**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES
(EUROSTAT)**

**ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT
(OECD)
STATISTICS DIRECTORATE**

Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Geneva, 9-11 February 2004)

Topic (ii): Metadata interchange

CLASET
AN EDI MESSAGE FOR EXCHANGING CLASSIFICATIONS

Contributed Paper

Submitted by Eurostat[1]

Abstract

CLASET is an EDI message designed to facilitate the exchange of tree structures such as classifications, code lists, organisational charts, concept definitions, catalogues, etc.

The message has been designed in a generic way and provides mechanisms to describe the nature of the information exchanged within it.

All the implementation and software tools are based on a unique Conceptual Data Model (CDM). This CDM can be used to define all information needed about the message and three main data structures: classifications, tables of links between two classifications and updates. This CDM can be used not only as an exchange message, but also as a database model.

CLASET is presently implemented in four formats: XML, SGML, UN/EDIFACT and HTML; this list is not at all finite: new syntaxes might come under consideration for inclusion in the message, with the move to yet unknown technologies.

---

[1] Prepared by Danny DELCAMBRE, danny.delcambre@cec.eu.int

I. WHAT IS CLASET ?

1.   CLASET is an EDI message designed to exchange any kind of **structured** metadata : classifications, tree structures, organisational charts, concepts, definitions, catalogues, code lists and links between classifications.

2.   The message has been designed in a generic way and provides mechanisms to describe the nature of the information exchanged within it.

II. WHY CLASET

3.   Together with methodological manuals, concepts, definitions and business registers, classifications belong to the very absolute prerequisites of all statistical systems, national as well as international. With the ever-increasing integration of these systems, the need has emerged for an efficient and reliable tool for exchanging these hierarchical structures. Sofar the lack of such a standardized tool has often led to burdensome reformating work of the information received prior to its inclusion into any database system.

4.   The objective of CLASET is to simplify the exchange procedures of classifications by making available a normalized EDI message allowing for a direct consultation and uploading of the information transmitted, regardless of the hardware and software used by the addressee and recipient bodies.

III. WHAT KIND OF INFORMATION CAN BE EXCHANGED ?

5)   The message covers the following exchange scenarios :

- General information on classifications (maintenance agency, contact addresses, version number, etc.) ;
- All or part of the content or structure of a classification ;
- Data maintenance operations on classifications (update of codes in case of revisions) ;
- All or part of the links between classifications ;
- Any combination of the above.

IV. MESSAGE IMPLEMENTATIONS

6)   CLASET is presently implemented into three syntaxes:

- SGML (Standard Generalised Mark-up Language) ;
- UN/EDIFACT (Status of United Nations standard obtained in September 1998[2]) ;
- XML (Extended Mark-up Language) ;

7)   The structure of the files in these syntaxes is always derived from the data model. This method brings exactness without ambiguity in data. This list is not at all finite: new syntaxes might come under consideration for inclusion in the message, with the move to yet unknown technologies.

---

[2] http://www.unece.org/trade/untdid/d00a/trmd/claset_c.htm (United Nations Directories for Electronic Data Interchange forAdministration, Commerce and Transport).

V.   CLASET DATA MODEL

8)   All the implementation and software tools are based on a unique Conceptual Data Model (CDM). This CDM can be used to define all information needed about the message and three main data structures: classifications, tables of links between two classifications and updates. This CDM can be used not only as an exchange message, but also as a database model.

9)   The UML model of CLASET was developed with the collaboration of several National Statistical Offices and therefore captures well the semantic of statistical classifications and associated artefacts.

VI.   HOW DOES IT WORK IN PRACTICE

10)   A toolbox is available for free in order to ease the implementation of CLASET. Its functionalities are:

- Conversion between each supported implementations;
- Generation of an HTML file for validation purposes;
- Configurable text import and export to/from a CLASET message.

11)   This toolbox has been written in Java.

12)   Ideally the CLASET tools should be integrated in the user's central system for the management of classifications. This necessitates to put in correspondence the user's data model with the CLASET data model. This integration will require only a few days/man work. This somehow modest investment will bring about large benefits as the CLASET tools offer almost automatic import and export functionalities.

13)   When interfacing CLASET with its own system, the user will opt for one of the proposed syntaxes, or the import / export of text files. The conversion tools provided with CLASET make it possible to go from any of the syntaxes to the other two.

14)   Parametrisation of text import/export can be done with an XML file that will generate the code to proceed the messages. This has been developed to allow the user to interface CLASET with export files from a database.
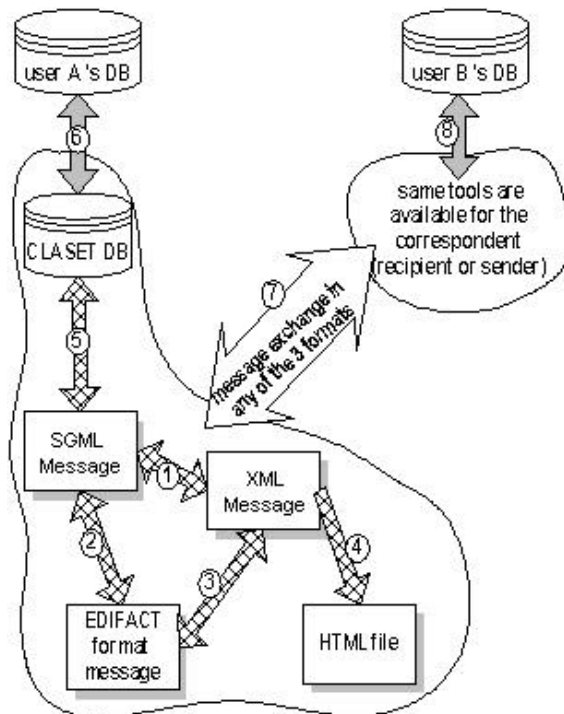
15) CLASET has proved itself in life use over a number of years in France and Finland. Eurostat has implemented CLASET in the download section of its server for classifications RAMON. In Belgium a regional office for unemployment is testing CLASET in a European project for employment ontology. In this scope, CLASET has been implemented in the "Protégé" project (http://protege.stanford.edu) for the building block of employment nomenclatures persistence.

VII. WHERE DO I FIND THE TOOLS AND THE DOCUMENTATION ?

16) The CLASET tools can be downloaded free of charge from the following address :

http://europa.eu.int/comm/eurostat/ramon/miscellaneous/index.cfm?TargetUrl=DSP_CLASET

17) The RAMON server is presently under heavy reconstruction, so this address might have changed by the time you try to access the Web site. If this is the case, please feel free to contact Danny DELCAMBRE (danny.delcambre@cec.eu.int) to get the right address.

18) Full documentation is also available from the above-mentioned address ; it includes the following documents :

- Concepts and terms related to classifications ;
- Conceptual data model ;
- Description of the UML model ;
- Description of the Merise model ;
- Quick implementation guide ;
- General implementation guide ;
- XML implementation guide ;
- SGML implementation guide ;
- EDIFACT message implementation guide ;
- Toolbox user guide ;
- CLASET XML schema documentation.
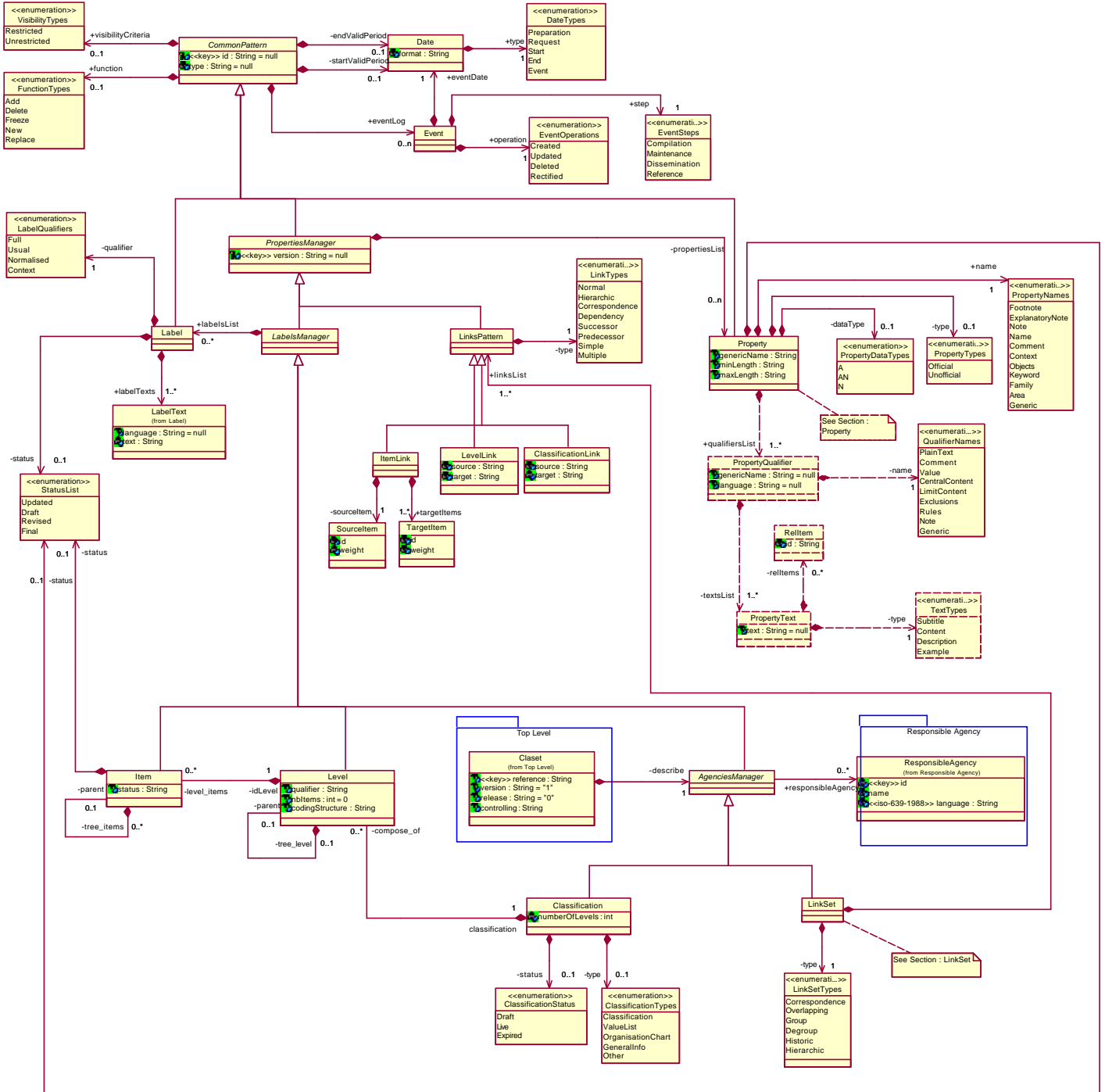
19)   For further information:

Danny DELCAMBRE
Eurostat, Unit B1 "Methodological coordination, structural indicators, classifications and registers"
danny.delcambre@cec.eu.int

ANNEX 1 – CLASET SUCCINCT UML MODEL

<<enumeration>>
VisibilityTypes
Restricted
Unrestricted

<<enumeration>>
FunctionTypes
Add
Delete
Freeze
New
Replace

CommonPattern
<<key>> id : String = null
type : String = null

+visibilityCriteria   -endValidPeriod   0..1
+function   -startValidPeriod   0..1

Date
format : String

+type   1

<<enumeration>>
DateTypes
Preparation
Request
Start
End
Event

+eventDate   1

+eventLog   **Event**   +operation   0..n

<<enumeration>>
EventOperations
Created
Updated
Deleted
Rectified

+step   1

<<enumerati..>>
EventSteps
Compilation
Maintenance
Dissemination
Reference

<<enumeration>>
LabelQualifiers
Full
Usual
Normalised
Context

-qualifier   1

PropertiesManager
<<key>> version : String = null

-propertiesList   0..n

+name   1

<<enumerati..>>
PropertyNames
Footnote
ExplanatoryNote
Note
Name
Comment
Context
Objects
Keyword
Family
Area
Generic

<<enumerati..>>
LinkTypes
Normal
Hierarchic
Correspondence
Dependency
Successor
Predecessor
Simple
Multiple

+labelsList   0..*

Label

LabelsManager

LinksPattern
-type   1

Property
genericName : String
minLength : String
maxLength : String

-dataType   0..1

<<enumeration>>
PropertyDataTypes
A
AN
N

-type   0..1

<<enumerati..>>
PropertyTypes
Official
Unofficial

+labelTexts   1..*

LabelText
(from Label)
language : String = null
text : String

+linksList   1..*

See Section :
Property

-status   0..1

<<enumeration>>
StatusList
Updated
Draft
Revised
Final

ItemLink

LevelLink
source : String
target : String

ClassificationLink
source : String
target : String

+qualifiersList   1..*

PropertyQualifier
genericName : String = null
language : String = null

-name   1

<<enumerati..>>
QualifierNames
PlainText
Comment
Value
CentralContent
LimitContent
Exclusions
Rules
Note
Generic

-sourceItem   1   1..*   +targetItems

SourceItem
id
weight

TargetItem
id
weight

RelItem
id : String

-relItems   0..*

-textsList   1..*

PropertyText
text : String = null

-type   1

<<enumerati..>>
TextTypes
Subtitle
Content
Description
Example

-status   0..1   -status
0..1   -status

Item
status : String

-parent   0.1
-tree_items   0..*

0..*

1   -level_items   -idLevel

Level
qualifier : String
nbItems : int = 0
codingStructure : String

-parent   0..1
-tree_level   0..*   1   0..*   -compose_of

Top Level

Claset
(from Top Level)
<<key>> reference : String
version : String = "1"
release : String = "0"
controlling : String

-describe   **AgenciesManager**   1

+responsibleAgencies   0..*

Responsible Agency

ResponsibleAgency
(from Responsible Agency)
<<key>> id
name
<<iso-639-1988>> language : String

classification   1   **Classification**
numberOfLevels : int

LinkSet

-status   0..1   -type   0..1

-type   1

See Section : LinkSet

<<enumeration>>
ClassificationStatus
Draft
Live
Expired

<<enumeration>>
ClassificationTypes
Classification
ValueList
OrganisationChart
GeneralInfo
Other

<<enumerati..>>
LinkSetTypes
Correspondence
Overlapping
Group
Degroup
Historic
Hierarchic

ANNEX 2 - EXTRACT OF A MESSAGE GENERATED IN THE XML FORMAT

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE Claset PUBLIC "-//EUROSTAT EG6 WG3//CLASET//EN" "clasetxml.dtd">
<Claset reference="1234" version="1" release="0">
 <Classification id="CC" version="1">
  <Label qualifier="Full">
   <LabelText language="DE">Klassifikation der Bauwerke </LabelText>
   <LabelText language="FR">Nomenclature des ouvrages de construction</LabelText>
   <LabelText language="EN">Classification of Types of Constructions</LabelText>
  </Label>
  <Level id="1" qualifier="1" nbItems="2">
   <Level id="2" qualifier="2" nbItems="6">
    <Level id="3" qualifier="3" nbItems="20">
     <Level id="4" qualifier="4" nbItems="46" />
    </Level>
   </Level>
  </Level>
  <Item id="1" idLevel="1">
   <Label qualifier="Full">
    <LabelText language="DE">Hochbauten</LabelText>
    <LabelText language="FR">Bâtiments</LabelText>
    <LabelText language="EN">Buildings</LabelText>
   </Label>
   <Item id="11" idLevel="2">
    <Label qualifier="Full">
     <LabelText language="DE">Wohngebäude</LabelText>
     <LabelText language="FR">Bâtiments d'habitation</LabelText>
     <LabelText language="EN">Residential buildings</LabelText>
    </Label>
    <Item id="111" idLevel="3">
     <Label qualifier="Full">
      <LabelText language="DE">Gebäude mit einer Wohnung</LabelText>
      <LabelText language="FR">Maisons individuelles</LabelText>
      <LabelText language="EN">One-dwelling buildings</LabelText>
     </Label>
     <Item id="1110" idLevel="4">
      <Label qualifier="Full">
       <LabelText language="DE">Gebäude mit einer Wohnung</LabelText>
       <LabelText language="FR">Maisons individuelles</LabelText>
       <LabelText language="EN">One-dwelling buildings</LabelText>
      </Label>
      <Property name="ExplanatoryNote">
       <PropertyQualifier name="CentralContent" language="DE">
        <PropertyText type="Content">- Einzelhäuser wie Bungalows, Villen, Chalets, Forsthäuser, Bauernhäuser, Landhäuser, Sommerhäuser, Wochenendhäuser usw.</PropertyText>
```

ANNEX 3 - EXTRACT OF A MESSAGE GENERATED IN THE SGML FORMAT

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE CLASET PUBLIC "-//EUROSTAT EG6 WG3//CLASET SGML//EN" "clasetSGML.dtd">
<CLASET>
 <HEAD VERSION="1" RELEASE="0">
  <REF>Claset1234</REF>
 </HEAD>
 <CLASSIF QUALIFIER="CLS">
  <ID>CC</ID>
  <VERSION>1</VERSION>
  <PROP NAME="LBL" DATATYPE="A">
   <QUAL NAME="USU">
    <P TYPE="CNT" LG="DE">Klassifikation der Bauwerke </P>
    <P TYPE="CNT" LG="FR">Nomenclature des ouvrages de construction</P>
    <P TYPE="CNT" LG="EN">Classification of Types of Constructions</P>
   </QUAL>
  </PROP>
  <LEVEL>
   <ID>1</ID>
   <QUALIFIER>1</QUALIFIER>
  </LEVEL>
  <ITEM>
   <ID>1</ID>
   <PROP NAME="LBL" DATATYPE="A">
    <QUAL NAME="USU">
     <P TYPE="CNT" LG="DE">Hochbauten</P>
     <P TYPE="CNT" LG="FR">Bâtiments</P>
     <P TYPE="CNT" LG="EN">Buildings</P>
    </QUAL>
   </PROP>
  </ITEM>
  <LEVEL>
   <ID>2</ID>
   <QUALIFIER>2</QUALIFIER>
  </LEVEL>
  <ITEM PARENT="1">
   <ID>11</ID>
   <PROP NAME="LBL" DATATYPE="A">
    <QUAL NAME="USU">
     <P TYPE="CNT" LG="DE">Wohngebäude</P>
     <P TYPE="CNT" LG="FR">Bâtiments d'habitation</P>
     <P TYPE="CNT" LG="EN">Residential buildings</P>
    </QUAL>
   </PROP>
  </ITEM>
```

ANNEX 4 - EXTRACT OF A MESSAGE GENERATED IN THE EDIFACT FORMAT

```
UNA:+,? '
UNB+UNOC:3+SENDER+RECIPIENT+031127:2153+Claset1234'
UNH+Claset1234+CLASET:D:96B:UN'
BGM+74'
VLI+CC:ZA++++++1:S08:142'
ATT+7+21:S06:142'
ELM+X+2'
CAV+5:S32:142'
FTX+ACN+++Klassifikation der Bauwerke +DE'
FTX+ACN+++Nomenclature des ouvrages de construction+FR'
FTX+ACN+++Classification of Types of Constructions+EN'
SCD+Z07+1+++1'
ATT+7+31:S06:142+::142:1'
SCD+Z08+1+++1+'
ATT+7+21:S06:142'
ELM+X+2'
CAV+5:S32:142'
FTX+ACN+++Hochbauten+DE'
FTX+ACN+++Bâtiments+FR'
FTX+ACN+++Buildings+EN'
SCD+Z07+2+++2'
ATT+7+31:S06:142+::142:2'
SCD+Z08+11+++2+1'
ATT+7+21:S06:142'
ELM+X+2'
CAV+5:S32:142'
FTX+ACN+++Wohngebäude+DE'
FTX+ACN+++Bâtiments d?'habitation+FR'
FTX+ACN+++Residential buildings+EN'
SCD+Z07+3+++3'
ATT+7+31:S06:142+::142:3'
SCD+Z08+111+++3+11'
ATT+7+21:S06:142'
ELM+X+2'
CAV+5:S32:142'
FTX+ACN+++Gebäude mit einer Wohnung+DE'
FTX+ACN+++Maisons individuelles+FR'
FTX+ACN+++One-dwelling buildings+EN'
SCD+Z07+4+++4'
ATT+7+31:S06:142+::142:4'
SCD+Z08+1110+++4+111'
ATT+7+21:S06:142'
ELM+X+2'
CAV+5:S32:142'
FTX+ACN+++Gebäude mit einer Wohnung+DE'
FTX+ACN+++Maisons individuelles+FR'
```