Implementation of profile in SGML project

1 Introduction

In the ITSIG exchange DTD for Standards the bibliographical, monitoring, and other information about a document (i.e. metadata) must be attached to the document instance. Roughly speaking, it is catalogue-like description[1] of a document. In standard development business such information appear in the several types of applications:

— electronic publishing of Standards;

— project monitoring databases (PMDB);

— electronic information services like PERINORM or ISO Online;

— catalogues production.

As a rule, this information resides in a particular PMDB and is requested for others PMDBs and applications. New ISONET manual describes more that hundred fields which may be potentially exchanged. The ISONET manual gives two representation — conventional and SGML. The least can be used in ITSIG exchanged DTD to encode metadata. Historically, we call document's metadata as "profile".

In this document we would like to describe the implementation of profile in the SGML project. As it was decided on the SGML group meeting on December 1997, we have to progress with the use of new ISONET manual for exchange of metadata. At regional and national SDOs there are procedures when an International Standard is endorsed as their standard. It is very important that profile of International Standard would be easily reused by Member bodies' SGML-based production system.

2 Architecture

2.1 General considerations

We already understood, that it is not convenient to use the same profile in the different applications. For example, ISO reference number is usually presented as "ISO 1234-1:1994", while for SGML publishing system it should be encoded as "ISO 1234-1 : 1994". Another example, special procedure as fast-track is a just a flag in PMDB, but it is a particular boilerplate text to be inserted on the cover page for publishing system.

We separate profile extracted from databases (database-oriented), which we have to encode with conformance with new ISONET manual and profile to be used by publishing system (document-oriented) which should be convenient for publishing needs. Easy extensibility of profile should be taken into account in the implementation.

It is important to emphasize that only database-oriented profile should be used for interchange of data between the different organizations involved. Document-oriented profile is internal choice of each organization. Of course, it may be coincident with database-oriented profile in the simple cases.

2.2 Database-oriented profile

In the example of database-oriented profile only four groups are used (see clause A.1 for the DTD):

---

1) Without any ideas how this information may be presented on a paper or screen.

00   data elements for the file label;

10   data elements for projects;

20   data elements for products (documents in our case);

40   data elements for committee information.

Example 1
```
<ISONET>
<E00>
<E00.003><DATE>1999-01-13</DATE></E00.003>
<E00.006><ORG.ID>A000xx</ORG.ID></E00.006>
</E00>


<E10>
<E10.001><REC.ID>A000xx:10:016543</REC.ID></E10.001>
<E10.202><ORG.ID>ISO</ORG.ID></E10.202>
<E10.204><COMM.ID>ISO/TC 176/SC 3</COMM.ID></E10.204>
<E10.300><HSTAGE>6060</HSTAGE></E10.300>
<E10.310><DATE>1998-01-01</DATE></E10.310>
</E10>


<E20>
<E20.001><REC.ID>A000xx:20:016543</REC.ID></E20.001>
<E20.104>
<REFNUM><CODE>ISO</CODE> <INT>9000</INT>-<CODE>1</CODE>:<INT>1994</INT></REFNUM>
</E20.104>
<E20.200>
<TITLE LANG=EN>
<COMP>xxx</COMP> -- <COMP>xxx</COMP> -- Part 1: <COMP>xxx</COMP>
</TITLE>
<TITLE LANG=FR>
<COMP>yyy</COMP> -- <COMP>yyy</COMP> -- Partie 1: <COMP>yyy</COMP>
</TITLE>
</E20.200>
<E20.208><LANG>EN</LANG></E20.208>
<E20.212><INT>0</INT></E20.212>
<E20.216><CODE>ST</CODE></E20.216>
<E20.220><INT>1</INT></E20.220>
<E20.228><CODE>ZX</CODE></E20.228>
<E20.310><DATE>1998-01-01</DATE></E20.310>
<E20.502><ICS>10.110.10</ICS></E20.502>
<E20.504><TEXT LANG=en>paints</TEXT><TEXT LANG=en>varnishes</TEXT></E20.504>
<E20.700><ORG.ID>ISO</ORG.ID></E20.700>
</E20>


<E40>
<E40.104><COMM.ID>ISO/TC 176</COMM.ID></E40.104>
<E40.130><TEXT lang=en>Quality management and quality assurance</TEXT></E40.130>
<E40.140><CODE>ISO/TC</CODE></E40.140>
<E40.145><INT>176</INT></E40.145>
<E40.150><COMM.ID>TMB</COMM.ID></E40.150>

<E40.104><COMM.ID>ISO/TC 176/SC 3</COMM.ID></E40.104>
<E40.130><TEXT lang=en>Supporting technologies</TEXT></E40.130>
<E40.140><CODE>SC</CODE></E40.140>
<E40.145><INT>3</INT></E40.145>
<E40.150><COMM.ID>ISO/TC 176</COMM.ID></E40.150>
```

```
</E40>
</ISONET>
```

There are several changes to the latest edition of ISONET manual:

a) element TITLE (used in E20.200) is structured in the way that is mixed text and title's components; from one hand, the generation of titles becomes simple — one has to remove all tagging only; from another hand there is access to all individual components;

b) element REFNUM (used in E10.104) is structured also in the mixed way for the same reasons as above;

c) element E40.145 was added to specify the number of a committee.

As a rule, database-oriented profile contains a lot of information for the main document and just a few elements for the referred documents. An example for referred document is below:

Example 2
```
<ISONET>
<E00>
<E00.003><DATE>1999-01-13</DATE></E00.003>
<E00.006><ORG.ID>A000xx</ORG.ID></E00.006>
</E00>
<E10>
<E10.001><REC.ID>A000xx:10:016547</REC.ID></E10.001>
<E10.300><HSTAGE>6060</HSTAGE></E10.300>
</E10>
<E20>
<E20.001><REC.ID>A000xx:20:016547</REC.ID></E20.001>
<E20.104>
<REFNUM><CODE>ISO</CODE> <INT>9002</INT>:<INT>1994</INT></REFNUM>
</E20.104>
<E20.200>
<TITLE LANG=EN>
<COMP>Quality systems</COMP> -- <COMP>Model for quality assurance in production,
installation and servicing</COMP>
</TITLE>
<TITLE LANG=FR>
<COMP>Syst&egrave;mes qualit&eacute;</COMP> -- <COMP>Mod&egrave;le pour
l&apos;assurance de la qualit&eacute; en production, installation et prestations
associ&eacute;es</COMP>
</TITLE>
</E20.200>
<E20.208><LANG>EN</LANG><LANG>FR</LANG></E20.208>
<E20.216><CODE>ST</CODE></E20.216>
<E20.220><INT>1</INT></E20.220>
</E20>
```

2.3 Document-oriented profile

Document-oriented profile contains the metadata about the document itself (i.e. the main document) and it may contain the metadata about the documents which are referred in the main document. The database-oriented profile for a particular document has to be embedded in the document-oriented profile. So, the document-oriented profile structure is look in the following way:

```
<PROFILE>
<MAINDOC ID="r0">
<ISONET>
... -- database-profile for the main document
</ISONET>
<DOCPROF>
```

```
... -- document-oriented profile for the main document
</DOCPROF>
</MAINDOC>
<REFDOC ID="r1">
<ISONET>
... database-profile for the referred document
</ISONET>
<DOCPROF>
... -- document-oriented profile for the referred document
</DOCPROF>
</REFDOC>
... more referred documents, if any
</PROFILE>
```

At present, document-oriented profile used at ISO/CS is very simple — only elements are added: FIELD to keep a single piece of information, ISOREF to keep references like "ISO 1234:1999 — *Dummy title*", and XP to reuse content of previous elements. An example of additional information for referred document is below.

Example 1
```
<DOCPROF REFID="cs016537" ID="R4">
<FIELD ID="R4.1">IS</FIELD>
<FIELD ID="R4.2">ISO</FIELD>
<FIELD ID="R4.3">ISO</FIELD>
<FIELD ID="R4.4">9002</FIELD>
<FIELD ID="R4.8">2</FIELD>
<FIELD ID="R4.6">1994-06-30</FIELD>
<FIELD ID="R4..M.E">Quality systems</FIELD>
<FIELD ID="R4..C.E">Model for quality assurance in production, installation and
servicing</FIELD>
<FIELD ID="R4..M.F">Syst&egrave;mes qualit&eacute;</FIELD>
<FIELD ID="R4..C.F">Mod&egrave;le pour l&apos;assurance de la qualit&eacute; en
production, installation et prestations associ&eacute;es</FIELD>
<ISOREF VARC="ISO 9002 : 1994" ID="R4.E">
<xp refid="R4..M.E"> &mdash; <xp refid="R4..C.E">
</ISOREF>
<ISOREF VARC="ISO 9002 : 1994" ID="R4.F">
<xp refid="R4..M.F"> &mdash; <xp refid="R4..C.F">
</ISOREF>
</DOCPROF>
```

3 Profile in XML and XLL

It is obvious that profile may be implemented in the XML (which is simplified SGML). XLL (extension of XML for linking) may be used for linking between database-oriented and document-oriented profiles.

4 Conclusions

The separation of profile into database-oriented and document-oriented has the following advantages:

a)   database-oriented profile is based on the agreed format, i.e. ISONET manual, and

b)   implementation of document-oriented profile is internal choice of an organization.

TODO:

a)   adopt the propose approach;

b)   finalize the DTD for database-oriented profile (some elements may be added), and

c)   implement it.

Annex A
(informative)

DTDs for profile

A.1 DTD for database-oriented profile (draft)

```
<!-- (C) International Organization for Standardization 1988, 1989,
1990, 1991, 1992, 1994, 1995, 1996, 1997, 1998
Permission to copy in any form is granted for use with
conforming SGML systems and applications as defined in
ISO 8879, provided this notice is included in all copies.

This is a DTD for ISONET manual. Version 0.01
-//ISO/CS//DTD isonet::0.01//EN

-->


<!--
simple data types
-->
<!--     integer value as in C       -->
<!ELEMENT int     - - (#PCDATA)          >
<!ATTLIST int
id       CDATA   #IMPLIED
>
<!--     real value as in C       -->
<!ELEMENT real     - - (#PCDATA)          >
<!ATTLIST real
id       CDATA   #IMPLIED
>

<!--     minimal literal as in SGML  -->
<!ELEMENT code     - - (#PCDATA)          >
<!ATTLIST code
id       CDATA   #IMPLIED
>

<!--     text (with entities) in SGML -->
<!ELEMENT text     - - (#PCDATA)          >
<!ATTLIST text
id       CDATA   #IMPLIED
lang     CDATA   "en"
>

<!--     Public identifier as in SGML -->
<!ELEMENT pubid     - - (#PCDATA)          >
<!ATTLIST pubid
id       CDATA   #IMPLIED
>

<!--     date in accordance with ISO 8601 -->
<!ELEMENT date     - - (#PCDATA)          >
<!ATTLIST date
id       CDATA   #IMPLIED
>
```

```
<!--     harmonized stage code -->
<!ELEMENT hstage    - - (#PCDATA)          >
<!ATTLIST hstage
id       CDATA   #IMPLIED
>


<!--     ICS classification -->
<!ELEMENT ics       - - (#PCDATA)          >
<!ATTLIST ics
id       CDATA   #IMPLIED
>


<!--     language code in accordance with ISO 639 -->
<!ELEMENT lang      - - (#PCDATA)          >
<!ATTLIST lang
id       CDATA   #IMPLIED
>


<!--     unique record indentifier -->
<!ELEMENT rec.id    - - (#PCDATA)          >
<!ATTLIST rec.id
id       CDATA   #IMPLIED
>


<!--     unique organization identifier -->
<!ELEMENT org.id    - - (#PCDATA)          >
<!ATTLIST org.id
id       CDATA   #IMPLIED
>


<!--     unique committee identifier -->
<!ELEMENT comm.id   - - (#PCDATA)          >
<!ATTLIST comm.id
id       CDATA   #IMPLIED
>


<!--
complex data types
-->

<!--     title which may contain from the several component separated by fillers -->
<!ELEMENT title     - - (#PCDATA|comp)+
>
<!ATTLIST title
id       CDATA   #IMPLIED
lang     CDATA   "en"
>

<!ELEMENT comp - - (#PCDATA)              >
<!ATTLIST comp
id       CDATA   #IMPLIED
>


<!-- reference number from the cover page may be presented as just a text
or in structured way for ISO reference number -->

<!ELEMENT refnum    - - (#PCDATA|code|int)+ >
<!ATTLIST refnum
id       CDATA   #IMPLIED
```

6

```
>


<!--
fields, group 00 (file label)
-->

<!ENTITY % e00 "e00.003, e00.006" >

<!ELEMENT e00.003   - - (date)              >
<!ATTLIST e00.003
id        CDATA   #IMPLIED
>
<!ELEMENT e00.006   - - (org.id)            >
<!ATTLIST e00.006
id        CDATA   #IMPLIED
>



<!--
fields, group 10  (projects)
-->

<!ENTITY % e10 "e10.001, e10.202, e10.204, (e10.300, e10.310)?, (e10.336, e10.338)?" >

<!-- project information -->
<!ELEMENT e10.001   - - (rec.id)            >
<!ATTLIST e10.001
id        CDATA   #IMPLIED
>

<!-- project owner -->
<!ELEMENT e10.202   - - (org.id+)           >
<!ATTLIST e10.202
id        CDATA   #IMPLIED
>

<!-- project technical committee -->
<!ELEMENT e10.204   - - (comm.id+)          >
<!ATTLIST e10.204
id        CDATA   #IMPLIED
>

<!-- project current stage code -->
<!ELEMENT e10.300   - - (hstage)            >
<!ATTLIST e10.300
id        CDATA   #IMPLIED
>

<!-- date of current stage code -->
<!ELEMENT e10.310   - - (date)              >
<!ATTLIST e10.310
id        CDATA   #IMPLIED
>

<!-- project next stage code -->
<!ELEMENT e10.336   - - (hstage)            >
<!ATTLIST e10.336
id        CDATA   #IMPLIED
>
```

```
<!-- target date of next stage code -->
<!ELEMENT e10.338   - -  (date)             >
<!ATTLIST e10.338
id        CDATA   #IMPLIED
>


<!--
fields, group 20  (products)
-->


<!ENTITY % e20 "e20.001, e20.104, e20.200, e20.208?, e20.212?, e20.216, e20.220?,
e20.228?, e20.310?, e20.502?, e20.504?, e20.700?" >

<!-- document information -->
<!ELEMENT e20.001   - -  (rec.id)          >
<!ATTLIST e20.001
id        CDATA   #IMPLIED
>

<!-- document reference number -->
<!ELEMENT e20.104   - -  (refnum)          >
<!ATTLIST e20.104
id        CDATA   #IMPLIED
>

<!-- document title -->
<!ELEMENT e20.200   - -  (title+)          >
<!ATTLIST e20.200
id        CDATA   #IMPLIED
>

<!-- document languages -->
<!ELEMENT e20.208   - -  (lang+)           >
<!ATTLIST e20.208
id        CDATA   #IMPLIED
>

<!-- number of pages type -->
<!ELEMENT e20.212   - -  (int)             >
<!ATTLIST e20.212
id        CDATA   #IMPLIED
>

<!-- document type -->
<!ELEMENT e20.216   - -  (code)            >
<!ATTLIST e20.216
id        CDATA   #IMPLIED
>

<!-- document edition -->
<!ELEMENT e20.220   - -  (int)             >
<!ATTLIST e20.220
id        CDATA   #IMPLIED
>

<!-- document price group -->
<!ELEMENT e20.228   - -  (code)            >
<!ATTLIST e20.228
```

```
id        CDATA   #IMPLIED
>

<!-- date of publication -->
<!ELEMENT e20.310   - -  (date)          >
<!ATTLIST e20.310
id        CDATA   #IMPLIED
>

<!-- document ICS -->
<!ELEMENT e20.502   - -  (ics+)          >
<!ATTLIST e20.502
id        CDATA   #IMPLIED
>

<!-- document TIT descriptor -->
<!ELEMENT e20.504   - -  (text+)         >
<!ATTLIST e20.504
id        CDATA   #IMPLIED
>

<!-- issuing body -->
<!ELEMENT e20.700   - -  (org.id+)       >
<!ATTLIST e20.700
id        CDATA   #IMPLIED
>


<!--
fields, group 40  (committees)
-->

<!ENTITY % e40 "e40.104, e40.130, e40.140, e40.145, e40.150?" >

<!-- committee identifier -->
<!ELEMENT e40.104   - -  (comm.id)       >
<!ATTLIST e40.104
id        CDATA   #IMPLIED
>

<!-- committee name -->
<!ELEMENT e40.130   - -  (text+)         >
<!ATTLIST e40.130
id        CDATA   #IMPLIED
>

<!-- committee type -->
<!ELEMENT e40.140   - -  (code)          >
<!ATTLIST e40.140
id        CDATA   #IMPLIED
>

<!-- committee number -->
<!ELEMENT e40.145   - -  (int)           >
<!ATTLIST e40.145
id        CDATA   #IMPLIED
>

<!-- reporting committee -->
```

```
<!ELEMENT e40.150   - -  (comm.id)            >
<!ATTLIST e40.150
id        CDATA   #IMPLIED
>


<!--
containers for the groups
-->

<!ELEMENT e00   - - (%e00;) >
<!ATTLIST e00
id        CDATA   #IMPLIED
>

<!ELEMENT e10   - - (%e10;) >
<!ATTLIST e10
id        CDATA   #IMPLIED
>

<!ELEMENT e20   - - (%e20;)+ >
<!ATTLIST e20
id        CDATA   #IMPLIED
>

<!ELEMENT e40   - - ((%e40;)+) >
<!ATTLIST e40
id        CDATA   #IMPLIED
>

<!--

whole ISONET manual

-->
<!ELEMENT isonet  - - (e00?, e10?, e20?, e40?) >
<!ATTLIST isonet
id        CDATA   #IMPLIED
>
```

A.2 DTD for document-oriented profile

to be provided...