Convergence of Semantic Naming and Identification Technologies?

What are the Choices and
What are the Issues?
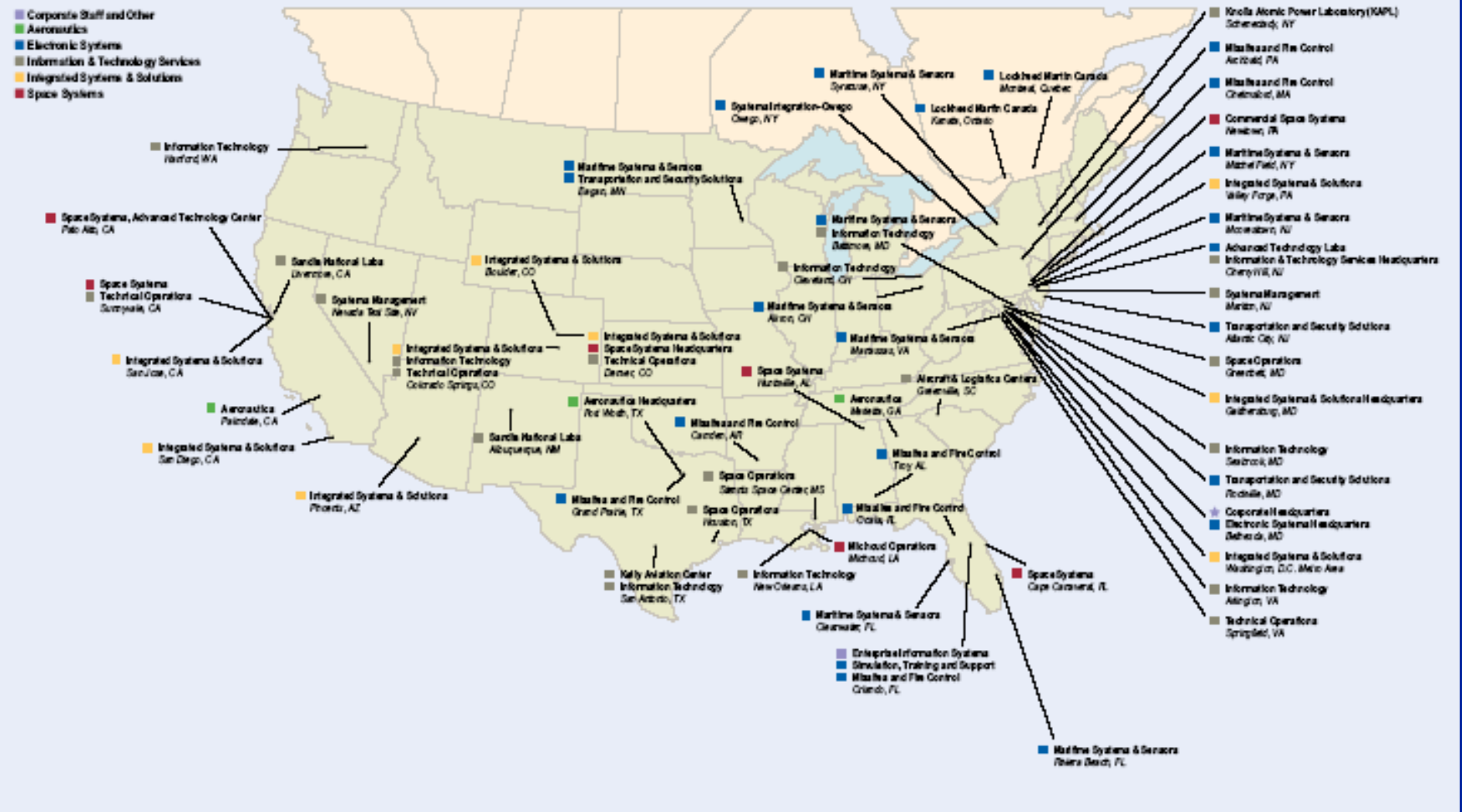
Arlington, VA
April 27, 2006

Ron Schuldt
Lockheed Martin
Enterprise Information Systems
Senior Staff Systems Architect

The IT Challenge – A Perspective

An Extensive IT Infrastructure

Agenda

- **The Semantics Problem**

- **Relevant Architectures and Standards**

- **Semantics Naming and Identification Choices**
  - **Semantic Web based**
  - **Metadata Registry - ISO/IEC 11179 based**

- **Universal Data Element Framework (UDEF) – A Semantic DNS for Structured Data**

- **Disaster Response Example Use Case**

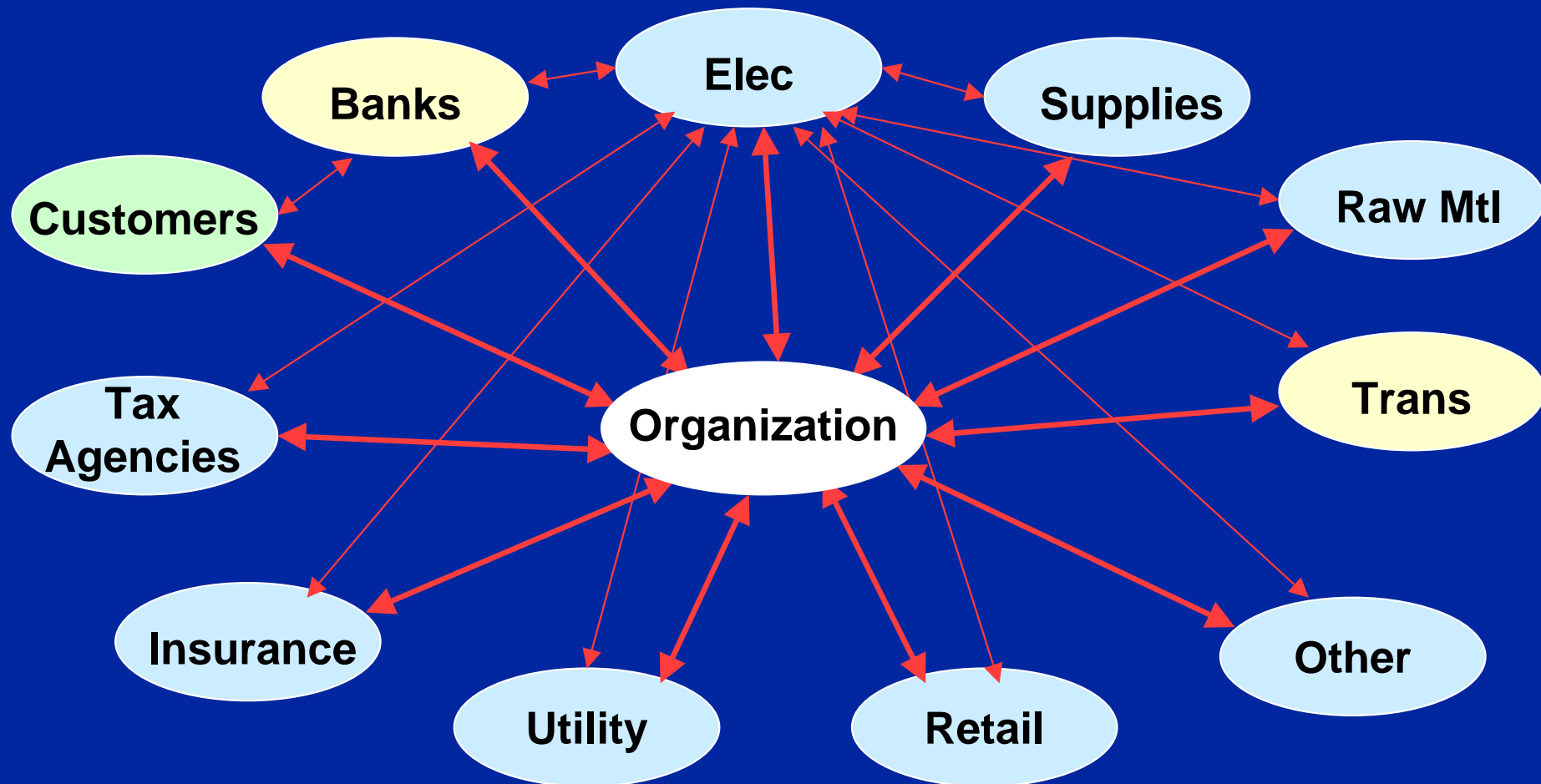- **Semantics Naming and Identification Issues**

The Semantics Problem

The Problem - Global Perspective

Many are attempting to set their own semantics standard

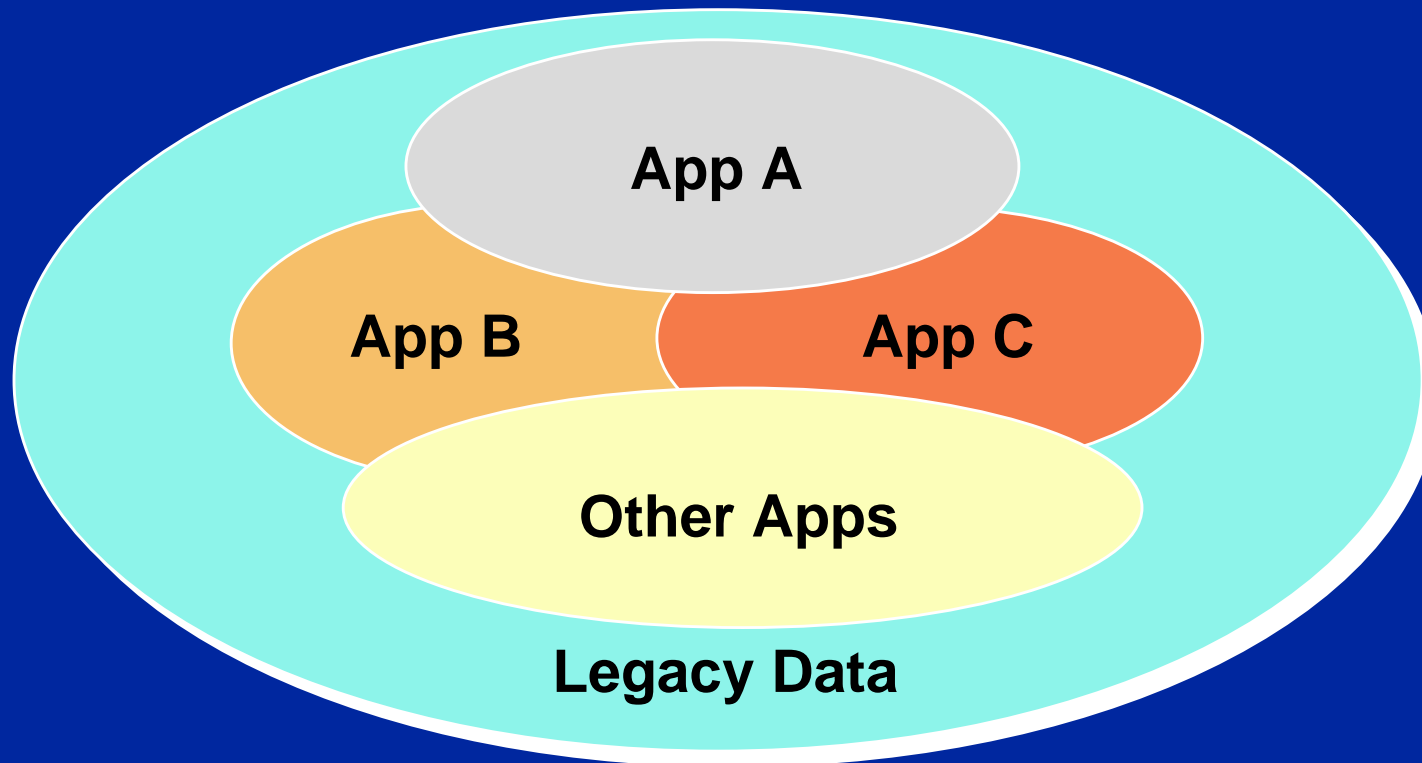**Each must interface with organizations they do not control**

The problem is the lack of common semantics and schema between organizations

The Problem - Enterprise Perspective

Conflicting semantic overlaps between back-office systems

Though semantically equal, the following are 4 different XML tag names

<PARTNUMBER>111-222-333</PARTNUMBER>

<partNumber>111-222-333</partNumber>

<PartNumber>111-222-333</PartNumber>

<partnumber>111-222-333</partnumber>

The Problem – Legacy Applications

- **Across the globe there are millions of legacy applications that will remain for many decades that need to be Web enabled – in preparation for Web Services and Service Oriented Architecture**

    - XML and associated W3C standards address the syntax requirements but an adopted content semantics standard does not exist yet that can transcend all functions of all organizations

- **Users of the legacy applications consistently resist changing the names of the fields**

    - The semantics solution needs to be non-intrusive to the application user

The Problem – Content Discovery

- **Content (Web pages, various documents in various formats, data in databases, etc.) resides on countless servers across the globe. Lack of standard names and their meaning makes it difficult to find the data objects of interest – both inter- and intra-enterprise.**

    - **W3C is attempting to address this through the Semantic Web suite of metadata standards (RDF, OWL, etc.) and URI for unique identification of instances.**

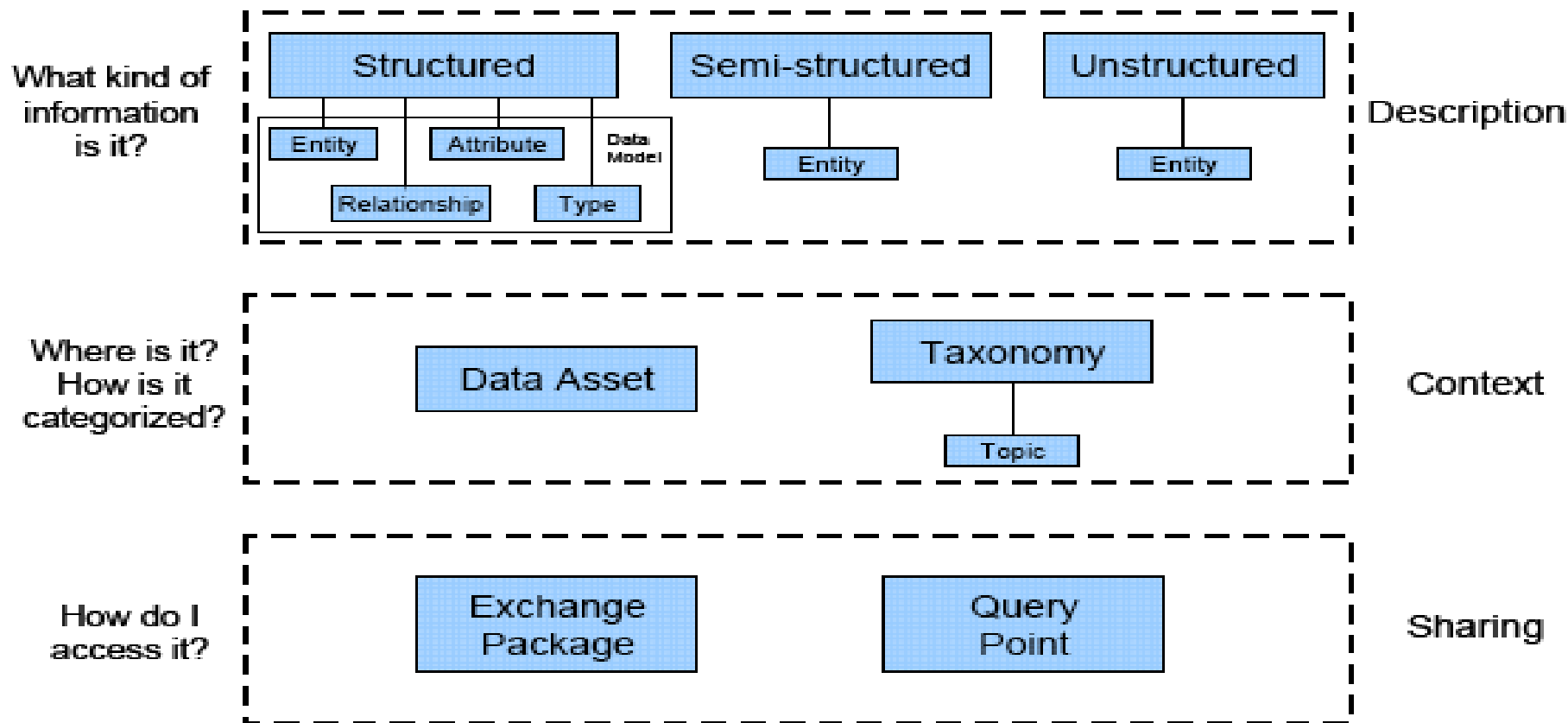# Relevant Architectures and Standards

OASIS Reference Model for SOA*
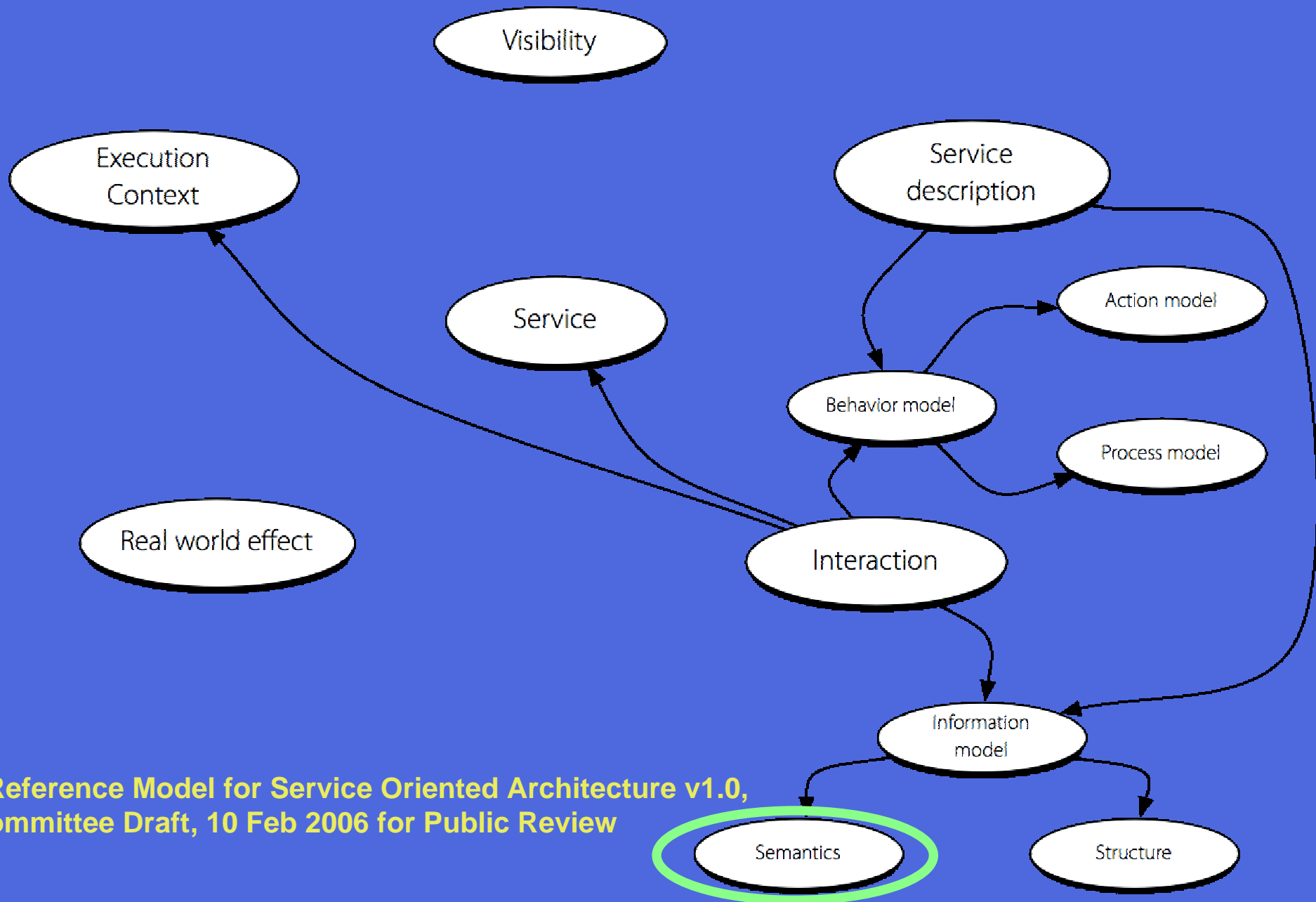
* Reference Model for Service Oriented Architecture v1.0,
Committee Draft, 10 Feb 2006 for Public Review

An Example Data Reference Model

The FEA Data Reference Model (DRM) 2.0

United States Federal Enterprise Architecture Data Reference Model

http://www.whitehouse.gov/omb/egov/a-2-EAModelsNEW2.html

A Semantics Reference Model

Reference Model by Andreas Tolk (2005)
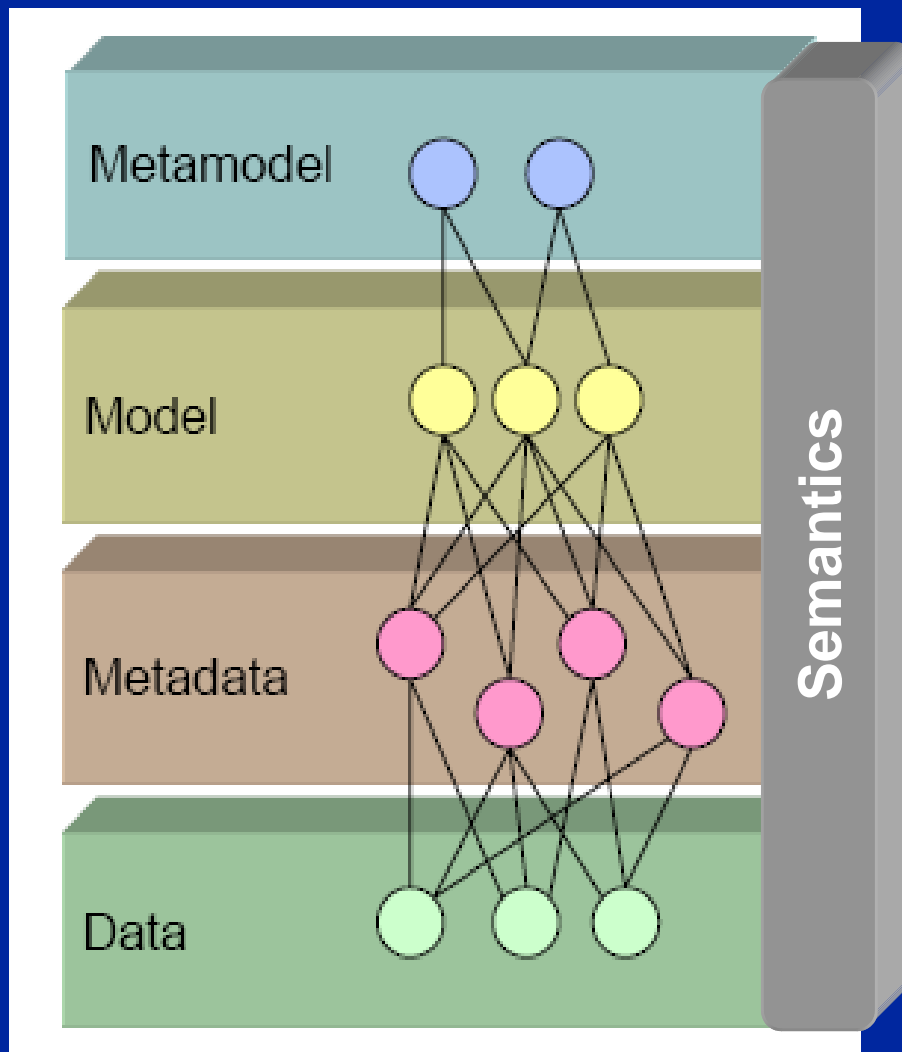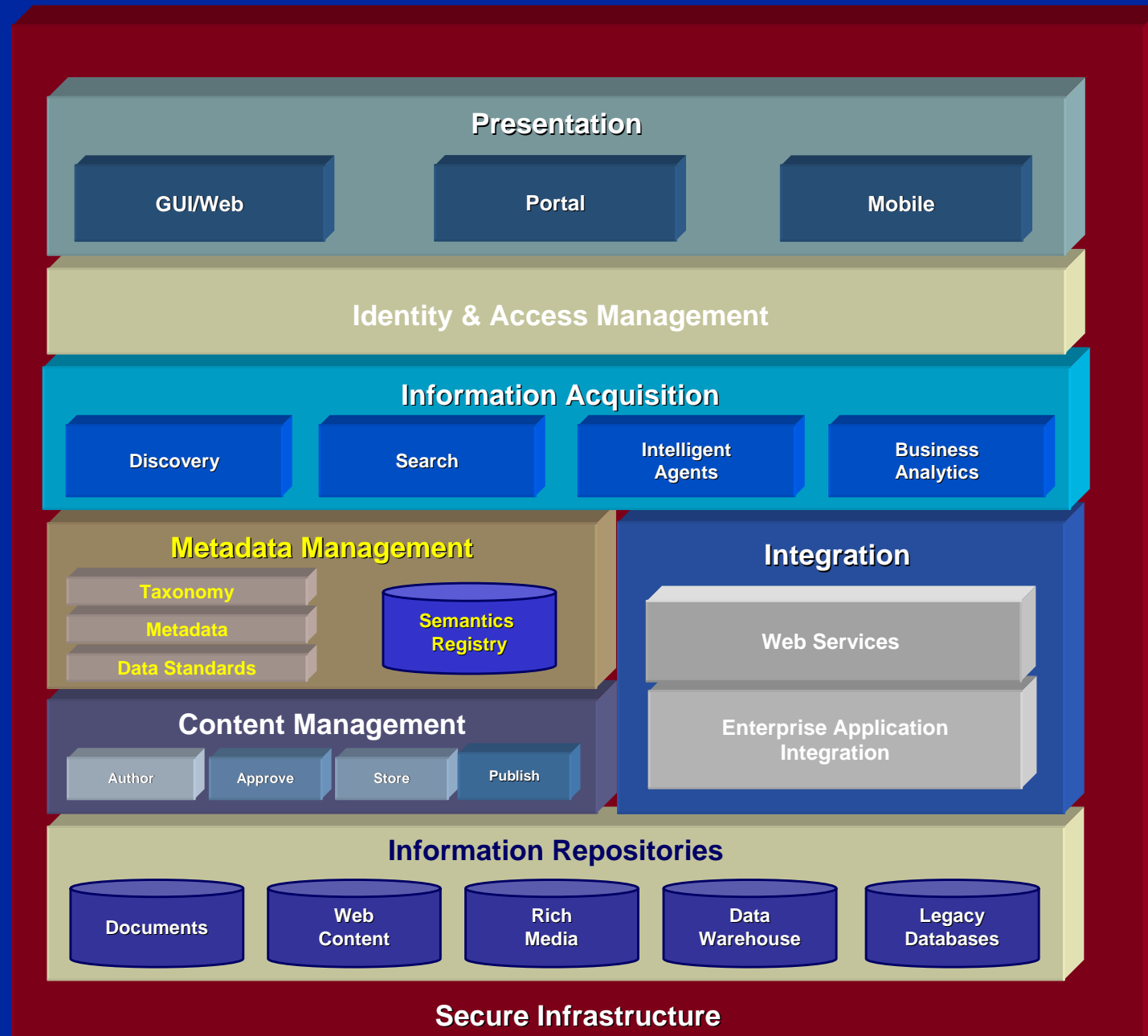
- *Understandable semantics transcend every aspect of inter- and intra-enterprise data exchange – whether machine-to-machine or machine-to-human or human-to-human.*

An Example Information Architecture

Presentation

GUI/Web | Portal | Mobile

Identity & Access Management

Information Acquisition

Discovery | Search | Intelligent Agents | Business Analytics

Metadata Management

Taxonomy
Metadata
Data Standards

Semantics Registry

Integration

Web Services

Enterprise Application Integration

Content Management

Author | Approve | Store | Publish

Information Repositories

Documents | Web Content | Rich Media | Data Warehouse | Legacy Databases

Secure Infrastructure
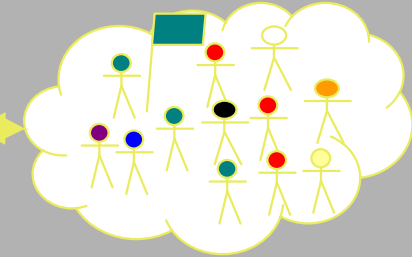
Example Metadata Use Cases

Finding People:

(Tacit Knowledge)

Social Networking → Semantic Linking →

Finding Content:

(Explicit Knowledge)

Content Discovery → Semantic Search → Content Repository, Content Repository, RSS Feeds
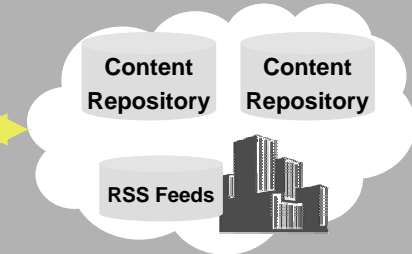
Achieving Visibility:

(Potential Knowledge)
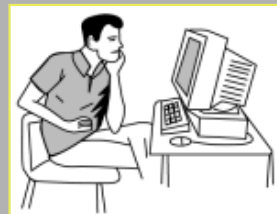
Dashboards & Business Intelligence → Semantic Aggregation → Map/Transform → desktop, database, web service, legacy
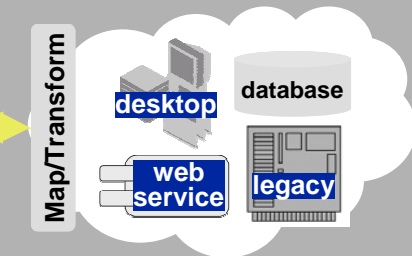
Building Applications:

Application GUI & Workflow → Business Logic → Persistent Data (RDB/XML/RDF)
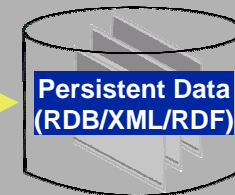
Interoperability:
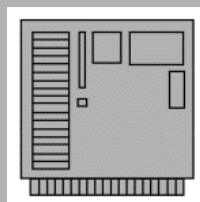
Service-Oriented Architectures → Semantic Mediation → Service-Oriented Architectures, Anything
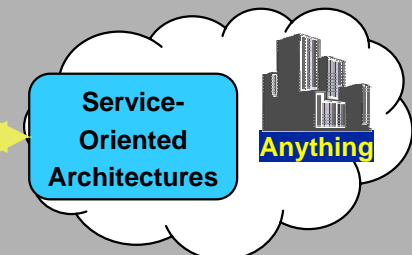
Sample Definitions of "Semantics"

- ### *Sample of Definitions from the Web:*
  - The relationships of characters or groups of characters to their meanings, independent of the manner of their interpretation and use. Contrast with syntax.

  - The science of describing what words mean, the opposite of syntax.

  - The meanings assigned to symbols and sets of symbols in a language.

  - The study of meaning in language, including the relationship between language, thought, and behavior.

  - The meaning of a string in some language, as opposed to syntax which describes how symbols may be combined independent of their meaning.

Proposed Definition and Standards

- **"Semantic Interoperability" Proposed Definition:**
  - *The shared meaning of a string of characters and/or symbols in some language within a context that assures the correct interpretation by all actors.*

"Semantic Interoperability" Standards

Cross Standard Semantics and Metadata Alignment – UDEF, RDF, OWL

Domain Specific "Semantic and Syntax Payload" Standards

Domain Specific Implementation Conventions (subsets & extensions)							
OAGIS	ACORD	XBRL	HL7	EIA-836	PLCS	....	Others

"Semantic Foundation" Standards

ISO/IEC 11179-5, ISO 15000-5, UN Naming and Design Rules

"Syntax Foundation" Standards

W3C – XML, XML Schema

Example Domain Specific Payload Standards

- OAGIS – Open Applications Group http://www.openapplications.org/
  - Participants - ERP and middleware vendors and end users
  - Example payload – purchase order
- HL7 - Health Care http://www.hl7.org/
  - Participants – health care providers across the globe
  - Example payload – health records
- ACORD – XML for the Insurance Industry http://www.acord.org/
  - Participants – insurance providers across the globe
  - Example payload – company insurance claim
- XBRL – Business Reporting - Accounting http://www.xbrl.org/
  - Participants – major accounting firms across the globe
  - Example payload – general ledger and company financial report to SEC
- EIA-836 – Configuration Management Data Exchange and Interoperability http://www.dcnicn.com/cm/index.cfm
  - Participants – DoD and aerospace and defense industry (AIA and GEIA)
  - Example payload – engineering change

ISO/IEC 11179 - Has Six Parts

Part 1: Metadata Registries - Framework

Part 2: Metadata Registries - Classification

Part 3: Metadata Registries - Registry Metamodel and Basic Attributes

Part 4: Metadata Registries - Formulation of Data Definitions

Part 5: Metadata Registries - Naming and Identification Principles

Part 6: Metadata Registries - Registration

http://isotc.iso.ch/livelink/livelink/fetch/2000/2489/Ittf_Home/PubliclyAvailableStandards.htm

Semantics Naming and Identification Choices

Comparing The Two Choices

Comparison Topic	Semantic Web	Metadata Registry
Key Standards	RDF & OWL each with variations	ISO/IEC 11179 – six parts
Domain Specific Payload Standards	Hundreds to thousands	Hundreds to thousands
Primary Scope	Unstructured content on servers	Structured data in databases and back-office applications
Naming Approach	Ontologies with controlled vocabulary (e.g., WordNet)	ISO/IEC 11179-5 based controlled vocabulary
Identification Approach	Definition instance URI	Data Element Concept unique identifier
Primary Benefit	Enable content discovery and inference relationships	Reduce costs of integrating multiple applications & Simplicity

UDEF – A Semantic DNS for Structured Data

Reduce Requirements and Design-Time Phase Semantics Analysis Time and Cost

Common Point-to-Point
Approach --- n(n-1)
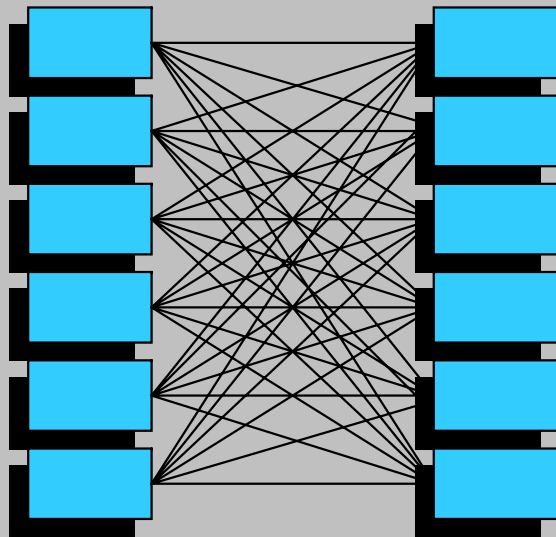
Adopt Global Semantics
Standard Approach --- 2n

Global Semantics Standard

$$\$\$$$

Savings
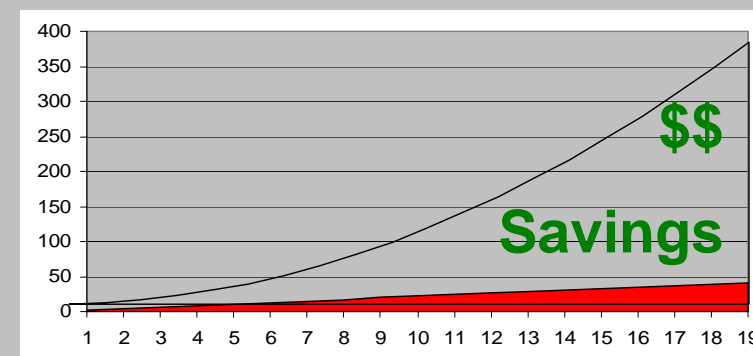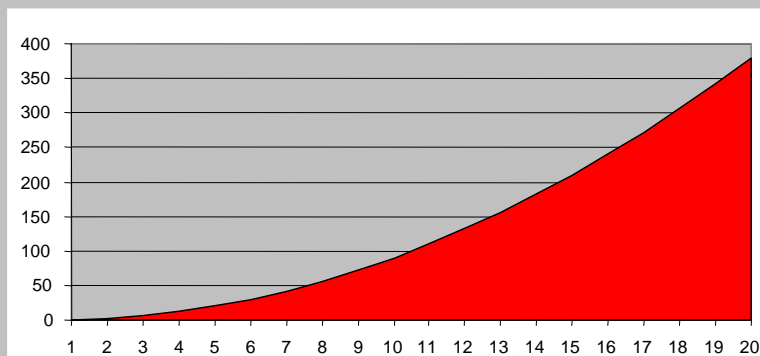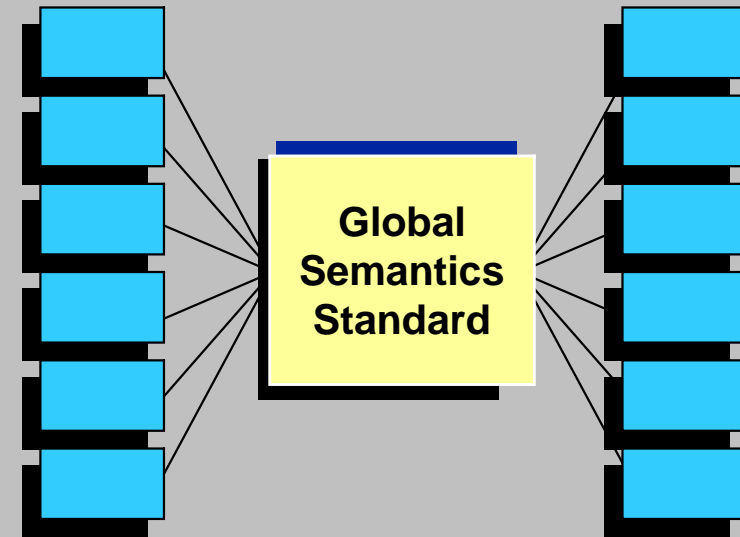
ISO/IEC 11179 Terminology

Object Class

Property

Representation

Data Element Concept

UDEF Maps Data Element Concepts - The Semantics

Data Element

Value Domain

Core Data Element

Application Data Element
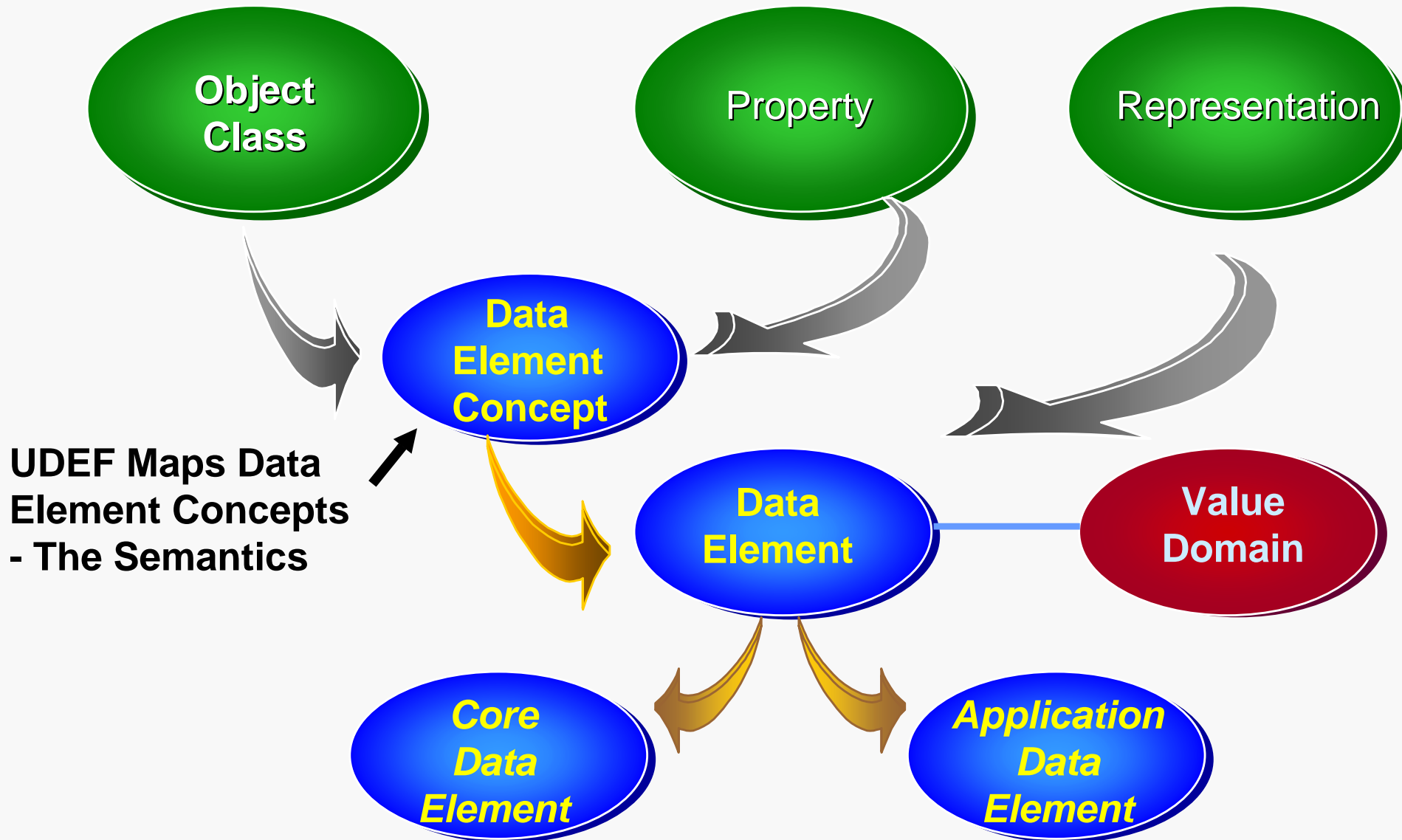
Universal Data Element Framework

UDEF is a proposed universal instantiation of ISO/IEC 11179-5

UDEF Object Class List
- **Entity**
- **Document**
- **Enterprise**
- **Place**
- **Program**
- **Product**
- **Process**
- **Person**
- **Asset**
- **Law-Rule**
- **Environment**
- **Condition**
- **Liability**
- **Animal**
- **Plant**
- **Mineral**
- **Event**

ISO/IEC 11179-5 Naming Convention

Data Element Name

Object Class Term

0...n qualifiers +
1 or more required
Object Class

+

Property Term

0..n qualifiers +
1 required Property
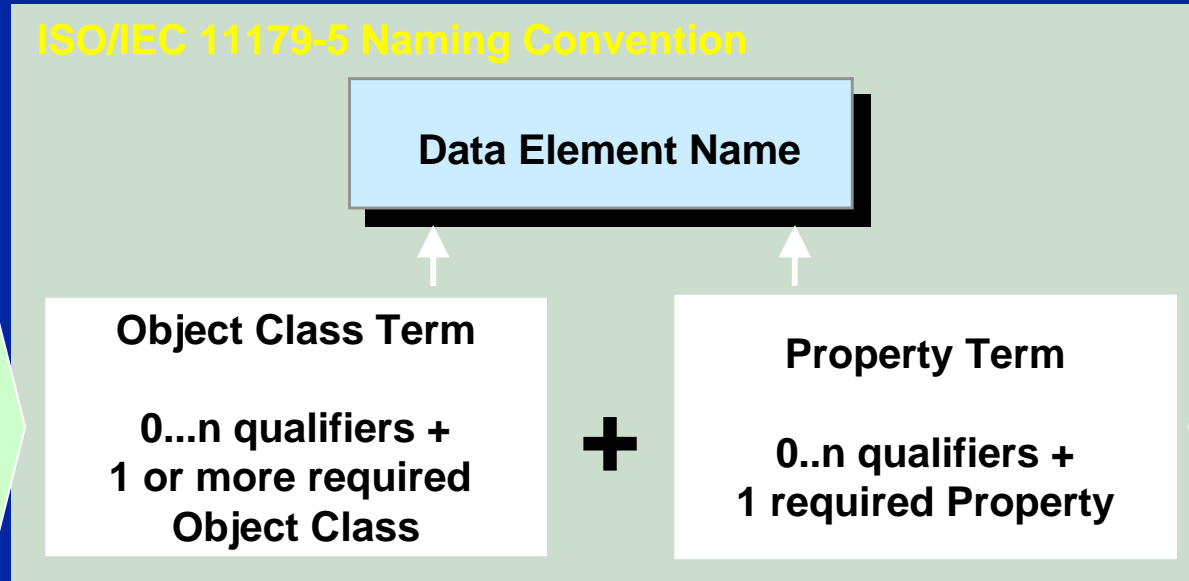
Property List*
- **Amount**
- **Code**
- **Date**
- **Date Time**
- **Graphic**
- **Identifier**
- **Indicator**
- **Measure**
- **Name**
- **Percent**
- **Picture**
- **Quantity**
- **Rate**
- **Text**
- **Time**
- **Value**
- **Sound**
- **Video**

Example UDEF-Based Data Element Concept Names

Document Abstract Text

Enterprise Name

Product Price Amount

Product Scheduled Delivery Date

Engineering Design Process Cost Amount

Patient Person First Name

UDEF names follow the rules of English – qualifiers precede the word they modify

* Based on Tables 8-1 and 8-3 in ISO 15000-5

Taxonomy Based Semantic DNS IDs

UDEF Trees

17 Object Class Trees

18 Property Trees

0
Entity

1
Asset

2
Document

…

t
Order

a
Work

b
Change

c
Technical

d
Purchase

1
Amount

…

4
Code

…

1
Region

…

33
Type

…

68
Language

…
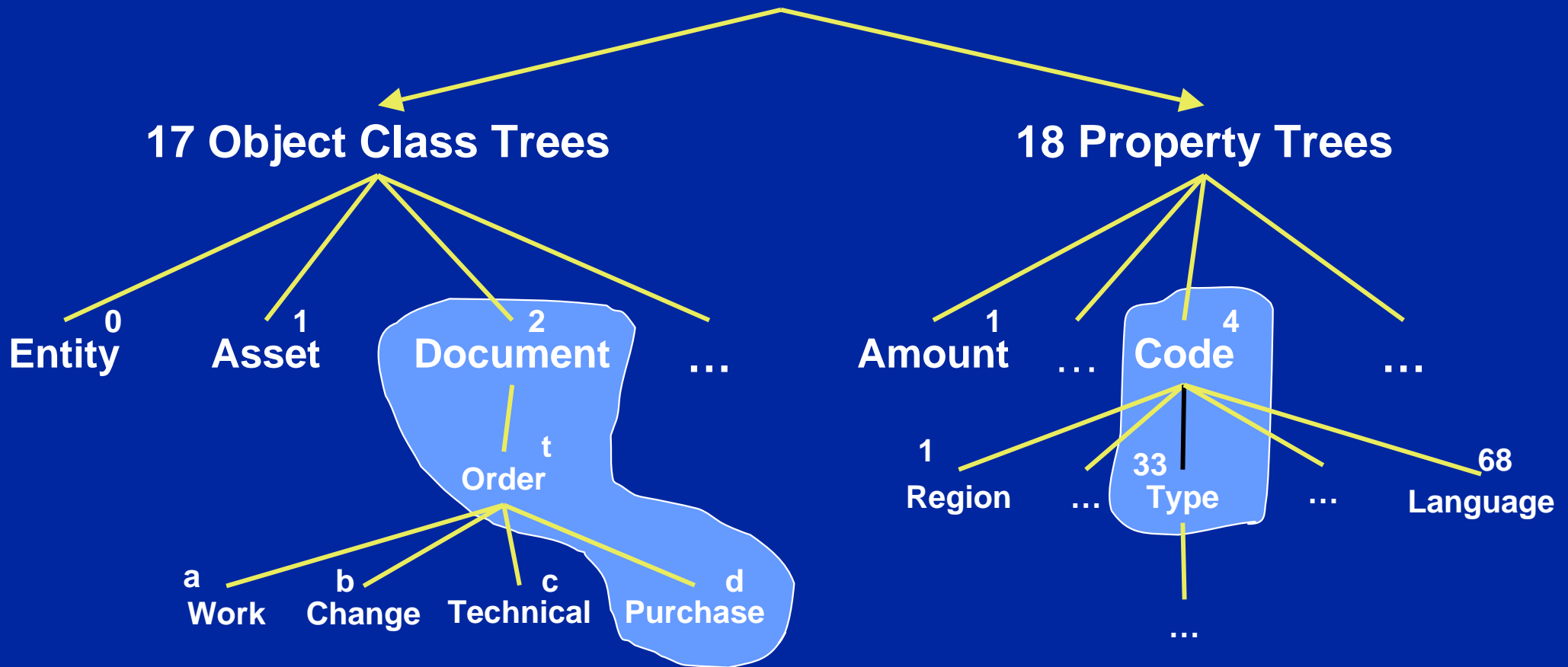
Purchase Order Document_Type Code has UDEF ID = d.t.2_33.4

See http://www.opengroup.org/udefinfo/defs.htm

Mapping Across Standards

Organizations cannot avoid multiple data standards
** Need global semantics standard **

PDM Sys A
Part No

OAGIS 7.1
ItemX

X12 (EDI)
Product/Service ID

STEP AP 203
Product ID

9_9.35.8

PDM Sys B
Part Num

RosettaNet
ProprietaryProductIdentifier

EDIFACT
Item Number

xCBL
PartID
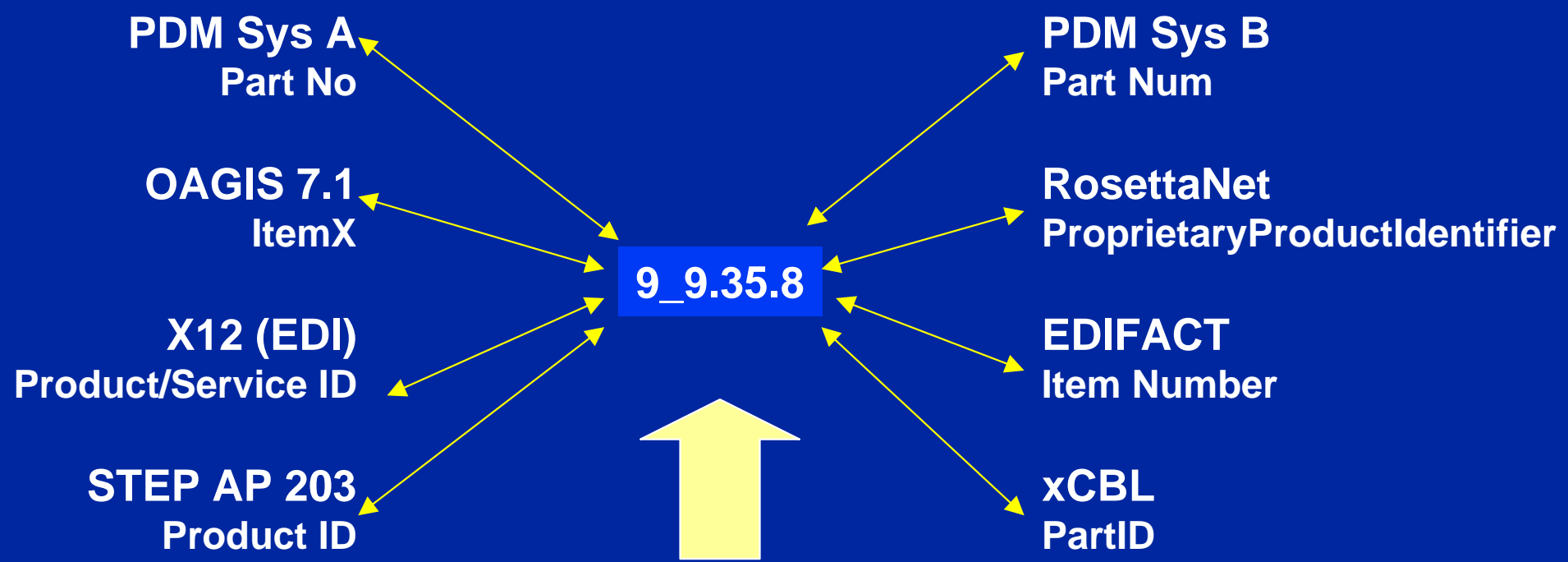
UDEF Universal Identifier

Product(9)_Manufacturer(9).Assigned (35).Identifier(8)

N (N-1) mapping effort instead becomes a 2N mapping effort

Enabling Discovery on Global Scale

Enterprise Metadata Management

Interfaces to Back-Office Systems

Run Time

EAI

Transformation Engines

Std Schema

Vendors with Canonical Models

UDEF-Indexed Metadata Registries

Global Semantics Registry

Use Matrices

Interface Developers

Design Time

- **Data Dictionary**
- **Mapping Matrices**
- **Std XML Schema**

Extend Matrices

Build/Extend Schema

Data Modelers And Apps Developers

UDEF-Indexed Metadata Registry/Repository

Software Vendors with UDEF ID APIs

UDEF Extension Board

Web Public

Centralized metadata registry/repository

- **Enables reuse to reduce costs**
- **Encourages standardization**

Value of Semantic Standard

Typical Interface Build Tasks

Analyze and document the business requirements.

Analyze and document the data interfaces (design time)

    Compare data dictionaries

    Identify gaps

    Identify disparate forms of representation

Perform data transformations as required at run time

    Transform those data that require it
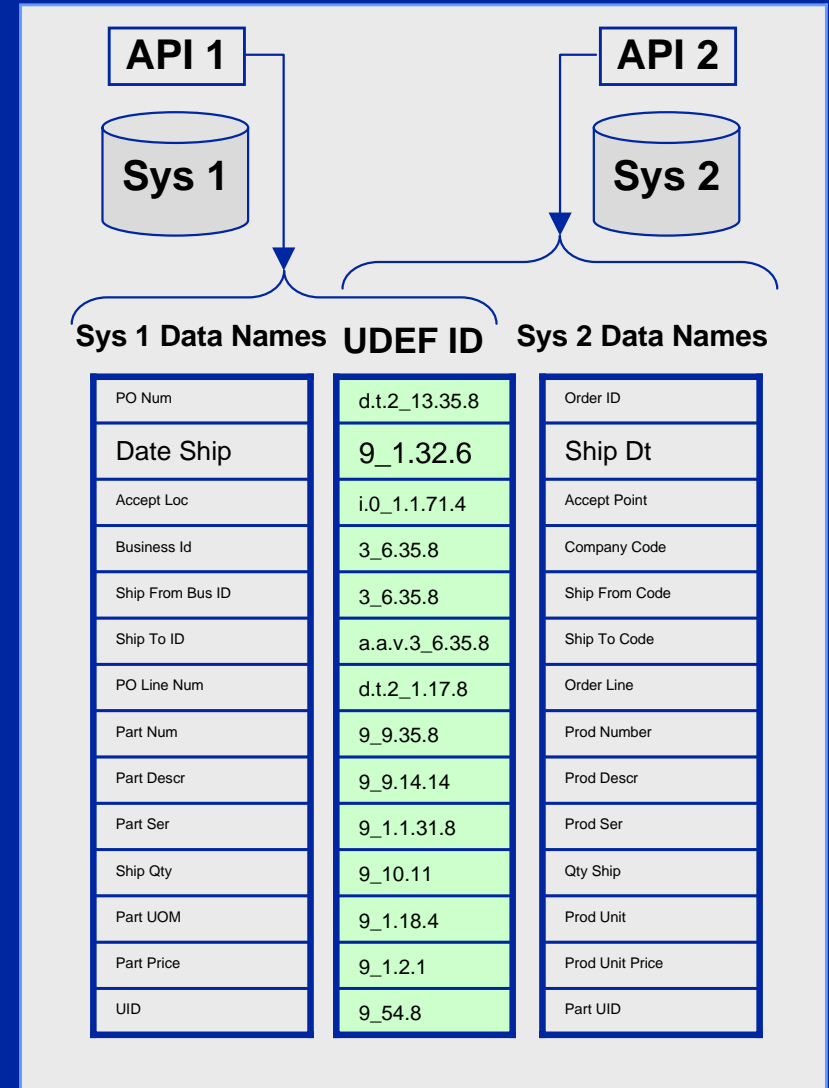
Business Value

Reduces dependency on system expert

Allows automated compare

Reduce design time labor

Step toward automated transform

API 1		API 2
Sys 1		Sys 2

| Sys 1 Data Names	UDEF ID	Sys 2 Data Names
PO Num	d.t.2_13.35.8	Order ID
Date Ship	9_1.32.6	Ship Dt
Accept Loc	i.0_1.1.71.4	Accept Point
Business Id	3_6.35.8	Company Code
Ship From Bus ID	3_6.35.8	Ship From Code
Ship To ID	a.a.v.3_6.35.8	Ship To Code
PO Line Num	d.t.2_1.17.8	Order Line
Part Num	9_9.35.8	Prod Number
Part Descr	9_9.14.14	Prod Descr
Part Ser	9_1.1.31.8	Prod Ser
Ship Qty	9_10.11	Qty Ship
Part UOM	9_1.18.4	Prod Unit
Part Price	9_1.2.1	Prod Unit Price
UID	9_54.8	Part UID

Like A Semantic DNS

UDEF IDs provide global semantic DNS-like indexing mechanism to discover services and data outside the firewall
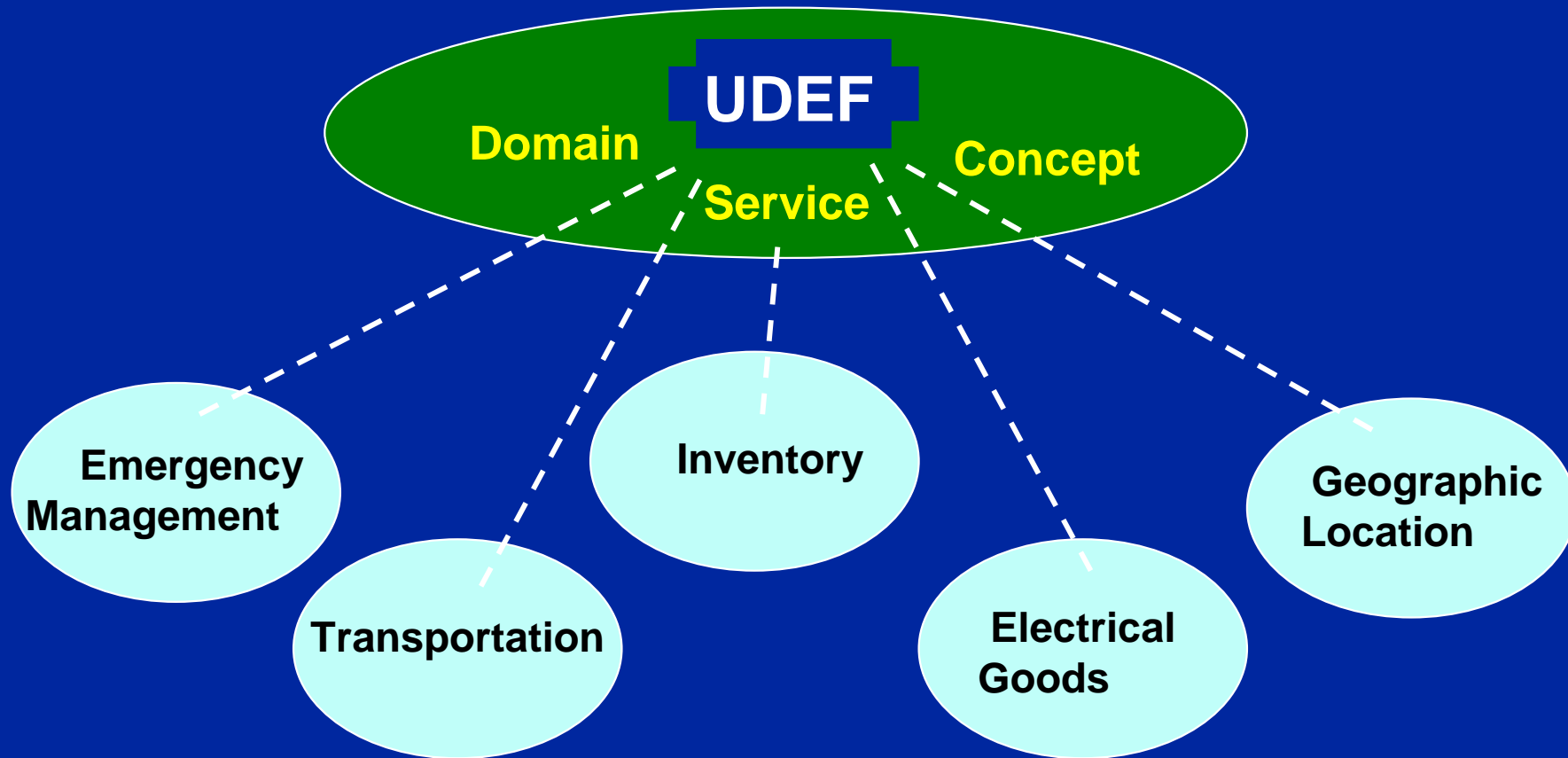
A Few Example Domain Taxonomies

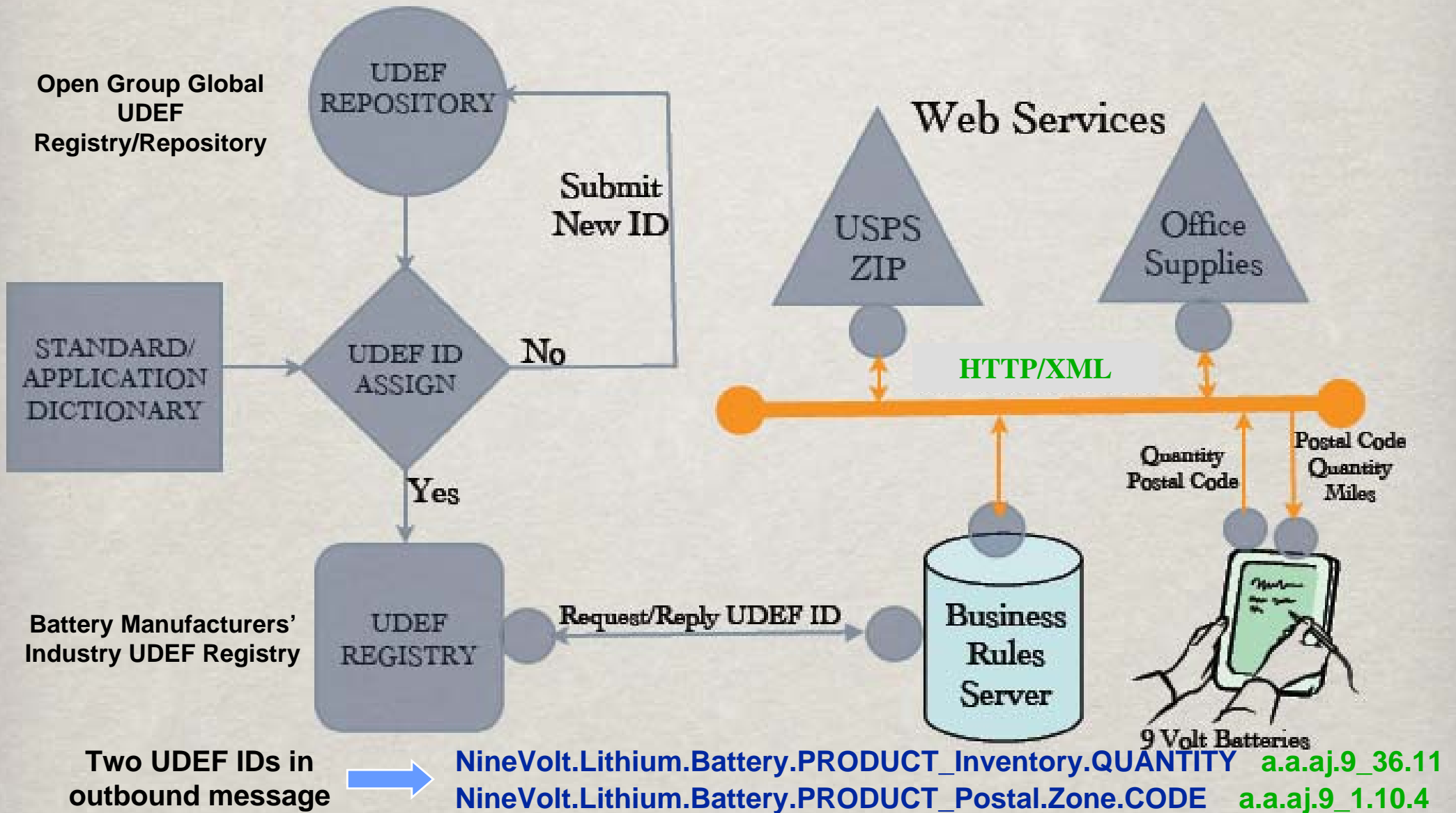# Disaster Response – Example Use Case

Disaster Response Scenario

Natural disaster response team shows up lacking batteries to operate GPS system and walkie-talkie for 200 search and rescue workers – need four hundred 9-volt batteries to even begin the search and rescue effort

- Assumes that UDEF has been adopted globally and that UDEF IDs are exposed at company portals
- Goal – determine if resources might be available nearby within a manufacturer's or supplier's inventory
- Uses two UDEF tags (IDs) to locate available resources in a battery manufacturer's inventory near the response team command center – an ad hoc query since formal interface not previously defined
- Use UDEF ID tags to support semantic integration of disparate procurement applications that use different purchase order semantics
- Two vendors participated – Unicorn and Safyre Solutions

Disaster Response Architecture

Open Group Global
UDEF
Registry/Repository

UDEF
REPOSITORY

Submit
New ID

STANDARD/
APPLICATION
DICTIONARY

UDEF ID
ASSIGN

No

Yes

Web Services

USPS
ZIP

Office
Supplies

HTTP/XML

Quantity
Postal Code

Postal Code
Quantity
Miles

Battery Manufacturers'
Industry UDEF Registry

UDEF
REGISTRY

Request/Reply UDEF ID

Business
Rules
Server

9 Volt Batteries

Two UDEF IDs in
outbound message

NineVolt.Lithium.Battery.PRODUCT_Inventory.QUANTITY   a.a.aj.9_36.11
NineVolt.Lithium.Battery.PRODUCT_Postal.Zone.CODE   a.a.aj.9_1.10.4

Disaster Response Video

Videos of Live Demos

http://www.opengroup.org/udefinfo/demo0511/demos.htm Oct 20, 2005

http://www.opengroup.org/projects/udef/doc.tpl?CALLER=index.tpl&gdid=9189 Dec 1, 2005

For Additional Information

The OPEN GROUP UDEF Forum Web Site

> http://www.opengroup.org/udef/

ISO/IEC 11179 – Specification and standardization of data elements

> http://isotc.iso.ch/livelink/livelink/fetch/2000/2489/Ittf_Home/PubliclyAvailableStandards.htm

Videos of live UDEF Disaster Response Pilot Use Case demo

> http://www.opengroup.org/udefinfo/demo0511/demos.htm **Oct 20, 2005**

> http://www.opengroup.org/projects/udef/doc.tpl?CALLER=index.tpl&gdid=9189 **Dec 1, 2005**

For Possible Follow-up Questions - Contact

Dr. Chris Harding – c.harding@opengroup.org

Ron Schuldt – ron.l.schuldt@lmco.com

Semantics Naming and Identification Issues

Convergence – What Are Some Issues?

Issues Topic	Semantic Web	Metadata Registry
Key Standards	RDF & OWL variations make it difficult to decide best match	Few vendors have adopted ISO/IEC 11179
Domain Specific Payload Standards	Too many overlapping payload standards	Too many overlapping payload standards
Primary Scope	Less suited to structured data in databases and back-office systems	Less suited to unstructured data
Naming Approach	Cross-domain terms that carry different meanings due to different context	Lacks rigor in defining terms
Identification Approach	URI does not help one find the same concept across multiple systems	
Primary Challenge	Each domain needs ontology based vocabulary	Metadata management is a technology that needs greater attention