

*Terminologies for the Multilingual
Semantic Web (MLSW) - Evaluation
of Standards in the light of current
and emerging needs*

Gerhard Budin
University of Vienna
Austrian Academy of Sciences

Multilingual Web Workshop
Limerick, September 2011

Overview

- Problem description and needs analysis → why terminologies for the MLSW?
- The role of standards, semantic interoperability, and terminological semantics
- Current situation of the emerging standards ecosystem
- Conclusion and outlook

Problem description and needs analysis → why terminologies for the MLSW?

- Multilingual Semantic Web – a strong current trend, given the rapidly rising number of workshop held, papers presented and published, R&D initiatives in funded projects, companies embarking on it
- In this context, terminologies – in the form of ontologies – have been a core topic in this context -> multilingual ontologies, localisation and translation of ontologies (i.e. of labels)
- Problem: the language-unaware and language-independent data modeling approach in the Semantic Web and for ontologies is in contrast to ubiquitous cultural differences that are expressed in multilingual terminologies in the form of asymmetries (partial equivalences), but not in ontologies and not in thesauri -> this creates huge problems in localisation & translation work that is increasingly using SMT, LOD, and semantic web technologies.
- What we need: a modeling approach to terminological semantics that includes the modeling of cultural differences (only some of which are expressed linguistically)

The role of standards, semantic interoperability, and terminological semantics

- -> a model of terminological semantics that integrates
 - domain logic and structures of domain knowledge (some domains are very culture-dependent – e.g. law, not everything is “logical” in there)
 - Conceptual knowledge expressed (sometimes in different ways) in lexical structures in each language (“langue”), including implicit knowledge
 - Texts that become corpora for computational analysis to understand the actual use of such terms (“parole”)
 - The socio-cultural contexts, the users, authors and audiences (pragmatics, in the form of annotations)
- -> this in turn is the pre-requisite for an integrative model of semantic interoperability in order to (semi-)automate cross-lingual linkages, inferences, translations, and other operations
- -> are our current and emerging standards able to satisfy these requirements? Are the different standards compatible with each other and interoperable in automated workflows?

Current situation of the emerging standards ecosystem

- **Stakeholders in standardization (across different levels and org. types):**
 - ISO, CEN, national standards institutes; W3C; TEI
 - ETSI, OASIS, GALA, TAUS, ELRA
 - CLARIN, FLARENET, META-NET, Multilingual Web,
 - Current projects such as MONNET, LISE, and many others
- **ISO standards (selected examples) (ISO TC 37)**
 - ISO 704 (term formation principles, building concept systems, writing definitions); ISO/TR 24156 (using UML in terminology); ISO 26162 Design, implementation, and maintenance of terminology management systems
 - ISO 30042 TBX – termbase exchange; ISO 16642 TMF – terminology markup framework; ISO 12620 Data categories -> ISOCat online registry
 - ISO 24612 LAF Linguistic Annotation Framework; ISO 24613 LMF Lexical Markup Framework; ISO 24616 Multilingual Information Framework (MLIF)
- **MLSW-specific models (potential standards candidates) (2 examples)**
 - Lemon - An Ontology-Lexicon model for the Multilingual Semantic Web
 - Ontology Integration and Interoperability (OntoIOp) (CEN/ISO)
- **-> next slide: our own involvement – pre-normative research**



universität
wien

Zentrum Sprachwissenschaften, Bild- und Tondokumentation
Institut für Corpuslinguistik und Texttechnologie (ICLTT)



OAW

Österreichische Akademie
der Wissenschaften



DARIAH

Digital Research Infrastructure
for the Arts and Humanities

TES4IP - Terminological Services for the Intellectual Property Domain

CLARIN

Common Language Resources and Technology Infrastructure



LISE

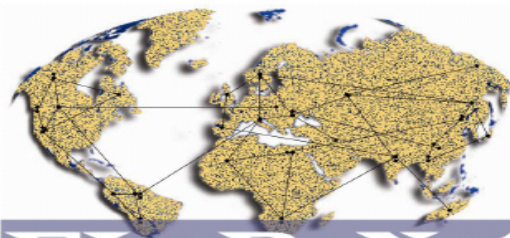
Legal Language
Interoperability
Service



META NET



Wide Information Network for Risk Management



F o s t e r i n g
L a n g u a g e
R e s o u r c e s
N e t w o r k

FLaReNet

Grant Agreement No.
ECP-2007-LANG-617001



MeLLANGE
Multilingual eLearning in LANGuage Engineering



Education and Culture

Leonardo da Vinci

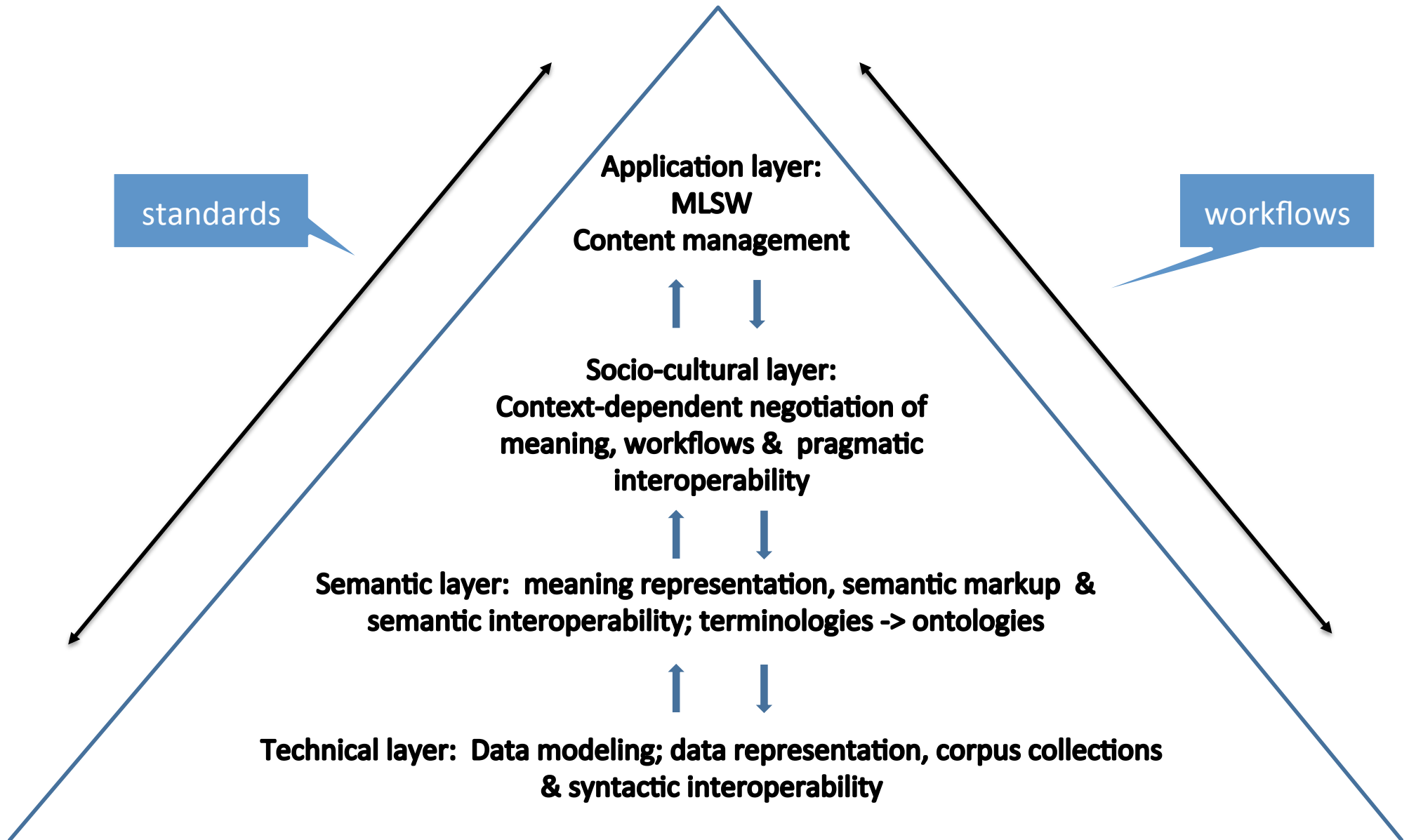
Evaluation of standards in pre-normative research

- by testing standards, i.e. by using them in research projects and by collecting experiences of other teams doing the same – different domains, MLSW application scenarios including user interaction
- LISE (Legal Language Interoperability Service)
 - Interoperability of legal terminologies -> legal ontologies; probably the most challenging domain in terms of cultural-systemic differences (within and across languages), workflows; from local to transnational levels
- TES4IP (Terminology Services for the Intellectual Property Domain)
 - Data mining, term extraction, semantic modeling of domain knowledge, terminology management for MLSW applications, thesauri, workflows
- MGRM (Multilingual Glossary of Risk Management)
 - Interoperability of different data representation forms for different purposes (from user-oriented glossaries and termbases via semantic frames up to fully axiomatised task ontologies for remote sensing applications) in a transdisciplinary, transcultural and multilingual domain

Conclusion and Outlook

- Networking among stakeholders is crucial to create standards that fulfill their purposes
 - how abstract or how concrete must a standard be
 - In addition to specifications on paper we need more real-life reference implementations of a standard for “instant” use and testing
 - Users to be involved right away when standards are being created
- MLSW scenarios differ across domains
 - Degree of cultural dependencies and their linguistic manifestations
 - Domain-specific expert cultures, best practices
 - Expectations of end-users according to their needs
 - Role of standards -> different combinations of standards for a scenario-specific standards eco-system (workflows, interoperability)
- LOD Linked Open Data
 - Important initiative that integrates different perspectives
 - Several terminologies have been transformed for LOD purposes (e.g. AGROVOC thesaurus on LOD, terminologies in eGovernment scenarios)

Integrated Terminology Resource Management for the Multilingual Semantic Web



Thank you for your attention

gerhard.budin@univie.ac.at