An Argument for Business Process Management in Localisation

Dr. David Filip, Dr. Eoin Ó Conchúir
Centre for Next Generation Localisation,
Localisation Research Centre,
CSIS Department,
University of Limerick,
Ireland
www.cngl.ie
www.localisation.ie
davidf@ul.ie, eoin.oconchuir@ul.ie

Abstract

Enterprise-level translation management systems cater well for their well-defined use cases. With the rise of user-generated content, the world of localisation is extending to include what we term as 'self-service' localisation. The localisation needs of this emerging environment may differ from more traditional enterprise-level scenarios. In this paper, we present an argument for using business process management (BPM) to help us better understand and define this emerging aspect of localisation, and we explore the implications of this for the localisation industry. Modelling a business process allows for that process to be managed and re-engineered, and the changes in efficiency quantified. It also helps to ensure that automated process aids and electronic systems are put in place to support the underlying business process, matching the real needs of its stakeholders. In this paper, we specifically look at emerging self-service localisation scenarios in the context both of the evolution of the traditional industry process as well as in the context of not-for-profit localisation.

Keywords: *: business process management, BPM, modelling, user-generated content, self-service localisation*

1. Acronyms Used and Basic Definitions[1]

BI - Business Intelligence. The process and technology of organising and presenting business process data and meta data to human analysts and decision makers to facilitate critical business information retrieval.

Bitext - a structured (usually mark up language based) artefact that contains aligned source (natural language) and target (natural language) sentences. We consider Bitext to be ordered by default (such as in an XLIFF file - defined below, an "unclean" rich text format (RTF) file, or a proprietary database representation). Nevertheless, unordered Bitext artefacts like translation memories (TMs) or terminology bases (TBs) can be considered special cases of Bitext or Bitext aggregates, since the only purpose of a TM as an unordered Bitext is to enrich ordered Bitext, either directly or through training a Machine Translation engine.

Bitext Management - a group of processes that consist of high level manipulation of ordered and/or unordered Bitext artefacts. Usually the end purpose of Bitext Management is to create target (natural language) content from source (natural language) content, typically via other enriching Bitext Transforms, so that Bitext Management Processes are usually enclosed within a bracket of source content extraction and target content re-importation.

Bitext Transformation - Similar to Bitext Management, but the Bitext is enriched with newly created or manually modified target content. The agents in Bitext Transformation may be both man and machine, or any combination of the two.

BOM* - Bill of Materials

BPM - Business Process Management

CAT* - Computer Aided Translation

---

[1] For standard localisation industry acronyms see MultiLingual 2011 Resource Directory (MultilLingual 2011). Such standard industry terms are marked with an asterisk (*). We also give short definitions for terms that may be considered commonplace to prevent misunderstanding.

ESB - Enterprise Service Bus, an open standards, message-based, distributed integration infrastructure that provides routing, invocation and mediation services to facilitate the interactions of disparate distributed applications and services in a secure and reliable manner (Menge 2007).

HB - Hand Back. This is being used systematically in two related meanings, either as the message/material conformant to a related HO BOM, leaving an organisation/swimlane as response to the HO, or the last process/subprocess that happens before the corresponding pool-crossing flow.

HO - Hand Off. This is being used systematically in two related meanings, either as the message/material leaving an organisation/swimlane to solicit a response conformant with its BOM, or the last process/sub-process that happens before the corresponding pool-crossing flow.

IS - Information System

LSP* - Language Service Provider

Man - used as synonymous with human, not male, such as for 'man-hours'.

Message - the token in an ESB facilitated workflow or generally any SOA driven workflow. Messages are being enriched as they travel through workflows.

MLV* - Multilanguage Vendor, a type of LSP.

NFP - Not-for-profit

Process - procedure consisting of logically connected steps with predefined inputs and outputs.

SLV* - Single Language Vendor, a type of LSP.

SMB* - small and medium-sized businesses

SOA - Service Oriented Architecture, an architecture concept which defines that applications provide their business functionality in the form of reusable services (Menge 2007).

Swimlane - Pool and Lane as used in BPMN not in sports.

TM* - Translation Memory

TMS* - Translation Management System

Token - whatever travels through a defined process or workflow. Each token instantiates the process or workflow. In this sense, multiple instances of a workflow are created not only as different tokens entering the predefined processing but also at any pre-defined point in the workflow or process where tokens are split according to business rules.

Workflow - an automated process. This is not a commonplace distinction, but we coin it for practical convenience.

XLIFF* - OASIS XLIFF, i.e. XML Localization Interchange File Format. We mention XLIFF in its capacity as a token in localisation processes and as a message being enriched in an ESB or SOA based workflow.

XOR - exclusive OR, logical connective. Used here to characterise the exclusive gate in modelling, as used in BPMN (2011).

2. Introduction

In its essence, localisation is driven by users' preferences to access information in their native language, and this is no different for information being presented online (Yunker 2003). In the corporate context, this has lead to companies providing localised versions of their websites, for example (Jiménez-Crespo 2010).

Meanwhile, with the widespread availability of 'Web 2.0' platforms, it is not only corporations themselves that are producing localisable and localised content (O'Hagan 2009). For example, fans of certain publications (in this case, comics) have produced unsolicited user-generated translations in a collaborative manner (O'Hagan 2009). Indeed, user-generated content (be it opinions or otherwise) is nothing new, although the possibility to work collaboratively online is relatively new. The involvement of online communities in translation has evolved to become solicited user-generated translations. This general concept of leveraging the latent talent of the crowd, particularly online, was coined as crowdsourcing (Howe 2006).

The shift in how content is being transformed in the localisation and translation world has been termed the "technological turn" (Cronin 2010). With respect to content distribution, Cronin argues that the most notable change has come in the form of electronic work station PCs being gradually replaced by the use

of distributed mobile computing. This transition is leading to Internet-capable devices becoming ubiquitous. Rather than localisation being driven by the need to produce static centrally-created content, the emergence of user-generated content is leading to the localisation of user-generated content into personalised, user-driven content. Internet-connected platforms present the potential of collaborative, community translations. This is in contrast to the commercial option of translation through employed translators, freelance translators, or the use of a localisation vendor to act as an intermediary.

While enterprise-based localisation of content and software, being produced in-house, is a mature process with quality assurance certifications available (Cronin 2010), the involvement of online communities (or the "crowd") in localisation is a relatively newer field. Similar to the concept of "open sourcing", the crowdsourcing of localisation is outsourcing the tasks involved to an "unknown workforce" (Ågerfalk and Fitzgerald 2008). We assume that in such a context, contractual agreements may not be in place with members of the community. Rather than being able to agree binding deadlines with paid translators, community members may offer to work on translation tasks on a whim (depending on the process put in place).

In this paper we argue that the evolved state of localisation is yet to be fully understood. Indeed, there is a constant evolution of how the concept of user-driven translation can be applied in real-world situations.

In the following sections, we argue that the activity of business process management (BPM) is a valuable tool for allowing us to understand the new requirements of information systems involving user-generated content and user-provided translations. In later sections, we present three case studies to illustrate how BPM may be applied, and what may happen if the underlying business processes are not correctly incorporated into a new information system. Finally, we conclude that given the advancement of self-service localisation, even in the corporate context, such emergent business processes can be better addressed through BPM.

3. New Business Processes, and Business Process Management

On the subject of newly-emerging business processes in localisation, we must define how a certain block of

content to be localised will be ultimately used. To illustrate this point, let us compare the difference in expectations between the localised version of a corporate brochure when contrasted with that same corporation's desire to localise its ongoing social media stream for different locales. With the former example, we may expect very formal and accurate language, whereas the latter may allow for a more informal approach. A further distinction may be made between relatively informal content being produced by a corporation and useful customer-generated content that may benefit other customers of different native languages. An example of this would be a descriptive forum message, posted online by a customer, providing a solution to an issue with a company's product. Indeed, translation quality is a multidimensional concept that can be approached in different ways including process-oriented international standards, or more community-based localisation (Jiménez-Crespo 2010).

To illustrate that point, we present Table 1. The table shows how content coming from different sources may be localised using different approaches. The upper-left quadrant may be seen as the traditional route taken in localisation. Such business processes are the main focus of translation management systems. The upper-right quadrant may be too costly compared to the value it produces, since a constant stream of user-generated content may overwhelm traditional localisation processes. Indeed, companies are presented with the emerging choice of facilitating their online community in localising content that has been produced by their peers. The lower two quadrants are of particular interest, as it is here that a community of translators (the "unknown workforce") may be asked to help with the localisation of content. It should be noted that volunteer translators are not necessarily individuals donating their free time, but also representatives of external organisations who would benefit from having the content made

	Traditionally-generated content	User-generated content
Traditional content localisation	Localisation of corporate-controlled content by a paid contracted entity (such as a localisation service provider).	Localisation of user-generated content by a paid contracted entity (such as a localisation service provider).
User-driven content localisation	Localisation of corporate-controlled content by volunteer community members.	Localisation of user-generated content by volunteer community members.

Table 1: Both in-house and community-generated content may be localised by either commercial localisation vendors or by the community itself.

available in their primary language.

Focusing on any of these four quadrants in Table 1 presents us with different business processes being represented. For example, a system allowing for ad-hoc volunteer translations of short social media messages may have quite different requirements to a system involving tightly-controlled contracted freelance translators. In the following sub-section, we argue that it is critical that the underlying business processes be closely matched by the functionality of electronic systems designed to support them. We explain how a mismatch in information technology (IT) strategy with information systems (IS) strategy along with business strategy may lead to practical failure of the system being produced.

3.1 Information Systems Perspective
In the localisation context, a "system" may be the socio-technical entity that supports traditional enterprise-based localisation, or a user-driven localisation scenario. To discuss how systems may be designed to cater for any particular permutation of the localisation process, we must first address the nature of a system itself. In information systems theory, the "system" does not merely refer to a computing machine such as a personal computer (PC). Neither does it refer simply to a software application (large or small, TMS, ESB etc.) designed to facilitate certain operations. Rather, we view an information system as a socio-technical entity, similar to Galliers (2004).

An information system is comprised of the information being processed and produced, along with the organisational context of its users and other stakeholders. An information system designed to encompass a socio-technical environment would combine information and knowledge sharing services that would facilitate both the exploration and exploitation of knowledge (Galliers 2006).

A long-standing view of information systems is that the activities falling under information communications technology (ICT) development must be closely aligned to the information system as a whole, which in turn must be aligned to the organisation's business strategy (Galliers 2006). A misalignment between these concepts or activities may lead to a failed system. A failure does not necessarily imply that the system itself does not function (Laudon and Laudon 1996). For example, a system may be perceived as failed if it has not been successfully adopted by its intended user base, even

if the system itself runs "as designed". In this paper, technology underlying localisation including CAT tools and Translation Management Systems (TMS) is discussed from this broader IS perspective. As such, they need to be aligned with business objectives.

3.2 Business Process Management (BPM)
A business process is a "set of partially ordered activities intended to reach a goal" (Hammer and Champy 1993). Relating this to localisation, a high-level business process may be taking a mono-lingual technical manual and all the steps required to adapting it to various target locales. Similarly, a business process may describe the activities required to produce a community-based localisation project. In localisation specifically, Lenker et al (2010) argue that by abstracting a localisation business process as a workflow, the process can be potentially automated and its efficiency improved. Business processes may be quite low-level, with a large organisation being comprised of thousands of such processes (Turban et al 1993).

Formally, a process is seeded with inputs, and it produces outputs. Thus, the output of a process can be measured. This is an advantageous approach, since measurements of process efficiency allow us to tweak the process and measure the consequences. BPM thus provides a structured framework for understanding the business process itself, and then optimising that process.

3.3 Modelling Business Processes
An information system may be developed to improve the current workings of an organisational unit, or it may be conceived to support an entirely new set of business activities. In either case, we may analyse the underlying business activities, producing conceptual models of the activities.

Modelling a business process is the act of formally describing the business processes at hand. Many businesses have process models of their systems (Cox et al 2005). Once contextual information has been elicited about the socio-technical system, and explicitly described through business process modelling, an understanding of what problems need to be solved should emerge (Cox et al 2005).

Business processes can be captured in a standard language, that being Business Process Model and Notation (BPMN, formerly also known as Business Process Modeling Notation). It is maintained by the Object Management Group (OMG). It offers an

extensive standard modelling framework, readily understandable by business people, including analysts and technical developers (BPMN 2011). Models recorded in this manner allow for the business processes to be modelled while abstracting from actual implementation details. This provides a standardised way of communicating process information to other business users, process implementers, customers, and suppliers. Requirements engineering approaches can be applied to BPM, such as employing role activity diagrams (Bleistein et al 2005).

By taking a set of models produced in a standard modelling language, BPM can let us carry out business process improvement through business process re-engineering. Software tools allow the analyst to work on the business process models in order to produce an optimised set of processes, ultimately improving the workings of the organisation.

4. Case studies

In this section, we present a number of case studies to demonstrate the concepts behind BPM, and how they may be applied to localisation. These case studies are then compared and contrasted in the following analysis and discussion section.

BPM, in essence, deals with understanding the business processes of an organisation. The concept of an organisation here is a socio-technical grouping of people and systems. In order to manage any business process, it is necessary to understand the participants in the system, the activities taking place in the system, and the message flow of information throughout the system (BPMN 2011). For example, Lewis et al (2009) analyse the set of activities and communication mechanisms involved in a traditional localisation workflow, and use this to understand newer community-based approaches to localisation. First, though, we present a simple example of a

system that supports the business logic of content creation.

4.1 Case Study 1: Content authoring business logic encapsulated by WordPress

With the advent of the World Wide Web in the early 1990s, content publishers (both individuals and organisations) were presented with a new opportunity to publish their content. At its most basic, text content can be published online as a hypertext mark-up language (HTML) document by uploading it to a web server. The document can contain static content, and so is limited in how it can encapsulate the business logic of a more complex content system. An information system may be represented somewhat by interlinking static HTML documents. More likely, however, is the need to support the business logic through dynamic server-side scripting which would output HTML documents generated on the fly.

By the late 1990s, a trend in personal web pages was to publish a 'log' of web sites found by the web page owner, in chronological order. Yet, by that stage, most web loggers (who became known as 'bloggers') hand-coded their web sites. No tools were publicly available that would support the requirement of dynamically publishing a series of links to a web page (Blood 2004).

In 1999, a free web logging system called Blogger (http://www.blogger.com) was launched with the tag "Push-button publishing for the people". The simplicity of the system made it very popular, with non-technical users beginning to use the web logging platform simply as a way to publish their thoughts and opinions online, without necessarily any links in the published post (Blood 2004). This was the birth of the blog post format.

At the time of writing this paper, WordPress (http://www.wordpress.org) is one of several popular open-source blogging systems, having first been released in 2003. Perhaps due to the platform's ease
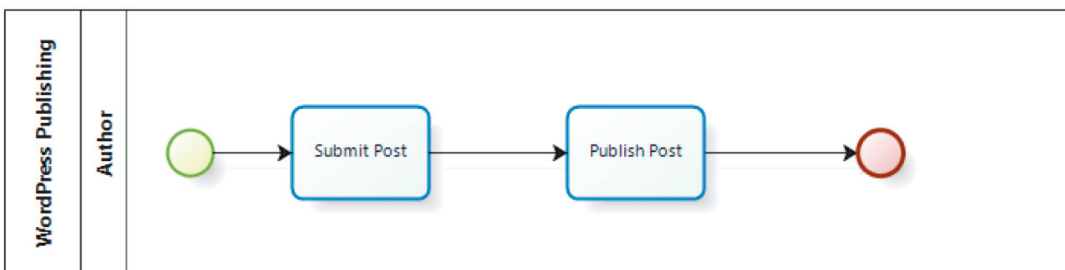

Figure 1: Single-user content authoring and publishing as supported by WordPress.

of use, but moreover its direct addressing of the business logic required by bloggers, the platform has gained a wide user base. WordPress has been adopted by individual bloggers and large organisations alike, such as the popular technology blog TechCrunch (http://www.techcrunch.com) and Forbes' blog network (http://blogs.forbes.com/) (WordPress 2011a).

Figure 1 illustrates the simplest content publishing workflow offered by WordPress. Note that we make use of Business Process Modelling and Notation (BPMN) for the illustrations in this paper. This allows for an abstracted understanding of the underlying business process.

WordPress is a dynamic server-side platform that encapsulates the business process of publishing and managing content online as an individual or as a team of content authors. It does so by supporting the activities of content creation, reviewing, editing, and publishing. WordPress supports the user roles of Super Admin, Administrator, Editor, Author, Contributor and Subscriber (WordPress.org 2011b). A team of content authors may assign these different roles to different people to manage the publishing process. For example, the Contributor role allows that person to author and edit their own content, but not publish it to the blog. An Author user has the same abilities, in addition to being able to publish their own content. Notably, the Editor role can create content, manage their content and others' content, and choose to publish others' content (it is beyond the scope of this article to further describe in detail the roles and capabilities offered by WordPress).

In summary, the system encapsulates the roles and activities required for publishing content online. The business process (the set of activities involved in authoring, editing and publishing online content) is closely matched by the action-centric functionality of the WordPress system. In this case, business process management may be used to understand the underlying business process, to model it, and to tweak it. By illustrating this specific case study of a content management system, we argue more generally that BPM is a worthy approach for understanding the underlying business process, and thus making it more likely that the system being developed will align more closely with actual requirements.

4.2 Case Study 2: The traditional industry localisation process in the industry, enterprise and SMB context

Figure 3 illustrates a high level model of the enterprise localisation process. Each of the high level processes represented by blocks in the figure would need to be defined in further levels of granularity in order to be relevant for real implementations. The model is nevertheless useful as a high-level representation. It is helpful for showing the most important process differences at the relevant level of complexity. In this paper we only include models that can be quickly understood at first glance, for several reasons:

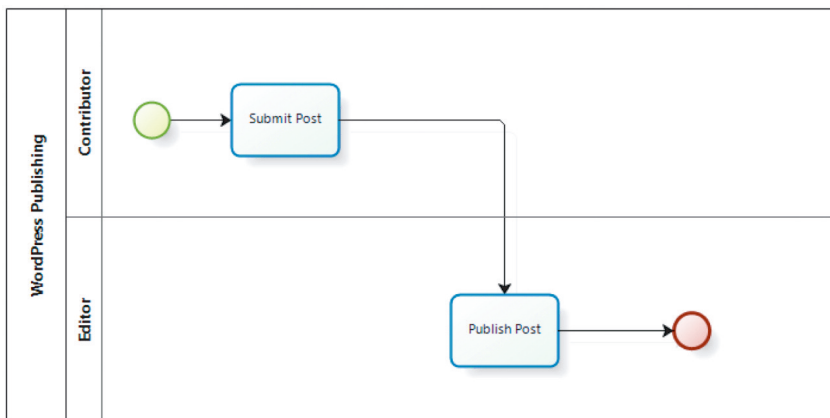1) To illustrate points made about process differences occurring in different localisation settings.

Figure 2: The business process of a Contributor submitting a post, and an Editor publishing that post, as supported by WordPress.

2) To illustrate how the BPMN standard can be used to create pictorial representations facilitating process discussion in a highly intuitive way.
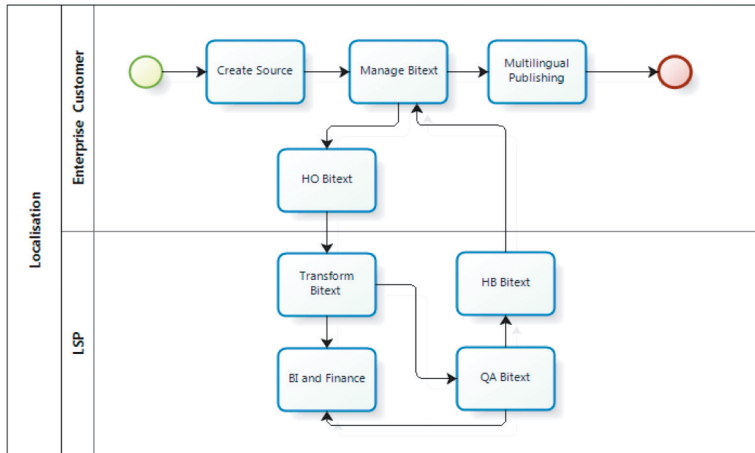
Figure 3: The localisation process in the enterprise context covering content management and publishing.
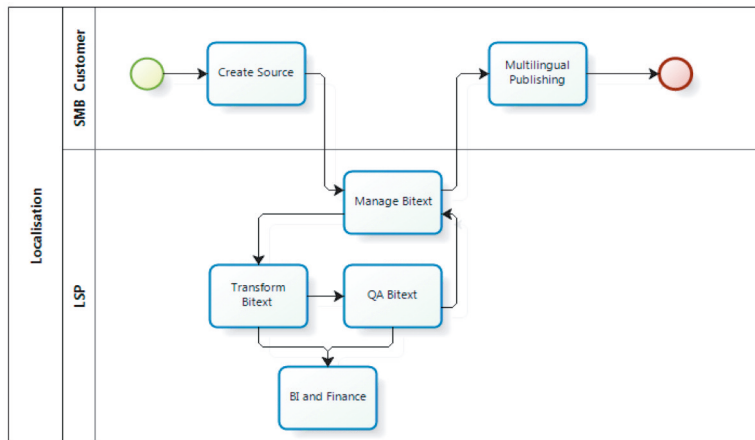

Figure 4: The management of Bitext is usually performed by an LSP partner for an SMB.

The model in Figure 3 anchors the localisation process in the broader context of multilingual content management and publishing. Content is being created specifically in one language, in the sense that a single piece of information can only be conveyed practically in one language at a time. The publisher, however, needs to publish its information in many languages. As the transitions from the creation in one language to multiple languages in publishing always include transformations specific to the language pair, we have labelled the intermediate steps as "Bitext Management". Bitext Management is the central piece of any localisation process. In fact, Bitext Management forms the fundamental distinction between localisation processes in different contexts in terms of whom, where, and how it is executed.

In contrast, Small and Medium Businesses usually lack the resources needed to take control of their translation memory leveraging. They are usually unable to manage their Bitext on their own. Therefore, although localisation customers legally retain rights to their bilingual corpora, in practice their Bitext Management is a black box for them which is managed by a long term LSP partner.

In summary, BPMN has allowed us to visually represent the high-level business processes of Bitext Management for enterprises (Figure 3) and SMBs (Figure 4). It helps to demonstrate that the primary distinction between both cases is whether the "Manage Bitext" activity happens in-house, or is the responsibility of an LSP.

4.3 Case Study 3: The localisation process in the Not-For-Profit context

Further to enterprise and SMB localisation, we would like to investigate whether not-for-profit (NFP) localisation is any different. At a first glance it may seem so. Again, we make use of BPMN to help answer this question.

Figure 5 illustrates a typical localisation process for a

to the translator who types a new document without using translation tools, and the hard copy of the translated document can be rubber-stamped (at a fee) as being translated correctly and accurately by a court-approved interpreter.

More generally, this is the low tech scenario of the localisation process typical for low Localization Maturity Levels (DePalma 2006; DePalma 2011;
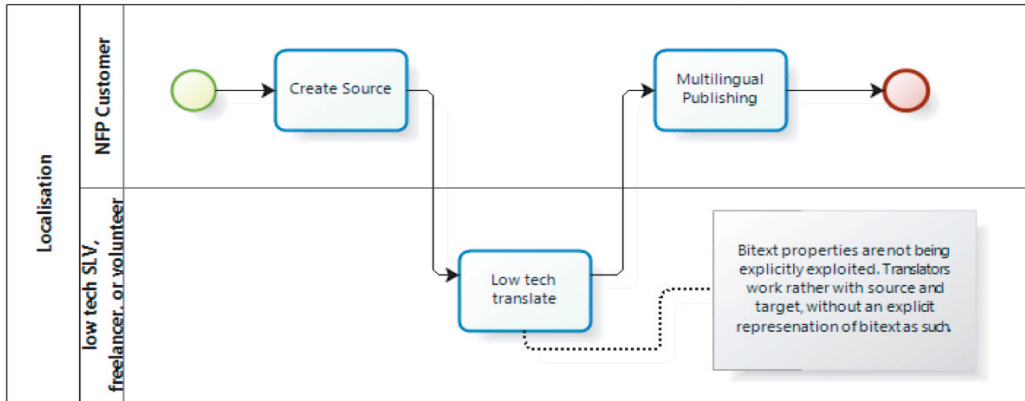

Figure 5: Modelling the localisation process in a not-for-profit scenario.

not-for-profit customer. It makes use of a low tech SLV, freelance or volunteer translators. While the source content is produced in-house by the NFP organisation, the translation process is performed externally (represented by the "Low tech translate" activity in the figure). "Low tech" is used here in the sense that this scenario does not make any explicit use of Bitext properties, due to an apparent, or real, lack of CAT tools in the process. In particular, the low tech SLV may be an over-the-street agency that only accepts content by fax, sends the content by fax

Paulk et al 1993). The business process is not specific to not-for-profit organisations. This has important implications for those building localisation solutions for not-for-profits that may have fewer resources in place to support the localisation process. Such service and technology solutions would need to address a certain level of effectiveness, and hence sophistication. As a result, the solutions would need to take responsibility for Bitext Management, as the typical NFP customer will not be able to manage their Bitext on their own. Organisations that are aiming to
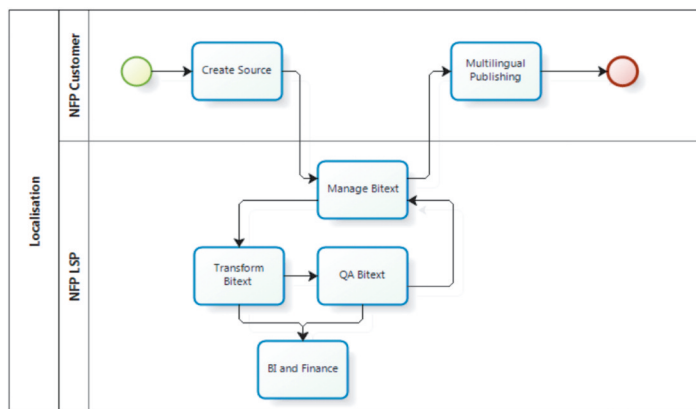

Figure 6: The localisation process in the not-for-profit context features Bitext Management outside of the organisation.

---

[2] CSA gave a preview of the 2011 TMS report on 8th September, 2011. However, the full report was still pending publication at time of writing.

support not-for-profit localisation may - in effect - emulate the SMB localisation model, at least at this high structural level. Figure 6 illustrates this finding.

One may therefore come to the conclusion that there is no difference between the traditional localisation process (Figure 4) and the not-for-profit model (Figure 6). However, in section 5.3 we describe why this is actually not the case.

5. Case Study Analyses

In the previous sections, we presented three case studies by modelling the relevant business processes. Some comparisons were made between the case studies. In this section, we discuss how the existing localisation solutions address the above described scenarios and present further conclusions arising from the analysis of these case studies.

Localisation platforms, such as CAT tools and Translation Management Systems (TMS), do currently exist and primarily address the traditional enterprise localisation process. We wish to understand the level and nature of impact of next generation localisation factors that we see arising with the inclusion of crowd sourcing concepts. To do so, we need to discuss the role of CAT tools and TMSs in the localisation-enabling Information Systems (IS).

5.1 The role of current platforms in addressing localisation business needs
Since 2006, Common Sense Advisory (CSA) has been publishing an authoritative comparison of translation management systems (TMSs) (Sargent and DePalma 2007 and 2008). As there has not been a comprehensive report since 2008 (only individual TMS scorecard additions have been published)[2], the 2008 report still serves to define classifications and groupings. Our classification in this paper draws loosely from the CSA classification.

The most prestigious category according to CSA is the Enterprise TMS (ETMS) or "cradle to grave" systems. These systems are expected to be enterprise-class information and automation systems. Many players have been trying their luck in this category. The initiator and long time leader of this category had been Idiom WorldServer (now SDL WorldServer), which, even today, remains unparalleled in the expressivity of its workflow engine within the class of ETMSs. However, this class of TMSs is being rendered largely obsolete due to the present-day

development of general enterprise architecture, in terms of business need and development.

It has been noted (Sargent and DePalma 2008; Morera et al 2011) that localisation automation systems have been successful in narrowing permissible workflow complexity in building a particular production workflow. Complexity here refers, roughly, to the number of the classical workflow patterns (van der Aalst et al 2003; Morera et al 2011).

TMSs can be considered as highly specific automation systems, and different categories of TMSs may be distinguished by their level of specificity for localisation workflow support. Part of their success is in simplification relative to traditional industry patterns.

For instance, most of the existing systems are hard wired for a single source language per project. This means that they will be challenged by multiple source languages scenarios that play an increasingly important role. The reason that current solutions have been built to cater exclusively for a single source language scenario is that most of the current enterprise-class localisation processes actually normalise to a single source language, very often English, especially in multinationals. Even Asian and German-based multinationals, that would often try to use their local languages as the source languages, are forced to use English due to outside forces. Such forces would include the present state of the market and procurement necessities such as economies of scale. If English is not used as a source language, it still tends to be used as a pivot language, through which all content is translated. In the following, however, we leave aside the complexities of managing multiple source languages.

The least capable, in terms of building complex automation workflows, would be the category of TM Servers. The capabilities of TM Servers in the area of automation can range from a simple automated segment pair lifecycle through to a predefined set of states that each pair can retain throughout its life, all the way from 'new', through to 'revised' and to 'deprecated'. Every product in this category manages to automatically search and retrieve relevant terminology, both for full and fuzzy matches.

However, this capability has been commonplace in our industry for so long that it is not even considered "automation". It is, indeed, a level of automation that

can be taken for granted thanks to the native functionality of computer aided translation (CAT) technology and is usually not enhanced to a great degree by server-level products (apart from the apparent advantages of committing to a regularly backed up well-resourced database, compared to a locally installed database or a local proprietary database file).

In fact, many tools that had been working without issue locally or through local area networks (LAN) had maturity challenges when introducing or perfecting their server-based product. The leader in this capability has, so far, been the Lionbridge Translation Workspace that is offered through the GeoWorkz.com portal (originally known as Logoport).

We see a tension between the interests of large LSPs in attempting to control the technology space, while customers seek to avoid technology lock-in. There are repercussions of this tension for the LSP world. An LSP may have a significant number of stakeholders. Various types of LSPs exist ranging from mutually-coordinated freelancers, to bricks-and-mortar SLVs, through to large multimillion so-called MLVs competing for a place on the CSA beauty contest ladder (Kelly and Stewart 2011).

The standardisation driven by enterprises will be exploited downwards and we expect that this will lead to the language industry becoming even more strategic, yet even more commoditised. We predict that there will be no differentiator for SLVs except

for resource management. MLV competition will become even fiercer as the standardised SOA and ESB based architecture will drive the cost of entry even lower. Cyclically, the MLVs will need to deal with large enterprises taking Bitext Management and other value added high margin services in house, forming specialised service units such as Oracle's Ireland based WPTG (Worldwide Product Translation Group**).**

5.2 Adoption of Crowdsourcing in Localisation
The democratisation of the Web has emerged through the power of the "crowd". This trend has also been increasingly applied to the localisation process where the concept of crowdsourcing has seen members of the crowd performing localisation tasks, such as translation and reviewing. There are two settings in which the stakeholders are ahead in embracing this relatively new trend:

1) Enterprises

2) Not-for-profit (NFP)

The crowd is important for both of these because of similar, yet distinct, reasons. In the not-for-profit (and potentially charitable) setting, accessing a crowd of volunteers would be attractive. Crowd-sourced translation may also be attractive for enterprise, but there are significant levels of investment required for supporting that through technology, oversight and management. In other words, the return on investment (ROI) must still be properly calculated even if engaging with an unpaid crowd.
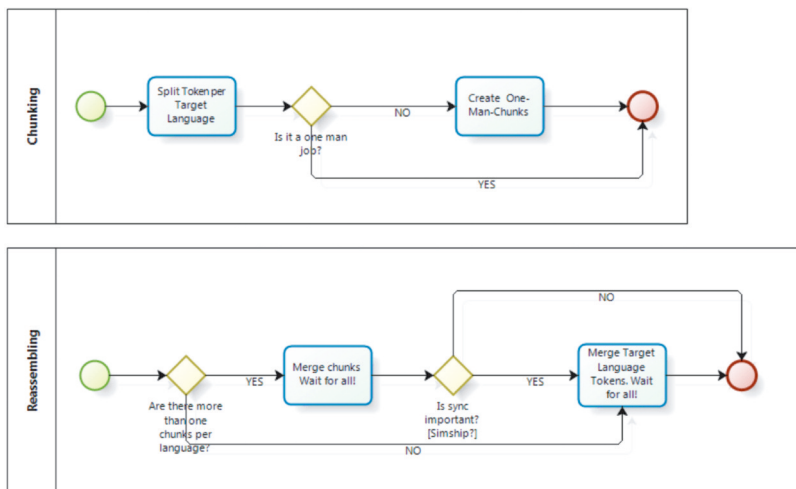

Figure 7: The chunking and reassembling activities in a typical localisation process.

We speculate that the motivation of the unpaid crowd may be a distinguishing factor in next generation localisation. This may not be such an issue in a more traditional paid translation context.

More specifically, volunteers may have little time to contribute to a localisation project. The implication of this is profound: the chunks of content presented to them as tasks need to be much smaller than those required in the traditional localisation workflow. We discuss this topic further in the next sub-section.

5.3 New Requirements for Bitext Chunking

Figure 7 shows the lower level models of chunking and reassembling that we have been using in previous models when referring to Bitext Management.

The chunking process multiplies the tokens that are travelling through the process in two steps. First, it creates a token per target language. Second, it creates a token per one-man-chunk.

A process that uses chunking must also contain reassembling further down the road to ensure that tokens are properly merged back (i.e. well handled). One may notice that the re-merging of target versions into one deliverable token is optional and more likely to occur in an industry setting than in a not-for-profit setting.

Using XLIFF as the message container provides benefits as XLIFF is capable of carrying a token in the size of thousands of files, or as small as a single translation unit (OASIS XLIFF 2008).

Figure 8 applies equally to the industry setting and the not-for-profit setting. There is, however, a very important parameter that governs the behaviour of the XOR gateway diagram. From a technical perspective, the decision is simply based on a single parameter.

Figure 8 represents the process of abstracting the steps that are needed to be taken to get a certain

output, given an input. The figure does not itself specify whether or not the workflow process needs to be automated in real life. The parameter is the size of a one-man-chunk. In the paid industry setting the one-man-chunk may easily comprise effort of up to five man-days (in case of relaxed schedules even ten man-days may count as one-man-chunks, and in the literary translations world one person routinely deals with effort in terms of man-months).

However not-for-profit organisations may have to deal with real life emergencies as they arise (such as tsunamis, earthquakes, famines, and many other less dramatic, yet time sensitive, issues). Therefore, they may have very tight schedules as in the translation industry, but seldom have the budgets to buy full-time resources.

Therefore, the one-man-chunk in the volunteering setting is better defined in terms of man-hours. The five-man-day chunk is not extraordinary for enterprise settings, but could take months for a volunteer to complete. As such, the content requires a much higher level granularity of chunking for fast turnaround of each chunk.

Assuming that a not-for-profit project needs to publish multilingual information within a week of the creation of the source text, and assuming that the crowd of highly-motivated volunteers have on average 20% of normal full-time employment to dedicate to the project, we conclude that a project should be chunked accordingly to blocks of four man-hours.

In the case of more stringent deadlines, or where the crowd is less disciplined, chunking may need to be set at two man-hours, or smaller.

Chunks smaller than one man-hour may not be effective in practice, unless the tasks are specialised, such as for user interface translation projects. Following this discussion, we can see the typical model for NFP localisation should be as illustrated in
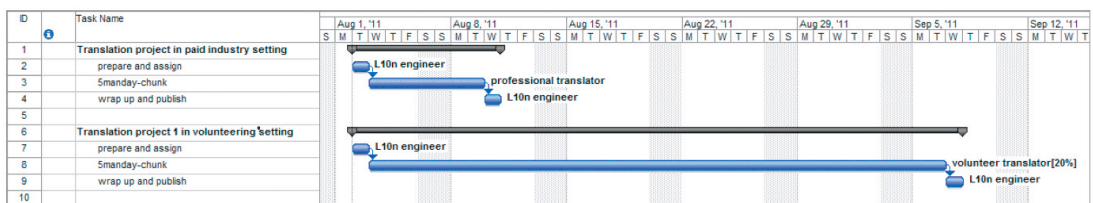

Figure 8: Industry chunking is not for volunteers

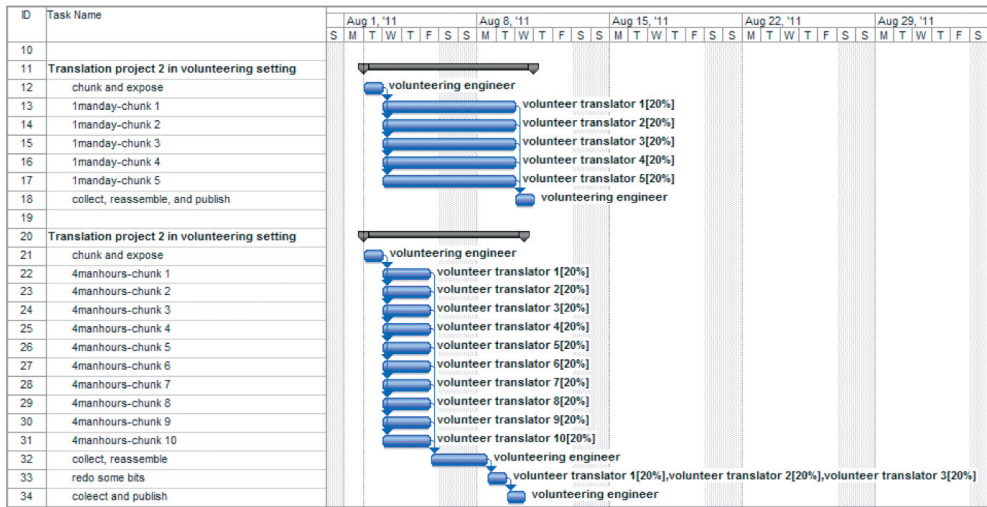[3] See classic discussion of workflow expressivity by Aalst et al. 2003.

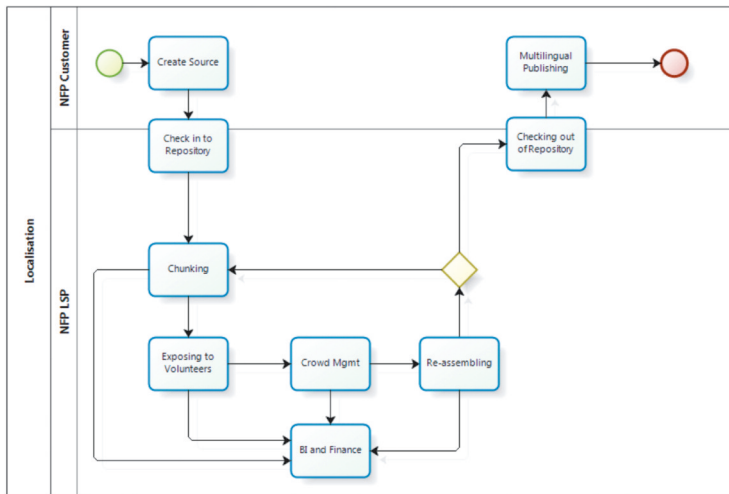**Figure 9:** Automated chunking in terms of man-hours is essential for volunteering settings

Figure 10: a model of not-for-profit localisation, with further detail provided for content chunking.

The process illustrated in Figure 10 is structurally similar to traditional models. Yet, there are different business needs for the supporting technology between the two different scenarios. There are radical differences, for example, in the availability of resources. In the self-service scenarios that leverage crowd-sourced translation, whether in an enterprise support or a charitable NFP scenario, automated chunking, pull-driven automated assignments, and automated reassembling are a must due to the demand for much finer granularity of chunking. In contrast, in the traditional bulk localisation scenario these are only tentative activities that are often simply performed manual**y.**

6. Conclusion

What is the token and/or the message in the localisation process? We have hinted that ideally the localisation ESB message should have the form of a flexibly chunkable and reassemblable Bitext. With OASIS XLIFF, the industry has such a standard, yet evolving, format to capture industry wisdom and address new business needs. It is capable of carrying payload and metadata with a wide range of granularities and process requirements. Through the business process management practices applied in this paper, we have found that the common denominator of all localisation processes may be as follows:

Parsing of source text -> routing Bitext -> enriching Bitext -> quality assuring Bitext -> exporting target text.

For performing the localisation processes in any organisational setting it is critical to be able to extract global business intelligence from most of the workflows and processes involved.

For an enterprise, managing Bitext has also traditionally meant enforcing process and technology. We argue that this is not a priori a consequence of including Bitext Management in the enterprise process. Rather, in the past, the enterprise may have had to take stringent control due to the lack of standardisation in the areas of both Bitext and Bitext Transformation processes.

Today many enterprise-level practitioners have seen that enforcing process and methodology is not sustainable and/or indeed very expensive. We can see two complementary trends:

1) Standardisation of Bitext message, both payload and metadata.

2) Reuse of available SOA architectures and extra-Localisation workflow solutions, namely the underlying ESBs.

What can be used as the ESB in this case? While most readily-available ESB specialised middleware comes to mind, it can, theoretically, be any sufficiently expres^{s}ive3 workflow engine. 'Theoretically' must be emphasised here, as clearly any Turing-complete engine can do what is needed, which is, however, far from claiming that the level of effort needed would be practically achievable or otherwise relevant. In real life situations, many factors play important roles in making this decision, including but not limited to:

1) Legacy investment into and the present state of the overall IS in the organisation

2) Level of fitness of the current IS for the business needs of the organisation

3) Legacy investment into and the present state of specialised localisation technology

4) Importance of unified BI on localisation within the organisation

5) Licensing models of legacy solutions
6) Long term vendor relationships

Enterprise users want to prevent lock-in and manage quality on an 'as needed' basis, which very often applies to string level. In fact, we see, from our case study analysis, the community workflow and the enterprise workflow converging.

The 21st century has seen an onslaught of service-oriented architectures, not only in the IT mainstream but also in the localisation and translation industry. Many an industry player has realised that they no longer wish to be locked in to a particular language technology stack, and some have found their Enterprise Service Buses relevant as potential backbones for what they need to achieve in the area of localisation and translation.

It seems clear that the challenge in the localisation and translation industry is not just of process modelling. It is rather a complex Change Management issue that cannot be properly addressed without applying mature Business Process Management technique**s.**

Acknowledgements

All BPMN models used in this paper were created using the free Bizagi Modeler Software, v2.0.0.2, which is BPMN 2 complian**t.**

References

Aalst, Van der, W.M.P., Hofstede, ter, A.H.M., Kiepuszewski, B., Barros, A.P. (2003) 'Workflow Patterns', Distributed and Parallel Databases, 14, 5-51.

Aalst, Van der, W.M.P. (2004) 'Pi calculus versus Petri nets: Let us eat "humble pie" rather than further inflate the "Pi hype"', unpublished.

Ågerfalk, P. J. and Fitzgerald, B. (2008) 'Outsourcing to an Unknown Workforce: Exploring Opensourcing as a Global Sourcing Strategy', MIS Quarterly, 32(2), 385-409.

Bleistein, S., Cox, K., Verner, J. and Phalp, K. (2006) 'Requirements engineering for e-business advantage', Requirements Engineering, 11(1), 4-16.

Blood, R. (2004) 'How Blogging Software Reshapes the Online Community', Communications of the ACM, 47(12), 53-55.

Brabham, D. C. (2008) 'Moving the crowd at iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application', First Monday, 13(6).

Business Process Model and Notation v2.0 (BPMN) (2011) Needham, Massachusetts: Object Management Group, Inc (OMG).

Cronin, M. (2010) 'The Translation Crowd', Tradumatica: traduccio i tecnologies de la informacio i la comunicacio, 8, December 2010.

Cox, K., Phalpc, K. T., Bleisteina, S. J. and Vernerb, J. M. (2005) 'Deriving requirements from process models via the problem frames approach', Information and Software Technology, 47(5), 319-337.

DePalma, D.A. (2006) Localization Maturity Model, Lowell: Common Sense Advisory, 16 August.

DePalma, D.A. (2011) Localization Maturity Model 2.0, Lowell: Common Sense Advisory, 28 March.

Galliers, R. D. (2004) 'Reflections on Information Systems Strategizing', in Avgerou, C., Ciborra, C. and Land, F., eds., The Social Study of Information and Communication Technology: Innovation, Actors, and Contexts, Oxford: Oxford University Press, 231-262.

Galliers, R. D. (2006) 'On confronting some of the common myths of Information Systems strategy discourse', in Mansell, R., Quah, D. and Silverstone, R., eds., The (Oxford) Handbook of Information and Communication Technology, Oxford: Oxford University Press.

Hammer, M. and Champy, J. (1993) Reengineering the corporation: a manifesto for business revolution, New York: HarperBusiness.

Howe, J. (2006) 'The rise of crowdsourcing', Wired, 14(6), available: http://www.wired.com/wired/archive/14.06/crowds.html [accessed 4 July 2011].

Jiménez-Crespo, M. A. (2010) 'Web Internationlisation strategies and translation quality: researching the case of "international" Spanish', Localisation Focus, 9(1), 13-25.

Kelly, N., Stewart, R.G. (2011) The Top 50 Language Service Providers, Lowell: Common Sense Advisory, 31 May.

Laudon, K.C. and Laudon, J.P. (1996) Management Information Systems: Organization and Technology, Englewood Cliffs: Prentice-Hall.

Lenker, M., Anastasiou, D. and Buckley, J. (2010) 'Workflow Specification for Enterprise Localisation', Localisation Focus, 9(1), 26-35.

Lewis, D., Curran, S., Doherty, G., Feeney, K., Karamanis, N., Luz, S. and McAuley, J. (2009) 'Supporting Flexibility and Awareness in Localisation Workflow', Localisation Focus, 8(1), 29-38.

Menge, F. (2007) 'Enterprise Service Bus', Free and Open Source Software Conference 2007 (FOSS4G), Victoria, Canada, 24-27 September.

Morera, A., Aouad, L, Collins, J.J. (2011) 'Assessing Support for Community Workflows in Localisation', accepted for 4th Workshop on Business Process Management and Social Software (BPMS2'11), August.

MultiLingual (2011) MultiLingual 2011 Resource Directory, Sandpoint: MultiLingual, available: https://www.multilingual.com/downloads/2011RDPrint.pdf [accessed 1 August 2011].

OASIS XLIFF 1.2 (2008) Oasis, available: http://docs.oasis-open.org/xliff/v1.2/os/xliff-core.html [accessed 2 August 2011].

O'Hagan, M. (2009) 'Evolution of User-generated Translation: Fansubs, Translation Hacking and Crowdsourcing', The Journal of Internationalisation and Localisation, 1, 94-121.

Paulk, M.C., Curtis, W., Chrissis, M.B., Weber, C.V. (1993) Capability Maturity Model for Software, Version 1.1, Technical Report CMU/SEI-93-TR-024 ESC-TR-93-177, Pittsburgh, Pennsylvania: Software Engineering Institute, available: http://www.sei.cmu.edu/reports/93tr024.pdf [accessed 2 August 2011].

Sargent, B.B., DePalma, D.A. (2008) Translation Management Systems, Lowell: Common Sense Advisory, 16 September.

Sargent, B.B., DePalma, D.A. (2007) Translation Management System Scorecards, Lowell: Common Sense Advisory, 26 February.

Turban, E., Leidner, D., McLean, E. and Wetherbe, J. (2007) Information Technology for Management: Transforming Organizations in the Digital Economy, ed. 6, Wiley.

WordPress.org (2011a) Showcase [online], available: https://wordpress.org/showcase/ [accessed 13 July 2011].

WordPress.org (2011b) Roles and Capabilities [online], available: https://codex.wordpress.org/Roles_and_Cap-abilities [accessed 5 July 2011].

Yunker, J. (2003) Beyond Borders: Web Globalization Strategies. Indianapolis: New Riders.