Ontologizing EDIFACT

www.tripcom.org

doug foxvog – NUIG

TripCom General Assembly Meeting

29 April 2008

Status

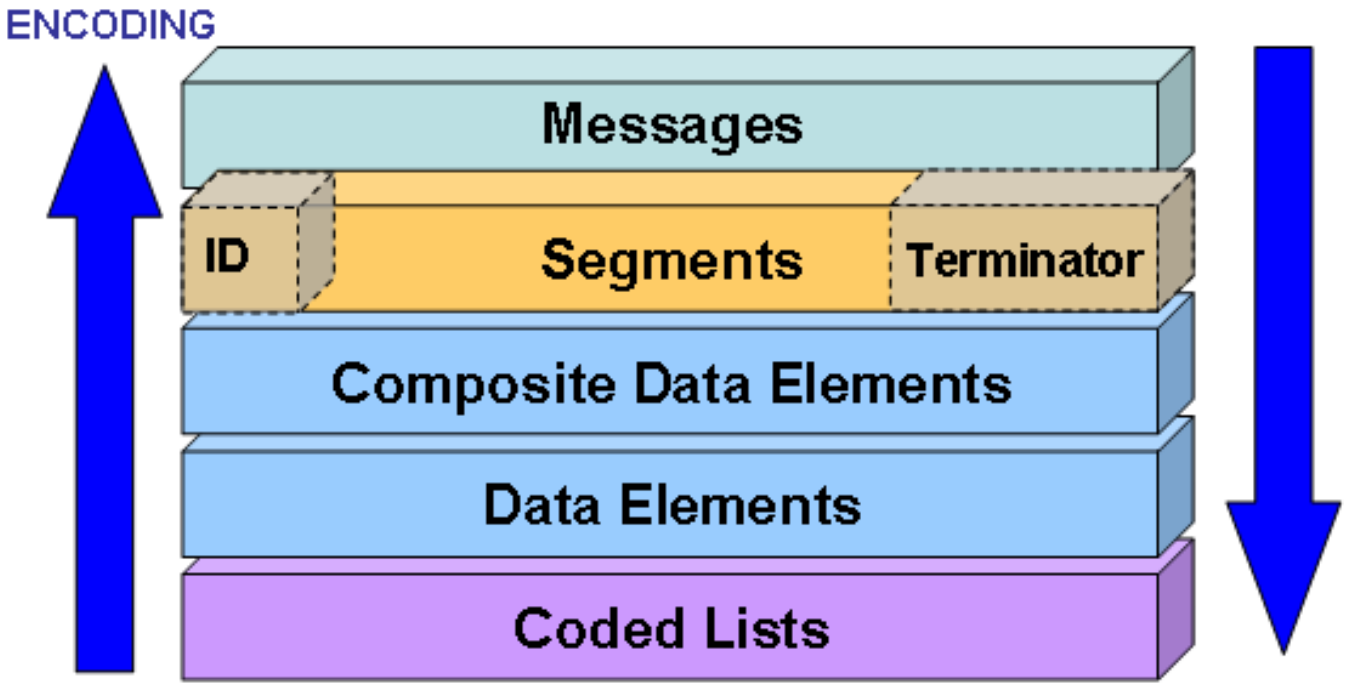
- Analyzed EDI versions (M6) — T 7.1, D 7.1
  - Reviewed different standards
  - Reviewed different EDIFACT versions
  - Selected EDIFACT subsets
- Encoded selected EDIFACT subsets' format semantically (M12) — T 7.2, D 7.2
  - Encoded syntax for EANCOM, EDIFICE, EMEDI, ETIS, and EDIFICAS subsets

Status (cont.)

- Encoded selected EDIFACT subsets' meaning (M24) — D 7.2, T 7.3
  - Encoded Code Sets for subsets
  - Encoded meanings for Data Elements & Segments
  - Encoding templates for implicit statements

- Relationships among ontology modules within EDIFACT and with external ontologies — D 7.3
  - Define EDIFACT internal dependencies (M24) — T 7.4
  - Map created ontologies to external ontologies (M30) — T 7.5

EDIFACT Message Structure



- Messages are composed of Data Segments and Data Segment loops.
- Data Segments are composed of Data Elements.
- Composite DEs are composed of Simple DEs.

Ontologizing EDI Syntax

- Used existing vocabulary for specifying
  - formats
    - Order, Optionality, Repeat Counts
  - inter-component restrictions
  - restricted subsets – through taxonomy of ontologies

- Described format of
  - 66 message types (93 versions)
  - 925 data segment loops
  - 91 data segment types (203 versions)
  - 116 complex data element (297 versions)
  - 2335 format restrictions (6197 uses)

Example EDIFACT Data Segment

DTM+137:19940121:102’

- DTM = segment tag, indicating a “Date/time” segment
- + = segment tag and data element separator
- 137 = date qualifier, code indicating that the date is the document/message date/time
- : = separator for component data elements within a composite (here, date qualifier and date)
- 19940121 = date, in the format specified by the date format qualifier
- : = separator for component data elements within a composite (here, date and date format qualifier)
- 102 = date format qualifier, code indicating the format of the date (here, CCYYMMDD)
- ’ = segment terminator

Syntax of DTM Data Segment

```
instance ds#DTM_DS memberOf edi#DataSegment
          edi#hasFormat hasValue fc#M1_Only_Format
          edi#hasFormatDescriptor hasValue DTM_DS_FD

instance DTM_DS_FD memberOf edi#FormatDescriptor
          nonFunctionalProperties
              dc#description hasValue "FormatDescriptor for EDIFACT DTM data segment"
          endNonFunctionalProperties
          edi#formats1stComponent hasValue cde#C507_CDE


instance cde#C507_CDE memberOf edi#ComplexDataElement
          edi#hasFormat hasValue C507_CDE_Format
          edi#hasFormatDescriptor hasValue C507_CDE_FD

instance C507_CDE_Format memberOf edi#Format
          nonFunctionalProperties
              dc#description hasValue "Format for EDIFACT C507 Complex Data Element"
          endNonFunctionalProperties
          edi#formatFor1stComponent hasValue fc#FC_M1
          edi#formatFor2ndComponent hasValue fc#FC_O1
          edi#formatFor3rdComponent hasValue fc#FC_O1

instance C507_CDE_FD memberOf edi#FormatDescriptor
          nonFunctionalProperties
              dc#description hasValue
                             "FormatDescriptor for EDIFACT C507 Complex Data Element"
          endNonFunctionalProperties
          edi#formats1stComponent hasValue de#DE_2005       // date qualifier
          edi#formats2ndComponent hasValue de#DE_2380       // date
          edi#formats3rdComponent hasValue de#DE_2379       // date format qualifier
```

Encoding EDIFACT Subsets' Meaning

- **T 7.3 (D7.2) EDIFACT ontology semantics definition (M24)**
  - Meanings of permitted/suggested codes for data elements ontologized
  - Meanings of data elements ontologized
  - Meanings of data segments ontologized
  - Meanings of implicit relations between message components determined and ontologized
  - Templates defined to encode statements based on implicit relationships in messages

Not very applicable to TripCom Use Cases
Priority given to improving ontologies.

Sixth Framework Programme Priority 2
Information Society Technologies (IST)
Specific Targeted Research Project

Example Definitions in Ontologies

```
concept HazardousCargo
      subConceptOf
            {cyc#DangerousTangibleThing, Cargo}
    nonFunctionalProperties
      dc#description hasValue "Cargo with
       dangerous properties, according to
       relevant dangerous goods regulations."
    endNonFunctionalProperties


instance cyc#IrishLanguage
      memberOf cyc#NaturalLanguage
```


Sixth Framework Programme Priority 2
Information Society Technologies (IST)
Specific Targeted Research Project

Example Relation Definition

```
relation dutyPreferenceOriginAgreement
     (ofType cyc#PartiallyTangibleProduct,
      ofType cyc#Agreement)
    nonFunctionalProperties
        dc#description hasValue
     "dutyPreferenceOriginAgreement (GOODS,
      AGREEMENT) means that duty charged on
      GOODS is based on their origin as
      specified by the agreement, AGREEMENT."
    endNonFunctionalProperties
```


Sixth Framework Programme Priority 2
Information Society Technologies (IST)
Specific Targeted Research Project

Example Code Set Definition

```
instance DE_2009CodeSet memberOf edis#CodeSet
    nonFunctionalProperties
      dc#description hasValue "Code Set for EDIFACT Data
       Element 2009, Terms time relation"
    endNonFunctionalProperties

instance DE_2009ECodeSet memberOf edis#CodeSet
    nonFunctionalProperties
      dc#description hasValue “EANCOM special Code Set for
       EDIFACT Data Element 7085, Terms time relation"
    endNonFunctionalProperties

instance ISO4217CodeSet memberOf edis#CodeSet
    nonFunctionalProperties
      dc#description hasValue "ISO 4217 three-letter
       currency codes"
    endNonFunctionalProperties
```

Example Code Set Usage

```
relationInstance edis#encodedAs
      (ct#HazardousCargo , ecs#DE_7085CodeSet, "11")

relationInstance edis#encodedAs
      (dr#dutyPreferenceOriginAgreement,
       ecs#DE_9213CodeSet , "2")

relationInstance edis#encodedAs
      (cyc#Euro, ecs#ISO4217CodeSet, "EUR")
```

Example Templates

```
relationInstance edis#componentOfType
      (ds#CUX_DS, cyc#UnitOfMoney)
// The CUX data segment represents a currency ...

relationInstance edis#hasSameMeaningAs (ds#CUX_DS,
       scp#ThisFAPosition, scp#FirstSubcomponentPosition)
// ... which is the same as that in its first data element

relationInstance edis#subComponentOfType
      (ds#CUX_DS, 1, cyc#UnitOfMoney)

relationInstance edis#subComponentOfType
      (ds#CUX_DS, 2, cyc#UnitOfMoney)
// The first two CUX slots represent currencies ...

relationInstance edis#directlyEncoded (ds#CUX_DS, 1)
relationInstance edis#directlyEncoded (ds#CUX_DS, 2)
// ... which are those encoded in their data elements
```

Example Templates

```
instance ds#CUX_DS memberOf edi#DataSegment
      decimalValuedSubcomponent hasValue 3
// The third slot of the CUX data segment is a number ...

relationInstance edis#directlyEncoded (ds#CUX_DS, 3)
// ... which is the same that encoded in its data element

relationInstance edis#subComponentOfType
      (ds#CUX_DS, 4, cyc#CurrencyExchange)
// The fourth slot of the CUX data segment represents a
currency exchange

relationInstance edis#directlyEncoded (ds#CUX_DS, 4)
// ... which is the same that encoded in its data element
```

Example Templates

```
instance CUXTemplate1 memberOf edis#ComponentTemplate
      edis#templateForComponent hasValue ds#CUX_DS
      edis#templateRelation hasValue
         cur#usedExchangeRateBetween
      edis#templateArg1Encoded hasValue
         scp#FirstSubcomponentPosition
      edis#templateArg2Encoded hasValue
         scp#SecondSubcomponentPosition
      edis#templateArg3Encoded hasValue
         scp#ThirdSubcomponentPosition

// The exchange rate used between the two specified
// currencies is the specified number
```

Terms in Ontologies

Focus was on defining & interlinking concepts

- Types/Classes
  - around 2800 created
  - over 1100 from external ontology (OpenCyc)
- Relations
  - around 850 created
  - over 100 from external ontology
- Instances (mostly countries, currencies, languages, cities)
  - over 1000 created
  - over 1600 from external ontology
  - Tens of thousands of cities/provinces not counted

Templates

- Defined
  - 241 Code Sets
    - Assigned 4670 Codes
  - 75 component template
  - 150 component type assignments

Sixth Framework Programme Priority 2
Information Society Technologies (IST)
Specific Targeted Research Project

Next Deliverable

- **D 7.3 Relationships among ontology modules within EDIFACT and with external ontologies (M30)**
  - T7.4 Definition of internal dependencies of EDIFACT (M24) Necessary part of T7.3
  - T7.5 Mapping of created EDIFACT ontologies to other ontologies (M30)
    - Mapping to OpenCyc Integrated into T 7.3
    - Mapping to other ontologies created for TripCom
    - Mapping to other external ontologies

T7.6 Evaluation and refinement of the ontologies and their applications (M36)

- Merging the disparate ontologies created in T7.3
- Evaluating the combined ontologies
- Refining the ontologies

Sixth Framework Programme Priority 2
Information Society Technologies (IST)
Specific Targeted Research Project

Questions?

www.tripcom.org

universität innsbruck

National University of Ireland, Galway
Ollscoil na hÉireann, Gaillimh

Universität Stuttgart
Germany

TU WIEN
Technische Universität Wien
Vienna University of Technology

Freie Universität Berlin

ontotex t
Semantic Technology Lab

profium

CEFRIEL
FORGING INNOVATION

Telefónica
Telefónica Investigación y Desarrollo