# CCT Representation Project Analysis and Results

## Prepared by

Kurt Conrad and Bob Smith, 2005.01.21

## Ontolog Overview

Ontolog (http://ontolog.cim3.net) is an open, international, virtual community of practice working on business domain ontologies. To date, it has over 100 members from more than 13 countries. The community was convened as an offshoot of the OASIS UBL effort in 2002 to:

1. Discuss practical issues and strategies associated with the development of both formal and informal ontologies used in business, and

2. Identify ontological engineering approaches that might be applied to the UBL effort (and by extension, to the broader domain of eBusiness standardization efforts).

## The CCT-Representation Project

Ontolog's CCT-Representation project (http://ontolog.cim3.net/cgi-bin/wiki.pl?CctRepresentation) was launched in March, 2004. It seeks to influence the adoption of ontologies and ontological engineering methodologies in eBusiness standards by demonstrating an ontological formalization of ebXML Core Component Types.

Project Contributors
- Pat Cassidy
- Kurt Conrad (Lead Editor)
- Peter Denno
- Scott L. Holmes
- Nenad Ivezic
- Holger Knublauch
- Tim McGrath
- Garret Minakawa
- Duane Nickull
- Adam Pease (Lead Ontologist)
- Sue Probert
- Nicolas Rouquette
- Bob Smith
- Alan Stitzer
- Evan Wallace
- Peter Yim (Project Manager)

Project Observers
- Lisa Colvin
- Mark Crawford
- Monica Martin
- Bo Newman
- John Yunker

## Approach and Results

To date, this project has employed a three-step approach.

- Mapping Core Component Types (CCTs) to concepts that already exist in standardized ontologies.
- Creating new formalizations, where needed.
- Evaluating the quality of the resulting mappings and formalizations.

We started with the definitions and specifications provided in sections 8.1 "Approved Core Component Types" and 8.2 "Approved Core Component Type, Content, and Supplementary Components..." of the Core Components Technical Specification V2.01.

For each approved CCT and their corresponding Supplementary Components, we attempted to identify an equivalent concept in SUMO (Suggested Upper Merged Ontology), MILO (Middle Level Ontology), or the QoSontology (an existing ontology of concepts used in the IT industry).

Where mappings to the existing ontologies were not identified, we extended the concepts in existing ontologies by creating new formalizations. A domain ontology – named CCT-Ont – was developed to hold these additional formalizations. Each of the new concepts and their associated axioms were encoded using KIF (Knowledge Interchange Format) syntax.

In conjunction with the initial mapping activity, areas of semantic ambiguity were briefly noted and recorded by the ontology engineering team. At a later time, the results were evaluated. For each CCT and supplementary component, we ranked the quality of the resulting mapping / formalization:

Category A: No issues, solid mapping
Category B: Minor issues
Category C: Major issues

Issues and knowledge gaps were briefly noted to serve as a starting point for further analysis.

Thus far, little has been done to identify specific drivers / root causes for the issues that the team has identified.

- One possibility is that we overlooked supplemental documentation that could have provided us with the additional information that we needed.
- Another possibility is that we were unable to interpret the intended semantics of the CCTS terminology.
- It is also possible that some of the CCTS terms, as defined, remain ambiguous and warrant further specification.
- Not to be discounted is the possibility that these issues and knowledge gaps point to alignment breakdowns between two divergent, but hopefully complementary sets of methodologies: Those developed and employed by UN/CEFACT to identify and define Core Components and those associated with the discipline of ontological engineering.

## Going Forward

While additional exploration of the previously identified issues could be done, independently, by the Ontolog project team, we feel that it would be beneficial to work more closely with subject matter experts within UN/CEFACT to identify the underlying patterns and most appropriate next steps for completing the Core Component ontology project. Because of the possible methodological underpinnings of these issues, we anticipate the likelihood that work processes on both sides could be improved through these discussions.

It is our hope, that our existing mapping and formalization work can serve as the basis for a complete ontology of the Core Component Types, which in turn, would enable us to formalize the semantics of Core Components, in general, and UBL, in particular. This is expected to result in both a solid set of ontological engineering methodologies and a baseline set of ontologies that could be used to deal with the complex problems of semantic harmonization, UBL customization, and more context-sensitive processing of eBusiness documents.

## Mapping the Core Component Types to the Suggested Upper Merged Ontology

Reference: ebXML Core Component Technical Specification v2.01c – dated 2003-11-15

Version 0.6 – dated 2005.01.20 – Quality Assessment comments expanded by Adam

| Ref# | CCT / SC | CCT Name | Definition | Primitive data-type | Remarks | SUMO Mapping | Quality Assessment* |
|---|---|---|---|---|---|---|---|
| 01-000 | CCT | Amount. Type | A number of monetary units specified in a currency where the unit of currency is explicit or implied. | | | this would be a CurrencyMeasure expressed using MeasureFn in the second argument being a RealNumber with the first argument being an instance UnitOfMeasure and subclass of CurrencyMeasure . The resulting CurrencyMeasure can be related to a thing with the relation monetaryValue | C<br>- Too many concepts are folded in to this one term. |
| 02-000 | CCT | Binary Object. Type | A set of finite-length sequences of binary octets. Shall also be used for Data Types representing graphics (i.e., diagram, graph, mathematical curves or similar representations), pictures (i.e. visual representation of a person, object, or scene), sound, video, etc. | | | ContentBearingObject | A |
| 03-000 | CCT | Code. Type | A character string (letters, figures or symbols) that for brevity and/or language independence may be used to represent or replace a definitive value or text of an Attribute together with relevant supplementary information. Should not be used if the character string identifies an instance of an Object Class or an object in the real world, in which case the Identifier. Type should be used. | | | SymbolicString , however, if the reason is just brevity, than it's an implementation detail, and the nformation should be represented. It would be up to an application to condense that information using a thesaurus or other approach, if necessary. | B<br>- We have some confusion about types<br>- Is it a pointer or an abbreviation?<br>- Is it an implementation detail or does the item have inherent semantics?<br>- How is a code different from an identifier? |

| Ref# | CCT / SC | CCT Name | Definition | Primitive data-type | Remarks | SUMO Mapping | Quality Assessment* |
|---|---|---|---|---|---|---|---|
| 04-000 | CCT | Date Time. Type | A particular point in the progression of time together with relevant supplementary information. Can be used for a date and/or time. | | | use time | B<br>- Clear mapping, but formalization doesn't resolve all problems (SUMO has more detail) |
| 05-000 | CCT | Identifier. Type | A character string to identify and distinguish uniquely, one instance of an object in an identification scheme from all other objects in the same scheme together with relevant supplementary information. | | | uniqueIdentifier | C<br>- If this is meant to be a universally-unique identifier (UUID) then rate an "A" |
| 06-000 | CCT | Indicator. Type | A list of two mutually exclusive Boolean values that express the only possible states of a Property. | | | true | A |
| 07-000 | CCT | Measure. Type | A numeric value determined by measuring an object along with the specified unit of measure. | | | this would be an expression using MeasureFn , if the value is not directly time dependent, such as Watt (Joule per second). If not a time-dependent quantity, the MeasureFn would have in the second argument being a RealNumber with the first argument being an instance UnitOfMeasure. The resulting ConstantQuantity can be related to a thing with the relation measure , which can also serve to relate a ConstantQuantity to a thing. | C<br>- Too many concepts are folded in to this one term. |
| 08-000 | CCT | Numeric. Type | Numeric information that is assigned or is determined by calculation, counting, or sequencing. It does not require a unit of quantity or unit of measure. May or may not be decimal | | | Number | B |

| Ref# | CCT / SC | CCT Name | Definition | Primitive data-type | Remarks | SUMO Mapping | Quality Assessment* |
|---|---|---|---|---|---|---|---|
| 09-000 | CCT | Quantity. Type | A counted number of non-monetary units possibly including fractions. | | | this would be an expression using MeasureFn , if the value is not directly time dependent, such as Watt (Joule per second). If not a time-dependent quantity, the MeasureFn would have in the second argument being a RealNumber with the first argument being an instance UnitOfMeasure. The resulting ConstantQuantity can be related to a thing with the relation measure , which can also serve to relate a FunctionQuantity to a thing. | C<br>- Too many concepts are folded in to this one term. |
| 10-000 | CCT | Text. Type | A character string (i.e. a finite set of characters) generally in the form of words of a language. Shall also be used for names (i.e. word or phrase that constitutes the distinctive designation of a person, place, thing or concept). | | | a Text | B+<br>- Too much ambiguity in definition |
| 01-001 | SC | Amount. Content | A number of monetary units specified in a currency where the unit of currency is explicit or implied | decimal | | this would be a RealNumber part of an expression using MeasureFn in the second argument being a RealNumber with the first argument being an instance UnitOfMeasure and subclass of CurrencyMeasure . The resulting CurrencyMeasure can be related to a thing with the relation monetaryValue | B<br>- Too many concepts are folded in to this one term. |

| Ref# | CCT / SC | CCT Name | Definition | Primitive data-type | Remarks | SUMO Mapping | Quality Assessment* |
|---|---|---|---|---|---|---|---|
| 01-002 | SC | Amount Currency. Code List Version. Identifier | The Version of the UN/ECE Rec.9 code list. | string | | uniqueIdentifier | B<br>- Need to clarify the distinction of Code vs ID, which seems questionable.<br>- Also, it appears that there are many different terms defined just because there is a really a relation with several different possible arguments. |
| 01-003 | SC | Amount Currency. Identifier | The currency of the amount | string | Reference UN/ECE Rec. 9, using 3-letter alphabetic codes. The UN/ECE Rec. 9 is also published as ISO 4217, but is available in electronic form and free of charge. | This would be a SymbolicString that is related to a subclass of CurrencyMeasure by the relation uniqueIdentifier . | B<br>- Need to clarify the distinction of Code vs ID, which seems questionable.<br>- Also, it appears that there are many different terms defined just because there is a really a relation with several different possible arguments. |
| 02-001 | SC | Binary Object. Content | A set of finite-length sequences of binary octets. | binary | | ComputerData | A |
| 02-002 | SC | Binary Object. Format. Text | The format of the binary content. | string | | ComputerData | B<br>- How is this different than text type? |
| 02-003 | SC | Binary Object. Mime. Code | The mime type of the binary object. | string | Reference IETF RFC 2045, 2046, 2047 | Relate a MimeEncodingScheme to ComputerData with mimeType | B<br>- Too many concepts are folded in to this one term. |
| 02-004 | SC | Binary Object. Character Set. Code | The character set of the binary object if the mime type is text. | string | Reference IETF RFC 2045, 2046, 2047 | Relate an instance of MimeText to ComputerData with mimeSubType | B<br>- Too many concepts are folded in to this one term. |

| Ref# | CCT / SC | CCT Name | Definition | Primitive data-type | Remarks | SUMO Mapping | Quality Assessment* |
|---|---|---|---|---|---|---|---|
| 02-005 | SC | Binary Object. Encoding. Code | Specifies the decoding algorithm of the binary object. | string | Reference IETF RFC 2045, 2046, 2047 | Procedure | B<br>- Not sure that we understand this term<br>- If it is an enumerated set, then rate an "A". Otherwise, rate "B"<br>- Also, there is confusion regarding how this element is different from 02-001 and 02-002 |
| 02-006 | SC | Binary Object. Uniform Resource. Identifier | The Uniform Resource Identifier that identifies where the Binary Object is located. | string | | UniformResourceIdentifier | B |
| 02-007 | SC | Binary Object. Filename. Text | The filename of the binary object. | string | Reference IETF RFC 2045, 2046, 2047 | filename | B<br>- Clear, but reference to an external document, should define succinctly |
| 03-001 | SC | Code. Content | A character string (letters, figures or symbols) that for brevity and/or language independence may be used to represent or replace a definitive value or text of an Attribute. | string | | SymbolicString | B<br>- Is it a pointer (implementation detail) or an actual thing?<br>- How is it different from other strings? |
| 03-002 | SC | Code List. Agency. Identifier | An agency that maintains one or more code lists. | string | Defaults to the UN/EDIFACT data element 3055 code list. | uniqueIdentifier | B<br>- Need to clarify the distinction of Code vs ID, which seems questionable.<br>- Also, it appears that there are many different terms defined just because there is a really a relation with several different possible arguments. |

| Ref# | CCT / SC | CCT Name | Definition | Primitive data-type | Remarks | SUMO Mapping | Quality Assessment* |
|---|---|---|---|---|---|---|---|
| 03-003 | SC | Code List. Agency Name. Text | The name of the agency that maintains the code list. | string | | SymbolicString | B<br>- Need to clarify the distinction of Code vs ID, which seems questionable.<br>- Also, it appears that there are many different terms defined just because there is a really a relation with several different possible arguments. |
| 03-004 | SC | Code List. Name. Text | The name of a list of codes. | string | | names | B<br>- Need to clarify the distinction of Code vs ID, which seems questionable.<br>- Also, it appears that there are many different terms defined just because there is a really a relation with several different possible arguments. |
| 03-005 | SC | Code List. Identifier | The identification of a list of codes | string | Can be used to identify the URL of a source that defines the set of currently approved permitted values | uniqueIdentifier | B<br>- Need to clarify the distinction of Code vs ID, which seems questionable.<br>- Also, it appears that there are many different terms defined just because there is a really a relation with several different possible arguments. |
| 03-006 | SC | Code List Scheme. Uniform Resource. Identifier | The Uniform Resource Identifier that identifies where the code list scheme is located. | string | | UniformResourceIdentifier | B<br>- Need to clarify the distinction of Code vs ID, which seems questionable.<br>- Also, it appears that there are many different terms defined just because there is a really a relation with several different possible arguments. |

| Ref# | CCT / SC | CCT Name | Definition | Primitive data-type | Remarks | SUMO Mapping | Quality Assessment* |
|---|---|---|---|---|---|---|---|
| 03-007 | SC | Code List. Uniform Resource. Identifier | The Uniform Resource Identifier that identifies where the code list is located. | string | | UniformResourceIdentifier | B<br>- Need to clarify the distinction of Code vs ID, which seems questionable.<br>- Also, it appears that there are many different terms defined just because there is a really a relation with several different possible arguments. |
| 03-008 | SC | Code List. Version. Identifier | The Version of the code list. | string | Identifies the Version of the UN/EDIFACT data element 3055 code list. | uniqueIdentifier | B<br>- It appears that there are many different terms defined just because there is a really a relation with several different possible arguments. |
| 03-009 | SC | Code. Name. Text | The textual equivalent of the code content | string | If no code content exists, the code name can be used on its own | SymbolicString | B<br>- It's not clear what this is and why a simple text type cannot be used.<br>- The contingent definition is also puzzling. |
| 04-001 | SC | Date Time. Content | The particular point in the progression of time | string | For times use an ISO 8601 compliant format that includes the UTC offset | time | A |
| 04-002 | SC | Date Time. Format. Text | The format of the date/time content | string | Reference ISO 8601 and W3C note on date time | SymbolicString | A |

| Ref# | CCT / SC | CCT Name | Definition | Primitive data-type | Remarks | SUMO Mapping | Quality Assessment* |
|---|---|---|---|---|---|---|---|
| 05-001 | SC | Identification Scheme Agency. Identifier | The identification of the agency that maintains the identification scheme. | string | Defaults to the UN/EDIFACT data element 3055 code list. | uniqueIdentifier | B<br>- Need to clarify the distinction of Code vs ID, which seems questionable. Also, it appears that there are many different terms defined just because there is a really a relation with several different possible arguments. |
| 05-002 | SC | Identification Scheme. Agency Name. Text | The name of the agency that maintains the identification scheme | string | | Agent | A |
| 05-003 | SC | Identification Scheme Data. Uniform Resource. Identifier | The Uniform Resource Identifier that identifies where the identification scheme data is located | string | | UniformResourceIdentifier | B?<br>- It's not clear whether this is an item or a relation<br>- In either case it appears redundant and differentiated only by the argument used if it's a relation. |
| 05-004 | SC | Identification Scheme. Identifier | The identification of the identification scheme. | string | | uniqueIdentifier | B<br>- What's difference with 05-005 |
| 05-005 | SC | Identification Scheme. Name. Text | The name of the identification scheme. | string | | uniqueIdentifier | B<br>- What's difference with 05-004 |
| 05-006 | SC | Identification Scheme. Uniform Resource. Identifier | The Uniform Resource Identifier that identifies where the identification scheme is located. | string | | UniformResourceIdentifier | B<br>- URI should be location independent |
| 05-007 | SC | Identification Scheme. Version. Identifier | The Version of the identification scheme. | string | Identifies the Version of the UN/EDIFACT data element 3055 code list. | version | B |

| Ref# | CCT / SC | CCT Name | Definition | Primitive data-type | Remarks | SUMO Mapping | Quality Assessment* |
|---|---|---|---|---|---|---|---|
| 05-008 | SC | Identifier. Content | A character string to identify and distinguish uniquely, one instance of an object in an identification scheme from all other objects within the same scheme | string | | uniqueIdentifier | B |
| 06-001 | SC | Indicator. Content | The value of the indicator | string | For example on, off, true, false | SymbolicString | B<br>- This appears to be any sort of enumerated type and therefore essentially a character string, meaning that it is redundant |
| 06-002 | SC | Indicator. Format. Text | Whether the indicator is numeric, textual or binary | string | | SymbolicString | B<br>- Appears redundant with other text types |
| 07-001 | SC | Measure. Content | The numeric value determined by measuring an object. | decimal | For example, 24.387 kilograms (24.387 is the Measure. Content) | measure | B<br>- The relationship to object is unclear. The measure of an object must include unit and numeric amount, but also the relevant dimension in the case of a linear measure. The amount is just a number so this item is redundant with any other bare number. |
| 07-002 | SC | Measure Unit. Code | The type of unit of measure | string | Reference UN/ECE Rec. 20 and X12 355. | ConstantQuantity | B<br>- References a big standard, prefer to see a definition included here<br>- Some overlap w/ 01-000 (unit of measure) |
| 07-003 | SC | Measure Unit. Code List Version. Identifier | The Version of the measure unit code list. | string | | version | B |

| Ref# | CCT / SC | CCT Name | Definition | Primitive data-type | Remarks | SUMO Mapping | Quality Assessment* |
|---|---|---|---|---|---|---|---|
| 08-001 | SC | Numeric. Content | Numeric information that is assigned or is determined by calculation, counting or sequencing. | as defined by Numeric. Format. Text | May be decimal | Number | B<br>- Redundant with other bare numbers. |
| 08-002 | SC | Numeric. Format. Text | Whether the number is an integer, decimal, real number or percentage | string | | subclasses of Number | B<br>- Unclear mapping since a percentage must be distinguished from a number less than 1. These two items shouldn't be covered by the same term. |
| 09-001 | SC | Quantity. Content | A counted number of non-monetary units possibly including fractions. | decimal | For example 7 bales (7 is the Quantity. Content) | ConstantQuantity | B<br>- Redundant with other numbers |
| 09-002 | SC | Quantity. Unit. Code | The unit of the quantity | string | May use UN/ECE Rec. 20 | ConstantQuantity | C<br>- Appears redundant with other unit types |
| 09-003 | SC | Quantity Unit. Code List Agency. Identifier | The identification of the agency which maintains the quantity unit code list | string | | uniqueIdentifier | C<br>- Redundant with 03-002 |
| 09-004 | SC | Quantity Unit. Code List. Identifier | The quantity unit code list. | string | Defaults to the UN/EDIFACT data element 3055 code list. | uniqueIdentifier | B |
| 09-005 | SC | Quantity Unit. Code List Agency Name. Text | The name of the agency which maintains the quantity unit code list. | string | | Agent | C |
| 10-001 | SC | Text. Content | A character string (i.e. a finite set of characters) generally in the form of words of a language. | string | | SymbolicString | B<br>- Redundant with other text types |

| Ref# | CCT / SC | CCT Name | Definition | Primitive data-type | Remarks | SUMO Mapping | Quality Assessment* |
|---|---|---|---|---|---|---|---|
| 10-002 | SC | Language. Identifier | The identifier of the language used in the corresponding text string | string | Reference ISO 639: 1998 | subclasses of Language | A<br>- References ISO standard |
| 10-003 | SC | Language. Locale. Identifier | The identification of the locale of the language. | string | | GeographicArea | B<br>- Language and geographical areas are not the same<br>- Specifications of locales have tremendous ambiguity<br>- Role of localize is unclear<br>- Assuming an ISO code list |

* Quality Rankings:

Category A: No issues, solid mapping
Category B: Minor issues
Category C: Major issues

Ontologies consist of:

- SUMO, which can be found at http://cvs.sourceforge.net/viewcvs.py/*checkout*/sigmakee/KBs/Merge.txt
- MILO, which can be found at http://cvs.sourceforge.net/viewcvs.py/*checkout*/sigmakee/KBs/Mid-level-ontology.txt
- QoSontology, which can be found at http://cvs.sourceforge.net/viewcvs.py/*checkout*/sigmakee/KBs/QoSontology.txt