Can I Use It, and If so, How?

Discussion of Consortium Proposal
for OLIF2 File Header

Christian Lieske

SAP AG – MultiLingual Technology

1. What is in the header and why?

> Can I/My
> Application
> Use This File?

2. How is header information represented?

```
<!ELEMENT olif (header, body) >

<!ATTLIST olif olifV CDATA #REQUIRED >
<!ELEMENT prop (#PCDATA) >
```

3. Discussion

> What If I
> Need It
> Differently?

What is in the Header and Why? (I)

You/your application need to know if it makes sense to process a certain OLIF2 file at all

1. Is the file **relevant** at all? → language(s), project, ...

2. Am I **allowed** to use it? → copyright, distribution, ...

3. Where can I turn to for **more information**? → contact person, additional resources, ...

4. **Who** created the data **when** and **how**? → creation tool, user, ...

5. Can I **handle** it? → encoding, size, special qualifiers, ...

The OLIF2 header aims at giving value to lexical and terminological data in a world of change and interchange

1. **Patterned after headers of formats for Translation Memory Exchange (TMX) and Corpus Encoding Initiative (CES)**

2. **Accomodates version tracking for formats and tools**

3. **Supports 80/20 solutions**

4. **Facilitates compression**

5. **Allows references to supplementary, external information**

A **lean format** like OLIF cannot include all possibly useful information in the OLIF data itself

■ **Example: Relationships** between semantic features and values (e.g. subgroups for a Subject Matter Code)

It would be helpful to have a **graphical view/explicit representation** of this information/these relationships

■ **Solution: Draw a picture of the hierarchy of the semantic features and values, and put it on the Web. Then put a URL for this picture into the OLIF data.**

How is Header Information Represented? (I)

SAP™

1. **In principle, header information can be represented by the standard Extensible Markup Language (XML) constructs *element, entity* and *attribute***

```
<!ELEMENT olif (header, body) >

<!ATTLIST olif olifV CDATA #REQUIRED >
<!ELEMENT prop (#PCDATA) >
```

2. **Header information is, however, *Metadata* (i.e. information about information) and for this special representation mechanisms exist**

   ■ **Example: Resource Description Framework (RDF)**

   ■ **Advantages:**

      1. **Metadata semantics is codified**

      2. **Metadata vocabularies can be shared**

      3. **Tools for using metadata are emerging**

How is Header Information Represented? (II)

SAP™

- **With RDF, metadata/header information is described by so-called *RDF Schemas***

- ***Dublin Core*, is one of the best-known schemas**

- **RDF uses XML namespaces to avoid naming conflicts in case metadata information is merged**

```
<RDF xmlns="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:dc="http://purl.org/dc/elements/1.1/"
     xmlns:oc="http://purl.org/oc/">
<Description about="http://www.olif.net/olif2example.xml">
<dc:title>A sample OLIF2 file</dc:title>
<dc:creator>Christian Lieske</dc:creator>
<dc:date>1999-03-10</dc:date>
<dc:subject>A file that demonstrates the OLIF metadata</dc:subject>
</Description>
<oc:creator>Xtool</oc:creator>
</RDF>
```

- **The current header presumably can be mapped onto an RDF schema easily**

Discussion

1. **Element fileDescription is missing**

2. **...**