

OLIF2 Consortium Review Meeting



December 13, 2001
Walldorf, Germany



Andrew Bredenkamp, *DFKI*

Jennifer Brundage, *SAP*

Deborah Coughlin, *Microsoft*

Daniel Grasmick, *SAP*

Michael Kranawetvogl, *Bowne*

Hubert Lehmann, *Linguattec*

Christian Lieske, *SAP*

Susan McCormick, *SAP*

Carlo Mergen, *EC*

Peter Quartier, *IBM*

Jean Senellart, *Systran*

V. Srinivasan, *SAP*

Gregor Thurmair, *Sail Labs*

Michael Wetzel, *Trados*

9.00 – 9.30	Welcome and introductory remarks:	<i>Daniel Grasmick</i>
	Review and approval of the agenda; Introduction of new consortium members:	<i>Susan McCormick</i>
9.30 – 10.45	Discussion of OLIF v.2 test results:	<i>Consortium</i>
10.45 – 11.00	Coffee break	
11.00 – 13.00	Presentation and discussion of OLIF v.2 applications; data exchange, supporting tools, etc:	<i>Sail Labs, Systran, Linguattec, Bowne, Trados, SAP</i>
13.00 – 14.00	Lunch	
14.00 – 15.00	Discussion of validation and certification:	<i>Christian Lieske</i>
15.00 – 15.30	Introduction of Asian languages:	<i>Susan McCormick</i>
15.30 – 15.45	Coffee break	
15.45 – 16:15	Update on SALT/OLIF collaboration:	<i>SALT/OLIF Working Group</i>
16:15 – Close	Future plans:	<i>Consortium</i>

esting OLIF v 2



dtd_mod.zip Systran - Modified dtd for morphological deeper description	jsenellart	09/11/2001
oGlossaryXSL.zip XSL stylesheet for creating glossary from OLIF file as HTML in Internet Explorer	christian_lieske	11/02/2001
oMonoXSL.zip XSL stylesheet for displaying terms from OLIF file as HTML in Internet Explorer	christian_lieske	09/11/2001
oMultiXSL.zip XSL stylesheet for displaying multilingual terms from OLIF file as HTML in Internet Explorer	christian_lieske	09/12/2001
olif2_sample_sap1.xml SAPterm entries in OLIF; entries in XML displayed	svwbspouse	09/11/2001
olif2_sample_sap2.xml SAPterm entries in OLIF; special OLIF stylesheet	svwbspouse	09/11/2001
olifExMonoCharEnc.xml OLIF - monolingual Example - character encoding	christian_lieske	09/11/2001
olifExMonoJavaHelp.xml OLIF - monolingual Example with definitions - JavaHelp	christian_lieske	11/02/2001
sapterm-terms1.txt SAPterm entries for OLIF conversion	svwbspouse	09/11/2001
systran1.xml Systran - Simple Olif2 sample	jsenellart	09/11/2001
systran2.xml Systran - Same Entries with more detailed morphological description (need dtd_mod package)		

Modelling SAPterm in OLIF

SAPterm

- Multilingual termbase
- Concepts associated with language equivalents
- Multidirectionality

```
<entry>
<Component>IS-IS-CD</>
<Comp description>Collections/Disbursements</>
<German>Stundungszeitraum</>
<Part of speech>noun</>
<Gender>m</>
<Czech>období odkladu</>
<Part of speech>noun</>
<Gender>n</>
<Hungarian>halasztás időszaka</>
<Part of speech>noun</>
<Italian>periodo di dilazione</>
<Part of speech>noun</>
<Gender>m</>
</entry>
```



For each multilingual SAPterm entry:

- **> 1 OLIF entry**
- **Each OLIF entry with 1 or more transfers**
- **Transfers bilingual, unidirectional**
- **All monos require key DC's**
- **Optionally, transfers have key DC's**



Increase in number and size of entries



- Upper/lower case for attribute names
- Error with *trTarget* attribute
- Key DC's in transfer were obligatory

Patch DTD available at:

<http://www.olif.net/olif2/formalization/errata/errata.htm>



Changes to Transfer Restrictions and Structural Changes

- *Statement* blocks for transfer restrictions, contexts, tests, and structural changes that are grouped with logical operators.
- Addition of *context* for structural change

Result: Greater consistency, more code



- Expansion of morphological description for *mono*
- Basic testing of *transfer restriction* formalization



Presentations from:

Sail Labs

Systran

Linguetec

Bowne

Trados

SAP



New OLIF2 Consortium member



www.basistech.com

Euclid (Encoding and Language Identifier)

Euclid is a high-performance engine for determining the encoding and language of unspecified text, including European and Asian languages/encodings.

Japanese Morphological Analyzer (JMA)

Korean Morphological Analyzer (KMA)

Chinese Morphological Analyzer (CMA)

Portable, robust, high-performance text analysis and word segmentation engines

Chinese Script Converter (C2C)

C2C automates the conversion of characters between the two modern Chinese scripts: Traditional and Simplified.



- Meeting of SALT/OLIF Working Group at MT Summit in Santiago, Spain
- OLIF data categories for Data Category Registry
- Current TBX specification:

<ftp://ftp.ttt.org/oscar/tbx/>

(is ussion) Future ! ns



- Contact with SIMPLE, PAROLE – Sail Labs
- Workflow module – Christian Lieske
- New data categories- frequency, etc., concept
- Certification – Christian Lieske
- Expansion for Asian languages – Basis, DFKI
- XML Schema
- Compression
- OLIF consultants – implementation support
- SALT-OLIF – „different flavors“; localization
- Intractable OLIF DCs – semantic type, subj field, infl, sem. Reading
- Using validated OLIF for data exchange, esp. with MT
- Sign off on OLIF v.2 in ? Months
- Advertisement