socio-technical concerns
  - Environmental, forensic, regulatory
- Content information standards not a silver bullet
- Accuracy is critical for product quality and manufacturability
- Product model may contain many information types
- Context in which the archive exists is complex
  - Multiple interacting systems
  - Preservation not the primary organizational mission

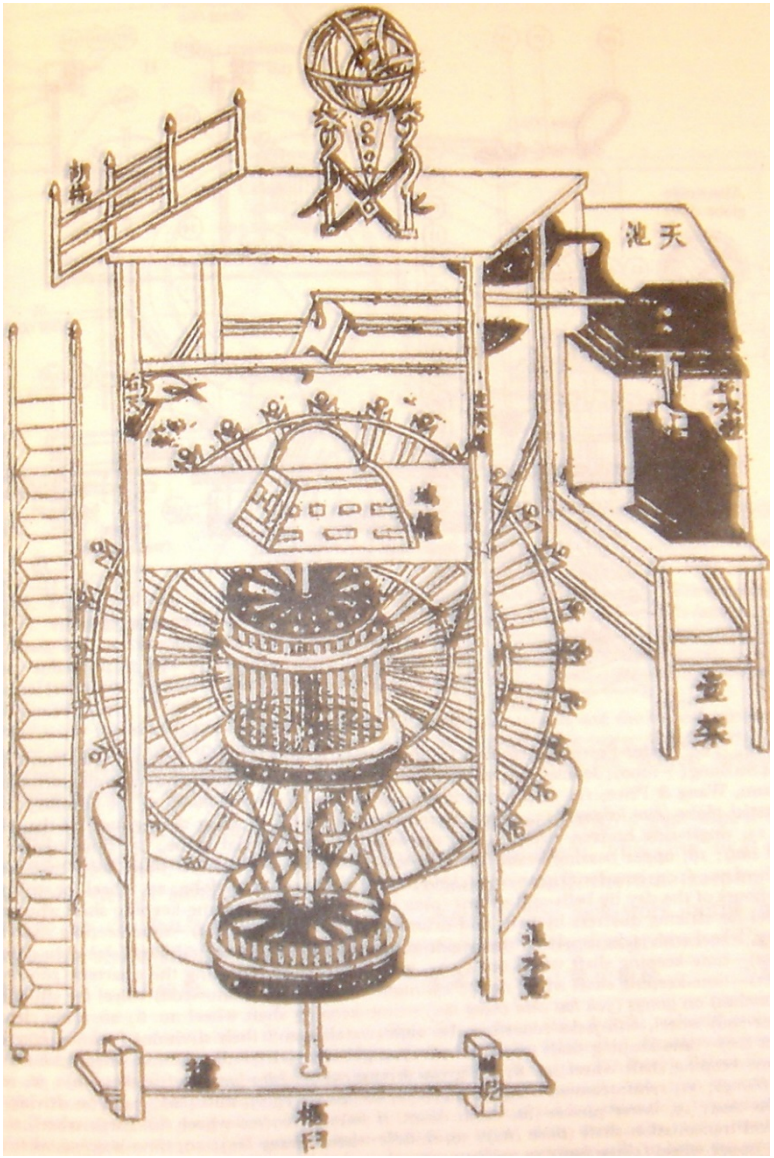
So What Exactly is Product Data?

From ISO 10303 (STEP), the **ST**andard for the **E**xchange of **P**roduct Model Data:

- **Product**
  - A thing or substance produced by an industrial, business, or natural process
- **Product data**
  - A representation of product information suitable for communication, interpretation, or processing by human beings or by computers

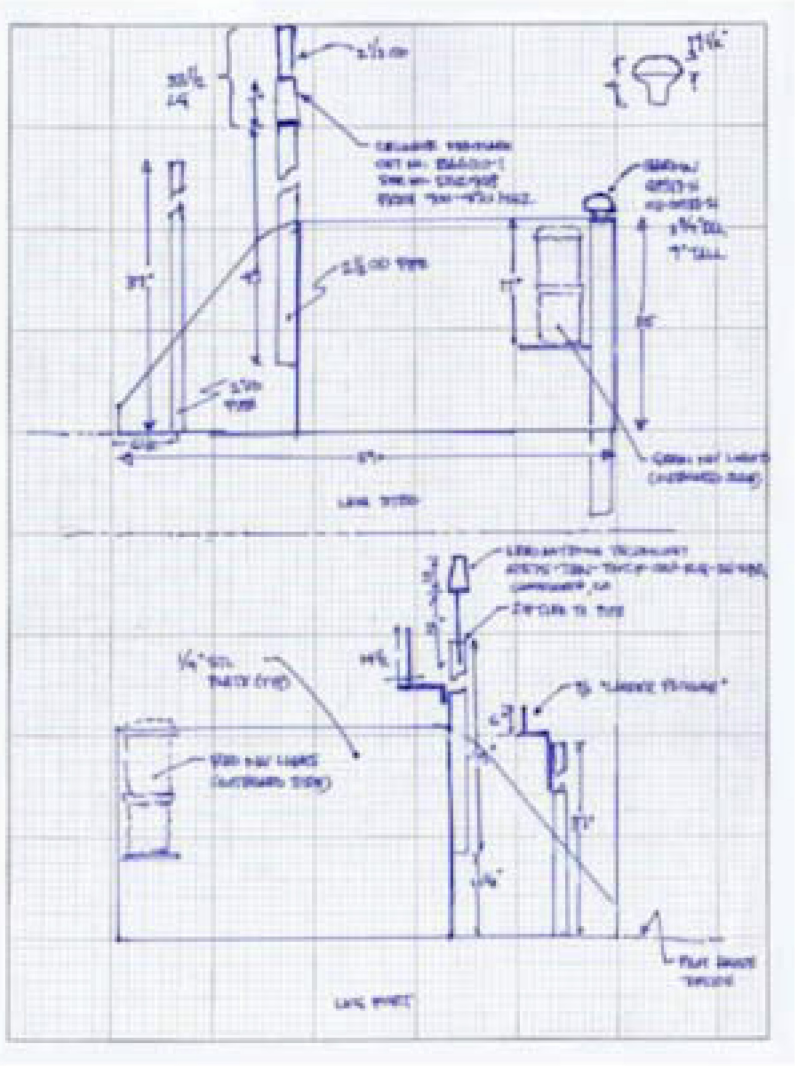
Our scope limited to industrial products

Product Data From the 11th Century



Su Song's clock tower built in Kaifeng (1092)

20th Century Example



B. Kassel and P. Rakow. Long Term Data Retention of Weapon System Information FY2010 Final Report. Naval Surface Warfare Center Carderock Division. NSWCCD-20-TR–2010/06. October 2010.

Non-Digital Product Data

- 2D drawings and related documentation
- Paper is dominant physical medium
- Document is dominant logical unit

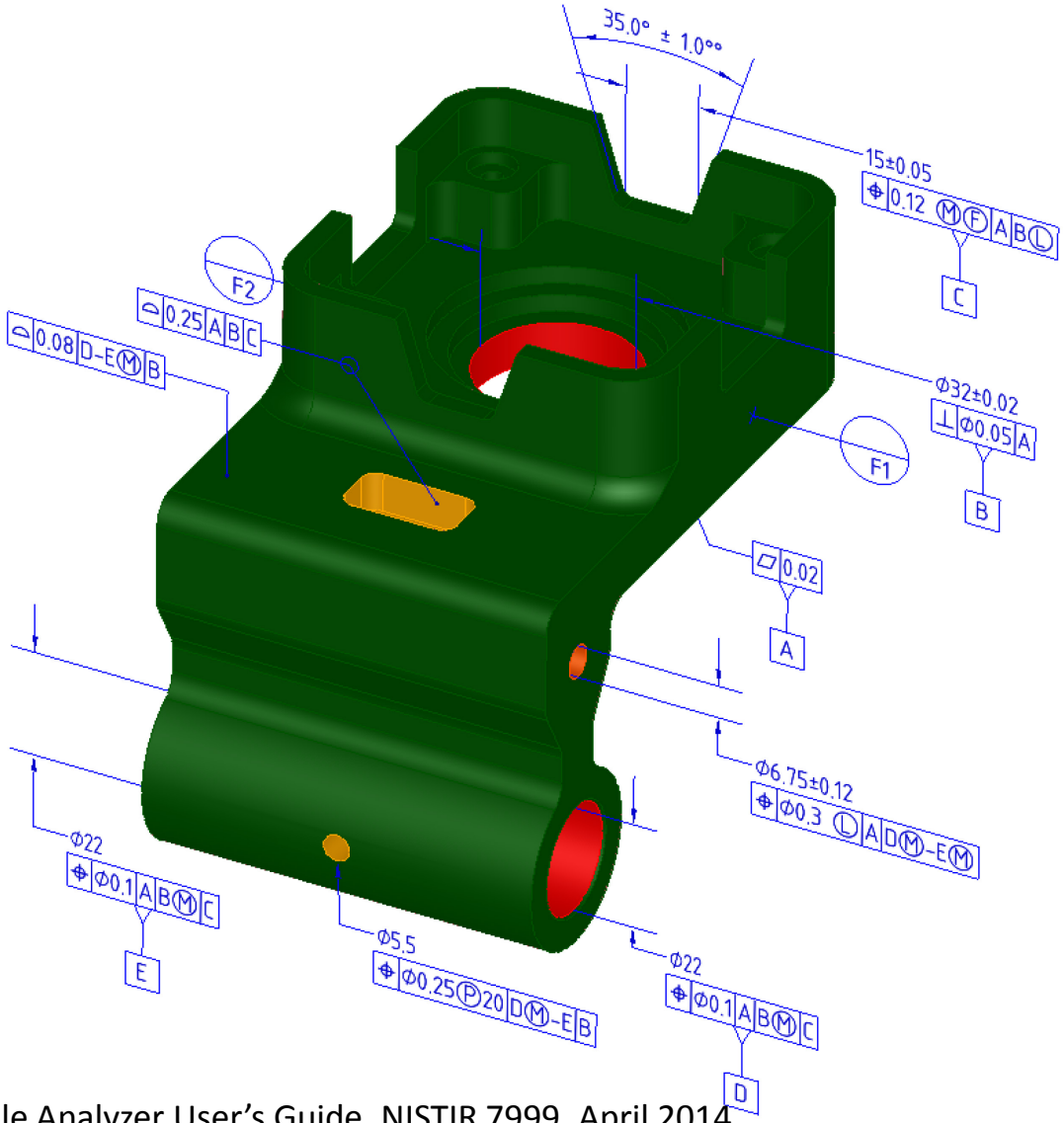
Advent of 3D Modeling (Late 20th Century)

- Product data created using solid modeling software tools
- But 3D not used for long-term storage
  - 2D projections of 3D models needed to capture non-geometrical information
  - An imperfect solution
- Management of digital formats becomes a big issue
- Document still dominant logical unit

Product Data Today

- 3D modeling systems are more sophisticated
  - Can represent and render Product and Manufacturing Information (PMI), including GD&T annotations
  - Can associate supporting documentation with model elements
- Model is the master
  - Document no longer dominant logical unit
  - Preservation no longer just about documents and formats

21st Century Example

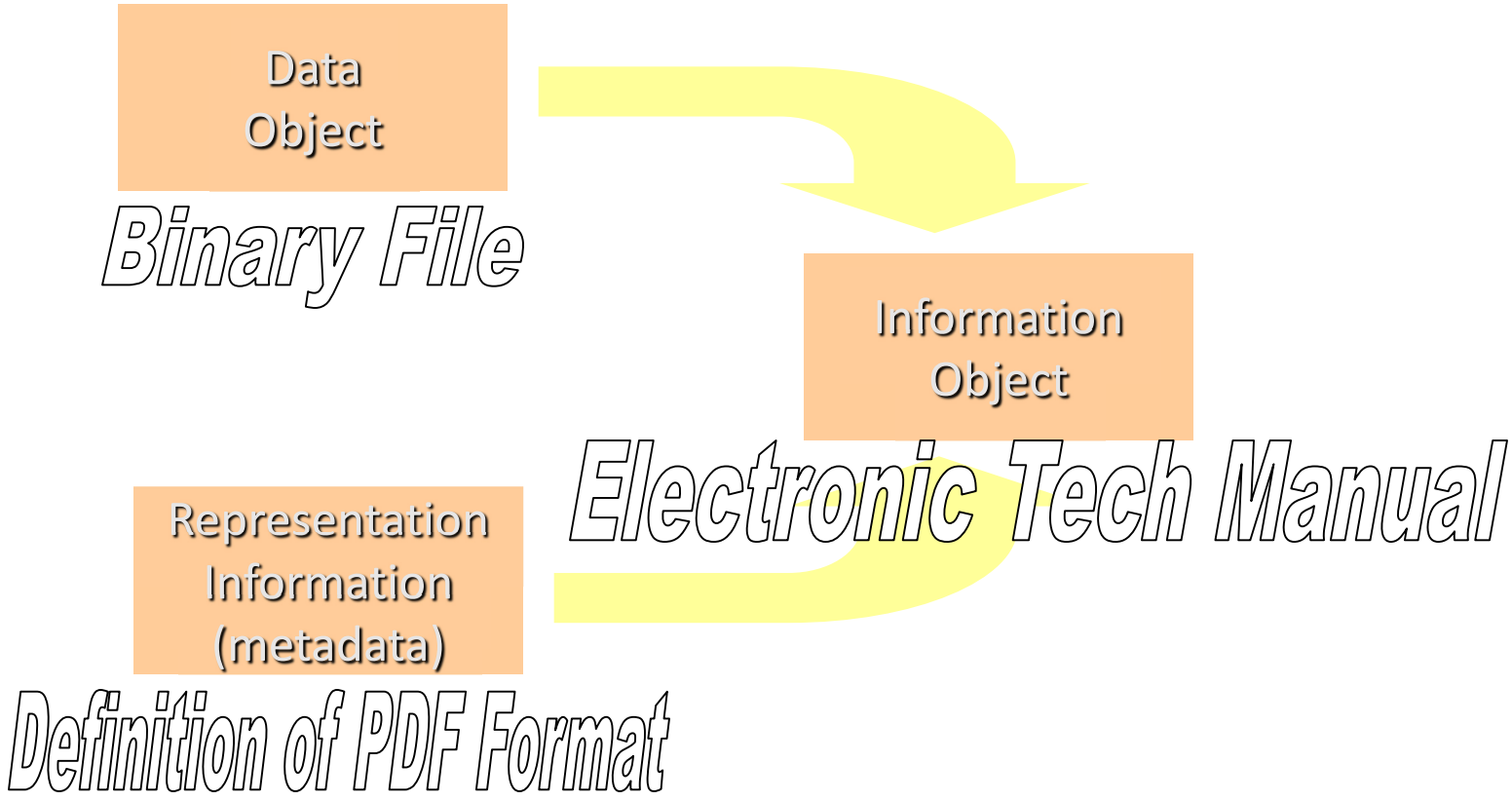


R. Lipman. STEP File Analyzer User's Guide. NISTIR 7999. April 2014.

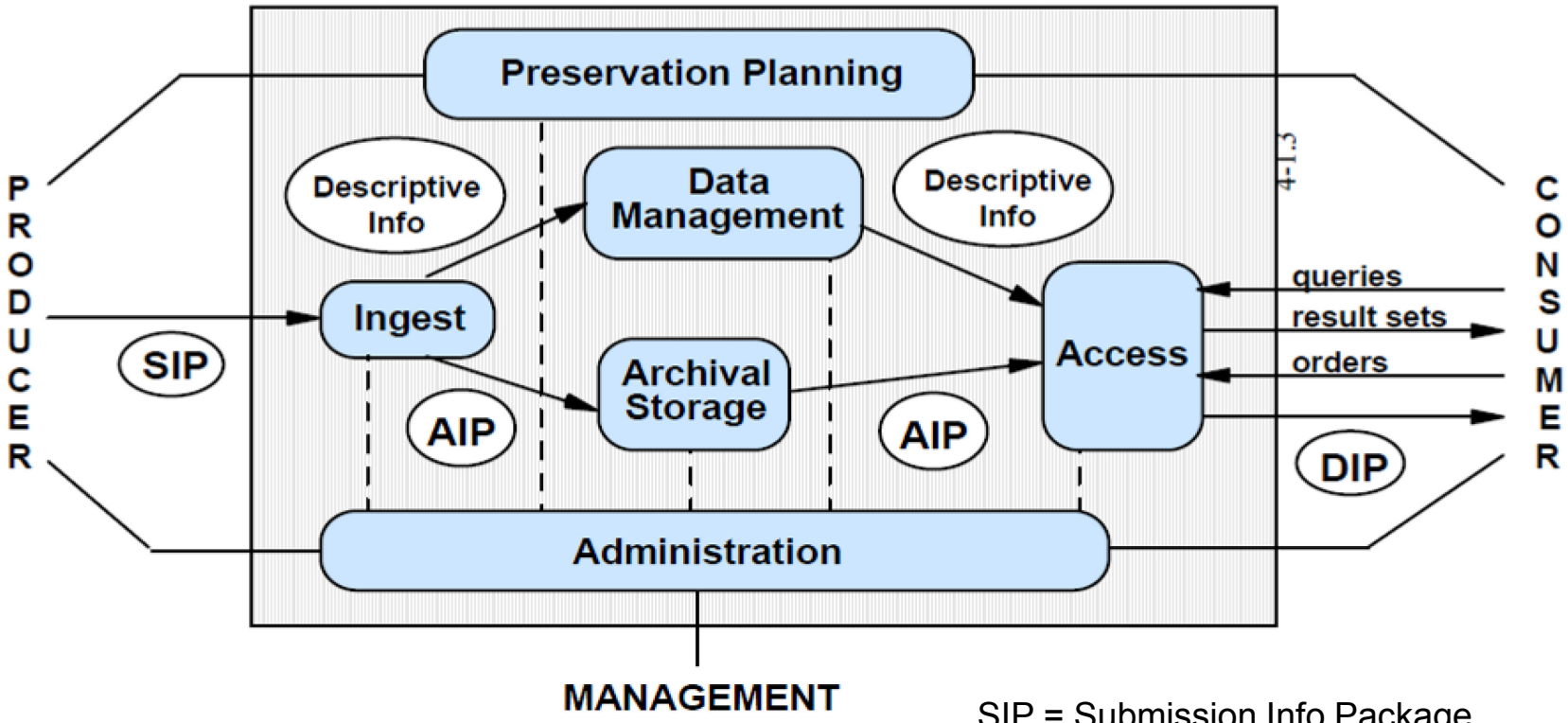
Existing Preservation Standard: OAIS

- Open Archival Information System (OAIS) reference model
- ISO 14721
- Provides common preservation terminology
- Covers information and functional aspects
- Core vocabulary reusable across all disciplines
- Neither content-specific nor business-specific
- Not an RA substitute
  - But a very useful building block

Information = Data + Interpretation



OAIS Functional Entities



SIP = Submission Info Package
AIP = Archival Info Package
DIP = Dissemination Info Package

Existing Preservation Standard: ACTDR

- Audit and Certification of Trustworthy Digital Repositories (ACTDR)
  - Also known as Trustworthy Repositories Audit & Certification (TRAC)
- Consultative Committee for Space Data Systems standard
- Lists evaluation criteria for a preservation solution
- Conforming implementation shall provide evidence to auditor that they meet the criteria
  - Standard provides examples of evidence
  - Examples are documentation-based
- Outside scope of ACTDR: association of properties of a repository implementation with evaluation criteria

ACTDR Example

4.1.6 The repository shall obtain sufficient control over the Digital Objects to preserve them.

Supporting Text

This is necessary in order to ensure that the preservation can be accomplished, with physical control, and is authorized, with legal control.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Documents showing the level of physical control the repository actually has. A separate database/metadata catalog listing all of the digital objects in the repository and metadata sufficient to validate the integrity of those objects (file size, checksum, hash, location, number of copies, etc.)

Discussion

The repository must obtain complete control of the bits of the digital objects conveyed with each SIP. Sufficient physical and legal control is necessary for the archives to make any changes required by their Preservation Implementation Plan for that data and to distribute it to their consumers. For example, in cases where SIPs only reference digital objects, the repository must also reference the digital objects or preserve them if the current repository is not committed to such preservation.

Building upon OAIS and ACTDR

- We propose a holistic approach to **formally** describe archival systems that is:
  - **Comprehensive** – represents preservation function, other (non-preservation) business functions, interactions
  - **Traceable** to evaluation criteria
  - **Suitable for product data archives**
- Our goal: a “good” system description

What is “Good”?

- Covers the essential aspects
- Uses common terminology across descriptions
- Uses a common notation
- Coherence
- Modularity
- Provides views showing same elements in different contexts
- Facilitates implementation

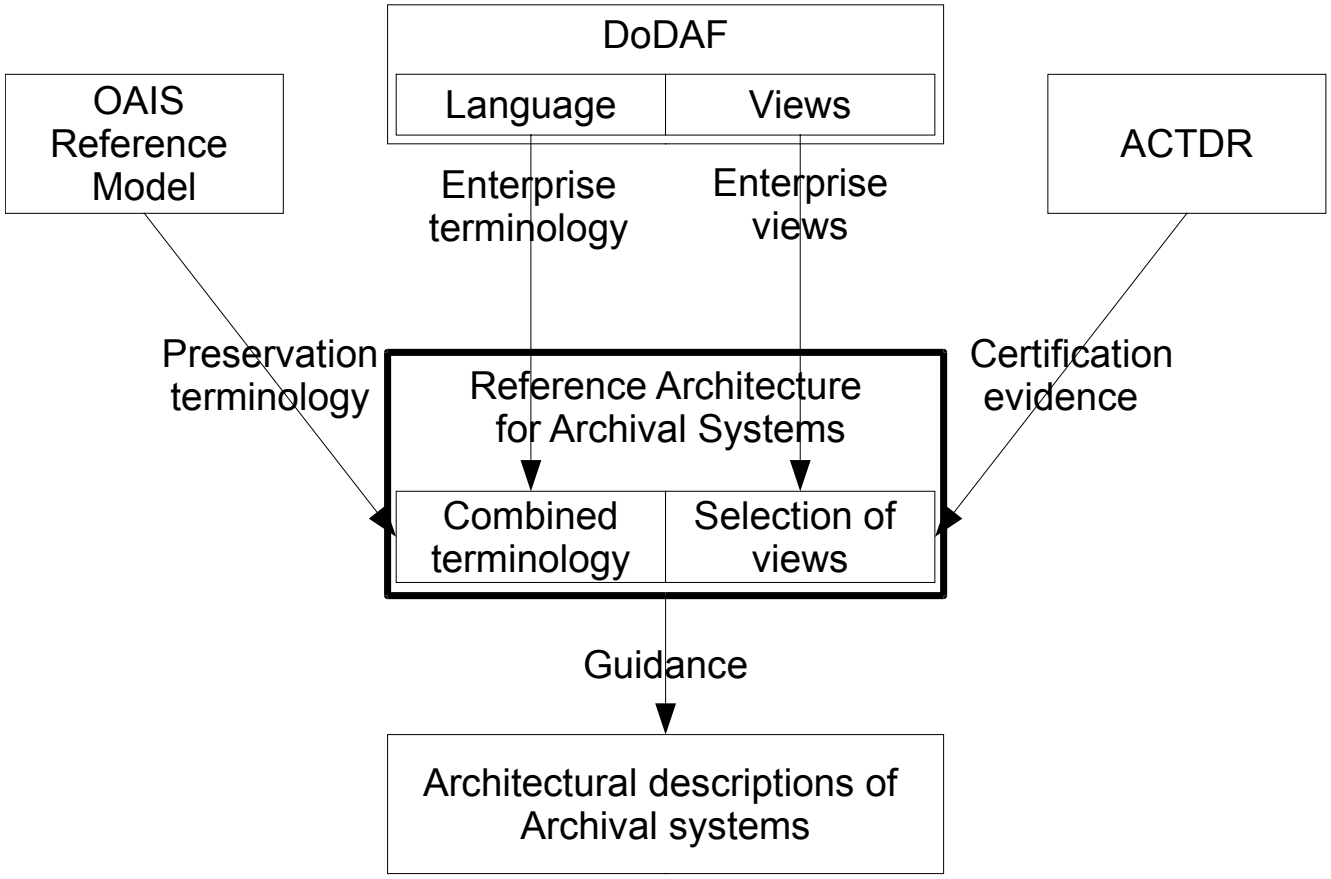
Enterprise Architecture (EA)

- Discipline for creating *architectural descriptions* of the elements of an enterprise, or one of its parts
- Improves transition from “as is” to “to be”
- Facilitates systems interoperability
- EA Frameworks usually focus on either
  - Method (steps to follow to describe a the current/future structure of an organization)
  - Language (formally concepts used to describe such structure) and views
- We use the language-based Department of Defense Architecture Framework (DoDAF)

DoDAF

- Represents enterprises using common terminology across descriptions
- Has implementation based on UML
  - Standard notation
  - Splits description into diagrams
  - Leverages model-driven and service-oriented architectures (MDA, SOA)

Approach: The Big Picture

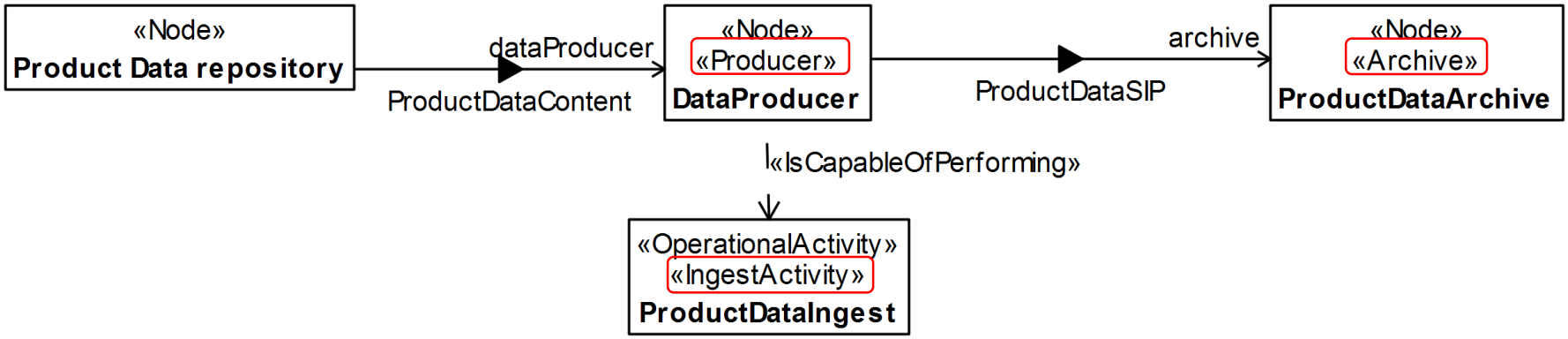


DoDAF Views

- Can highlight a particular aspect of the description, and serve as certification evidence
- View types used in our approach
  - Data and Information viewpoint (DIV)
  - Services viewpoint (SvcV)
  - Systems viewpoint (SV)
  - Operational viewpoint (OV)
  - Capability viewpoint (CV)

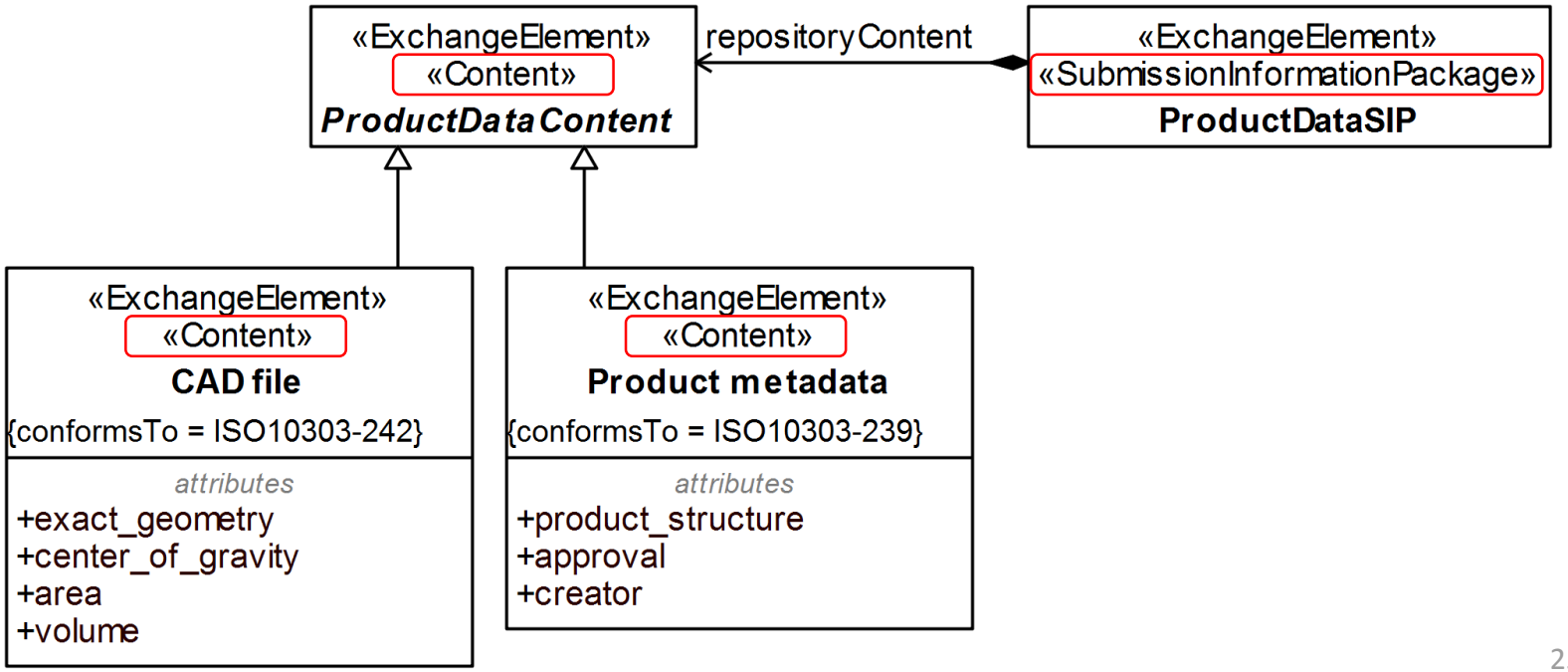
RA Example: Information flows

- Represent resource flows between nodes (OV-2)
- Node: logical entity, performer
- Producer, Archive, IngestActivity are OAIS terms



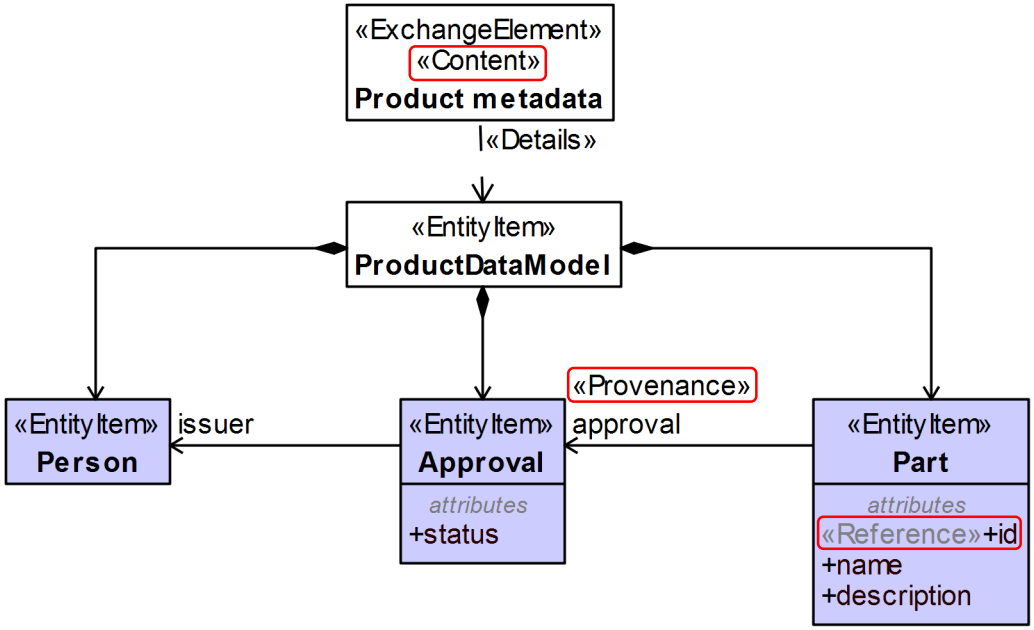
RA Example: Content definition

- Conceptual view of exchanged information (DIV-1)
- SIP and Content are OAIS terms
- ACTDR 4.1.1: “The repository shall identify the Content Information and the Information Properties that the repository will preserve.”



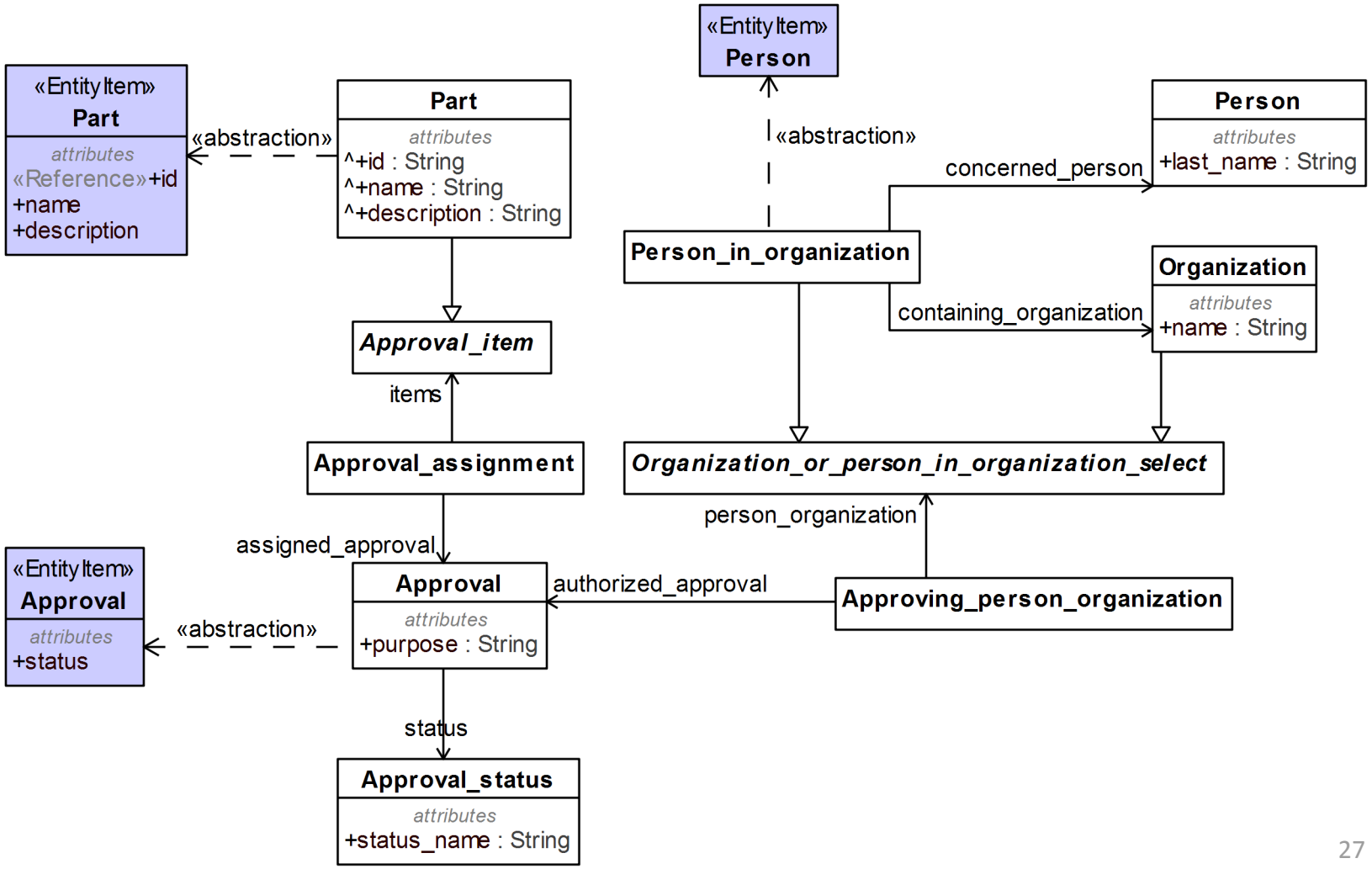
RA Example: Metadata definition

- Conceptual view of metadata (DIV-1)
- Provenance and Reference are OAIS terms
- ACTDR 4.5.1: “The repository shall specify minimum information requirements to enable the Designated Community to discover and identify material of interest.”



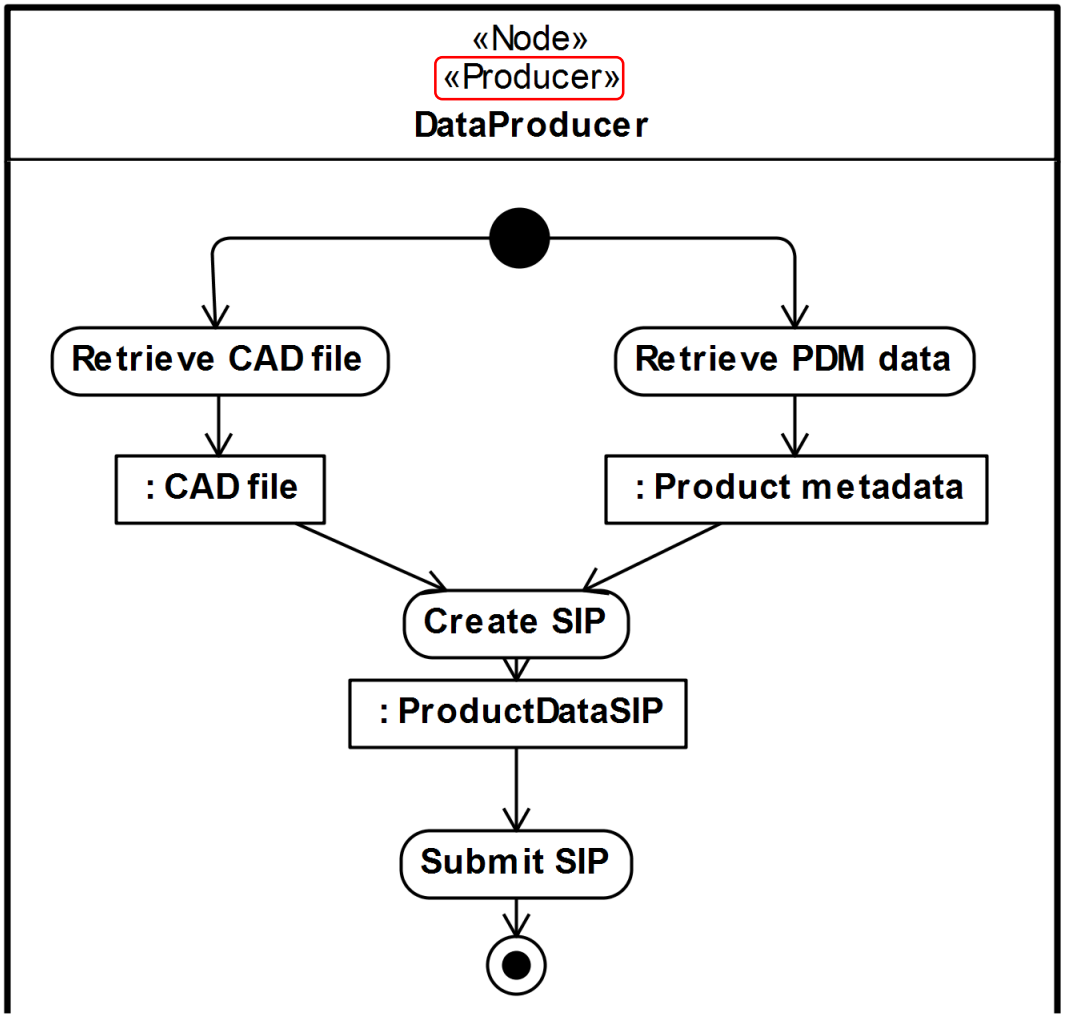
RA Example: Metadata definition

- Logical view of metadata (DIV-2)
- Connects logical and conceptual entities

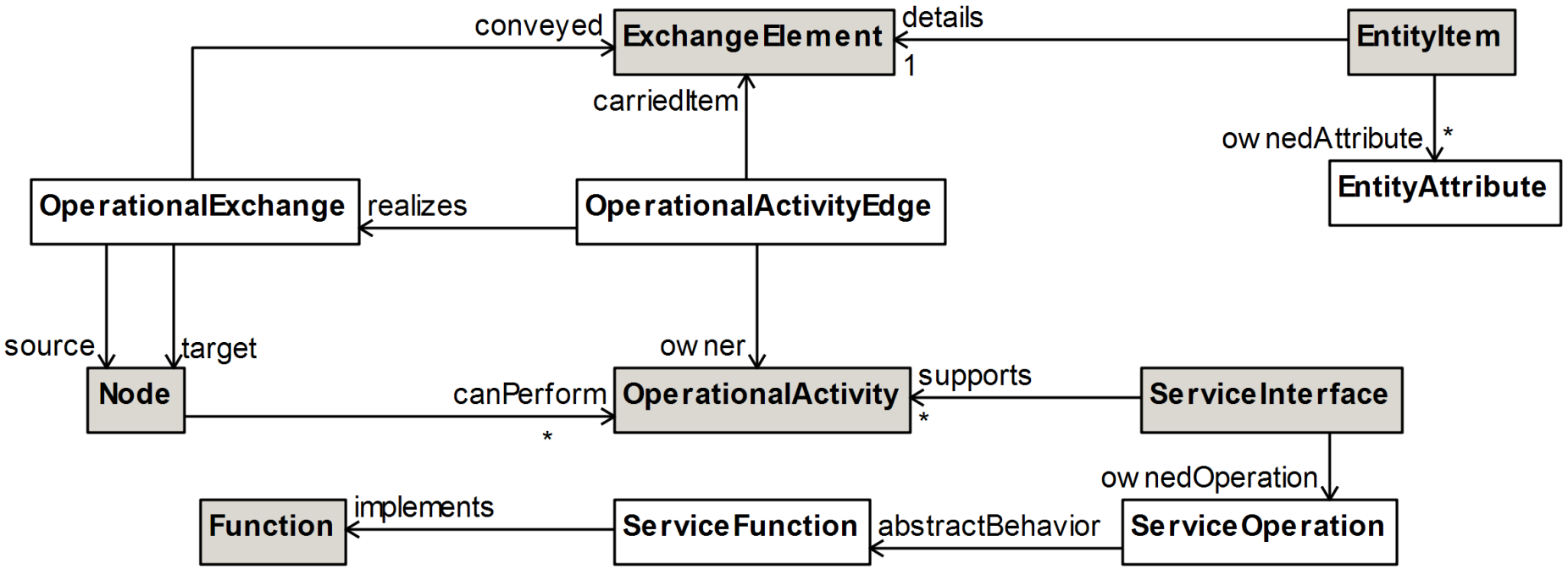


RA Example: Ingest activity

- Description of an ingest activity (OV-5b)

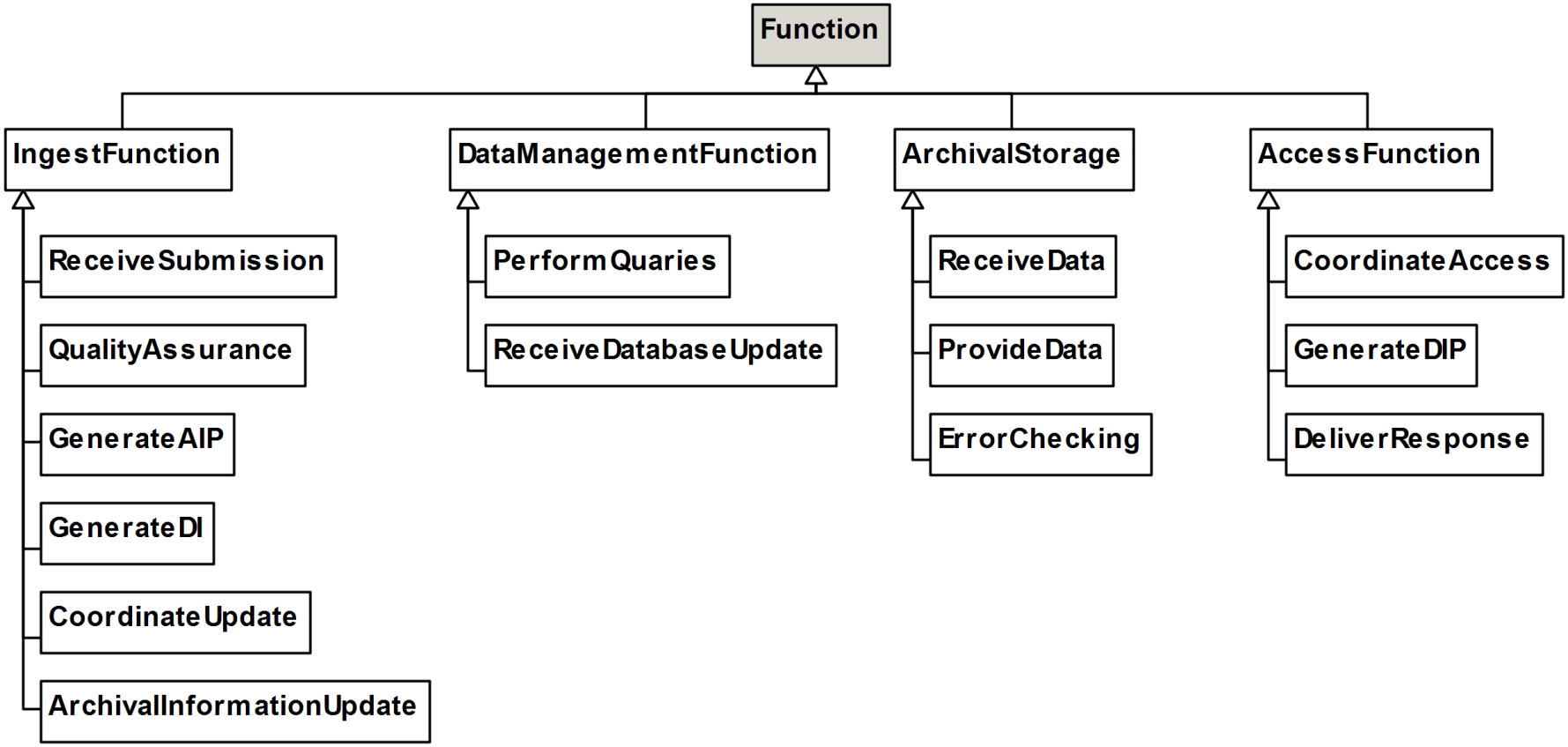


Subset of DoDAF used in previous examples



OAIS terminology extension

- Example specializing DoDAF Function concept



Summary of Approach

- DoDAF-based
  - OAIS Reference Model inspires preservation terminology
  - Views describe specific aspects, provide traceability to certification criteria
- Product data ingest examples
  - Formally specify activities, information
  - Use OAIS terminology to describe activities
  - Use international standards to describe product geometry and management content information

Potential Application: LOTAR

- LOTAR = “Long Term Archiving and Retrieval”
- http://lotar-international.org
- Family of standards defining an auditable process for the long-term archiving of digital product model data
- Driven by US and European aerospace industry requirements
- Harmonized with German automotive industry standard Verband der Automobilindustrie (VDA) Recommendation 4958
- Based on OAIS Reference Model

Why LOTAR?

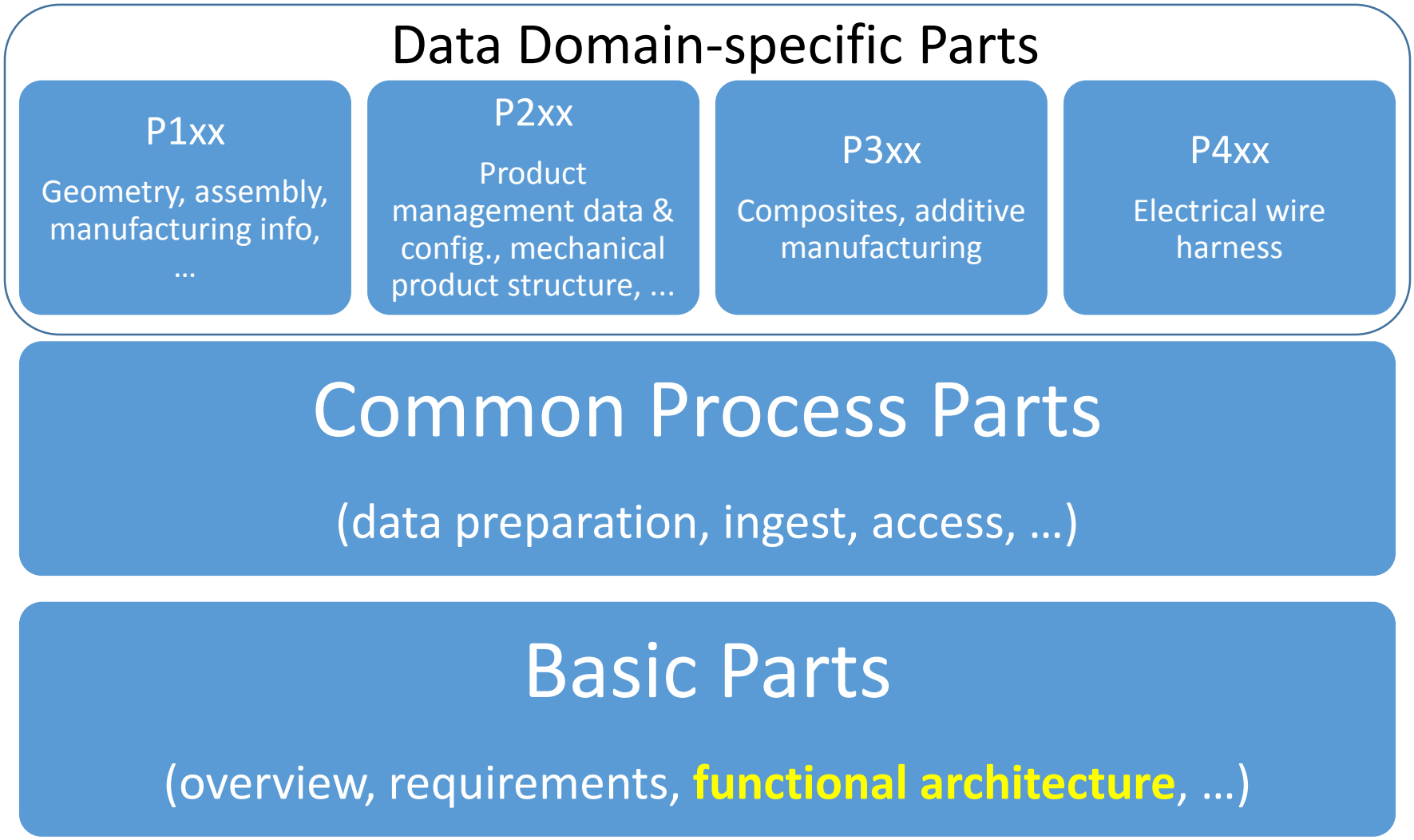
- Legal requirements
  - Product liability
  - Type certification
- Business requirements
  - Support in operation
  - Design reuse
- Existing **document-based** digital preservation standards/technologies are insufficient

A lot of digital product data…

Dave Sizer derivative work. Altair78, via Wikimedia Commons. December 2009

…from multiple suppliers

LOTAR Overview



LOTAR Will Extend OAIS by Adding…

- Guidance on representing content information
  - 3D geometry
  - assemblies
  - Manufacturing information
  - Product metadata
- Industry-specific guidance for ingest and access
  - Requires a functional architecture
    - LOTAR basic Part yet to be developed

LOTAR Functional Architecture could use an RA-based approach to rigorously define…

- Traceability to business/legal requirements (certification, liability, design reuse)
- Use cases
- Aerospace-specific specializations of OAIS preservation concepts
- Interactions with business functions external to preservation

Conclusion

- Standardized terminology is useful for better understanding archival systems
- Architectural descriptions can:
  - Formally use that terminology to describe aspects of the archival system
  - Integrate preservation functions with key business functions
- Model-based approach to developing archival systems, with precise semantics
  - Provides traceability
  - Enables a Model-Driven Architecture