



The Integration of Taxonomies Using Ontology Structures

Presented by:

Gerald M. Radack, Ph.D.

Principal Computer Scientist

Concurrent Technologies Corp. (*CTC*)

Project Team and Responsibilities



- ◆ Lehigh University
 - Project Management
 - Ontology Development
 - Taxonomy Integration
 - Compiler Development
- ◆ ECCMA
 - Development of terms and definitions
 - Inclusion of taxonomy terminology in eOTD
 - Facilitation of OWL output from eOTD
- ◆ *CTC*
 - Technical guidance on data modeling, Federal Catalog System (FCS), eOTD core model, and ISO 22745
 - ISO/IEC JTC1/SC32 metadata standards
 - ISO TC37 terminology standards

Outline



- ◆ Problem Statement
- ◆ Overall Project Objective
- ◆ Approach
- ◆ Progress to Date
- ◆ Integration of Taxonomies
- ◆ Challenges
- ◆ Remaining Tasks
- ◆ Example and Demonstration

Problem Statement



- ◆ Relate Government/Commercial Item Descriptions/Taxonomies to Supplier Capabilities
 - DLIS must be able to integrate DOD item taxonomies with commercial systems
 - Commercial cataloging systems have domain-specific taxonomies that are not interconnected
 - Information is stored in different, often proprietary, and incompatible formats



The Situation

- ◆ Data elements have to be translated and transferred manually among different systems
- ◆ The result is a significant cost of time and money



Objective

- ◆ To align the FCS taxonomy with supplier taxonomies so as to provide DLIS with full access to suppliers and their products



Approach

- ◆ Use the ECCMA Open Technical Dictionary (eOTD) as the basis for the integration of taxonomies
- ◆ The eOTD is the industry version of the FCS and seeks recognition as the international standard for e-catalogs via the ISO 22745 designation



Other Taxonomies

- ◆ United Nations Standard Products and Services Code (UNSPSC)
- ◆ Standardized Material and Service Classification (ecl@ss)
- ◆ Common Procurement Vocabulary (CPV)
- ◆ North American Industry Classification System (NAICS)
- ◆ ISO 13584 Parts Library (PLIB) Part 511 Fasteners Dictionary

Mapping between Taxonomies



- ◆ Mapping keywords is insufficient
 - one-to-one correspondences aren't always possible
 - overlapping classes
 - » functional vs. compositional taxonomies
 - » e.g. grinding machine:cutlery (0161-1#01-007071#1) vs. grinding machine:carbide tool bit (0161-1#01-007098#1)
- ◆ Need deeper analysis
- ◆ Focus on systemic classification of attributes that connects communities of information
- ◆ This connection can be established with the Semantic Web



The Semantic Web

◆ Definition

- The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.
(Berners-Lee et al., Scientific American, May 2001)

◆ Key International Standards

- World Wide Web Consortium (W3C) recommendations
 - » Resource Description Framework (RDF)
 - » Web Ontology Language (OWL)
- Same status as HTML / XML

Usage of the Semantic Web



- ◆ 2.4 million RDF/OWL documents on the Web (as of March 2008)
 - LiveJournal automatically creates data using the popular Friend-of-a-Friend (FOAF) schema for every user
 - Yahoo! Search announced support for Semantic Web standards
- ◆ Open source Semantic Web tools
 - from IBM, Hewlett-Packard, Nokia, etc.
- ◆ Commercial software vendors
 - Oracle 11g DBMS supports RDF and OWL
 - Adobe's products use RDF to provide metadata for documents, photos
 - Semantic Web specific companies: TopQuadrant, Aduna Software, etc.
- ◆ Companies that have invested in Semantic Web solutions
 - Boeing: integrate work of partners involved in airplane design
 - Chevron: manage the life cycle of power plants and oil refineries
 - British Telecom: online service to help vendors work together
 - Harper's Magazine: web site that connects articles to events on a timeline
 - Many more...

Ontology



◆ Definition

- A logical theory that accounts for the intended meaning of a formal vocabulary (Guarino 98)
- Has a formal syntax and unambiguous semantics
- Usually more than just a hierarchy / taxonomy
- Inference algorithms can compute what logically follows

◆ Relevance to Web

- Identify context
- Provide shared definitions
- Eases the integration of distinct resources

OWL



- ◆ **Web Ontology Language**
 - data is linked to ontologies that provide semantics
 - syntax based on Resource Description Framework (RDF)
 - » but richer semantics
 - » can describe more complicated relationships
 - semantics based on description logic (DL)
- ◆ **Description logics**
 - form of knowledge representation
 - useful for formally defining classes
 - studied extensively in 1990's
 - mature reasoning software (FaCT, RACER, etc.)
- ◆ **Benefits**
 - optimized computation of subsumption
 - » calculate implicit subClassOf relations
 - ontology integration
 - » if two ontologies use class expressions to define their vocabularies in terms of a third ontology, then subsumption can be used to compute an integrated ontology

OWL Class Constructors



Constructor	DL Syntax	Example
intersectionOf	$C1 \cap C2$	GasTurbine \cap AircraftPart
unionOf	$C1 \cup C2$	Door \cup Airframe \cup TailSection
complementOf	$\neg C$	\neg Aircraft
oneOf	$\{x1, \dots, x2\}$	{F15, F16}
allValuesFrom	$\forall P.C$	\forall partOf.Airframe
someValuesFrom	$\exists P.C$	\exists hasPart.Fastener
maxCardinality	$\leq nP$	≤ 10 hasPart
minCardinality	$\geq nP$	≥ 2 hasPart

OWL Axioms



Axiom	DL Syntax	Example
subClassOf	$C_1 \subseteq C_2$	Aircraft \subseteq System \cap Aerospace
equivalentClass	$C_1 \equiv C_2$	Helicopter \equiv Aircraft \cap RotaryWing
disjointWith	$\neg C_2$	RotaryWing \subseteq \neg FixedWing
sameIndividualAs	$\{x_1\} \equiv \{x_2\}$	{washer} \equiv {shim}
differentFrom	$\neg\{x_2\}$	{bank:Blood} \subseteq \neg {bank:Capacitor}
subPropertyOf	$P_1 \subseteq P_2$	hasMaximumWeight \subseteq hasWeight
equivalentProperty	$P_1 \equiv P_2$	weight \equiv displacement
inverseOf	$P_1 \equiv P_2^{-1}$	partOf \equiv hasPart ⁻¹
transitiveProperty	P^+	predecessor ⁺ \subseteq predecessor
functionalProperty	$T \leq 1P$	$T \leq 1$ hasNSN
inverseFunctionalProperty	$T \leq 1P^{-1}$	$T \leq 1$ hasPartNumber ⁻¹

Is-a Hierarchies



- ◆ OWL defines strict **is-a** hierarchies
 - A `rdfs:subClassOf` B is interpreted as “Every A is a B ”
- ◆ Many product taxonomies are not is-a hierarchies
 - They were created to support purchasing
 - eCl@ss example:

27 Electric engineering, automation, process control engineering

27-05 Accumulator, battery

27-05-01 Station. batt., accum.

27-05-02 Traction battery, starter battery

27-05-04 Portable battery

27-05-06 Battery charger

27-05-90 Accumulator, battery (other)

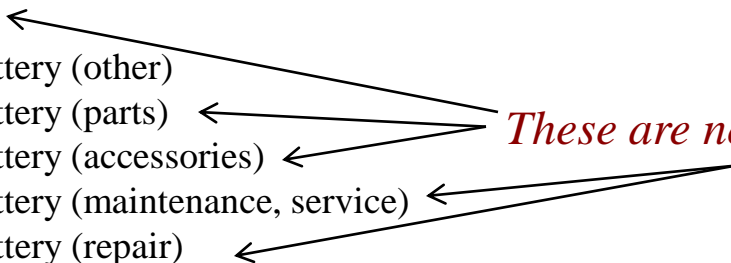
27-05-91 Accumulator, battery (parts)

27-05-92 Accumulator, battery (accessories)

27-05-98 Accumulator, battery (maintenance, service)

27-05-99 Accumulator, battery (repair)

These are not batteries!



Overall Project Plan



- ◆ Select some representative domains
- ◆ Select some representative target taxonomies
- ◆ Create an OWL ontology for the eOTD
- ◆ Create an OWL ontology for the FCS
- ◆ Create OWL ontologies for the target taxonomies
- ◆ Demonstrate ontology integration
- ◆ Create OWL wrappers for the target taxonomies
- ◆ Develop ontology translation compiler

Project Scope



- ◆ Selected Domains for Taxonomy Mapping
 - DLIS specified a set of domains for the project consisting of:
 - » 71 INCs/AINs
 - ◆ selected Batteries, Bearings, Bushings, Fasteners, Gaskets and Microcircuits
 - » 17 FIIGs
 - » 14 FSCs applicable to 490,825 NSNs
 - later added FIIG A003B (BOLTS AND SCREWS) for 57 more INCs/AINs
- ◆ Selected taxonomies
 - United Nations Standard Products and Services Code (UNSPSC)
 - Standardized Material and Service Classification (eCl@ss)
 - Common Procurement Vocabulary (CPV)
 - PLIB Part 511 Fastener Taxonomy
 - » Replaced North American Industry Classification System (NAICS) at DLIS request

Breadth and Depth Analysis



- ◆ Purpose: Examine how the 71 INCs map to each of the four target taxonomies
- ◆ Analysis of each target taxonomy
 - Sorted by descending taxonomic levels
 - Counted number of matches per level
 - Showed how many instances (members) there were in each class.
- ◆ In the form of Excel spreadsheets and Word documents

Primary Ontologies



- ◆ FCS ontologies
 - 71 AINs and their characteristics
- ◆ eOTD ontology
 - same set of AINs
 - use eOTD identifiers and terminology
- ◆ Baseline versions delivered
- ◆ Both ontology sets continue to be refined during the project

Target Ontologies



- ◆ eCl@ss
- ◆ CPV
- ◆ UNSPSC
- ◆ ISO 13584-511 Fasteners dictionary

FCS Ontologies



- ◆ fcs-top.owl
 - covers all FSGs and FSCs
 - 735 OWL classes
 - » with fsgCode or fscCode, rdfs:subClassOf, rdfs:label , rdfs:comment
 - » derived from Tables 076 and 316
- ◆ fcs-ont.owl
 - covers 71 selected AINs, their mandatory MRCs
 - 71 OWL classes
 - » with inc, rdfs:subClassOf, rdfs:label , rdfs:comment
 - » minCardinality = 1 for mandatory MRCs
 - » derived from Tables 98 , 99 and 124
 - 385 OWL properties
 - » with mrc, rdfs:label , rdfs:comment
 - » derived from Tables 127 and 347

OWL in RDF/XML format



```
<owl:Class rdf:ID="MicrocircuitDigitalLinear">
  <fcs:inc rdf:datatype="&xsd:string">20574</fcs:inc>
  <rdfs:label>MICROCIRCUIT,DIGITAL-LINEAR</rdfs:label>
  <rdfs:comment rdf:datatype="&xsd:string">
    A microcircuit having the dual functions of a MICROCIRCUIT (1),
    DIGITAL and a MICROCIRCUIT (1), LINEAR.
  </rdfs:comment>
  <rdfs:subClassOf rdf:resource="&fsc;MicrocircuitsElectronic"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:minCardinality
        rdf:datatype="&xsd;nonNegativeInteger">1</owl:minCardinality>
      <owl:onProperty rdf:resource="outputLogicForm" />
    </owl:Restriction>
  </rdfs:subClassOf>
  ...
</owl:Class>
```

The RDF ID cannot have spaces or commas

Item Name Code

Approved Item Name

Definition

Subclass of FSC 5962

MRC CQWX is mandatory

OWL File in Protege



The screenshot displays the Protege 3.3.1 interface. The title bar shows the file path: (file:\C:\Documents%20and%20Settings\Donald%20Hillman\My%20Documents\fcs-ont.pprj, OWL / RDF Files). The menu bar includes File, Edit, Project, OWL, Code, Tools, Window, and Help. The toolbar contains various icons for file operations and editing.

The main workspace is divided into several panes:

- SUBCLASS EXPLORER:** Shows the asserted hierarchy for the project 'fcs-ont'. The root is 'owl:Thing', which has several subclasses listed, including 'fcs:BatteriesNonrechargeable', 'fcs:BatteriesRechargeable', 'fcs:BearingsAntifrictionUnmounted', 'fcs:BearingsPlainUnmounted', 'fcs:BushingsRingsShimsAndSpacers', 'fcs:FasteningDevices', 'fcs:HardwareCommercial', 'fcs:HosePipeTubeLubricationAndRailingFittings', 'fcs:KnobsAndPointers', 'fcs:MicrocircuitsElectronic', 'fcs:O-Ring', 'fcs:PackingAndGasketMaterials', 'fcs:ProductionJigsFixtures.AndTemplates', and 'fcs:Screws'.
- CLASS EDITOR:** Currently editing the class 'owl:Thing'. It shows a table for properties and values:

Property	Value	Lang
rdfs:comment		

Below the table, there are sections for 'Asserted Conditions' (with radio buttons for 'NECESSARY & SUFFICIENT' and 'NECESSARY') and 'Disjoints'.

The bottom status bar shows the current view is 'Logic View' and the time is 5:10 PM. The Windows taskbar at the very bottom shows the Start button and several open applications, including FaCT++-win-v1.1.11, Protege.exe, and fcs-ont Protégé 3.3....

eOTD Ontology



- ◆ Covers
 - 71 (original) + 57 (from FIIG A003B) AINs
 - espis for corresponding mandatory MRCs
- ◆ 31 new abstract classes
 - eOTD has no taxonomy, but some grouping (e.g. Bolt, Microcircuit, etc.) is helpful
- ◆ 159 OWL classes
 - esci, rdfs:label, rdfs:comment, rdfs:subClassOf
 - Properties with minCardinality = 1
 - Strict is-a classification, multiple parents allowed
 - Disjointness is used sparingly
- ◆ 161 OWL properties
 - espi, rdfs:label, rdfs:comment

Abstract Classes



- ◆ ASSEMBLY, BATTERY can be regarded as a subclass of ASSEMBLY (i.e. ASSEMBLY, BATTERY is a ASSEMBLY)
- ◆ Leads to the introduction of abstract classes, i.e. classes that strictly speaking do not exist in the scheme under consideration, but which are required for taxonomic purposes. Its major difficulty is that it cannot be instantiated, i.e. it does not refer to anything concrete in the domain.

Guidelines for Abstract Classes



- ◆ Remove one or more modifiers from AINs
 - “Bearing,Roller,Self-Aligning” → Bearing
 - “Bearing,Roller,Self-Aligning” → BearingRoller
- ◆ Identify “foundational” classes from FSGs and FSCs
 - Foundational classes focus on a single common feature
 - Examples

FSCs	Abstract Classes
“Bearings, Antifriction, Unmounted” (3110) “Bearings, Plain, Unmounted” (3120) “Bearings, Mounted” (3130)	BearingsAntifriction BearingsPlain BearingsMounted BearingsUnmounted
“Bushings, Rings, Shims, and Spacers” (5365)	Bushings Rings Shims Spacers

FCS-eOTD Ontology



- ◆ Links the FCS to the eOTD
 - Each AIN has a corresponding eOTD class
 - » owl:equivalentClass
 - Each MRC has a corresponding eOTD property
 - » owl:equivalentProperty
 - Expressions to map FSGs and FSCs (if applicable)
 - » $\text{fsc:BearingsPlainUnmounted} \equiv \text{eotd:BearingPlain} \cap \text{eotd:BearingUnmounted}$
 - » $\text{fsc:BushingsRingsShimsAndSpacers} \equiv \text{eotd:Bushing} \cup \text{eotd:Ring} \cup \text{eotd:Shim} \cup \text{eotd:Spacer}$

Mapping Issues



- ◆ FCS is not always a strict is-a hierarchy
 - “Battery Assembly” (AIN:00152) has FSC 6140 (Batteries, Rechargeable)
 - But every BATTERY ASSEMBLY is not a BATTERY (it’s two or more batteries plus a common mounting!)
 - **Solution:** $eotd:BatteryRechargeable \equiv fsc:BatteriesRechargeable \cap \neg ain:BatteryAssembly \cap \dots$
- ◆ FCS sometimes uses a general term when a subset is intended
 - “Gasket” (AIN:04522) and “Gasket, Spiral Wound” (AIN: 37840)
 - “Gasket” is disjoint with “Gasket, Spiral Wound”
 - AIN 04522 refers to Gaskets otherwise unclassified
 - **Solution:** $ain:Gasket \subseteq eotd:GasketAll$, $ain:GasketSpiralWound \subseteq eotd:GasketAll$, $ain:Gasket \subseteq \neg ain:GasketSpiralWound$

Mapping to Target Taxonomies



◆ Guidelines

- Find an eOTD expression for each term in the target taxonomy
- Find a target expression for each term in the eOTD
- Prefer owl:equivalentClass to owl:subClassOf
- Treat classes in procurement taxonomies as *x-and-related-things*
 - » To maintain strict is-a relationships
- Use $A \cap \neg B$ pattern to exclude things that are not strictly is-a

UNSPSC Ontology



◆ Overview

- v. 11.1
- Source: Excel spreadsheet
 - » only codes and labels, no properties
- Four level taxonomy (21536 concepts)

◆ OWL File Coverage

- All top level concepts
- Full details in relevant areas
- 152 classes
 - » unspsc:code, rdfs:label, rdfs:subClassOf

eCl@ss Ontology



◆ Overview

- v. 6.0
- Source: website, eCl@ss refused to provide a machine-readable copy
- Four level taxonomy

◆ OWL File Coverage

- 88 classes
 - » segment, mainGroup, group, or commodityClass
 - » rdfs:label, rdfs:subClassOf
- Only includes classes relevant to project
- Taxonomy is procurement related

CPV Ontology



◆ Overview

- Source: Excel spreadsheet
 - » only codes and names, no properties or definitions
- Variable depth taxonomy (up to 7 levels)
- 3224 concepts

◆ OWL File Coverage

- 76 classes
 - » code, rdfs:label, rdfs:subClassOf
- Only includes classes relevant to project
- Taxonomy is procurement related

ISO 13584-511 Ontology



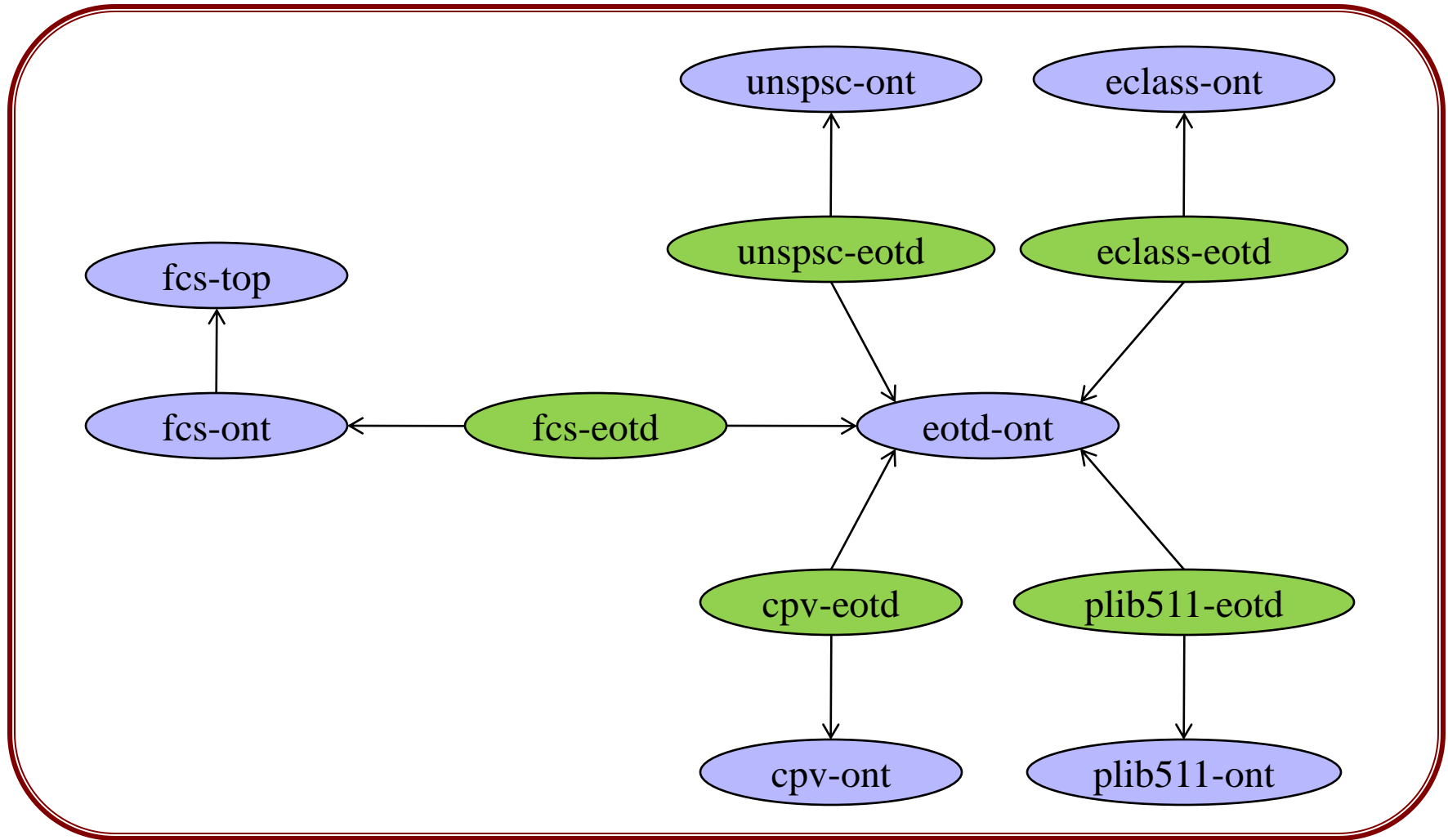
◆ Overview

- ISO 13584-25 is a data model for parts libraries (PLIB); ISO 13584-511 is the Fasteners dictionary, maintained by China National Institute of Standardization (CNIS)
- Source: PLIB file, converted to XML by *CTC*
- 180 classes, 69 feature classes, 200 properties

◆ OWL File Coverage

- All classes except feature classes
 - » code, rdfs:label, rdfs:comment, rdfs:subClassOf
- All properties
 - » code, rdfs:label, rdfs:comment,

Ontologies Overview



Challenges



- ◆ Taxonomies differ in scope and purpose
- ◆ Naming conventions differ across classifications
 - e.g. “bearing, roller” versus “roller bearing
- ◆ Target taxonomies have deficiencies:
 - lack of definitions or inaccurate definitions
 - lack of freely available electronic version
 - lack of sample data
 - poor superclass/subclass structures
 - inconsistent modeling
 - failure to state/observe modeling conventions

Integration Demonstration



- ◆ Use a DL reasoner to integrate FCS ontology with the target ontologies
 - Compute which target classes are implicit subclasses of FCS classes (subsumption)
 - Automatically “merges” two taxonomies
- ◆ Intermediate validation
 - Quality of integration depends on quality of definitions in Tasks 2 – 4
 - May lead to refinement of earlier DL descriptions

Wrappers for Targets



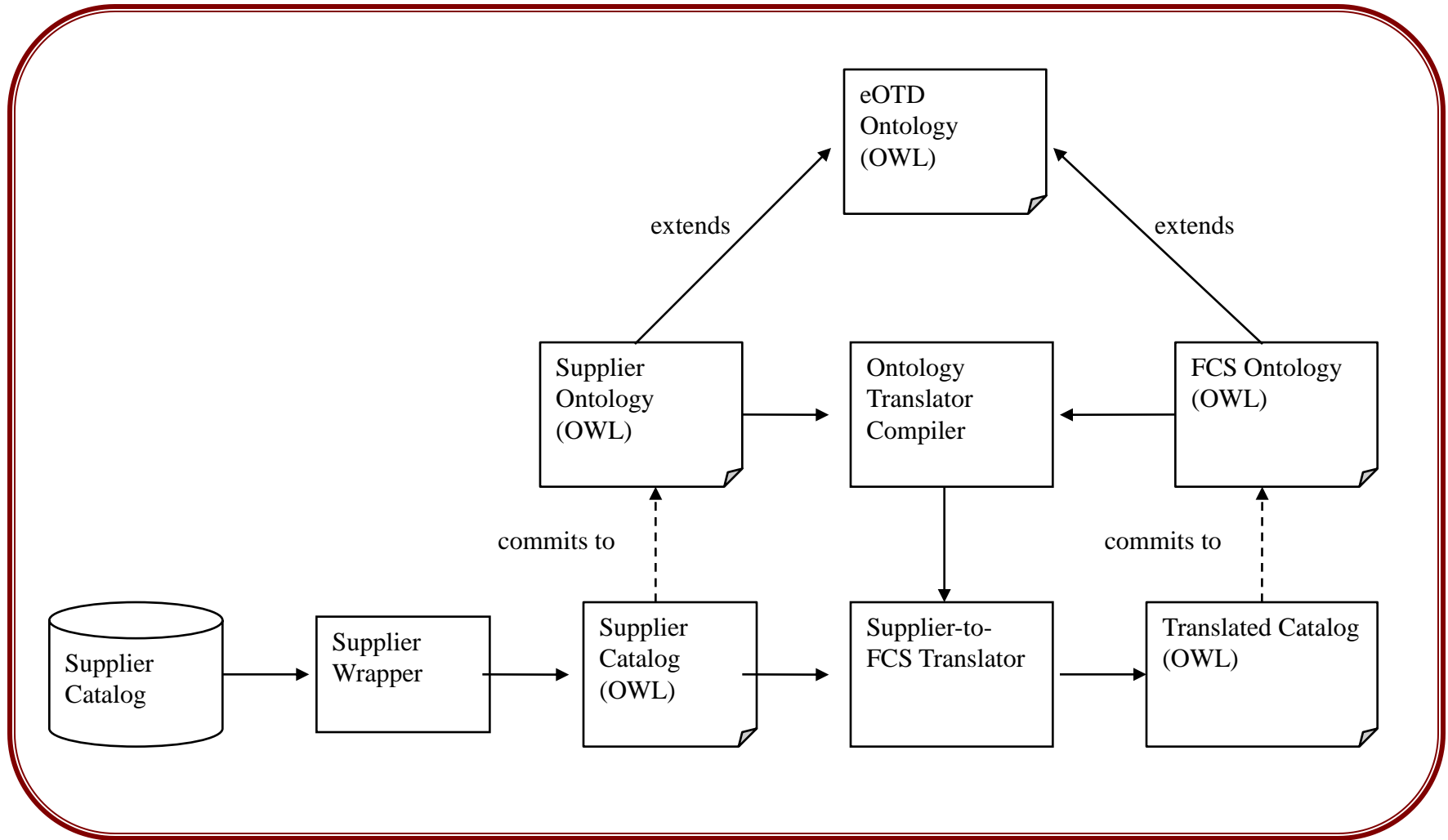
- ◆ Goal: automatically integrate a supplier catalog into the FCS
- ◆ Wrapper
 - software module that provides a standard interface to a legacy data format
- ◆ Create wrappers that convert actual data to corresponding RDF/OWL
 - May require creating new properties
 - Some taxonomies (e.g. UNSPSC, CPV) don't have properties

Translation Compiler



- ◆ Automatically generates translation code
 - Input: source and target ontology
 - Output: translator that classifies source data according to target taxonomy
 - » Classification may not be most specific possible
 - » Translating all data will require more detailed ontologies than provided by this project
 - uses DL reasoning to determine relationships
- ◆ Compiler can be rerun any time source / target ontologies change
 - Infrequent, so speed isn't critical
- ◆ Translator is fast
 - Can be run as frequently as data changes

System Architecture



Cataloging Scenario



- ◆ Cataloger receives new request
 - **Barrel Roller Bearing:**
eCl@ss code: **23.05.09.12** (*Barrel Bearing*)
Barrel roller bearings are single row, self-aligning roller bearings. They comprise solid outer rings with a concave raceway, solid inner rings with two ribs and a cylindrical or tapered bore and barrel rollers with cages. The bearings cannot be dismantled.



Barrel roller bearings are particularly suitable where high radial shock type loads occur and misalignments must be compensated, see Compensation of angular misalignments. Their axial load carrying capacity is small.

- ◆ How should this be classified in the FCS?

Barrel Bearing Inference



eClass Ontology

$eClass:BarrelBearing \subseteq eClass:RadialRollerBearing$
 $eClass:BarrelBearing \subseteq \neg eClass:InnerRing$
 $eClass:BarrelBearing \subseteq \neg eClass:NeedleRollerAndCageAssemble$
 $eClass:RadialRollerBearing \cap \neg eClass:InnerRing$
 $\cap \neg eClass:NeedleRollerAndCageAssemble \subseteq$
 $eotd:RollerBearing$

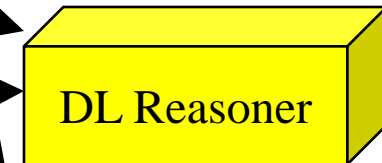
eOTD Ontology

$eOTD:RollerBearing \subseteq eOTD: BearingAntifriction$
 $eOTD:RollerBearing \subseteq eOTD: BearingUnmounted$

FCS Ontology

$fsc: BearingAntifrictionUnmounted \equiv eOTD: BearingAntifriction \cap eOTD: BearingUnmounted$

$eClass:BarrelBearing \subseteq$
 $fsc: BearingAntifrictionUnmounted$



Barrel Roller Bearing in FCS



- ◆ We were only able to infer an FSC, not an AIN
- ◆ However, from the definition, this is most likely INC 00016: “Bearing, Roller, Self-Aligning”
 - eCl@ss did not provide enough information to discover this initially, but such things can be added to the eclass-eotd ontology to improve future mappings
 - » `eclass:BarrelBearing` \subseteq
`eotd:BearingRollerSelfAligning` \equiv
`ain:BearingRollerSelfAligning`

Software Used



- ◆ Protégé
 - Ontology viewer and editor
 - Open source from Stanford
 - TGViz tab to view ontology graphs
 - Connect to DL Reasoner for advanced queries
- ◆ Fact++
 - DL Reasoner
 - Open source from Univ. of Manchester
- ◆ Java

Expected Benefits



- ◆ Creating ontologies for the FCS and commercial taxonomies provides semantically correct mappings for interoperability
- ◆ Sharing of information via ontologies provides greater item visibility
- ◆ Automatic translation from FCS ontology to target ontology expands supplier base
- ◆ Catalog consistency is promoted by “committing” to an ontology
- ◆ Can help with automated classification of NSNs

Any Questions?

