Managing Industry Wisdom as a Portfolio of Technical Standards

Examined in context of the Globalization, Internationalization, Localization and Translation (GILT) Industry

David Filip, Ph.D.

LRC/CNGL at CSIS University of Limerick, Ireland

davidf@ul.ie

Search Key Words

Technical Standards, Standardization, Technical Committee Process,

Consensus Building, Body of Knowledge

Author Background

David Filip is the Liaison Officer of OASIS XLIFF Technical Committee and Voting Member of XLIFF Technical Committee; Co-Chair of W3C MultilingualWeb-LT; W3C Advisory Committee Representative of University of Limerick, Ireland; XLIFF Technical Committee Liaison at Unicode Localization Interoperability Technical Committee; TAUS Standards Advisory Board member etc. David works as Senior Researcher at LRC/CNGL at University of Limerick, Ireland.

Disclaimer

This paper represents David's own expert opinion on the matter of Open Standards in Localization. Nothing in this article, except when explicitly stated, can be construed as opinion or recommendation of any of the standardization bodies, associations, consortia etc. that David works with. However, all of the opinions propounded here are well backed with publicly accessible facts and data. Discussion on interpretation is welcome.

Abstract

We examine Knowledge Management at the level of industry wisdom, as a core set of technical standards adopted by friends and foes across a particular vertical, in the sense of the best practice or constitutional law. Industries create standardization ecosystems based on several approaches to intellectual property rights, transparency, commitment, consensus building etc. Successful technical standards are so by solving real business issues. Thus technical standardization workers need mandate from key business decision makers and their

communication is critical. Political work on building committees with wide and varied industry representation is a necessary prerequisite of successful technical work on standards.

Introduction

Knowledge management is usually examined as an organizational issue; that is an organizational issue within one organization (specifically or generally). We are looking at the issue at the level of industry wisdom, in the shape of a core set of technical standards adopted by friends and foes across a particular vertical industry. Such a body of industry wisdom can be seen from two complementing points of view: 1) as a body of constitutional law for a given area, i.e. in clear analogy to state science, or 2) body of managed knowledge in the sense of best practice (normative or informative) that can have profound influence on actual industry practice. Interestingly and importantly, the standardization communities tend to create knowledge management ecosystems of varied efficiency and effectivity, based on several types of approaches to intellectual property rights, transparency, commitment, consensus building etc.

Terminology

BCP

Best Current Practice – a normative track of RFCs at IETF. (IETF Network Working Group 1995)

Bitext

Text in two natural languages organized as an ordered succession of ordered and aligned source and target pairs. Such a structure is key for localization transformations, thus bitext standardization is a key to localization interoperability.

Essential Claim

Claim or even a potential claim made by any IPR owner against a standard specification or draft asserting Intellectual Property that is currently not available for licensing under the IPR mode conditions required by the given IPR policy that cannot be avoided implementing the specification (Rosen 2004 chap.13; W3C Patent Policy WG 2004a sec.Definition of Essential Claims; OASIS Board of Directors 2010; ETSI General Assembly 2011). See also Essential Patent.

Essential Patent

[used essentially in a standard] – A patent that cannot be avoided making a standard's implementation. See Essential Claim.

ETSI

European Telecommunications Standards Institute, "the mother of GSM" (Global System for Mobile communication, originally Groupe Spécial Mobile). One of the world's most influential producers of World Class (yet proprietary) Telecommunications standards. ETSI has huge base of members in the Telecommunication Industry. Although ETSI has been successfully penetrating IT and IT Security standardization areas, its membership fees (and voting weight) calculation for Enterprises is still based on their annual telecommunications turnover (ETSI General Assembly 2011; ETSI 2011, 2012).

ETSI ISG LIS

An Industry Specification Group that has been formed within ETSI to take over LISA OSCAR standards portfolio including related LISA intellectual property (ETSI ISG LIS Founding Members 2011).

FRAND

See RAND

ICU

International Components for Unicode (ICU Project 2012). An IBM driven open source project, in fact the most important reference implementation of both Unicode and CLDR. As ICU provides internationalization libraries, it actually consists of two subprojects ICU4C and ICU4J that provide libraries for C and C++ and Java respectively.

Spectacularly, ICU is the common denominator of both Android and iPhone Operating Systems.

IETF

Internet Engineering Task Force is a non-membership standardization body (IETF Secretariat 2012a). Contributors are strictly individuals that commit implicitly themselves (and their employers) simply by contributing without signing any formal contract. IETF creates Internet related Technical standards, protocols, processes and non-normative informational content as so called RFCs. IETF is institutionally and (partially) financially backed by the Internet Society.

IPR Mode

[of a standardization body or TC] – Intellectual Property Right rules under which a body or TC has been chartered. As a general rule a Technical Committee's Identity cannot possibly survive IPR mode change. I.e. usually Technical Committees that want to change their IPR mode must re-charter, which requires dissolution of the original TC. This is generally used as a guarantee that standards related Intellectual Property Right such as the right to use Essential Patents will be granted to standard implementers under legally stable conditions.

LISA

Localization Industry Standards Association, declared insolvent on February 28, 2011. (GALA Standards Initiative and LISA OSCAR 2011)

OASIS

Organization for Advancement of Structured Information Standards (formerly called SGML Open) (OASIS 2012). One of the most important IT standardization Consortia, based in Massachusetts USA, its foundational sponsors including IBM, Microsoft. Localization buy side, toolmakers and service providers are also well represented. OASIS concentrate on development of XML based specification that target interoperability and automation in specific areas, such as authoring, business transactions, web services, and localization.

Okapi Framework

Set of localization engineering transformation tools, an Open Source Reference Implementation of the most influential Localization Open Standards, such as XLIFF, TMX, TBX, SRX, ITS and by virtue of the previous also of the OAXAL reference architecture model (Savourel 2012). Okapi team is largely driven by ENLASO's Yves Savourel, the father of OpenTag and hence also XLIFF.

OSCAR

LISA's Technical Committee (Special Interest Group) for actual standardization work.

Explanation of the acronym is somewhat strained, meaning Open Standards for Container/Content Allowing Reuse. Dissolved along with LISA.

RAND

(also FRAND) – (Fair) Reasonable and Non-Discriminatory. IPR mode that allows owners charging for use of Essential Patents provided that the charge is reasonable and non-discriminatory (Rosen 2004 chap.13; Festa 2005; OASIS Board of Directors 2010; ETSI General

Assembly 2011). Unfortunately, both “reasonable” and “non-discriminatory” are relatively vague, so that their practical extent often needs to be ascertained by legal courts. OASIS TCs can be chartered as RAND, Unicode’s sole IPR mode is RAND, as FRAND is the sole IPR mode of ETSI.

RF

Royalty Free. IPR mode that mandates and guarantees Royalty Free use of Essential Patents in order to implement a standard. OASIS TCs can be chartered as RF (as is XLIFF TC) (OASIS Board of Directors 2010) and RF is the sole IPR mode of W3C (W3C Patent Policy WG 2004b).

RFC

Request for Comments – IETF Recommendation of a technical or process solution that can be normative or informative (IETF Secretariat 2012b).

TC

Technical Committee – Standardization bodies (consortia) usually own, create, maintain, and update technical standards through purpose specific Technical Committees. Many of them indeed do call them Technical Committees in their organizational structures (e.g. OASIS, Unicode, ISO), other don’t (e.g. W3C). Here we use Technical Committee in the generic sense, as any formal group that owns (creates/maintains/updates) a technical standard, even though it might be called something else in its home organization (Industry Specification Group, Working Group, Special Interest Group etc.)

Technical Committee Process

Again as in the Case of Technical Committee (TC), we use this term in a generic sense, as any set of policies, rules, habits etc. that is driving standard development and decision making within a Technical Committee in the generic sense.

Unicode

Unicode Consortium. Home of the Unicode Standard (Unicode TC 2012) and CLDR. Unicode also publishes specialized Technical Reports, UTR#. Some of these reports are Unicode Annexes, UAX#, i.e. essential parts of the current Unicode Standard release, however published as separate documents. Unicode also produces Unicode Technical Standards (UTS#). These are independent of the Unicode Standard and currently specify Unicode Standard or CLDR related algorithms, formats, mark up languages etc.

ULI

Unicode Localization Interoperability TC. The third currently youngest Unicode Consortium Technical Committee, established in May 2011 (ULI TC 2012).

W3C

World Wide Web Consortium (W3C 2012). W3C owns many key standards including XML, HTML and CSS. HTML 5 is clearly the coolest standardization stuff that is currently going on, probably the only standardization activity that makes it into general media. Most relevant for the localization industry is the Internationalization Activity, which – apart from working as an Internationalization gatekeeper for other W3C groups – hosts ITS Interest Group and MultilingualWeb-LT WG.

WG

Working Group, W3C and IETF term for Technical Committee that is mandated to produce W3C recommendations (i.e. standards in W3C speak) or RFCs (that can be normative or informational in IETF).

Open Standards

Before we start explaining about openness of standards we should briefly note that real working standards (open or not) must have been driven by a representative industry consensus. In fact the industry representativeness is one of the main competitive characteristics in standards bodies (consortia) in general and in the localization industry in particular.

Let us start with the assumption that Openness of Standards is an intrinsically positive property and that Localization Industry should better be keeping its standards open rather than proprietary. In order to argue for open standards we must first explain the conceptual difference between open standards and just standards (or proprietary standards) and also between open standards and open source, which is a quite important difference that is however popularly confused.

[Insert Figure 1: Open - Open - Closed – Proprietary here]

Closed source solutions can still be implementations of Open Standards (case 2). Interoperability is truly barred in case 4, where a Closed Source solution is at the same time Proprietary (fully ad hoc or based on a proprietary standard). The best case from the point of view of Interoperability clearly is case 1, i.e. an Open Source implementation of Open Stanadards; perfect examples of this case are ICU and Okapi framework. To make the whole

difference entirelly clear, we can say that the opposite of Open in “Open Source” is Closed, whereas the opposite of Open in “Open Standards” is Proprietary.

In order to explain the fine distinction between open and proprietary standards, two essential characteristics of Open Standards need to be explained, i.e. Transparency and guaranteed Royalty Free use.

[Insert Figure 2: Open and Proprietary Standards here]

Transparency

Although there unfortunately is no general consensus on which standards can be called open. All current accounts do agree that open standards must have been created via a transparent formal process. The requirement of formality for a standard’s transparency might not seem obvious, still any process that needs to be publicly verified as transparent better be formal. We speak about transparency of the Technical Committee Process. The process is usually codified on the consortium level; still various Technical Committees and types of standardization groups within a single consortium might meaningfully differ in their processes, which may be largely driven by their subject matter, objectives, scope etc. All of them however need to create standards through representative consensus. The consensus in turn, needs to have been driven by a transparent formal process. There are some implementation characteristics that should generally help you recognize if a body’s Technical Committee Process is a transparent one or not. There are many transparency marks that we should look for. At the very basic level, mailing lists (discussion groups, forums etc.) and archives; proceedings of meetings; and requirements gathering, review process and bug tracking should be all publicly visible an accessible (open to

general input and specific feedback from non-member public). Work in progress should be marked with open and well documented discussions; ideally with help of collaborative authoring tools with “blame” assignment and audit trail. Formal standard approval must be underpinned by verifiable (possibly open source) implementations. Fair quorum and consensus sensing or ballot rules must be ensured at all levels. Broad and varied industry representation should be the result of the above.

Royalty Free use of Essential Patents

IPR modes and policies are in fact the source of disagreement on what standards actually can be called open. All competing definitions would agree on the above specified criterion of Transparency, there are however at least three distinct levels of IPR liberty that can be possibly required to call a standard open. Therefore I would say that openness of standards comes in three levels: 1) The weakest that allows for (F)RAND licensing of Essential Patents (ISO, Unicode, some OASIS committees); 2) Relatively strong that enforces RF access to essential patents (W3C, most OASIS committees, including XLIFF TC); 3) The strongest that requires an Open Source reference implementation. Clearly RF is a prerequisite here, as Open Source cannot deal with “ugly” code, i.e. code making use of paid patents (Festa 2005). W3C is quite unique in not only setting the explicit RF requirements but also in having specific mechanisms in place how to enforce RF patents licensing conditions within the consortium (W3C 2004) (W3C Patent Policy WG 2004b) (W3C WARP PAG 2009). RF and enforcing RF is arguably the essential differentiator of W3C in standardization area. This was stressed by Adobe (Carl Cargill) and IBM (Steve Holbrook) Advisory Committee representatives at the last W3C AC meeting in Sophia Antipolis (May 2012). Although there was the RF ethos with W3C since the early days (1993

Grant of IPR by CERN, 1994-1995 formal incorporation at MIT) W3C actually learned about the need to have a bullet proof Patent Policy the hard way and the current policy crafted by membership corporate lawyers is only in force since 2003.

About the same time the biggest patent factories including IBM and Nokia embraced the benefits of RF patent licensing. The main benefit clearly being that RF drives technology adoption and hence market growth.

That is why IBM changed their general approach to RF around 2003. IBM employees who want to join a TC Chartered with RF IPR mode no longer need to prove that this does not endanger IBM's revenue stream from its patent portfolio. Instead the IBM patent lawyers are obliged to check the chartered scope for conflicts with IBM's top 100 revenue generating patents. However, all new IBM strategic stuff is cleared for RF upfront. The employee who wants to take part on an RF standardization activity only needs to show that he or she has bandwidth to work on the TC and that the standardization effort of the TC is strategic for IBM, so that they benefit from the RF driven adoption rather than from collecting petty fees from something that gets never broadly adopted.

Thus we can see that the biggest corporations who are popularly considered greedy and incapable of altruistic action are led to collaborate in an altruistic way by a force that can be fairly described as enlightened self-interest. Relationship of altruism and self-interest in this respect was discussed in the context of the state theoretical concept of social contract in the latest AC meeting in a session lead by Nokia's Ora Lassila (W3C Advisory Board). AC reps (one of them chaals aka Charles McCathieNevile with Opera) engaged in the social contract discussion

asked why do the corporations spend their time in organizations such as W3C instead of spending their time killing each other. The thing is that corporations like growing markets and among the growing markets the fastest growing are the emerging ones. Some emerging markets are defined geographically, but in the context of Web standardization we are obviously talking also about emerging vertical markets that are being erected across geographies due to technologies emerging ready to be productized. The big guys have simply realized that mere killing in the competitive landscape is not as beneficial as extending the landscape itself. In other words, strong position on a market growing due to RF driven adoption is generating more revenue long term than a dominant position on a stagnating or even waning market.

One standard or many?

Every now and again people from various corners in the localization industry (LISA, toolmakers etc.) propose the intriguing idea of one Superstandard, or a related one, of a Super-standardization Body.

I want to discuss the ideas of a Superstandard and a Super-body on a conceptual level, as it is hard to track a single source of this idea and it may well be that nobody propagates the idea so wholeheartedly as to take credit.

A Superstandard naturally sounds great at first. It seems to be the ultimate thing in standardization. After all, standardization is about doing things in a uniform way, and overcoming proprietary differences. However, as always, the devil hides in the details; standards must be adopted and implemented to be true standards. The smartest specification cannot be reasonably called a standard if it is not being used throughout its target market.

Standardization bodies sometimes inadvertently blur this difference. Such a blurring is always connected to a standardization body's failing or missing its goal. Published specification will fail if the producing Technical Committee was not representative or otherwise significant for the targeted market, and if the reference implementations are feeble or extremely limited in number. Once important standards will collapse if the publishing standardization body fails to manage its specification's later lifecycle. Failure to maintain once influential spec leads to a slow unmanaged death of the standard.

So a standard must address issues in a sufficiently discreet area, so that it can be not only developed, but also implemented, maintained, and updated. For instance OASIS IPR policy and the persistent character of its TCs are well suited for long term maintenance work on standards. On the other hand organizations like IETF or W3C who strictly mandate their WGs to dissolve after the binding version of the standard in question is published may have IPR issues with maintaining specifications or producing newer versions.

As one of the main objectives of technical standards is interoperability and no area of standardization can be conceived as an isolated one, no true standard must aspire to becoming a Superstandard. Where one standard starts assuming functions outside of its scope, it is starting to violate not only the realm of its neighboring standard, but more importantly the very constitutive principles of the true open standards creation. Since it is impossible to gather in one Technical Committee all the expertise needed to produce a working Superstandard, any aspiration to a Superstandard would necessarily lead sooner or later to deficiencies. In other words, a super-specification cannot ever make a working standard. Thus a Superstandard is

conceptually impossible. Therefore, setting out the proper scope – at the time of a Technical Committee creation, charter clarification, or re-chartering – is the first gating factor of standardization success.

It is no doubt tempting for standard developers and implementers to extend their own provisions into other standardization areas. Nonetheless, whenever they do it they act in a proprietary rather than in a standard way.

Localization Standards Ecosystem (Case Study)

Localization is a relatively new industry and academic topic, but it has always been tightly connected with Translation and Internationalization, since it is a service area mitigating the clash between human language and Information Technology (IT). There are thousands of languages on the planet that can be organized in many dozens of groups from different points of view, and Unicode is the formidable effort that currently does not miss its target to ensure that *scripts* needed to write all human natural languages are supported in IT infrastructures, that is in IT infrastructures that do support Unicode 6.1 standard.

Unicode Consortium

Unicode is of utmost importance for Internationalization, i.e. the upstream activities that make the localizer's life easier. The most explicitly Internationalization related Unicode project is CLDR – Unicode Common Locale Data Repository. CLDR is not a standard in a classical sense; it is – as the name suggests – a repository that is being constantly updated and released on a rolling basis following its Data Release Process. CLDR is a Standard Repository of Internationalization building blocks; such as date, time and currency formats, sorting rules etc.

To illustrate the importance of Unicode and CLDR for upstream internationalization it is good to mention ICU, i.e. International Components for Unicode – an IBM driven open source project, in fact the most prominent reference implementation of both Unicode and CLDR. Spectacularly, ICU is the common denominator of both Android and iPhone Operating Systems.

ULI

It happened only very recently (May 2011) that Unicode Consortium had formed a Technical Committee that is explicitly dedicated to Localization Interoperability, i.e. ULI (Unicode Localization Interoperability Technical Committee). ULI is at the first glance a formidable buy-side driven initiative, as its convenor is IBM's Helena Shih Chapman and her co-chairs are Kevin Lenzo from Apple along with Uwe Stahlschmidt from Microsoft's Windows team.

ULI vision is a pragmatic one driven by the personal shrewdness of its founding Chair. Look at real life friction points and fix them in an effective yet robust way (as opposed to the quick and dirty fix that has been all too common in the localization industry).

ULI applies itself first to the cluster of issues that are connected to segmentation. Segmentation is generally governed by UAX #29 (Unicode Annex # 29), and LISA OSCAR made once an attempt to address this issue with SRX (that is however now in prolonged limbo following LISA's death). SRX relationship to UAX #29 is not a transparent one. SRX can be considered incomplete from the engineering point of view; however, its proclaimed goal was NOT to provide a set of segmentation rules for a number of languages, but rather to provide a mechanism to exchange the rules to improve TMX interoperability. It is a known issue of TMX that it often fails to guarantee its proclaimed lossless transfer of Translation Memory data due to segmentation differences. Whoever will revive work on SRX will need to face the new ULI developments.

ULI's first practically achievable goal is to update UAX #29 rules with real life production data. The ULI message to SRX developers is that Localization should take UAX#29 and its ICU implementation for granted, and use SRX only for exchange of rules that differ from the standard UAX #29 behavior. Connected to the UAX #29 normalization is the ULI driven new character pair (segment separator and joiner) proposal. This new pair of characters should be only legal in plain text environment and should facilitate pipelined execution of UAX #29 rules. This proposal was presented to Unicode TC in May 2012 and is now on hold waiting for better motivation. After ULI fulfills its UAX #29 mandate it should apply itself onto one or more of the following localization interoperability topics: 1) Solid wordcounting (scoping) based on unambiguous character and word delimiting rules that will depend on GMX-V status under ETSI ISG LIS. 2) Authoring and translation memories standardization that will depend on TMX status under ETSI ISG LIS and status of memory exchange within XLIFF 2.0. 3) XLIFF Profiling for Interoperability, especially with respect to segmentation and memory exchange. 3) Lemmatization for CLDR and terminology exchange. Lemmatization is language specific and thus huge. Hence ULI might end up only scoping the project and prepare groundwork for a successor TC. Unambiguous and reliable UAX #29 is to some extent a prerequisite of all of the above.

W3C

As the Web used to be largely a Western (or "first world") centric phenomenon, W3C standards have been suffering from all sorts of internationalization deficiencies, and this was the reason to set up in 2006 the W3C Internationalization Activity. As the World Wide Web still has a lot of multilingualism issues to address, the original Internationalization Core Working Group Charter

period has been extended three times so far, and its current mandate will expire by the end of 2013, having been prolonged for the third time by virtue of the MultilingualWeb-LT Charter approval. The Working Group and Activity continue playing a critical role in ensuring internationalization readiness of other core W3C standards such as HTML 5, XML, CSS etc.

ITS

W3C Internationalization Tag Set is a mechanism how to provide (mainly) XML content with metadata facilitating localization or cultural adaptation. The most important data categories include the translate flag, term identification mark up, or directionality information. The specification is currently maintained by ITS Interest Group that is however not mandated to produce a new normative version. New development should be expected from the MultilingualWeb-LT Working Group that has chartered a new ITS version (Sasaki *et al* 2012). The working group has now published a Requirements Document Draft for ITS 2.0 (MultilingualWeb-LT WG 2012).

MultilingualWeb-LT

DFKI (Deutsches Forschungszentrum für Künstliche Intelligenz) and Centre for Next Generation Localization, along with other academic and industry partners, joined forces to forge a strong representative EC funded consortium with the goal to develop metadata categories facilitating interoperability across domains (Content Management and Localization) and between web layers (deep Web and surface Web). The EC funded seed group that has formed an open W3C Working Group called MultilingualWeb-LT (MultilingualWeb – Language Technology) that has in the meantime attracted attention of many more W3C members. MultilingualWeb-LT has chartered external dependency with OASIS XLIFF and has had from the start a strong liaison

relationship with the XLIFF TC. It is critical that the two groups work in sync to ensure semantic and functional matches among XML, HTML 5, and XLIFF 2.0 internationalization and localization metadata. On top of ITS 1.0 MultilingualWeb-LT will provide a normative recommendation for implementation of ITS and other data categories in XML and HTML 5. The group concentrates on three main success scenarios and exchange of ITS metadata categories in those scenarios: 1) Deep Web (such as DITA or Docbook content) exchanging internationalization metadata with a generalized Translation Management System (no matter what type of resources arrayed within the TMS). 2) Surface Web (the generated HTML 5 content) exchanging internationalization metadata with a Real Time Statistical Machine Translation Service. 3) Deep Web (such as DITA or Docbook content) exchanging internationalization metadata with Statistical Machine Translation engine training.

The common denominator clearly is internationalization metadata to span Content Creation and Management and Bitext Transformation standards. In all cases the vehicle in the Bitext Transformation standards area is supposed to be XLIFF. Long standing XLIFF DITA relationship is instrumental, as DITA Translation Subcommittee works mainly with XLIFF profiling.

OASIS XLIFF – the pivotal standard

All localization transformations are in fact instances of Bitext management; thus bitext standardization is the key to localization interoperability. XLIFF is the state of the art bitext standard format, superior in many respects (legal and technical) to proprietary and legacy bitext formats such as “unclean” rtf, ttx, and PO gettext.

The current version XLIFF 1.2 was published as an OASIS standard in February 2008 (OASIS XLIFF TC 2008), and ever since 2008 the TC has been thinking of a brave new major release numbered 2.0. XLIFF 1.2 publication date – February 1, 2008 – coincides in a significant way with an important acquisition that was publicly disclosed only 10 days later, i.e. SDL's acquisition of Idiom Technologies (Gilbane NewsShark 2008). This is significant for more than one reason. By this acquisition SDL proved to the industry buyers – finally and beyond any reasonable doubt – that they cannot afford to rely on a de facto standard driven by an independent (i.e. independent of a single services company) technology supplier. This dream should have been shattered by 2005, when SDL acquired Trados (DePalma 2005), the de facto standard Computer Aided Translation (CAT) Workbench by that time along with two associated (however unfortunately not fully compatible between themselves) bitext formats, i.e. "unclean" rtf and XML based ttx.

Nevertheless, it seems that buy side and toolmakers interest in XLIFF and its future incarnation has been steadily growing ever since the Idiom acquisition. It is natural for all TCs that after they publish a standard, they enter a prolonged calm period (unless they are outright dissolved like W3C or IETF WGs). For XLIFF the calm period lasted approximately till 2010.

In 2010, XLIFF Technical Committee started holding annual International Symposia. The first Symposium took place as a preconference activity of LRC XV annual conference in Limerick (Lieske 2010). In Limerick, Oracle's Niall Murphy explained how XLIFF is critical for unification of localization processes throughout Oracle's numerous acquisitions (Murphy 2010). Both Symposia held so far were a huge success and facilitated meaningful discussion of the TC with

its customer base. One of the highlights of the Warsaw Symposium in September 2011 was the presentation by Uwe Stahlschmidt and Kevin O'Donnel from Microsoft Windows Redmond team (Stahlschmidt and O'Donnell 2011). Microsoft presenters explained how XLIFF plays a key role in their vision of Operating System localization with service providers free choice of tools that will all fulfill descriptive (rather than prescriptive) engineering quality requirements, and how they are actually using the current version of the standard for a Windows Operating System release localization. The Microsoft XLIFF story continues steadily as 4 members of the Redmond team lead by Uwe joined XLIFF TC in late February 2012 and started without much delay playing a very active role in the standard's development.

At the publication of this paper, the XLIFF 2.0 development will be driven by more than 20 voting members. Voting members means people who regularly work on the standard; most of standardization committees usually have far more inactive members than active ones, so has XLIFF TC. Upwards of 20 people actively working on a single standard's development is actually a significant momentum for any standardization area not just the localization industry. Moreover this small crowd is currently composed from a nicely representative mix, as you can see on the pie chart.

[Insert here Figure 3: Representativeness of OASIS XLIFF Technical C.]

Currently about a third of the committee represents large enterprise buyers (SAP, IBM, Oracle, Microsoft) and this sound proportion seems to be sustainable looking at our voting membership growth pipeline. Tool Vendors are represented by just under a quarter of the committee (SDL, Multicorpora, Maxprograms, ENLASO, Lionbridge). Service providers with less

than 10% might seem underrepresented at the first glance (SDL, Lionbridge, ENLASO), however the associations represented on the committee in turn represent their service side membership (Polish Association of Translation Agencies and GALA more so than TAUS), so that it can be said that service side is represented by about a fifth of the committee. Individuals and academics add just the right amount of independence to the mix with their 19% combined.

Apart from technical development of the next generation standard the TC does other things such as nurturing liaison memberships or exploring state of the art of XLIFF implementations. Information from Symposia, collection of implementers' extension points and state of the art analysis (Morado Vázquez and Filip 2012) feeds back into the loop of the next generation standard's development. The above described role model representativeness is a result of conscious representativeness building activity performed by the TCs Promotion and Liaison Subcommittee.

ISO TC 37 (and Subcommittees)

ISO TC 37 is a co-host of two ex-OSCAR standards as described below. ISO TC 37 also looks into co-publishing of current XLIFF 1.2 and next XLIFF 2.0 with OASIS. OASIS (along with several other standardization consortia such as W3C or ETSI) has a special privileged relationship with ISO that allows the so called fast track co-publishing. ISO is important as a dissemination channel for many standardization areas. ISO adoption effectively leads to government enforcement of standards through ISO's tight relationship (ISO is actually a federation of the national standardization agencies) with national standardization bodies such as ANSI (US) or DIN (Germany) to name but a few.

Recent developments in Localization Interoperability Standards

Open Standards have a long tradition in Localization. LISA has been set up as early as 1990, which is, in the terms of IT standardization, ancient to use the same word as Scott McGrath, COO of OASIS, the host of the second oldest Localization Standardization body, i.e. XLIFF TC that was officially set up in OASIS 2001/2002 (OASIS XLIFF TC 2006) after a period of legal clarifications. Unfortunately, LISA died (February/March 2011), and we may say that it died after a protracted illness, or long standing failure to address its proclaimed goals. LISA was producing its standards via its Special Interest Group called OSCAR. Now, what is the current situation of OSCAR Portfolio, and how heavy a blow this death was for the industry? I am going to argue that the blow was in fact not too heavy. One of the reasons being that XLIFF – the pivotal standard of the Localization industry – has in early 2012 significant momentum towards wide consensus based next generation 2.0 version, and importantly excellent alliance forming potential with other relevant standardization activities within W3C, Unicode, etc.

A snapshot of the OSCAR standards portfolio – as it was at the point when LISA died – was released under a Creative Commons License and is currently being hosted on at least two reasonably independent servers: (LISA OSCAR and Melby 2011) and (GALA Standards Initiative and LISA OSCAR 2011). However the names of the standards, their logos, LISA logo etc. were transferred as separate Intellectual Property to a standardization body that LISA Management selected after the insolvency was declared. The chosen standardization body is ETSI, specifically its ISG LIS (Industry Specification Group Localisation Industry Standards). This basically means that although the OSCAR portfolio was technically released under a very liberal Creative Commons License (Creative Commons 2012), official successor versions of the standards that

are called the same (e.g. TMX, GMX) can only be created by ETSI ISG LIS, or someone who has concluded an agreement about the specific intellectual property items (such as names and logos) with ETSI. It should be said that, in September 2011, ETSI initialized inclusion of ISG LIS and its standards in the Memorandum of Understanding (MOU) that exists between ETSI and OASIS since 2007. This memorandum was expanded on April 20, 2011 to cover areas of electronic signatures, emergency management and other areas. Areas Mapping (an Annex to the MOU), was further expanded in September 2011 to cover localization standards as well. Explicitly mentioned are OAXAL (Open Architecture for XML Authoring and Localization Reference Model) and XLIFF TCs for OASIS and ISG LIS for ETSI (ETSI Secretariat 2012).

Two of the more important LISA OSCAR standards had become effectively co-owned by ISO TC 37 before LISA achieved co-publication of TBX 2.0 with ISO in 2007/2008, as ISO 30042:2008. The SRX is under development in SC4, TC37 as ISO/CD 24621, although SRX has not been co-published with ISO so far. I assume that ETSI ISG LIS will be able to continue LISA's collaboration on ISO/CD 24621, as the latest version of this item is dated after LISA died but before the initial ETSI ISG LIS meeting was held.

As the oldest localization standardization organization officially stopped working, localization industry reacted with a splutter of activity, starting as early as the Danvers Standards Summit that was called by dying LISA and concluded its second and closing day only after LISA had been officially declared insolvent. A number of industry stakeholders gathered in Danvers, Massachusetts (near Boston), including (but not limited to) Arle Lommel (LISA OSCAR), Kara Warburton (LISA OSCAR, ISO TC37 Chair), Henry Dotterer (ProZ.com), Jaap van der Meer (TAUS

and TDA), Helena Shih Chapman (IBM), Andrzej Zydroń (XTM), Alan Melby (Brigham Young University), Smith Yewell (Welocalize), Joachim Schurig (Lionbridge) and a few dozen more. A number of delegates including Jaap van der Meer joined Helena Chapman for an improvised meeting at IBM premises the day after to discuss localization standardization future. Helena also disclosed that Unicode Localization Interoperability Committee is in the process of creation since late 2010 (It was finally kicked off in May 2011). This is also where Jaap first pledged to become the industry's standards watchdog. The ensuing power game produced an interesting and in my opinion positive side effect. The standardization is perceived after a long time as something more than just a box to tick off on a complex RFP form. Number of stakeholders from across the industry, technical as well as business users, realized that the standards are not being imposed unto the community by some sort of godlike powers that be and that the standards can be only as good as the community engagement that drives them.

XLIFF TC as a body with a working interchange standard with credible vision of a next generation version benefited nicely from this new interest. Apart from XLIFF TC attracting IBM, Lionbridge, Oracle etc. to join/rejoin the specification effort, all bigger picture standardization efforts set up within the rolling 12 months before and after Danvers have dependencies with XLIFF (Interoperability Now!, ULI, TAUS Watchdog, ETSI ISG LIS, GALA Standards Initiative, Linport, MultilingualWeb-LT) and for most of them XLIFF is instrumental for reaching their set goals.

Interoperability Now!

Interoperability Now! (IN!) a quasi-standardization effort driven by Andrä/Ontram's owner and CEO Sven Andrä that was announced at Localization World Seattle in October 2010. IN! goal

was to produce an XLIFF profile with extensions ("pushing standards over the edge if necessary" to produce real machine to machine interoperability). Although I sympathize with IN!'s goals and regard their XLIFF:doc profile as a valuable input for XLIFF 2.0 development, their endeavor cannot be assessed as a complete success. Two of the three involved toolmakers (Ontram and Welocalize) were able to make a reference roundtrip implementation prototype of their interoperability package and protocol (including the mentioned XLIFF:doc profile as the payload of the package) within one year of the announcement of the initiative. Their goal was basically to create a standard without a standardization body, but following the principles of collaboration driven consensus and transparency (good principles per se) they ended up with creating a mock standardization body with restricted participation and hence limited representativeness (XTM has recently joined the IN! effort as the first new entrant since inception).

I cannot but agree with Joachim Schurig's assessment and must say that their time would have better been invested on the OASIS Technical Committee than in their limited representation meetings. XLIFF Promotion and Liaison Subcommittee recently organized an "infocall" to allow for transparent if informal discussion between IN! and XLIFF Technical Committee (OASIS XLIFF TC and Interoperability Now! 2012). IN! members have been repeatedly invited to join the committee and push their business critical extensions into the standard proper.

TAUS and TDA

Although TDA (TAUS Data Association) (TDA General Assembly 2009) is legally distinct from TAUS (Translation Automation User Society) (TAUS B.V. 2012), both activities are owned by the industry veteran Jaap van der Meer and have closely related agendas. Since this distinction is

largely obscured and not understood even by industry insiders, it might be good to explain. TDA (the industry's "super cloud") is the not-for-profit association founded in summer 2008, whose members pool Translation Memory and financial resources to create an industry relevant collection of Translation Memories that can be used for (Statistical) Machine Translation Training (apart from a few other more or less related purposes, such as Terminology search throughout the meta-corpus). On the other hand, TAUS has been characterized as the industry "think tank" lobbying for automation technologies within Translation and Localization industry since November 2004. It added the new epithet of "Interoperability Watchdog" during and following the LISA dissolution aftermath.

The TMX issue

Before TAUS became the "watchdog" and formed its Standards Advisory Committee (which I happily joined along with Alan Melby, Klemens Waldhör, Helena Chapman, and Sharon O'Brien), Jaap had been resisting the call of the joint memberships of both of his industry associations to join Technical Committees and to start influencing the standardization. TDA has been fully a bet on TMX, LISA OSCAR's Translation Memory Exchange format, without ever making an attempt to influence TMX development. But there are numerous issues with this standard. Under-engineered and by now far behind industry developments, Level 1 is the lowest common denominator that is widely implemented, but this level only exchanges plain text segments, ergo very limited use in industry. Level 2 is being ignored by vendors who do not provide inline mark up semantics. TMX contains no standard mechanism for in context matching information. Without SRX that has issues of its own, interoperability is very limited, leading to mighty leverage losses on tool migrations. Last but not least there is no

interoperability in the scoping area (GMX-V spec cannot be considered a standard due lack of reference implementations).

In one of the MultilingualWeb Workshops (in Pisa April 2011) (MultilingualWeb 2011) (Filip 2011) I was so insolent as to say that TMX is a legacy (i.e. dead) format that had been failing to develop to a next generation version since 2004 (1.4b was published as standard in 2005, but the draft was completed in autumn 2004, and 1.4b was just a hotfix of 1.4a that had been around since 2002). In a following discussion it has been noted that TMX is not dead as CDs are not (Remember Compact Discs, and LP vinyls?). I think the audio/storage external media do offer a suitable comparison here. I agreed in the discussion in Pisa that TMX is as dead as CDs (not yet as dead as vinyls; not that sort of vintage feeling yet; surely – as McLuhan has it – surviving legacy media lend a certain feeling of artistry if they continue to be used).

TMX is a legacy format that still plays an important role in collecting aligned corpora, migration and sanity check type of projects. But the importance of the format will continue dropping as industry has developed multiple differentiators that fail to be properly captured by the format, so that TMX falls shorter and shorter of reaching its postulated goal of lossless translation memory exchange.

In fact LISA OSCAR made an attempt to release a 2.0 version, but the spec had unfortunately several issues. I should like to name but the few gravest ones: The draft was effectively created by two active members of OSCAR, i.e. the Technical Committee failed to represent industry consensus. OSCAR failed to draw industry attention to the review because there was no consensus to represent and in fact the standard fell so far behind the standard features of CAT

tools that industry even did not care to form a consensus. Furthermore, the draft proposed to break backwards compatibility with no real business benefit in exchange; the proposed inline markup solution just pushed the <bpt>[begin pair tag] <ept> [end pair tag] semantics into attributes.

As LISA OSCAR had been working on the TMX 2.0 draft that later failed as we described above, Arle Lommel had joined the XLIFF Technical Committee on behalf of LISA and proposed to the XLIFF Technical Committee a collaboration of the two bodies on inline markup standardization. The ensuing discussion led to formation of XLIFF Inline Markup Subcommittee (Chaired by Yves Savourel). Few XLIFF Technical Committee members at that point (including Yves Savourel, the editor of TMX 1.4b) were of the opinion that TMX is too far behind the industry developments to catch up.

There had always been some basic goodwill between the two bodies, and in fact the lack of real and tight standardization of inline markup in XLIFF was largely due to XLIFF TC willingness to support inline codes specified by LISA OSCAR for TMX. The inline markup situation both in TMX and XLIFF has been often described by Rodolfo Raya (CTO of Maxprograms, XLIFF TC Secretary, and one of the very few main contributors to TMX 2.0 Committee Draft) as “markup salad”. Rightly so, but unfortunately the TMX 2.0 Draft had not brought much of a breakthrough to the scene.

I believe that the current XLIFF Inline Markup committee has indeed outlined the right options for a minimal tight set (the indicative Subcommittee decision is for 4 basic elements; it is

indicative as it needs to be confirmed or changed by the parent Technical Committee) to be used for inline markup coding in various localization interchange scenarios.

Container Efforts

I already mentioned Interoperability Now! that can be partially characterized as localization container effort, since – apart from their ingenuous XLIFF profile – they standardized a ZIPped interoperability package that carries their XLIFF:doc as its chief payload.

In Danvers, many people including Arle Lommel and Alan Melby agreed (inspired by a metaphoric speech given by Smith Yewell) that the Localization Industry needs a Container Standard. The analogy with a real container used in logistics has been freely used by the proponents. One of the main Linport drivers is the GALA Standards Initiative coordinated by Arle Lommel.

I would agree with fellow Danvers delegates that a container standard may be potentially important for Localization Interoperability; my priorities however currently are with the payload standards rather than the container. Also my strategy for the container or wrapper would be different from both Linport and Interoperability Now!. Together with SAP's Christian Lieske, I believe that the business wrapper for localization payload should be specified within a to-be-created Subcommittee of OASIS UBL TC (OASIS UBL TC 2012).

I hope that this case study provides some useful insight into the interoperability standardization issues in the localization industry. It should serve as insight both from the material point of view and as a generic study in power politics behind successful and not so successful specifications. One of the most interesting developments is that the localization standardization landscape

saw in 2011 a new entrant, ETSI, which is one of the strongest traditional (F)RAND players. This is a mark of rapidly growing importance of multilingualism for IT standardization, and at the same time a backdoor for introducing patent fees collection into localization standardization.

Wrapping Up

The notion of a super-standard is fundamentally flawed; standards can only work in ecosystems. Successful technical standards are so by virtue of solving a real business issue. Standards can only address business issues if technical workers in standardization have a clear mandate from key business decision makers in a representative mix of stakeholder organizations. Effective communication among technical and business users is a key standardization success factor.

Political Work on building standardization consortia/committees with wide and varied industry representation is a prerequisite of successful technical work on standards. A working Industry Standardization Body of Wisdom can be only created in diversified industries with sufficient competition, where (even) the (fiercest) competitors have realized the importance of altruistic behavior in standards creation (although the internal motivations of key corporate players is better described as enlightened self-interest). The specialized body of wisdom must have working ties to broader standards ecosystem. This study presents an explicit approach to technical standardization as an altruistic knowledge/wisdom management activity that transcends particular organizations and job descriptions. It is also a study in politics and power anatomy behind successful and failing standards; it should serve as a blueprint and guidance for

standardization lynchpins who wish to drive successful standards creation and adoption not just by chance but as a result of conscious and targeted activity.

References

Creative Commons (2012) Creative Commons — Attribution 3.0 Unported — CC BY 3.0 [online], available: http://creativecommons.org/licenses/by/3.0/ [accessed 31 May 2012].

DePalma, D.A. (2005) SDL Acquires Trados [online], available: http://www.commonsenseadvisory.com/Default.aspx?Contenttype=ArticleDetAD&tabID=63&Aid=1259&moduleId=391 [accessed 31 May 2012].

ETSI (2011) ETSI Membership Fees [online], available: http://www.etsi.org/WebSite/Membership/fees.aspx [accessed 28 May 2012].

ETSI (2012) ETSI Directives [online], available: http://portal.etsi.org/directives/home.asp [accessed 30 May 2012].

ETSI General Assembly (2011) ETSI Intellectual Property Rights Policy [online], available: http://www.etsi.org/WebSite/document/Legal/ETSI%20IPR%20Policy%20November%202011.pdf [accessed 30 May 2012].

ETSI ISG LIS Founding Members (2011) ISG LIS Proposal [online], available: http://docbox.etsi.org/HF/Open/ISGLIS/ETSI%20ISG%20LIS-Terms-of-Reference-v9c.pdf [accessed 30 May 2012].

ETSI Secretariat (2012) Mapping of Active Relationship with OASIS [online], available: http://webapp.etsi.org/AgreementView/Mapping_tables/Mapping_OASIS.htm [accessed 31 May 2012].

Festa, P. (2005) OASIS Patent Policy Sparks Boycott | TechRepublic [online], available: http://www.techrepublic.com/article/oasis-patent-policy-sparks-boycott/5587008 [accessed 29 May 2012].

Filip, D. (2011) Multilingual Transformations on the Web via XLIFF Current and via XLIFF Next [online], available: http://videolectures.net/w3cworkshop2011_filip_lrc/ [accessed 31 May 2012].

GALA Standards Initiative, LISA OSCAR (2011) LISA OSCAR Standards | Globalization and Localization Association [online], available: http://www.gala-global.org/lisa-oscar-standards [accessed 30 May 2012].

Gilbane NewsShark (2008) SDL Acquires Idiom Technologies - Information Technology News [online], available: http://gilbane.com/news/2008/02/sdl_acquires_idiom_technologie.html [accessed 31 May 2012].

ICU Project (2012) ICU - International Components for Unicode [online], available: http://site.icu-project.org/home [accessed 30 May 2012].

IETF Network Working Group (1995) RFC 1818 - Best Current Practices [online], available: http://www.ietf.org/rfc/rfc1818.txt [accessed 30 May 2012].

IETF Secretariat (2012a) IETF Overview [online], available: http://www.ietf.org/old/2009/overview.html [accessed 30 May 2012].

IETF Secretariat (2012b) Getting Started in the IETF [online], available: http://www.ietf.org/newcomers.html#whither [accessed 30 May 2012].

Lieske, C. (2010) Consolidated Takesaways from First XLIFF Symposium - XLIFF Wiki [online], available: http://wiki.oasis-

open.org/xliff/Consolidated%20Takesaways%20from%20First%20XLIFF%20Symposium [accessed 30 May 2012].
LISA OSCAR, Melby, A. (2011) LISA OSCAR Standards [online], available: http://www.ttt.org/oscarstandards/ [accessed 31 May 2012].
Morado Vázquez, L., Filip, D. (2012) *XLIFF Support in CAT Tools*, Subcomittee Report 1, XLIFF State of the Art, OASIS XLIFF TC, available: http://www.localisation.ie/resources/XLIFFSotAReport_20120210.pdf [accessed 31 May 2012].
MultilingualWeb (2011) Pisa Program [Proceedings] [online], available: http://www.multilingualweb.eu/documents/pisa-workshop/program [accessed 31 May 2012].
MultilingualWeb-LT WG (2012) Requirements for Internationalization Tag Set (ITS) 2.0 [online], available: http://www.w3.org/TR/its2req/ [accessed 31 May 2012].
Murphy, N. (2010) 'The Route to XLIFF Adoption in Oracle', Presented at the 1st International XLIFF Symposium, LRC at University of Limerick: Limerick, available: http://www.localisation.ie/xliff/resources/presentations/xliff-adoption.pdf [accessed 31 May 2012].
OASIS (2012) About Us | OASIS [online], available: http://www.oasis-open.org/org [accessed 30 May 2012].
OASIS Board of Directors (2010) Intellectual Property Rights (IPR) Policy | OASIS [online], available: http://www.oasis-open.org/policies-guidelines/ipr [accessed 30 May 2012].
OASIS UBL TC (2012) OASIS Universal Business Language (UBL) TC | OASIS [online], available: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=ubl [accessed 31 May 2012].
OASIS XLIFF TC (2006) OASIS XML Localisation Interchange File Format TC [Charter] [online], available: http://www.oasis-open.org/committees/xliff/charter.php [accessed 31 May 2012].
OASIS XLIFF TC (2008) XLIFF Version 1.2 OASIS Standard [online], *XLIFF 1.2 Specification*, available: http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html [accessed 28 May 2012].
OASIS XLIFF TC, Interoperability Now! (2012) [minutes of Record] Re: Draft Minutes from XLIFF P&L SC Infocall with Interoperability Now! [online], available: http://lists.oasis-open.org/archives/xliff/201202/msg00068.html [accessed 31 May 2012].
Rosen, L. (2004) Open Source Licensing Software Freedom and Intellectual Property Law [online], available: http://www.rosenlaw.com/oslbook.htm [accessed 29 May 2012].
Sasaki, F., Filip, D., Lewis, Dave (2012) MultilingualWeb-LT Working Group [online], available: http://www.w3.org/2011/12/mlw-lt-charter [accessed 31 May 2012].
Savourel, Y. (2012) Okapi Framework [online], available: http://okapi.opentag.com/ [accessed 30 May 2012].
Stahlschmidt, U., O'Donnell, K. (2011) 'Adopting XLIFF: The Microsoft Story', Presented at the 2nd International XLIFF Symposium, TM-Europe: Warsaw.
TAUS B.V. (2012) Terms & Conditions [online], *TAUS - Enabling better translation*, available: http://www.translationautomation.com/index2.php?option=com_virtuemart&page=shop.tos&pop=1&Itemid=95 [accessed 31 May 2012].

TDA General Assembly (2009) 'TAUS Data Association - Deed of amendment to the articles of accociation', available: http://www.tausdata.org/images/stories/pdf/tda-statutes-nov-2009.pdf.

ULI TC (2012) Unicode Localization Interoperability Technical Committee [online], available: http://uli.unicode.org/home [accessed 30 May 2012].

Unicode TC (2012) Unicode 6.1.0 [online], available: http://www.unicode.org/versions/Unicode6.1.0/ [accessed 30 May 2012].

W3C (2004) Business Benefits of the W3C Patent Policy [online], available: http://www.w3.org/2004/03/pp-points-20040210.html [accessed 23 May 2012].

W3C (2012) About W3C [online], available: http://www.w3.org/Consortium/ [accessed 30 May 2012].

W3C Patent Policy WG (2004a) W3C Patent Policy [online], available: http://www.w3.org/Consortium/Patent-Policy-20040205/#def-essential [accessed 30 May 2012].

W3C Patent Policy WG (2004b) W3C Patent Policy [online], available: http://www.w3.org/Consortium/Patent-Policy-20040205/ [accessed 29 May 2012].

W3C WARP PAG (2009) WARP: Widgets Access Request Policy: PAG Report [online], available: http://www.w3.org/2009/11/widgets-pag/pagreport [accessed 30 May 2012].

Figures

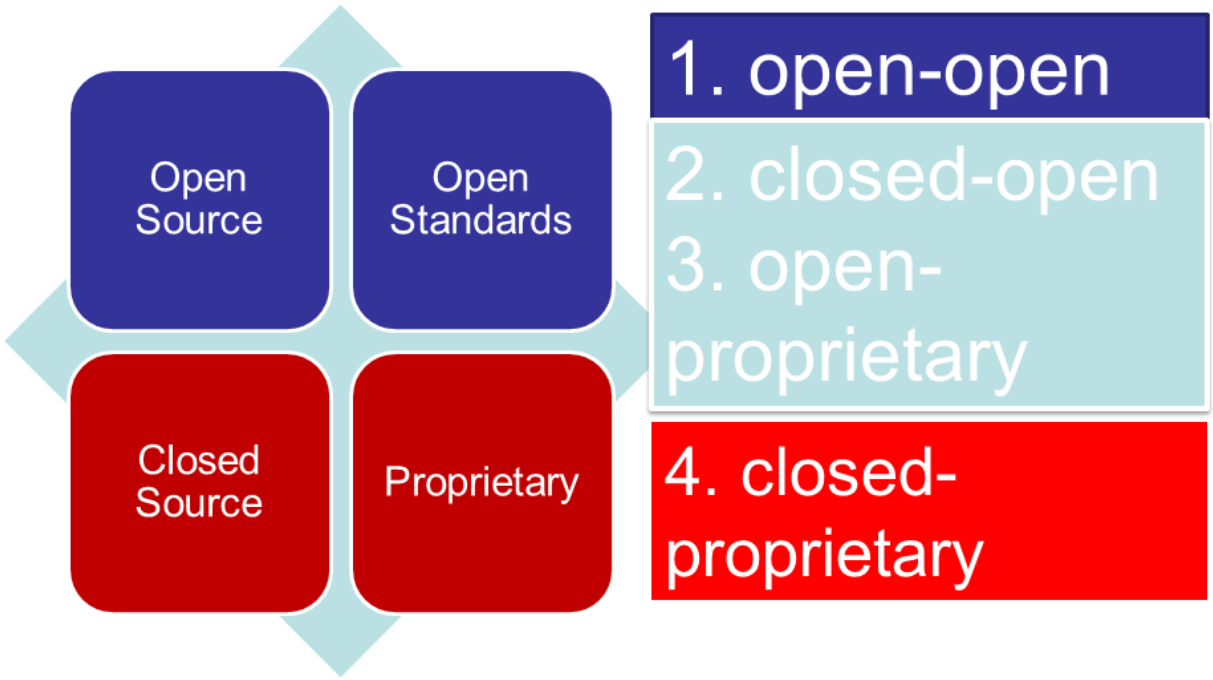

Figure 1: Open - Open - Closed – Proprietary

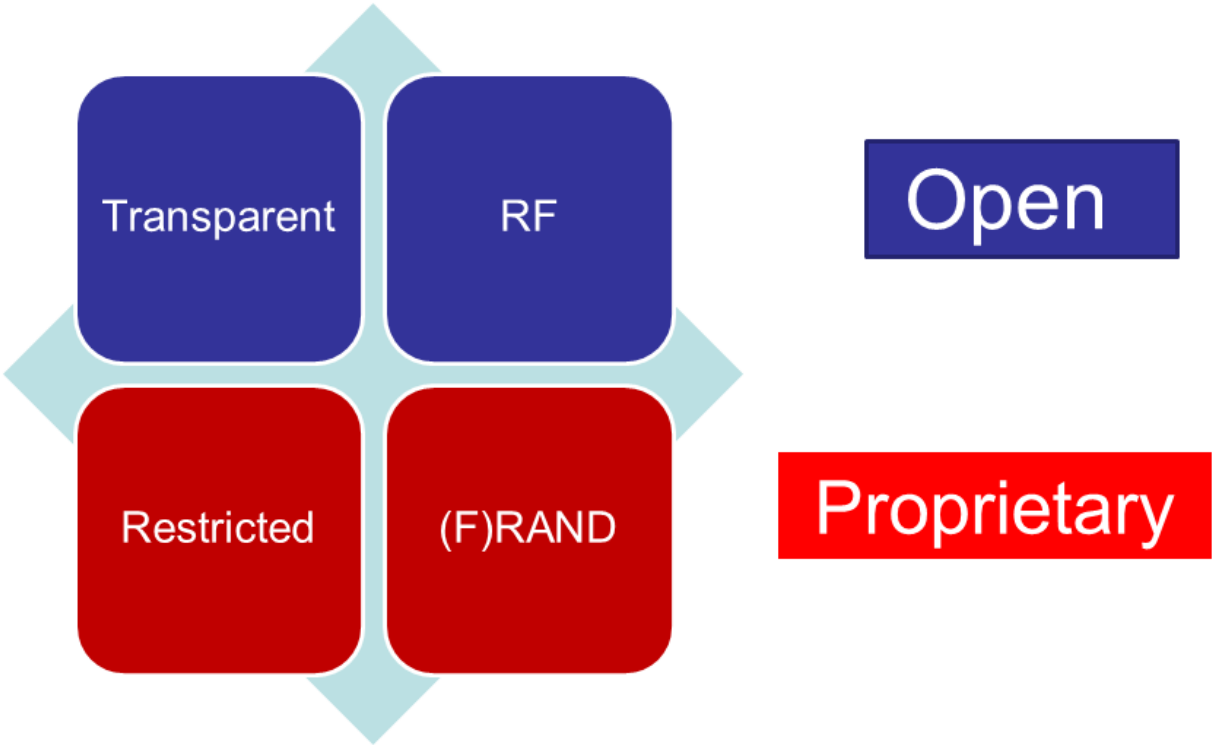

Figure 2: Open and Proprietary Standards

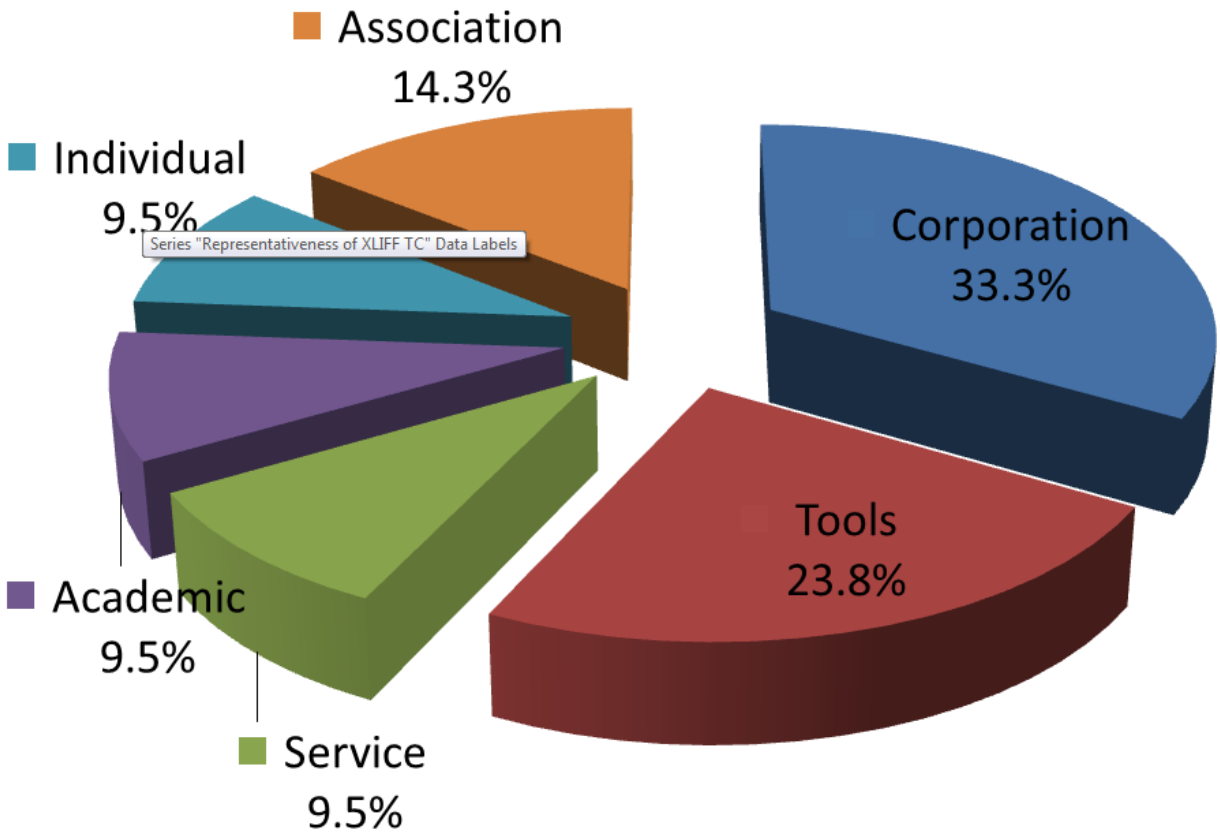

Figure 3: Representativeness of OASIS XLIFF Technical Committee