Descriptive metadata formats

The following standards should be added in the controlled vocabulary of the MDTYPE attribute:

- *"LIDO" (Lightweight Information Describing Objects, see http://lido-schema.org/)*
- *"EN15744" (CEN TC 372 Metadata standard EN 15744:2009 : Film identification - Minimum set of metadata for cinematographic works, see http://filmstandards.org/)*
- *"EN15907" (CEN TC 372 Metadata standard EN 15907:2010 : Film identification - Enhancing interoperability of metadata: Element sets and structures, see http://filmstandards.org/)*

Background:

The National Digital Library (NDL) of Finland is an entity within the remit of the Ministry of Education and Culture, which basis is formed by libraries, archives, museums, and other organizations in Finland storing cultural heritage. The NDL aims to create a nationally unified structure for contents and services ensuring the effective and high-quality management, dissemination, and digital preservation of cultural digital information resources. METS will be used as a base metadata document format in hundreds of independent background systems maintained by the memory organizations. All of the data will be preserved in a single preservation system, which includes several preservation services, such as data validation and migration. See http://kdk.fi/en/

Arguments:

- In the memory organizations of NDL, different descriptive metadata formats are used for different types of digital objects. A few more formats are needed in the METS specification: LIDO, EN15744 and 15907.
- LIDO schema is widely adopted in the museum field, superseding the CDWA Lite v1.1 schema and the museumdat v1.0 schema.
- EN 15744 and EN 15907 are standards of the European Committee for Standardization used for the cinematographic works and therefore will be widely used in the EU in the near future.

Example:

```
<dmdSec ID="dmd-id" CREATED="2013-01-28T14:36:00">
  <mdWrap MDTYPE="LIDO">
    <xmlData>
        <!-- Descriptive metadata in LIDO format -->
        <!-- ... -->
    </xmlData>
  </mdWrap>
</dmdSec>
```

Contact:

Juha Lehtonen, CSC – IT Center for Science, juha.lehtonen@csc.fi

METS version

The element <mets> should have a VERSION attribute. This attribute would state the METS version, which was used when the METS document was created or updated.

Background:

The National Digital Library (NDL) of Finland is an entity within the remit of the Ministry of Education and Culture, which basis is formed by libraries, archives, museums, and other organizations in Finland storing cultural heritage. The NDL aims to create a nationally unified structure for contents and services ensuring the effective and high-quality management, dissemination, and digital preservation of cultural digital information resources. METS will be used as a base metadata document format in hundreds of independent background systems maintained by the memory organizations. All of the data will be preserved in a single preservation system, which includes several preservation services, such as data validation and migration. See http://kdk.fi/en/

Arguments:

In NDL, METS will be used for several years ahead in hundreds of independent background systems. Different background systems use different METS versions, and all the METS documents are collected together for digital preservation. When validating (or migrating in the future) a METS document, it is necessary to distinguish, what was the original METS version that was used. For example, if a certain element or attribute is missing from a document, the reason may be that it did not exist in the METS specification when the document was created. This version information in the document itself would explain the issue.

Example:

```
<mets OBJID="mets-id" xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd" VERSION="1.9.1"
PROFILE="mets-profile">
  <!-- ... -->
</mets>
```

Contact:

Juha Lehtonen, CSC – IT Center for Science, juha.lehtonen@csc.fi

PIDs for different METS sections

The elements <dmdSec>, <amdSec>, <techMD>, <rightsMD>, <sourceMD>, <digiprovMD>, <fileSec>, <fileGrp> and <structMap> should have attributes PID and PIDTYPE. PID would be an attribute for a unique identifier, which globally defines a section of the METS document. PIDTYPE would be the identifier type.

Background:

The National Digital Library (NDL) of Finland is an entity within the remit of the Ministry of Education and Culture, which basis is formed by libraries, archives, museums, and other organizations in Finland storing cultural heritage. The NDL aims to create a nationally unified structure for contents and services ensuring the effective and high-quality management, dissemination, and digital preservation of cultural digital information resources. METS will be used as a base metadata document format in hundreds of independent background systems maintained by the memory organizations. All of the data will be preserved in a single preservation system, which includes several preservation services, such as data validation and migration. See http://kdk.fi/en/

Arguments:

In NDL, a METS document needs to be transformed in a database format (or vice versa), either in the digital preservation system or in a separate independent background system maintained by a memory organization. But sometimes it is necessary to update certain sections of certain METS documents. This will give us two issues:

- If certain parts in a database are updated, the respective sections in the METS document need to be updated.
- If certain sections of a METS document are updated, the respective parts in the database need to be updated.

PIDs would define direct links between the database parts and the METS document sections, giving the correct positions where updating is required.

Technically, attribute ID cannot be used as PID, since colons are not allowed in the ID datatype. However, colons are used e.g. in URNs.

Example:

```
<techMD ID="internal-id" CREATED="2013-01-28T14:36:00"
PID="urn:nbn:fi-1234567890" PIDTYPE="URN">
  <!-- ... -->
</techMD>
```

Contact:

Juha Lehtonen, CSC – IT Center for Science, juha.lehtonen@csc.fi

Internationalization

The element <mdWrap> should have (at least in the descriptive metadata section) a language code attribute LANGUAGE, where the used (human) language of the metadata is stated e.g. according to ISO 639-1 specification.

Background:

The National Digital Library (NDL) of Finland is an entity within the remit of the Ministry of Education and Culture, which basis is formed by libraries, archives, museums, and other organizations in Finland storing cultural heritage. The NDL aims to create a nationally unified structure for contents and services ensuring the effective and high-quality management, dissemination, and digital preservation of cultural digital information resources. METS will be used as a base metadata document format in hundreds of independent background systems maintained by the memory organizations. All of the data will be preserved in a single preservation system, which includes several preservation services, such as data validation and migration. See http://kdk.fi/en/

Arguments:

In the countries with several official languages, the same descriptive metadata may need to be represented with all of the official languages. There are two different official languages in Finland (i.e. Finnish and Swedish), and several descriptive metadata formats are allowed in the METS documents in NDL. We need a simple mechanism to distinguish the languages of descriptive metadata automatically via a common attribute, regardless of the used descriptive metadata type. It is true that some descriptive metadata formats contain internationalization, but this attribute would unify the language structure and give the possibility of internationalization for any metadata format.

Example:

```
<dmdSec ID="dmd-id" CREATED="2013-01-28T14:36:00">
  <mdWrap MDTYPE="DC" LANGUAGE="fi">
    <xmlData>
        <dc:title>Tuntematon sotilas</dc:title>
        <!-- ... -->
    </xmlData>
  </mdWrap>
  <mdWrap MDTYPE="DC" LANGUAGE="sv">
    <xmlData>
        <dc:title>Okänd soldat</dc:title>
        <!-- ... -->
    </xmlData>
  </mdWrap>
  <mdWrap MDTYPE="DC" LANGUAGE="en">
    <xmlData>
        <dc:title>Unknown soldier</dc:title>
        <!-- ... -->
    </xmlData>
  </mdWrap>
</dmdSec>
```

Contact:

Juha Lehtonen, CSC – IT Center for Science, juha.lehtonen@csc.fi

Date and time

The attribute CREATED (which occurs in various elements) should also accept inaccurate date and time in Extended Date/Time Format (EDTF).

Background:

The National Digital Library (NDL) of Finland is an entity within the remit of the Ministry of Education and Culture, which basis is formed by libraries, archives, museums, and other organizations in Finland storing cultural heritage. The NDL aims to create a nationally unified structure for contents and services ensuring the effective and high-quality management, dissemination, and digital preservation of cultural digital information resources. METS will be used as a base metadata document format in hundreds of independent background systems maintained by the memory organizations. All of the data will be preserved in a single preservation system, which includes several preservation services, such as data validation and migration. See http://kdk.fi/en/

Arguments:

Currently, CREATED attribute requires an accurate metadata creation time (incl. seconds). Organizations in NDL have collected the metadata of the objects and also digital objects for years. However, the known creation date and time of the descriptive or administrative metadata (or even creation date/time of a file) may be inaccurate. It should be possible to provide the best estimate available.

Example:

```
<dmdSec ID="dmd-id" CREATED="2010?">
  <mdWrap MDTYPE="DC">
    <xmlData>
      <!-- This descriptive metadata was probably created in 2010 -->
      <!-- ... -->
    </xmlData>
  </mdWrap>
</dmdSec>
```

Contact:

Juha Lehtonen, CSC – IT Center for Science, juha.lehtonen@csc.fi