

# OLIF2 Consortium: Organizational Meeting

*April 6, 2000  
SAP AG  
Walldorf, Germany*

A blue L-shaped graphic element in the bottom right corner of the slide.



# Agenda

---

9.00 – 9.15	Welcome and introductory Remarks: <i>Daniel Grasmick</i>
9.15 – 9.45	Structure of the OLIF2 Consortium: <i>Daniel Grasmick, Susan McCormick</i>
9.45 – 10.30	Time frame for OLIF2: <i>Daniel Grasmick, Susan McCormick</i>
	Financial issues for the consortium: <i>Daniel Grasmick</i>
10.30 – 10.45	Coffee break
10.45 – 12.00	Current status of OLIF: <i>Gregor Thurmair</i>
12.00 – 13.00	Discussion of changes to OLIF currently envisaged for OLIF2: <i>Susan McCormick</i>
13.00 – 14.00	Lunch
14.00 – 14.30	Review of current level of support for OLIF among tool vendors: <i>Daniel Grasmick, Susan McCormick</i>
14.30 – 15.30	Review of other interchange formats and initiatives: <i>all participants</i>
	Discussion of interaction of OLIF2 Consortium with SALT and/or OSCAR: <i>all participants</i>
15.30 – 15.45	Coffee break
15.45 – 17.00	Task descriptions for work groups to review current OLIF and suggest changes/additions in linguistic, terminology, and technical specifications; recommendations to be completed in April/May, 2000: <i>all participants</i>





# Consortium Participants

---

*Gregor Thurmair, Sail Labs*

*Johannes Ritzke, Sail Labs*

*Alex Muzarku, Logos*

*Pierre-Yves Foucou, Systran*

*Yves Mahe, Xerox*

*Paolo Martins, EU*

*Chris Pyne, L10NBRIDGE*

*Jörgen Danielsen, L10NBRIDGE*

*Nils van der Laan, Trados*

*Peter Quartier, Lotus*

*Ulrike Irmeler, Microsoft*

*Daniel Grasmick, SAP*

*Susan McCormick, SAP*

*Jennifer Brundage, SAP*

*Christian Lieske, SAP*

*Christoph Pahlke-Lerch, SAP*

# Welcome and Introductions

---

- *Company*
- *Professional background*
- *Terminology volume*
- *Languages supported*
- *Organization of terminology management in your company*
- *Terminology database(s) used*
- *Other tools related to terminology*
- *Any exchange formats?*
- *Future plans for terminology/lexicon management*





# Purpose of OLIF2

---

*To upgrade the current OLIF standard  
so that it can be supported by tool  
vendors and applied by users in 2001*

# Why a New Consortium?

---

- *OLIF was developed in the OTELO project as a prototype, but is not usable in its current form*
- *The SALT project plans to use the OLIF format as part of its XLT standard, but will not edit OLIF1 for content*
- *LISA TBX will be based on SALT XLT*
- *None of the other formats supports MT requirements*
- *Thus, usable OLIF is required*

*e.g., SAP will double its terminology volume by the end of 2000 and add additional NLP tools needing term data*



# Structure of the Consortium

---

- *OTELO participants*
  - *SAIL Labs, Logos, Lotus, SAP*
- *New MT representative*
  - *Systran*
- *Term Management representatives*
  - *Trados, Xerox*
- *Service (and tool) providers*
  - *L10NBRIDGE, L&H via SAIL Labs*
- *Users*
  - *EU, Microsoft...*
- *... And open to interested parties*

# Time Frame for OLIF2

---

## *Phase I: Specification*

- *Working groups make recommendations for changes to OLIF format by May 31, 2000*
- *Specifications for OLIF2 complete by September, 2000*

## *Phase II: Implementation*

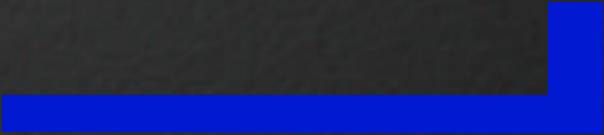
- *Tool vendors support new format in 2001*
- *Maintenance tools developed by end of 2000/beginning of 2001*



# Changes to OLIF for OLIF2



*Review current OLIF format for changes to:*

- *technical structure*
  - *linguistic analysis*
  - *terminology handling*
- 
- A blue L-shaped graphic in the bottom right corner of the slide.



## *Make OLIF compliant with XML:*

- *well-supported industry standard*
- *extensible - new element types easily defined*
- *well-suited for data exchange formats*
- *SALT project already working on XML-based standard in which they want to embed OLIF*

# Achieving XML-Compliance

---

- *OLIF entry structure remains basically the same for OLIF2*



- *OLIF2 is primarily 'rewrite' of OLIF, but with XML-compliance*



# XML-Driven Design Changes

---

*Use some of the features of XML to make design changes for OLIF2:*

- *reanalyze some current tags as attributes of XML element types, e.g.,*

*<LINK="synonym">*

- *allow for more embedding of structure*

*Current OLIF: ISO-Latin-1*

*OLIF2 functionality:*

- *double-byte characters*
- *bidirectionality*

*XML supports ISO/IEC 10646, which is similar to unicode*



# Changes to the OLIF Concept

---

*Make substantive changes to the structure*

- *company-code as part of central entry base*
- *formally distinguish bilingual from monolingual links*
- *develop protocol for user-defined fields*

# Converging with other Standards

---

*Coordination with other standardization initiatives such as SALT*

- *Achieve as much overlap as possible with, e.g.,*
  - » *names of element types*
  - » *structure of entries*



# Review of Linguistic Features

---

*Comprehensive review of linguistic features*

- *are features in correct feature groups?*
- *are all of the features that are essential for the different vendors covered?*
  - » *transitivity for Logos*
  - » *Systran requirements*
  - » *Xerox*
- *what about other NLP products or users?*

## *Review the current morphological analysis*

- *currently includes only German, Danish and English*
- *theoretical underpinnings of analysis are inconsistent*



*Special attention to:*

- *selectional restrictions (transfer conditions) - representation should be improved*
- *syntactic frames - currently for German, Danish and English only*
- *semantic types - should be reviewed and expanded*

# Features and Values

---

- *Make sure feature names and values conform to general practice*
- *Make sure all element types that we want to cover are actually in DTD*



## *Conventions for formulating canonical forms*

- *defined for formulation of entry string in given language*
- *necessary for optimal convergence of entries from different systems*
- *based on language-specific lexical conventions*
- *published as part of formal specification*

# Structure of Terminology

---

*Expand current structure?*

- *allow for deeper structure, more embedding (in line with MARTIF?)*
- *expand on feature/value pairs to allow more admin detail*



*Add unique entry identifier*

- *current OLIF does not support a unique identifier for each entry, although many termbanks require this*

# Review of OLIF Support Among Tool Vendors





# Overview of Other Exchange Formats and Initiatives



## *ISO 12200:1999 Standard*

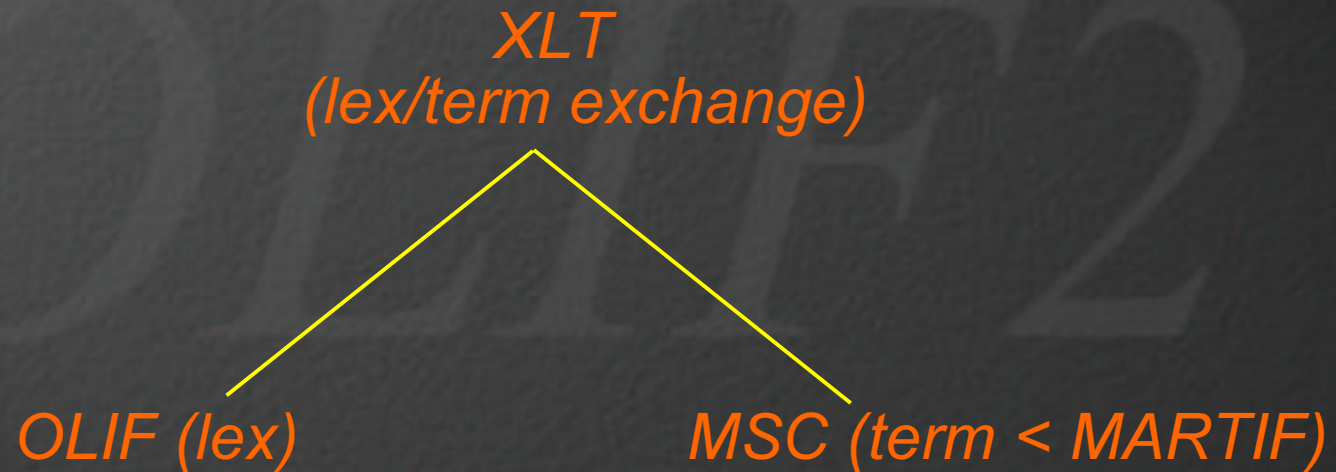
- *SGML-based*
- *strictly terminology*
- *formal concept-orientation*
- *extensive DTD*
- *lots of administrative information*
- *relatively complex embedding in structure*



## *Proposal ISO/TC 37/SC 3 N 318*

- *extended MARTIF - attempt to coordinate with TMX and OLIF*
- *adapted to XML*
- *extends MARTIF to include NLP some features*

*SALT Project - Currently funded by the EU*





## *Group within LISA Organization*

- *TMX - format for re-use of translation memory data*
- *TBX - lex/termbase exchange (subset of XLT)*

---

*“Generic model for the distribution and reuse of heterogeneous terminological data”*

- *for DB management*
- *compatibility with internet*
- *fairly complex hierarchical structure*
- *reworked to allow multiple word senses alongside concept model*



# Meeting Results: Participation of all companies invited working in 3 action groups ...



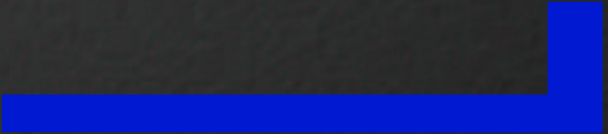
# TG1: Technical Structure

---

Goal: provide formal structure of the format

- Review for XML compliance
- Redundancy
- Links representation
- Definition of the header
- Incorporation of user-defined fields

= Output: OLIF DTD

A blue L-shaped graphic located in the bottom right corner of the slide.



# TG2: Linguistic Analysis

---

Goal: provide a “final” list of feature-value pairs for the linguistic component

- Canonical form formulation
- Morphology, syntax and semantics
- Transfer conditions and transformations
- Cross-references (based on ISO)

# TG3: Terminology Handling

---

Goal: to provide a “final” list of feature-value pairs for terminology

- Concordance with other standards
- Administrative information



# Languages Supported in OLIF2

---

## Priority 1

- EN
- DE
- DA
- FR
- ES
- PT
- JA

## Priority 2

- RU
- IT
- NL
- Other priorities...
- EL
- HU
- ZH
- ZF
- KO
- AR

- Terminology samples from all participants
  - at least 100 entries
  - incl. description
  - at least 2 languages and different categories