UDEF – Answering the Unasked Questions

EAPC Johannesburg 2008

Clive Hatton
Real IRM

Based on a presentation by
Ron Schuldt, Lockheed Martin
Chair of the UDEF Project
Agenda

• What is UDEF?
• Why do we need a UDEF?
• Who should care about UDEF?
• How do I use UDEF?
• When will UDEF become mainstream?
• Where can I find out more?
What is UDEF?

- A Universal Data Element Framework
- A global initiative
- A project of the Open Group’s Semantic Interoperability work group
- A common, extensible vocabulary for data
- An instantiation of the naming convention specified by the International Standard for Metadata Registries, ISO/IEC 11179.
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

ISO/IEC 11179 Terminology

Object Class

Property

Representation

Data Element Concept

Data Element

Value Domain

Core Data Element

Application Data Element

UDEF Maps Data Element Concepts - The Semantics
What does UDEF do?

• By design, the UDEF provides a loose coupling semantic DNS-like approach that is infinitely extensible.

• It provides rich semantics for the human users and at the same time provides structured semantic IDs that computers can “interpret” in the same manner as the Internet’s DNS.
What problem is UDEF trying to solve?

- **Semantic interoperability** – exchange of meaning
- **Semantics** – some definitions
  - The science of describing what words mean, the opposite of syntax.
  - The study of meaning in language, including the relationship between language, thought, and behavior.
  - 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.
Primary Purpose of UDEF Standard

- To share common meaning of data across the disparate data standards and applications found throughout the enterprise
- To reduce the design-time cost of building and maintaining applications interfaces
Why do we need a UDEF?

• *To establish a common vocabulary for data* (TOGAF Principle 13)
Chapter 29, Common Vocabulary and Data Definitions

Statement
“Data is defined consistently throughout the enterprise, and the definitions are understandable and available to all users”

Rationale
“…A common vocabulary will facilitate communications and enable dialogue to be effective. In addition, it is required to interface systems and exchange data.”

Implications
“The enterprise must establish the initial common vocabulary for the business. … Ambiguities … must give way to accepted enterprise-wide definitions and understanding. Multiple data standardization initiatives need to be coordinated.”
Who should care about UDEF?

• You should care if you:
  – Are interoperability challenged
  – Are interested in information / data architecture
  – Think you may meet this problem in the future
The Problem - Global Perspective

Each organization is attempting to set its own common vocabulary
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 vocabularies between applications

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
  - XML and associated W3C standards address the syntax requirements but a globally adopted semantics standard does not exist yet

• 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
What is the value of UDEF?

- **UDEF reduces the cost of semantic analysis and data mapping, when implementing data exchange between systems**
To interface two or more applications either within the enterprise or between enterprises ----

- Middleware vendors build and sell “adaptors” for interfaces between major commercial applications (e.g., SAP - Oracle)
  - As the version of a given application changes it typically drives the need for changes to the adaptors

- Application interface developers (data architects) spend considerable time and effort analyzing the semantics of data that need to be exchanged between applications
  - Semantic analysis requires up to 35-40% of the requirements and design-time phase labor cost of building a given interface
  - To minimize the semantic analysis effort typically requires support (availability) from the “experts” of each application since most applications are poorly documented
Goal of Global Semantics Standard

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

Savings
How does UDEF relate to other Semantic Interoperability standards?

- **“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, EDIFACT
Example Domain Specific Payload Standards

  - Participants - ERP and middleware vendors and end users
  - Example payload – purchase order
  - Participants – health care providers across the globe
  - Example payload – health records
  - Participants – insurance providers across the globe
  - Example payload – company insurance claim
  - 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](http://www.dcnicn.com/cm/index.cfm)
  - Participants – DoD and aerospace and defense industry (AIA and GEIA)
  - Example payload – engineering change description
SBVR – a competing standard?

**Semantics of Business Vocabulary and Business Rules**

- **SBVR is a part of the OMG’s Model Driven Architecture**
- **SBVR is intended to formalize complex business rules.**
- **SBVR defines the vocabulary and rules for documenting the semantics of business vocabularies, business facts, and business rules.**
- **Such vocabularies and rules can be interpreted and used by computer systems.**
- **SBVR defines an XMI schema for the interchange of business vocabularies and business rules among organizations and between software tools.**
(When) will UDEF succeed?

• **Something is needed to fill the standards gap for data semantics**
• **Will it be UDEF?**
• **Or something based on UDEF?**
• **Or something else?**
Current Status - UDEF Standard

In May 2005, The Open Group established the UDEF Forum

- Goal - establish the UDEF as the universally-used categorization system for data
  - Develop and maintain the UDEF as an open standard
  - Advocate and promote it
  - Put in place a technical infrastructure to support it
  - Implement a registry for it, and
  - Set up education programs to train information professionals in its use.

Actual UDEF Trees are Viewable Online

http://www.opengroup.org/udefinfo/defs.htm
http://www.opengroup.org/udefinfo/review/defs.htm Draft Baseline

UDEF based Disaster Response Pilot presented to several government audiences – promoted by the Federal SICoP

The Open Group UDEF Forum approved the Baseline UDEF plus the Specification for Extending the UDEF
How do I use UDEF?

• Object vocabulary
• Property vocabulary
• Qualifier vocabulary
Universal Data Element Framework

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

<table>
<thead>
<tr>
<th>UDEF Object Class List</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity</td>
</tr>
<tr>
<td>Document</td>
</tr>
<tr>
<td>Enterprise</td>
</tr>
<tr>
<td>Place</td>
</tr>
<tr>
<td>Program</td>
</tr>
<tr>
<td>Product</td>
</tr>
<tr>
<td>Process</td>
</tr>
<tr>
<td>Person</td>
</tr>
<tr>
<td>Asset</td>
</tr>
<tr>
<td>Law-Rule</td>
</tr>
<tr>
<td>Environment</td>
</tr>
<tr>
<td>Condition</td>
</tr>
<tr>
<td>Liability</td>
</tr>
<tr>
<td>Animal</td>
</tr>
<tr>
<td>Plant</td>
</tr>
<tr>
<td>Mineral</td>
</tr>
<tr>
<td>Event</td>
</tr>
</tbody>
</table>

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

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

* Property List
  - Amount
  - Code
  - Date
  - Date Time
  - Graphic
  - Identifier
  - Indicator
  - Measure
  - Name
  - Percent
  - Picture
  - Quantity
  - Rate
  - Text
  - Time
  - Value
  - Sound
  - Video
Taxonomy Based Semantic DNS IDs

UDEF Trees

17 Object Class Trees
- Entity
- Asset
- Document
  - Order
    - Work
    - Change
    - Technical
    - Purchase
- Amount
- Code
  - Region
  - Type
  - Language

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

See http://www.opengroup.org/udefinfo/defs.htm
Organizations cannot avoid multiple standards & systems

** Need global semantics standard **

N (N-1) mapping effort instead becomes a 2N mapping effort
Enabling Discovery on Global Scale

Enterprise Metadata Management

Run Time
- EAI Transformation Engines

Design Time
- Data Dictionary
- Mapping Matrices
- Std XML Schema

Centralized metadata registry/repository
- Enables reuse to reduce costs
- Encourages standardization

Interfaces to Back-Office Systems

Vendors with Canonical Models
- UDEF-Indexed Metadata Registries

Software Vendors with UDEF ID APIs

Global Semantics Registry

UDEF Extension Board

Web Public

Data Modelers And Apps Developers

Interface Developers

Use Matrices

Build/Extend Schema

Extend Matrices

Build/Extend Schema
**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

### Example Data Mapping

<table>
<thead>
<tr>
<th>Sys 1 Data Names</th>
<th>UDEF ID</th>
<th>Sys 2 Data Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>PO Num</td>
<td>d.t.2_13.35.8</td>
<td>Order ID</td>
</tr>
<tr>
<td>Date Ship</td>
<td>9_1.32.6</td>
<td>Ship Dt</td>
</tr>
<tr>
<td>Accept Loc</td>
<td>i.0_1.71.4</td>
<td>Accept Point</td>
</tr>
<tr>
<td>Business Id</td>
<td>3_6.35.8</td>
<td>Company Code</td>
</tr>
<tr>
<td>Ship From Bus ID</td>
<td>3_6.35.8</td>
<td>Ship From Code</td>
</tr>
<tr>
<td>Ship To ID</td>
<td>a.c.v.3_6.35.8</td>
<td>Ship To Code</td>
</tr>
<tr>
<td>PO Line Num</td>
<td>d.t.2_1.17.8</td>
<td>Order Line</td>
</tr>
<tr>
<td>Part Num</td>
<td>9_9.35.8</td>
<td>Prod Number</td>
</tr>
<tr>
<td>Part Ser</td>
<td>9_1.13.8</td>
<td>Prod Ser</td>
</tr>
<tr>
<td>Ship Qty</td>
<td>9_10.11</td>
<td>Qty Ship</td>
</tr>
<tr>
<td>Part UOM</td>
<td>9_1.18.4</td>
<td>Prod Unit</td>
</tr>
<tr>
<td>Part Price</td>
<td>9_1.2.1</td>
<td>Prod Unit Price</td>
</tr>
<tr>
<td>UID</td>
<td>9_54.8</td>
<td>Part UID</td>
</tr>
</tbody>
</table>
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
A Typical Semantic Interoperability Challenge and a UDEF Solution Approach
### Semantic Alignment Matrix - Example

<table>
<thead>
<tr>
<th>UDEF ID</th>
<th>EIA-836</th>
<th>X12 (EDI)</th>
<th>Application A</th>
</tr>
</thead>
<tbody>
<tr>
<td>3_6.35.8</td>
<td>CAGE Code</td>
<td>Entity Identifier Code + Identification Code Qualifier</td>
<td>Company ID</td>
</tr>
<tr>
<td>9_9</td>
<td>Product Name</td>
<td>Product/Service Name</td>
<td>Supplier</td>
</tr>
<tr>
<td>y.3_9</td>
<td></td>
<td>Entity (Supplier) Name</td>
<td></td>
</tr>
<tr>
<td>e.2_8</td>
<td>Contract Document Identifier</td>
<td>Buyer’s Contract Number</td>
<td>Contract No</td>
</tr>
<tr>
<td>9_11</td>
<td>Product Quantity</td>
<td>Quantity Ordered</td>
<td>Item Qty</td>
</tr>
<tr>
<td>2_33.4</td>
<td>Document Type Code</td>
<td>Report Type Code</td>
<td>Doc Type</td>
</tr>
</tbody>
</table>

&lt;ContractDocumentIdentifier GUID="e.2_8">12345&lt;/ContractDocumentIdentifier&gt;

&lt;BuyersContractNumber GUID="e.2_8">12345&lt;/BuyersContractNumber&gt;

&lt;ContractNo GUID="e.2_8">12345&lt;/ContractNo&gt;
Substantial time and effort is required to integrate a new supplier’s order management system with a customer’s procurement system. This vocabulary gap problem results in a labor-intensive effort that is typically measured in days or weeks.
The UDEF Approach

Map Both Systems and/or Both Standards Independently to the UDEF

OAG Purchase Order:

```xml
<POID GUID="d.t.2_8">ZT8923</POID>
<POTYPE GUID="d.t.2_33.4">224</POTYPE>
<ACKREQUEST>1</ACKREQUEST>
<CONTRACTB GUID="e.2_8">94-A17</CONTRACTB>
<TAXEXMT GUID="q.3_1.4.9.4">FOR RESALE</TAXEXMT>
- <PARTNER>
  <NAME GUID='q.3_10' index='1'>RUNNER UNLIMITED</NAME>
  <ONETIME>0</ONETIME>
  <PARTNRID GUID='q.3_8'>A123</PARTNRID>
  <PARTNRTYPE GUID='a.at.3_1.37.4'>SoldTo</PARTNRTYPE>
  <TAXID GUID='e.2_8'>NT132765</TAXID>
- <ADDRESS>
  <ADDRTYPE>BUYER</ADDRTYPE>
  <CITY GUID='q.3_1.1.10.10'>PINACLE</CITY>
  <POSTALCODE GUID='q.3_1.1.10.4'>87345</POSTALCODE>
  <STATEPROVN GUID='q.3_1.1.37.10'>NB</STATEPROVN>
</ADDRESS>
```

xCBL Purchase Order:

```xml
<?xml version="1.0" encoding="UTF-8" ?>
- <Order>
  - <OrderHeader>
    - <OrderNumber>
      <BuyerOrderNumber GUID="d.t.2_8">ZT8923</BuyerOrderNumber>
    </OrderNumber>
    <OrderIssueDate GUID='d.t.2_5'>19991108T23:59:59</OrderIssueDate>
  </OrderHeader>
  - <OrderLine>
```

Page 33
The Online Vocabulary Gap Solution

Prototype Demonstrated live May 2003 at an EIDX Conference

Vocabulary Gap Software
On-the-Wire
Java and XSLT

Source XML Instance
Doc with UDEF IDs

Target XML Instance
Doc with UDEF IDs

SOAP

DHTML
On-line Gap Analysis Report

Source System

Target System
Vocabulary Gap Report < 1 Second

<table>
<thead>
<tr>
<th>OAGIS Standard</th>
<th>xCBL Standard</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>9_13.11</td>
<td>QUANTITY</td>
</tr>
<tr>
<td>9_2.13.11</td>
<td>QuantityValue</td>
</tr>
<tr>
<td>a.v.3.1.37.4</td>
<td>PART:NTYPE</td>
</tr>
<tr>
<td>a.v.3.1.10.10</td>
<td>CITY</td>
</tr>
<tr>
<td>a.v.3.1.10.4</td>
<td>POSTALCODE</td>
</tr>
<tr>
<td>a.v.3.1.37.10</td>
<td>STATEPROV:</td>
</tr>
<tr>
<td>a.v.3.10</td>
<td>NAME</td>
</tr>
<tr>
<td>a.v.3.14.35.8</td>
<td>null</td>
</tr>
<tr>
<td>a.v.3.12.14</td>
<td>ADDR:LINE</td>
</tr>
<tr>
<td>a.v.3.36.4</td>
<td>null</td>
</tr>
<tr>
<td>a.v.3.4.12.14</td>
<td>null</td>
</tr>
<tr>
<td>a.v.3.5.37.4</td>
<td>PART:NRID</td>
</tr>
<tr>
<td>a.g.3.1.10.10</td>
<td>CITY</td>
</tr>
<tr>
<td>a.g.3.1.10.4</td>
<td>POSTALCODE</td>
</tr>
<tr>
<td>a.g.3.1.37.10</td>
<td>STATEPROV:</td>
</tr>
<tr>
<td>a.g.3.10</td>
<td>NAME</td>
</tr>
<tr>
<td>a.g.3.12.14</td>
<td>ADDR:LINE</td>
</tr>
<tr>
<td>a.g.3.36.4</td>
<td>null</td>
</tr>
<tr>
<td>a.g.3.4.12.14</td>
<td>null</td>
</tr>
<tr>
<td>a.g.3.5.37.4</td>
<td>null</td>
</tr>
<tr>
<td>d.t.2.11.17.8</td>
<td>POL:LINE:NUM</td>
</tr>
<tr>
<td>d.t.2.33.4</td>
<td>F:OTYPE</td>
</tr>
<tr>
<td>d.t.2.5</td>
<td>DATE:TIME</td>
</tr>
<tr>
<td>d.t.2.8</td>
<td>POID</td>
</tr>
</tbody>
</table>
Sample UDEF Mapping Matrix Online

Maps UDEF to
X12, EDIFACT, OAGIS, RosettaNet and xCBL
for a Typical Purchase Order

http://eidx.comptia.org/guidelines/xref_download.aspx
For Additional Information

The OPEN GROUP UDEF Forum Web Site
http://www.opengroup.org/udef/

ISO/IEC 11179 – Specification and standardization of data elements

Join the Semantic Interoperability Stream

For Possible Follow-up Questions - Contact

Dr. Chris Harding – c.harding@opengroup.org
Ron Schuldt – ron.l.schuldt@lmco.com
Clive Hatton – clive.hatton@realirm.com