Toward an Open-Source Foundation Ontology Representing the Longman's Defining Vocabulary: The COSMO Ontology OWL Version

Patrick Cassidy
MICRA, Inc., Plainfield, NJ
cassidy@micra.com

Presented at OIC2008, Fairfax Virginia December 4, 2008

A Foundation Ontology for the IC:

- Why does IC need a foundation ontology?
- Motivation for work on COSMO
- Structure and status

Foundation Ontology and the IC

- The IC has many parts
- The parts develop their own databases, terminologies, and ontologies. Local communities want to do their own thing, not be forced to conform.
- To function effectively, the parts need to transfer information accurately – i.e., to **interoperate at the semantic level**
- **Semantic Interoperation requires a common standard of meaning**
- Automated mapping is too inaccurate
- Semiautomated mapping is too expensive; order of n^2 effort

The Problem

- Locally developed applications can use small, specialized ontologies, idiosyncratic ontologies, or no ontology at all and still perform their work perfectly.
- **BUT** When local applications need to share complex information, a common standard of meaning is essential for communication.
- The Solution – a common Foundation Ontology to provide a standard for **Content** to complement the existing standards for **Format**.
- There is a widespread assumption that getting some broad agreement on a common Foundation Ontology is impossible. There is no technical, social, or psychological barrier – what has been missing is adequate funding.

Integration of Diverse Information

- Multiple Diverse views of the same information will always be present
- Integration requires a method to translate from one terminology to others
- **Overlap of meanings requires dissection of complex meanings into component primitives**
- A foundation ontology having representations of all of the primitives is required; a common syntax is insufficient to resolve ambiguity.

Foundation Ontology

- Generically, a Foundation Ontology is an ontology containing logical representations of the most general (abstract) entities (types, relations) that are used in constructing more specialized or domain-specific representations. Existing examples are OpenCyc, SUMO, BFO, DOLCE, ISO15926 and others.
- The COSMO ontology is intended to develop into a comprehensive open-source Foundation Ontology, having all of the ***primitive*** ontology elements required to create useful representations of entities in any domain.
- For practical convenience, more specific extensions can be maintained to avoid unnecessary recreation of existing ontology elements.

The Foundation Ontology . . .

. . . is not required to be used *in toto* in every application; individual applications will only use as much as is needed to support the reasoning for that application. Redundancy will not cause computational inefficiency in the applications. A utility should be part of a common FO to extract only the needed parts.

. . . is required when separately created ontologies, applications, or databases need to transfer information. The FO supports **translation** of data from one local terminology into the other by having **a complete inventory of primitive elements** into which complex domain entity representations can be analyzed.

. . . Will not break existing applications or databases if used only for translating data transferred from one system to another.

Distracting Terminology Issues

- Concept: a unit of thought or of automated information processing – **not** necessarily an abstract mental object. Ontologies are composed of ontology elements ("ontelms") that represent such entities: see next slide.
- Definition: A description of the meaning – **not** necessary and sufficient conditions; to specify the meaning of (in words or logic)
- Meaning: an interpretation that approximates human-level understanding (see later slide)
- Understanding : conversion to a logical representation of the meaning

Words, Concepts, Ontelms

- Words are not Concepts. The elements in an ontology (types, relations, functions, axioms, instances) are neither "concepts" nor words, but language-independent logical structures. The meanings of the ontology elements do not change, but the words used to refer to them may change rapidly and vary with user.
- To avoid distracting terminology discussion, these are referred to as "***ontelms***" (ontology elements) in this presentation.

Meaning: Procedural Semantics

Meaning and Links: William A. Woods, AI Magazine 28(4) Winter 2007

– "In this theory the meaning of a noun is a procedure for recognizing or generating instances, the meaning of a proposition is a procedure for determining if it is true or false, and the meaning of an action is the ability to do the action or to tell if it has been done."

Meaning via Human Interpretation

- Nirenburg and Raskin: *Ontological Semantics* MIT Press, 2004
  - "Meaning should be studied and represented"
  - Meaning needs to be "anchored in extralinguistic reality" but the "verificationist premise" of Procedural Semantics is not shared
  - In Ontological Semantics meaning is intensional. "… meaning is a statement in the Text-Meaning Representation (TMR) language "
  - "The connection between the outside world … and Ontological Semantics … is carried out through the mediation of the human acquirer of the static knowledge sources."

Meanings for the Foundation Ontology

- Whether meanings are interpreted intensionally (as equivalent to their ontological representations) or extensionally (by use of verification procedures), the ontology itself serves to construct the meanings used by the computer for reasoning and deciding.
- Evidence that database meanings have been properly interpreted will require human evaluation of the correctness of inferences.
- Evidence that text meanings have been properly represented can be obtained from (1) question-answering or (2) conversation (the Turing test).
- For robotic systems, recognizing objects and object types, performing actions and recognizing when actions have been performed will be additional tests.

What Does it Mean to "Specify the meaning of a term"?

"**The biological mother of a person is a woman who has given birth to that person**"

{{**?Mother** isTheBiologicalMotherOf **?Child**}
impliesThat
(ThereExists {((exactly one) **?Event**) and
((exactly one) **?Date**) and ((exactly one) **?Location**)} suchThat
{{**?Event** isa BirthEvent} and
{**?Event** occurredOn **?Date**} and
{**?Event** occurredAt **?Location**} and
{**?Mother** is (The Mother in **?Event**)} and
{**?Child** is (The Baby in **?Event**)} and
{(The BirthDate of **?Child**) is **?Date**} and
{(The BirthPlace of **?Child**) is **?Location**}})}

Primitive Concepts

- Primitives: the most basic units of thought (such as the part-of relation) that are used in combination to create more complex units of thought (such as an Automobile).
- A concept or ontelm that cannot have its meaning specified solely by use of some combination of independently described primitives
- No consensus on how many primitives there are
- The COSMO project aims to provide an estimate of the upper limit (if any) on necessary primitives

How Many Primitives?

- Wierczbicka's "universal core" contained 60 primitives common among multiple languages (see Cliff Goddard Bad Arguments Against Semantic Primitives, *Theoretical Linguistics*, Vol. 24 (1998), Available at: http://www.une.edu.au/bcss/linguistics/nsm/pdfs/bad-arguments5.pdf)
- The Longman Dictionary of Contemporary English (LDOCE) uses 2148 words to define its over 64000 terms.
- Cheng-Ming Guo analyzed the Longman defining vocabulary (Ph.D. Thesis, 1989) and determined that there are 1433 actual "basic" words (representing 3200 word senses) that can be used, recursively, to define all of the words in the Longman dictionary

How Many Primitives? (continued)

- The Japanese Toyo Kanji contain 1850 characters – those required to be learned by completion of secondary education. Some basic words are represented phonetically, not as characters.
- Sign language (AMESLAN) dictionaries contain from 2000 to 5000 signs.
- The first representation of the Longman defining vocabulary plus associated basic concepts in COSMO will contain at least 7000 types and 600 relations, but probably fewer than 10000 types (in progress). Many of these may not be primitive.
- Doug Lenat speculates that as many as 15,000 primitive concept representations may be needed to serve as a "Conceptual Defining Vocabulary" (personal communication).

Longman Definitions: "obligation"

- See: http://www.ldoceonline.com/
- **Obligation**: a moral or legal duty to do something
- Duty: something that you have to do because it is morally or legally right
- Have to: if you have to do something, you **must** do it because it is necessary or because someone makes you do it.
- Must: to **have to** do something because it is necessary or important, or because of a law or order
- Necessary: something that is necessary is what you need to have or **need to** do

COSMO: “obligation”

- <owl:Class rdf:ID="Obligation">
- <rdfs:comment>A MentalObject that refers to some FutureSituation that the Agent having the Obligation may cause to happen or may refrain from doing; if the Agent does not perform an Action to cause the FutureSituation to occur, then some negative consequence is likely to be incurred for failure to perform the Obligation. . The type of negative consequence (legal punishment, social condemnation, eternal damnation, pangs of conscience, being grounded by one's parents) will be characteristic of different types of Obligation.. Each Obligation is assigned by some Authority, which could be a person’s own conscience (reflecting learned social mores), or the mores of the community. In the case of a Debt, the Authority may be the person owing the debt and the person to whom the debt is owed, if the debt arises from some agreement or transaction.
- An Obligation may be created in an ObligationCreatingEvent (which see).

  The notion of 'Obligation' is too primitive to be easily described by simple relations. In essence, an 'obligation' is a relation of an Agent to an Event that is derived from a belief about what kind of behavior is best in a situation. The exact formalization of this notion is still incomplete as of 0.49. See also 'ResponsibilitySituation' for a closely related concept.

  Linguistically an Obligation is expressed in several ways:
- 'Tom has an obligation to do X'
- 'Tom is obliged to do X'
- 'Tom has a duty to do X'
- 'Doing X is Tom's (obligation/duty).'
- 'Tom ought to do X'
- 'Tom must do X'
- 'Tom should do X'
- 'Tom is responsible for doing X'
- Similar phrases may be used to express an action that is not an Obligation, but is a prerequisite for some desired situation: 'In order to get into college, Tom must get good grades.'
- The linguistic analyzer must recognize the discourse relations that distinguish obligations from prerequisites. The type 'Obligation' in COSMO represents only true Obligations.
-
- Each instance of Obligation will represent an Action that the agent with the Obligation is obliged to perform or refrain from. When expressed linguistically, that action will be prefaced by the word 'to', e.g. 'to drive no faster than 60 miles per hour'.
-
- Cyc: A collection of microtheories; a subcollection of #$SupposedToBeMicrotheory. Each instance of the collection #$Obligation is a microtheory which contains assertions describing what some agent (the #$obligatedAgents) is obliged to do, or make true, for one or more other agents, possibly including society in general. An obligation is the most general case of some agent owing something to another. Obligations may be undertaken in conjunction with various kinds of #$Agreements. Unlike an agreement, however, an obligation need not have a second known party (though some do). An obligation can exist and be understood without identifying another particular agent as the 'holder' of the obligation - and that may be true, even if the beneficiary (#$obligationOwedTo) can be identified. For example, assuming that parents have an obligation to care for their children, it is not clear with whom a parent has 'agreed' to take care of his or her child. Some common ways to incur an obligation are through social transactions (e.g. family duties, friendship, favors) or through financial transactions (e.g. a #$PaymentObligation). In addition, obligations may be imposed on those who are subject to one or more instances of #$CodeOfConduct, e.g. # $SportsRulesOf-BoxingSportsEvent or #$OfficeCodeOfConductMt.
-
- Corresponds to senses 2 and 3 and part of sense 1 of 'obligation' and sense 2 of 'duty' in WordNet:
- NOTE that sense 2 is a state, and linguistically would be expressed by a phrase like 'under an obligation', rather than the word 'obligation' itself. Sense 3 should be a subtype,but is not yet represented.
- 1. (14) duty, responsibility, obligation - (the social force that binds you to the courses of action demanded by that force; 'we must instill a sense of duty in our children'; 'every right implies a responsibility; every opportunity, an obligation; every possession, a duty'- John D. Rockefeller Jr)
- 2. obligation - (the state of being obligated to do or pay something; 'he is under an obligation to finish the job')
- 3. obligation, indebtedness - (a personal relation in which one is indebted for a service or favor) </rdfs:comment>
- (continued . . . .)

COSMO: “obligation” (continued)

```
 <guid>bd58bfd0-9c29-11b1-9dad-c379636f7270</guid>
  <rdfs:subClassOf rdf:resource="#Responsibility"/>
  <rdfs:subClassOf rdf:resource="#RulesForConduct"/>
  <rdfs:subClassOf rdf:resource="#NeedMicrotheory"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#wasAssignedByAuthority"/>
      <owl:someValuesFrom rdf:resource="#Authority"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#refersToExternalEntity"/>
        <owl:someValuesFrom rdf:resource="#FutureSituation"/>
      </owl:Restriction>
   </rdfs:subClassOf>
   <wordnet>obligation</wordnet>
   <wnsense>obligation1n</wnsense>
   <wnsense>obligation2n</wnsense>
   <wnsense>obligation3n</wnsense>
   <wordnet>duty</wordnet>
   <wnsense>duty2n</wnsense>
</owl:Class>
```

What Makes a Concept Primitive?

- If it cannot be represented in the ontology by use of pre-existing ontelms.
- If two ontelms can only be represented by mutual reference (direct or transitive) to each other, they are considered as primitive.
- If the meaning of an an ontelm can only be described by reference to example instances, rather than by necessary conditions, it is considered as primitive.

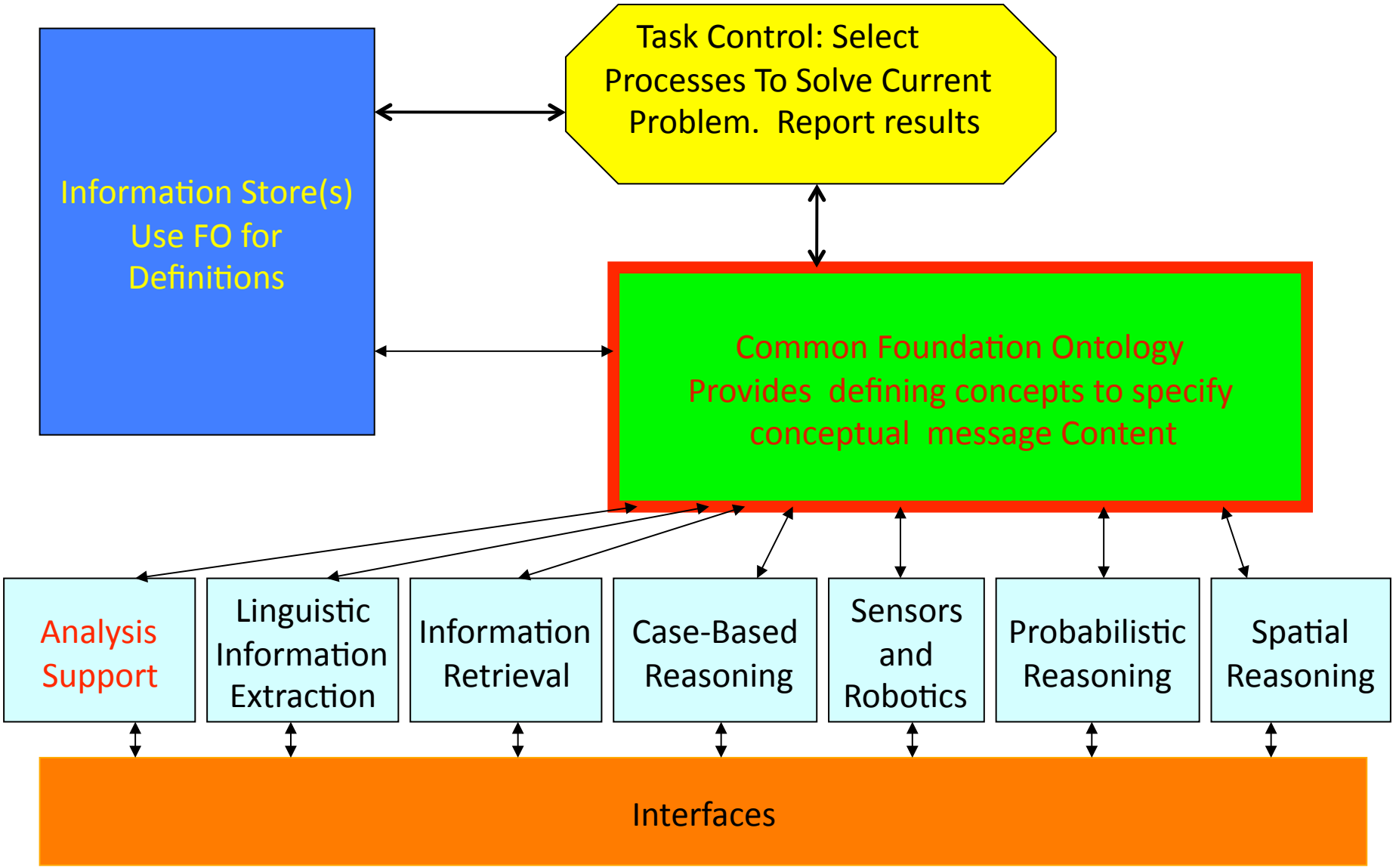
Integration of Knowledge Sources Via Semantic Interoperability

- Automated reasoning that is reliable enough to be trusted to make important decisions without human intervention requires accurate information.
- Information transferred from other systems can be used reliably only if the information is interpreted accurately. 99% accuracy is insufficient.
- Accurate interpretation requires a common foundation ontology among information sources.

How Is Semantic Interoperability Achieved by a Common Foundation Ontology?

- The elements of domain ontologies or databases are represented as combinations of ontelms already present in the Foundation Ontology.
- When information is to be communicated between systems using different domain ontologies, each system communicates, in addition to the data, the ontelms not already in the Foundation Ontology (or public extensions) that are required to understand the meanings of the data.
- Each system, able to interpret the ontelms used to describe the meanings, will be able to produce the same inferences from the same data.

The Foundation Ontology for Integrating Applications



Alternatives to a Common Foundation Ontology? Mapping *post-hoc* vs. *ab initio*

Problems mapping ontologies developed completely independently:

- Representations often combine fundamental components of meaning in different ways
- Elements of different ontologies may overlap, rather than map directly or be in a hierarchical relation
- Dissecting the components of each such representation requires human-level intelligence
- The documentation rarely has sufficient information even for a human to resolve the ambiguities
- Mapping legacy ontologies to a common Foundation Ontology will reduce the problem from order of n^2 to n.

The COSMO Project

- Motivated by an absence of a widely accepted Foundation Ontology that can serve as a standard of meaning
- Initiated in in 2005 [13] as a project of the Ontology and Taxonomy Coordinating Working Group (ONTACWG) , a working group of the Federal Semantic Interoperability Community of Practice.
- Continued by Patrick Cassidy

The COSMO Project (continued)

- Since late 2007, the objective has been to create an initial version that contains representations of all of the words in the Longman Defining Vocabulary.
- This version will be tested to determine if it contains all of the primitives needed to represent terms in specialized fields.
- The number of new primitives required for each increment of new representations will indicate whether there is an asymptotic limit to the number of primitives required to represent all fields.
- This criterion of sufficiency is probabilistic.

What's New in the COSMO?

- A little less than half of the ontelms in COSMO are not also present in OpenCyc or SUMO
- BUT the goal is to make it as small as possible while still having all of the semantic primitives needed to describe entities in any domain
- Keeping it small will make it easier for multiple developers to agree on the structure, and make it easier to learn and to use
- "A theory should be as simple as possible, but no simpler" -- Einstein

COSMO Phasing

- Phase 1 will develop an OWL ontology with representations of all of the Longman defining words. (est. completion early 2009)
- Phase 2 will test that version for adequacy in specifying meanings of at least 1000 specialized terms
- Phase 3 will convert the OWL version to a Common-Logic compatible version
- Phase 4 will develop a Natural Language interface to the ontology to make use easier

Open Source, Open Method

- To serve as a widely used standard, any ontology needs input from many different developers and users with differing views and preferences. COSMO is fully open to input from any source, provided that it is logically consistent with existing content.
- If funding becomes available for a collaborative development of a Common Foundation Ontology by a similarly open method, that project will supersede COSMO.

Multiple Viewpoints

- An important function of a Foundation Ontology is to serve as a means to ***translate*** other, specialized knowledge representations into each other.
- Different ways to represent the same entity can be accommodated, provided that they are logically consistent and can be translated into each other.
- A given application may use only a small part of the COSMO, extracted as needed for its own purposes; therefore redundant alternative representations will not reduce the computational efficiency of applications

Criterion for Evaluation

- The question to be determined is whether new primitives are required to represent knowledge in specialized domains, and if so, how many?
- The rate of increase of the number of ontels in the COSMO for each increment (e.g. of 1000 term representations) will provide evidence whether there is a limit (an asymptote) in the number of terms required to represent many other fields.
- If no asymptote is suggested, a small rate of increase may still allow use of a common Foundation Ontology as a means of semantic interoperability, but with more careful attention to versioning.
- When mature, the need to add new primitives should rarely occur

COSMO: Current Status

OWL version

- Types (classes): 5710
- Relations (OWL properties): 620
- Restrictions: 1023
- Longman Terms remaining to be added: 966
  - (out of 2148)

History of COSMO

- Started by including all common terms in OpenCyc and SUMO (+MLO)
- Added parents of the common terms
- Added terms from DOLCE and BFO not in either
- Added types and relations to map database tables to the ontology
- At revision 589 started supplementing with missing Longman defining terms; beginning statistics:
  - 3659 types (OWL classes)
  - 362 relations (OWL properties)
  - 414 restrictions
- Current revision (747) 5710 types, 620 relations:

  Added 2051 types and 258 relations to COSMO while adding representations of 950 Longman defining words

Conclusion

- Representation of the Longman defining vocabulary in OWL will likely require fewer than 10,000 ontology elements.
- Some of those are not primitive elements, and can be specified as combinations of other elements
- Adding rules in a CL-compliant format will increase the number of elements, by at least the number of relations

Toward the Future

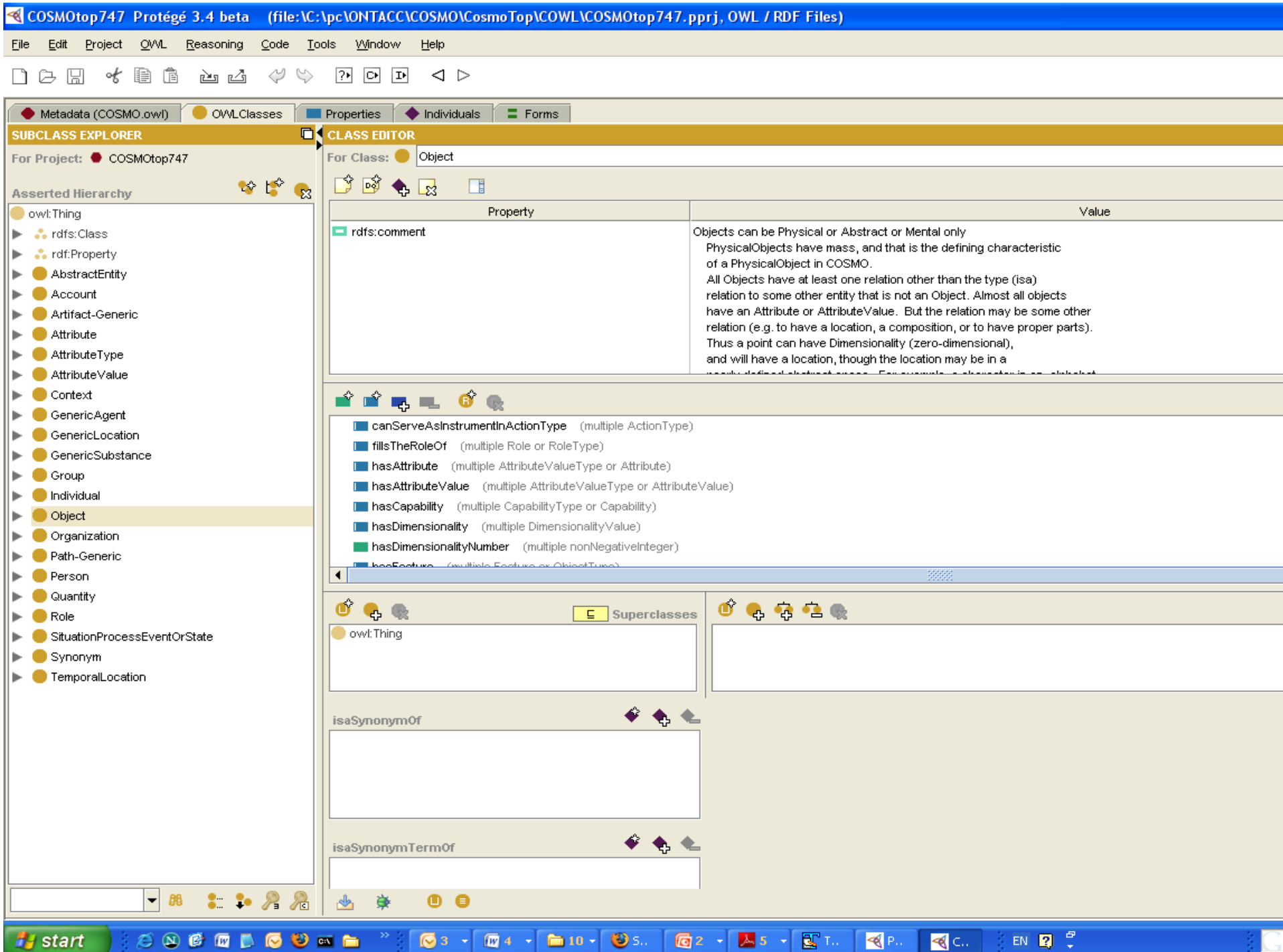
- The potential for widespread agreement on a common Foundation Ontology presents an opportunity to develop a tool that can substantially accelerate progress in the use of computers for intelligence applications.
- The need for a common standard of meaning within the Intelligence community is too important to be delegated to other federal agencies (e.g. NSF): such research should be supported by the IC itself.

END

- COSMO ontology:
  - http://micra.com/COSMO
  - Email: cassidy@micra.com

Skepticism

- "We cannot get everyone to agree on a single foundation ontology"
  - We don't need everyone, just a self-sustaining community
- "We don't need another foundation ontology"
  - The fact that none has gained a critical mass of users demonstrates that we ***do*** need another one, but one that is constructed by a very wide community of users.
  - The COSMO is not a common FO, but is being used to demonstrate that a common Foundation Ontology is feasible, if funding is available.
- There is no 'conceptual defining vocabulary'
  - Implies an unlimited number of primitive concepts; this is susceptible to experimental refutation, and the COSMO project is designed to test this question


COSMOtop747 Protégé 3.4 beta (file:\C:\pc\ONTACC\COSMO\CosmoTop\COWL\COSMOtop747.pprj, OWL / RDF Files)
File Edit Project OWL Reasoning Code Tools Window Help
Metadata (COSMO.owl)
OWLClasses
Properties
Individuals
Forms
SUBCLASS EXPLORER
For Project: COSMOtop747
Asserted Hierarchy
owl:Thing
rdfs:Class
rdf:Property
AbstractEntity
Account
Artifact-Generic
Attribute
AttributeType
AttributeValue
Context
GenericAgent
GenericLocation
GenericSubstance
Group
Individual
Object
Organization
Path-Generic
Person
Quantity
Role
SituationProcessEventOrState
Synonym
TemporalLocation
CLASS EDITOR
For Class: Object
Property
Value
rdfs:comment
Objects can be Physical or Abstract or Mental only
PhysicalObjects have mass, and that is the defining characteristic
of a PhysicalObject in COSMO.
All Objects have at least one relation other than the type (isa)
relation to some other entity that is not an Object. Almost all objects
have an Attribute or AttributeValue. But the relation may be some other
relation (e.g. to have a location, a composition, or to have proper parts).
Thus a point can have Dimensionality (zero-dimensional),
and will have a location, though the location may be in a
canServeAsInstrumentInActionType (multiple ActionType)
fillsTheRoleOf (multiple Role or RoleType)
hasAttribute (multiple AttributeValueType or Attribute)
hasAttributeValue (multiple AttributeValueType or AttributeValue)
hasCapability (multiple CapabilityType or Capability)
hasDimensionality (multiple DimensionalityValue)
hasDimensionalityNumber (multiple nonNegativeInteger)
Superclasses
owl:Thing
isaSynonymOf
isaSynonymTermOf
start
EN

```
<owl:Class rdf:ID="ConceptualWork">
   <rdfs:subClassOf rdf:resource="#AbstractSymbolicObject"/>
   <rdfs:comment>In COSMO a 'ConceptualWork' (a MentalObject) is classified as an
    AbstractSymbolicObject,  since such works are always  created in symbols, though the symbols may
    have information content – the  'meaning'.  COSMO differs somewhat from the  Cyc description in
    that we consider Codes to be included,  but  have a different usage of the term 'Code'.

    Cyc: OPENCYC 1: MAY 23, 2002
     The collection of abstract works which are the deliberate creations  of one or more individuals
     working in  concert, have instantiations  [#$instantiationOfCW] which are
     #$InformationBearingThings, and  associated  #$AbstractInformationStructures. This is a
     specialization  of #$DevisedPracticeOrWork [q.v.]. For works with  propositional content ; see the
     more specific collection, #$PropositionalConceptualWork (PCW).  Positive  examples include:
     #$MobyDickNovel (as opposed to any instances of #$BookCopy such that  (#$instantiationOfCW
     #$MobyDickNovel BOOK_COPY)),  Beethoven's 9th Symphony (as opposed to any  performance of
     this symphony  or any copy of its score).
     Negative examples include:  games (performances are not IBTs), awards (they do not have
     associated #$AbstractInformationStructures),  paintings (not abstract), customs (not deliberate
      creations), natural languages (not a deliberate creation),  and codes (their uses, not instantiations,
     are IBTs).
   </rdfs:comment>
</owl:Class>
```

Guo's Longman Analysis

- **Guo, Cheng-ming (1989)** *Constructing a machine-tractable dictionary from "Longman Dictionary of Contemporary English"* **(Ph. D. Thesis),** New Mexico State University.

- **Guo, Cheng-ming (editor)** *Machine Tractable Dictionaries: Design and Construction,* Ablex Publishing Co., Norwood NJ (1995)

- **Yorick Wilks, Brian Slator, and Louise Guthrie**, *Electric Words: Dictionaries, Computers, and Meanings*, MIT Press, Cambridge Mass (1996).

Words, Concepts, Representations

- **Words are not Concepts**
- ***Concept***: a unit of thought or reasoning
  - **(from Random House Webster)**
  - 1. a general notion or idea; conception.
  - 2. an idea of something formed by mentally combining all its characteristics or particulars; a construct.
  - **3. a directly conceived or intuited object of thought.**
- In an ontology a "concept" is only **that which is represented** by the elements of the ontology (types, relations, instances, rules, functions). These are the things that are manipulated by a reasoning system
- **The "representandum"**
- **Words are not representanda.**

Words Label Concepts

- Ambiguity: the same word labels multiple concepts
- Synonym: more than one word labels the same concept
- Context-sensitive usage: the same word in different contexts can label different concepts
- An ontology organizes representations of concepts – mapping to words is a different task.

Some Primitives fromWierzbicka

- I, YOU, SOMEONE, SOMETHING, THIS,
- THE SAME, THINK, WANT, KNOW, SAY, DO, HAPPEN, GOOD, BAD, WHEN/TIME,
- WHERE/PLACE, BECAUSE, NOT, MAYBE, LIKE, KIND OF, PART OF.