White Paper

Web Content within a Global Information Management Strategy

A global website can be a very effective vehicle for reaching large audiences, especially those in foreign markets. How do you design, implement and manage this resource to its best advantage – ensuring it dovetails with your overarching Global Information Management strategy?

This SDL white paper describes effective design and implementation strategies that will help marketing and localization managers design and implement a truly global website. Web managers will also find many valuable tips for multilingual websites. Take advantage of these proven strategies and best practices to make your global website more successful and to reduce your implementation, localization and maintenance costs.

Sophie Hurst
Senior Product Marketing Manager

www.sdl.com
www.sdltridion.com

1.0 Executive Summary

It is easy to forget the impact of the Internet. But scroll back fifteen years and imagine trying to research an obscure organization on the other side of the world, book a flight online and reserve your seat and meal, manage your mobile phone in a self-service environment, or bid in an online auction. All actions we take for granted today. We have become Internet connoisseurs too. We are quick to click to an alternative site if the one we are viewing is poorly structured, tricky to navigate – or most importantly, hard to read and understand.

When designed and implemented correctly, a corporate website can be one of the most cost-effective tools in an organization's sales and marketing inventory. Done incorrectly though, it becomes a nightmare of cost overruns, internal strife, fragmented messages and customer alienation.

This paper addresses Global Information Management best practices that will help you to design and implement a truly global website. Design the site with your users in mind to ensure that their Web experience supports – not undermines – your marketing message. Your design strategy should consider the use of a data format which enables straight forward content reuse – such as global XML. It should address which languages should be used and to what depth the website will be localized. Will the site support all the requirements of local markets, such as double-byte or bi-directional languages; different currencies and address formats and different number formatting? Will internal or external resources be responsible for translation, localization and review? How will international users access the localized content?

Companies use various metrics to measure the success (or return on investment) of their website implementations. Effective solutions allow content to be localized and posted quickly with minimum effort. These solutions centralize and standardize technology platforms and translation workflow processes across the entire enterprise. They can help you to strike the right balance between global and local content and present it to your customers in a culturally appropriate way.

Internationalize the site so that it supplies local information according to the user who is looking at your pages. A properly internationalized site will cost less to localize, enable you to publish content more quickly and provide a better user experience.

Websites can be implemented using centralized, decentralized or hybrid structures. Each structure has advantages for supporting specific target markets. Different types of Web content, such as text, graphics, active content, search engine keywords and multimedia assets, require different approaches to localization. Take advantage of new technologies, such as Translation Management Systems to gain control of these assets.

The Web is a volatile place, and website information needs to change frequently. How effectively you manage change and updates can determine the long-term success or failure of your implementation.

2.0 Global Websites in the Context of Global Information Management

When you decide to purchase a new vehicle, where does your research begin? Most commonly, it will be the websites of the automotive companies. It's a similar situation for a multitude of other purchase and after-sales support decisions. Although a corporate website can have a significant impact on customer satisfaction and revenues, it needs to be designed and implemented correctly. This is not easy to achieve though. The global information lifecycle – from content creation through to the publishing of information in multiple languages on the Web – involves complex processes that span geographical, organizational and technological boundaries. All too often, responsibility for it is devolved to local operations, with the resultant fragmented corporate messaging, prolonged product introduction cycles and varying content quality.

Global Information Management (GIM) solutions meanwhile are the only solutions that unify and automate fragmented information lifecycle processes across your global operations. Unifying and automating their GIM process in the Web environment, enables organizations to:

- Improve the quality of the customer experience on the Internet
- Protect and enhance brand quality and value
- Accelerate product and message time-to-market
- Realize increased revenues from synchronized shipping
- Communicate with relevance and consistency to local markets
- Address previously non cost-effective markets / customer groups
- Drive out unnecessary process costs and delays

So how do you begin creating a Global Information Management Strategy around your Web environment? A strategy which ensures your Web content is synchronized with content emerging from other channels, such as products and services themselves, the point-of-sale and printed collateral? This is the central theme we will explore in this SDL white paper.

3.0 Preamble: the parts and how they fit together

3.1 Content, structure, delivery and display

Whenever users access your website, the requested content is retrieved from where it is stored, delivered to their browser and displayed with the help of formatting tags, character encodings and fonts. There are many different types of content, storage structures, delivery protocols and display devices. Collectively, they determine the strategies and tactics necessary to create, manage and maintain a global website.

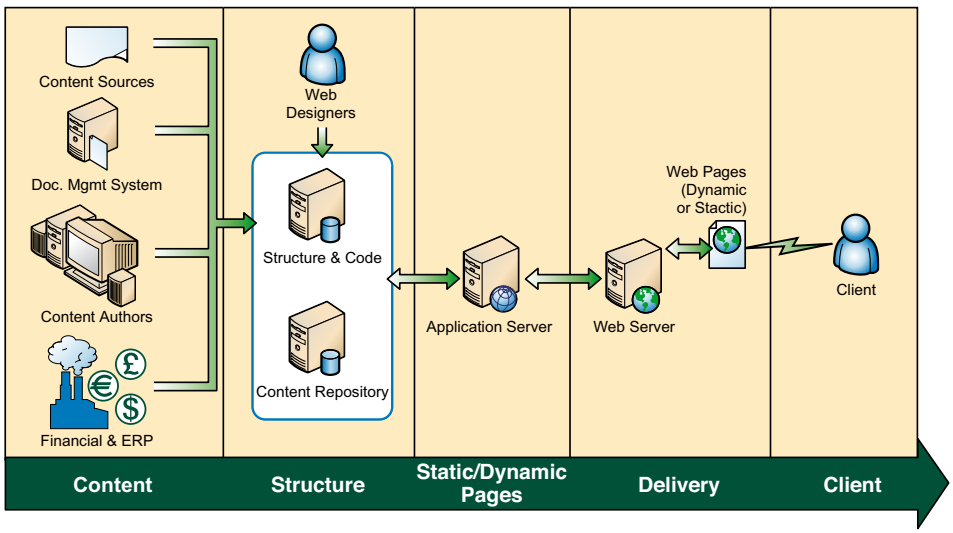

Figure 1: Content, structure, delivery and display

Content

Whatever flows from the server to a user's browser is generally referred to as output. This is delivered in the form of text, graphics, scripts, applets and multimedia; there are different file formats for each type of content. We will examine the content types in more detail in the localization section.

Structure

Content must be stored and managed. Most Web environments rely on databases and a combination of off-the-shelf and home-grown applications to manage content. A Content Management System (CMS) is often used to manage content, text and visual assets that will be used to generate Web pages. The CMS generally provides authoring and publishing functionality. The authoring element allows the content manager or author to manage the creation, modification and storage of chunks of content in a database. The publishing element uses and compiles that information to publish the content in one or more forms: Web, print, online help, etc. The features of CMS systems vary, but most include Web-based publishing, version- and revision control, indexing, search, and retrieval. All data within a CMS is indexed. Individuals use keywords to search for and retrieve data managed by the CMS.

Delivery

Basic websites are stored as static pages on the Web server; they are sent to the user in response to browser requests.

Most websites today, however, generate their pages with the help of application servers, such as Microsoft's IIS platform or Sun's J2EE, or equally popular open-source equivalents like Apache and PHP. These systems offer an execution platform for applications which essentially store and retrieve data from a database, generating 'dynamic' pages in response to browser requests by applying formatting to retrieved content.

These systems work interactively with complex databases, such as Oracle and SQL Server, as well as enterprise portal solutions, like IBM's WebSphere or Microsoft's SharePoint. This enables the delivery of sophisticated e-commerce sites, interactive technical support applications and extensive online catalogs, to name just a few possibilities. This paper will focus primarily on this type of dynamic, database-driven site.

Display

Most information on the Web is communicated in text format. How the text displays on a user's screen is controlled by two things: character set encoding and mark-up such as HTML tags. The character set encoding tells the browser which characters to display. There are many different character encoding schemes for different writing systems and computer platforms. The Unicode encoding scheme is the most universal and contains the important characters in all relevant business languages. Unicode is an international standard for characters and symbols (ISO-10646). It has two main character set encodings UTF-8 and UTF-16 (UTF=Unicode Transformation Format). UTF-8 uses one, or up to four bytes for the character encoding; whereas UTF-16 uses two or four bytes for the character encoding.

Tip: Unicode is the best choice for problem-free character encoding for current and future browsers, although it is not supported by older browsers and operating systems.

Official Name	Common Name	Languages Covered
ISO 8859-1	Latin-1	Western European
ISO 8859-2	Latin-2	Eastern European
ISO 8859-3	Latin-3	Southeastern European
ISO 8859-4	Latin-4	Northern European
ISO 8859-5	Cyrillic	English and Cyrillic
ISO 8859-6	Arabic	English and Arabic
ISO 8859-7	Greek	English and Greek
ISO 8859-8	Hebrew	English and Hebrew
ISO 8859-9	Latin-5	Western European and Turkish
ISO 8859-10	Latin-6	Scandinavian
JIS X0208		Japanese
SHIFT-JIS		Japanese
KS C5601		Korean
GB 2312	Chinese	Simplified Chinese (PRC)
BIG5		Traditional Chinese (Taiwan)
Unicode		Universal

Table 1: Common character sets

4.0 XML and other mark-up languages

Mark-up languages are used to place tags throughout the content. The browser interprets the tags to determine how to format the content on the user's screen. Tags may also contain other information, which does not directly affect how the content is displayed by the browser, but determines how the content is interpreted by other applications.

HyperText Markup Language (HTML) is a tagged-text formatting language that uses a set of pre-defined tags to describe the structural elements of a page, such as tables, rule lines, style sheets, text and graphics. The browser parses the HTML, which tells it how to format the Web page on the screen, fetches from the server any images or forms referenced in the HTML and displays the content accordingly.

Extensible Markup Language (XML) is a data format for structured document interchange on the Web. XML is not a replacement for HTML because each language has a different purpose: HTML is about formatting information; XML is about structuring information. At a day-to-day level, XML enables authors to write once, reuse sections of content and then apply these to the Web or any other channel. It minimizes the duplication of content – reducing time-to-market. It reduces localization costs by increasing productivity and accelerating Web content delivery. And it ensures consistent brand communication is maintained in the Web environment. At a technical level, XML's main strength is that it allows enterprises to define their own tags and content structures, to address their specific business logic and applications.

XML by itself does not solve the problem of Web content for global markets though. To effectively leverage XML for global markets, organizations need a repository that can remember connections between source chunks of Web content and translated chunks; they need a way to store translation assets at the sentence level and leverage them as required; and they also require a means of managing Web content terminology consistently.

Organizations that create global Websites and who are moving to an XML strategy must think globally as part of the process of implementing XML. They need to include global XML as part of their overall Global Information Management strategy. Global XML is the vision to store information in an XML language – and then replicate it in multiple languages across the Web or any other channels of communication.

TIP: Including global XML as part of an overall Global Information Management strategy enables an organization to store information in an XML language – and then replicate it in multiple languages across the global network of web sites.

Standard versions of HTML, XML and XHTML are published by the World Wide Web Consortium (W3C) at www.w3.org.

5.0 Business strategies for global websites

The best measure of success for a website implementation is its return on investment (ROI). Companies use different metrics to determine ROI, but three factors generally determine whether a website implementation is successful or not:

- Speed to deployment of content
- Quality of end-user experience in all locations (usability)
- Cost to implement and maintain the website

As sites grow in complexity and the number of target markets increases, successful sites will adopt the following strategies: centralize technology, centralize processes, enforce platform consistency and provide a process to promote local diversity.

5.1 Global versus local content

Many corporate websites consist of global content that is managed by head office and local content that is created and managed in-country. There is an inherent tension here. Depending on the degree of autonomy in the in-country offices, local sites may have quite a different look and feel than the corporate parts of the site. On the other hand, companies that emphasize a head-office approach may undermine the effectiveness of local sites. The key is to find the right balance between global and local content.

Company marketing strategy is the most important factor affecting how much to invest in local sites. What revenue potential and growth targets exist for the particular market? What will be the return on investment for money spent on website localization? The best practice here is to develop a business case for each specific market, to be able to quantify the return on investment for website development spending.

Such a business case is not always easy to develop though. Companies often base localization decisions on how vigorously the in-country people demand localization. Local sales and country managers may be very vocal about the positive impact that an increase in localized material would have on their sales and marketing efforts. Yet, a systematic approach based on real market research and independent evaluation will yield the most dependable results.

Local-only content

Some content may only be relevant to a particular locale, such as local sales promotions or country-specific programs. In addition, product availability and configuration may vary from region to region. For these reasons, relevant content will most likely be authored in-country and will not require translation or localization.

Translation versus localization versus globalization

Translation is the process of adapting meaning from one language into another. This is not a literal, word-for-word process. Rather, the translator must first understand the meaning communicated by the source language and then author words in the target language that convey the same meaning.

Localization is the process of adapting a product or service for a particular country or region. This does not only refer to translation. Dates, times, numbers and currencies must all be displayed appropriately (i.e. the site must be properly internationalized). Additional changes may also be required. For example, generic contact information must be replaced with local contacts for that particular country. Some kinds of general content that are valid everywhere may only need to be translated; whereas, content that is more local in nature must be localized for each locale.

Globalization – often confused with localization – involves the wider strategy of selling into global markets. It incorporates many complex business decisions that make corporations truly international in scope, such as purchasing product supplies around the world, dealing with trade regulations in the different countries in which they buy and sell and supporting customers around the world, in whichever language, country or culture they need.

5.2 Localization issues

Cultures

Locale and cultural issues become especially relevant for e-commerce sites. An e-commerce site, because it is conducting business transactions over the Internet, requires a high degree of interactivity with the user. Products are investigated, compared, ordered and paid for. Product specifications must appear in the proper units (Imperial versus Metric), prices must be quoted in the appropriate currencies and shipping charges and taxes must be calculated according to where the user lives.

Data protection issues

Most e-commerce transactions involve the exchange of personal data, such as credit card numbers, login user identifications and passwords. This highly sensitive data is protected by law in different ways in different countries.

In European Union countries, for example, there is a specific directive that regulates the movement of personal data across the national borders of the EU member countries and also sets a baseline of security around personal information wherever it is processed, stored or transmitted. Directive 95/46 of the European Parliament came into effect in 1998. Companies are subject to these regulations whenever they serve customers in the areas governed by these laws. One software company was fined $60,000 for improperly sharing data with its U.S. head office. The data had been gathered in Europe.

> **Tip: Make sure that you are aware of and comply with the data protection legislation that applies in your target countries.**

On a technical note, two different protocols are used to transmit sensitive data securely. Protect your customers' data with one of the following tools. Secure Sockets Layer (SSL) is a protocol for transmitting private information via the Internet. SSL uses a private key to encrypt data and creates a secure connection between the client and server. Any amount of data can be sent securely over this connection. Many websites use this protocol to obtain confidential user information, such as online banking services. URLs requiring an SSL connection start with https: instead of http:. Secure HTTP (S-HTTP) is another protocol used for safe data transfer. S-HTTP is designed to transmit individual messages securely, rather than unlimited amounts of data.

Localization workflow: centralized versus decentralized
Should all your content, no matter which language, be managed by head office, or should it be controlled by different locations around the world? Who decides which content should appear on individual country or language sites? Here are some things to consider:

The centralized model has head office manage all content for the local offices. This may work well for websites that are informational and global in nature. The localized portions are more language-dependent than country-dependent and the localized sites are usually just a translated subset of the English pages.

The decentralized model pushes all content responsibility into the local markets. This may work well for sites that provide mainly country-specific information and interaction. The content is maintained autonomously by company representatives in each country. The pages are authored in-country and are specific to that country (or language) and probably do not have localized equivalents on the parent site or on any of the other localized sites. For example, the events page might only contain local events and sales promotions relevant to the specific country or region.

A hybrid workflow model combines the best of a centralized and decentralized approach. Head office is responsible for issues such as the business rules, content guidelines, international style guide and general content. The individual markets determine which of the general content they need to have localized for their market, and they also author original content specific to their market. This approach will work best for companies that follow the best practices outlined in this paper. With a high degree of standardization and automation, they can better focus on the marketing needs of individual markets.

5.3 Depth of localization

In its simplest form, a localized website might have an identical copy of the English site in each of the target languages; however, this is rarely necessary. In order to get the best return on investment (ROI) for website localization, it may make more sense to localize just a subset of the site into specific target languages. Market analysis and revenue goals in each target market should determine how much localized content needs to be provided. Another important factor to consider is the cost to re-translate portions of the site that change frequently. A news page, for example, may change daily. Is it commercially viable to localize that content? Localizing a website may require a significant commitment of time and money; the strategic goals for each target market and content type must determine the depth of localization.

Content

Corporate information
The corporate mission, values, history and culture, as well as information about the company leaders, are important marketing tools. This information also tends to be fairly static, so it makes sense to localize as much as possible.

Products and Services
Product and service information is likely to be the most relevant material to your users. Helping users educate themselves about your products in their own language demonstrates your commitment to their market and can positively influence buying decisions. But be careful how you present your product; in many countries, overly aggressive or comparative marketing techniques are culturally inappropriate or even illegal.

Service and Support
Providing local service and support information, such as in-country contact points, demonstrates a commitment to the market. But how much of the technical support material can you afford to localize? Complex products may have extensive support literature that, in the past, would have been cost-prohibitive to translate. The traditional solution was to localize the main support pages, but not to localize downloadable PDF files, technical support articles and error-message databases. Today, however, new technologies – such as Knowledge-based Translation (KbT) – are reducing costs to a point where large-scale localization of technical support and online support materials has become a realistic alternative. It also provides excellent support to customers online, reducing the number of support calls that come into a call center.

Primary, secondary and tertiary languages

Another method to help standardize the depth of localization across multiple languages is to divide the languages into tiers. Each language in the group is localized to the same depth. This approach works well with centrally managed sites. Two to four language tiers are generally sufficient. An advantage of this method is that the company only needs to provide one set of source material to the localization vendor for each tier.

Source-only content

Content may remain in the source language and not be localized for several reasons. In the case of highly specialized technical information, for example, the content may be in English. People accessing this information may have sufficient English skills to understand the material.

There are other reasons not to localize that apply to all languages - the sheer volume of information may make it too costly to localize. Or the content may be so short-lived (news, for example) that it does not make sense, or the content may have been posted by the in-country team and is only relevant for that local market.

5.4 Access to localized pages

The way in which users navigate to and from localized content affects the quality of their experience. Here are some things to consider:

Country flag or language name?

Do users need to indicate a country preference or a language preference or both? On most websites today, users either click on a flag or on the name of their language (in their language) to indicate their language preference. This is not as straight forward as it seems.

Assume you are selling products into Switzerland, which has three official languages: French, German and Italian. The financial and tax implications are determined by the country, not the language; yet, you will want to communicate with your buyers in their language. Thus, your users must indicate both the country and the language.

Assume you provide online training in desktop productivity tools. Users throughout South America and Europe will want to take the course in Spanish. Here the language, not the particular country, is more important.

Another interesting aspect is the increasing mobility between countries. People may prefer to do business in their native language, regardless of which country they are currently living in – so the more flexible you can be in offering preferences, the better.

Three ways to access localized pages

Good: country-specific URL
If your business is more country-specific than language-specific you might publish distinct URLs for each country. The advantage is that users go directly to the relevant information and navigate in that content; the country site is self-contained. However, usability issues may arise if the localized site is not extensive enough and the user often links to non-localized content.

Better: link from the home page
The most common method is to assume that all users around the world will first arrive at the parent site, which is usually in English. Place a globe or map icon in the top right quadrant of the screen. Clicking on the icon should bring up a page where the user can select country and language.

Best: global gateway
The first time users access your site, they are presented with a global gateway page. After they select their country and language, a cookie containing that information is stored on their computers and they are directed to the localized pages. The next time they access the site, they go straight to the localized material without having to go through the gateway again.

> **Tip:** **Allow users to change their preferences at any time by returning to the home page and clicking on the globe.**

Handling links and navigation

Many sites pay great attention to the detail and drill-down of the navigational elements displayed to the user. However, in the case of a partially localized site, it is important to maintain the integrity of the site's look-and-feel by making any transitions between localized and English content as seamless as possible. Many sites fall at this barrier; the user is transferred with little or no warning from one language to another. The same applies to links. The ideal strategy here is to provide a localized experience that is as self-contained and complete as possible and to set proper user expectations about the extent of localization.

> **Tip:** **Warn users whenever they are about to switch languages.**

5.5 Platform consistency

Another factor that complicates the global-versus-local debate is consistency. In many cases, the Web pages developed in-country will have been created using different tools, platforms and processes; they may also be hosted on different servers. With each country left to its own devices, chaos and redundant effort can easily spread across website implementations. Company brand integrity is often the most obvious casualty. The best practice here is to centralize and optimize content management and translation workflow. By centralizing the infrastructure and standardizing the platforms, companies can achieve the best balance of globally consistent, properly managed content with locally relevant information, and, at the same time, reduce redundancies and hidden management costs.

The proliferation of increasingly mature Content Management Systems and Translation Management Systems allows companies to have both greater control and greater flexibility. Workflow automation shows the greatest promise in this area. Implementing a Web-based Translation Management System – such as SDL Translation Management System – enables companies to enforce a common platform and process throughout all levels and locations of the organization and still provide a high degree of local autonomy. By automating the translation management workflow, resources are freed up for more strategic utilization and expensive redundancies are eliminated.

6.0 Internationalize before you localize

Internationalization is the practice of creating source material in such a way that does not limit its use to one locale. Internationalization is usually a one-time, up-front investment. The advantage of starting with a properly internationalized website is that no matter how many localized versions you subsequently create, you can derive them all from one source. There are several important aspects to creating an internationalized website. Follow these design principles to save time and money when you are ready to adapt your site into other languages.

6.1 What seems intuitive may not be

Some languages read from left to right and speakers of such languages learn to process information from left to right. They tend to place things on a page to create emphasis based on how they process (read) the page. However, in a culture that reads right to left or top to bottom, what left-to-right readers take for granted will not have the same effect.

> **Tip:** **Move your navigation and action buttons to the right-hand side for right-to-left languages.**

There are many cultural norms – such as colors and gestures – that are taken for granted, but which often have very different meanings in other cultures. The following should be avoided in any graphics used: hand gestures and body parts, graphics with multiple meanings, religious or astrological symbols (stars, crosses, planets), shapes tied to a specific culture (stop signs, mail boxes, sports equipment). Stock photos of people should also be avoided, unless you are prepared to provide appropriate photos for each target market.

> **Tip:** **Use words, instead of graphics, if there is any doubt about the cultural meaning of a visual representation.**

6.2 Locale-specific information

Websites are usually developed using the conventions for the locale in which the developers live. Assumptions are made about how to display dates (day-month-year or month-day-year) in which calendar (Gregorian (Western), Hebrew, Japanese, Chinese, etc.), times (a.m. and p.m. or 24-hour clock), numbers (comma, space or period as thousands separator), currency (dollars, euros, yen, etc.) and fonts (names, sizes, etc.). A better way is to use standardized routines to supply this information in a format based on the system settings for the user's environment. That way no matter who accesses the site, the information will automatically appear in the proper format for that particular user.

> **Tip:** **Build pages using Cascading Style Sheets (CSS) so that localizers can change the fonts for all pages in one place.**

Websites may also contain other information whose format changes in other parts of the world: address formats (postal codes, states, number of lines, etc.), name formats (salutations, order of given name versus surname), telephone number formats (number of digits, country and area codes), units of measure (Imperial versus Metric) and paper sizes (letter/legal versus A4/A3).

6.3 Text and graphics issues

Text expansion

Text tends to expand when it is translated. The general rule is that large sections of text expand by about 30%, whereas single words and terms, depending on the language, can expand by as much as 400%.

> **Tip:** **Allow sufficient space, especially in buttons, graphics and tables, to absorb text expansion.**

Sort order

Sort order is not the same for all languages, particularly for languages that do not use the Western alphabet. In Swedish, for example, some extended characters (e.g. å) are sorted after the letter z. In many Asian cultures, characters are composed using a prescribed order of brushstrokes; characters are sorted by the brushstroke order. Also, after localization, the first letter of the word might change, which affects its position in the sort order list. Thus, it is difficult to sort automatically.

> **Tip: Ask your localizer to sort any alphabetized lists.**

Know who is accessing the site

Hundreds of millions of people use the Web every day; although only about one third of them are surfing in English. Some users may be using old versions of browsers and operating systems. If you overuse the latest features supported by the newest browser versions and operating systems, potential users may not be able to fully access your content. In addition, some users may prefer to access the Web using their mobile phone or personal data assistant. Graphics take longer to download than text: putting too many graphics or active features on your page may frustrate and alienate some users.

Another issue is how companies provide contact information. For example: an 800 number that is only valid in the U.S. will not be of any help to a visitor viewing your site from Japan. Another example here is in the use of web forms: often for instance 'State' is a mandatory field, or the ZIP needs to be a country specific format. Thus, it is important during the website design phase to make all such locale-dependent information available according to the user, rather than hard-coding it into the pages. Internationalize before you localize.

7.0 Localization strategies for global websites

7.1 Localizing text

All text localization should be managed with the help of Computer-Aided-Translation (CAT) applications. CAT tools, such as SDL Trados, provide human translators with a user-friendly environment in which to perform the translation. Among many other things, the tools protect the HTML and XML formatting tags, while allowing the translators to work freely within the translatable content. CAT tools are also known as Translation Memory (TM) tools. As content is translated for the first time, the tool captures the source segment and its translation into a translation memory database. The next time the same or a similar segment is encountered, the tool suggests the previously translated segment. The real benefit of these tools grows over time as the database of translated material expands. This is crucially important in scenarios where certain parts of the website are updated frequently: previously translated material is leveraged and translators can concentrate on new and modified text. This reduces the cost and time required to update translated content.

Global XML has a crucial role to play here too. As part of a global XML strategy, translators need to be able to visualize the context of XML chunks. Translating chunks of content is often harder than chapters because the content is often out of context. Translation memories and style guides help the translator get a fuller picture of the chunk meaning. Using appropriate online editing and review functionality, for example, terms can be highlighted automatically, ensuring that the translator and reviewer are always aware of previously translated terms. Additionally, by integrating content management into a Translation Management System, SDL can enable the preview of content, so that translators can see the wider context of chunks that have been sent for translation.

Another important tool for localizing text is a Web-based terminology manager, such as SDL MultiTerm. With the help of such systems, terminology issues can be effectively managed across the whole enterprise. The tool provides easy access to company-standard terminology for translators and business people alike, which enforces consistent use of key terms in the source language and also streamlines processes in all languages. A consistent message in every language helps promote brand integrity in every market. A further advantage is the seamless integration of terminology managers with CAT tools to improve localization quality and reduce cost.

For a dynamic website, the Web pages are created in real time when the user's browser interacts with the application. The software pulls chunks of content together and inserts formatting tags to create the pages that are

sent to the browser. These pieces of content, which reside in one or more databases, must be localized up front and stored in a database structure that provides content in the correct language upon request. Many components work together to facilitate this localization process. These structures can be difficult to manage manually. A better way is to implement a Translation Management System, such as SDL Translation Management System. It is fully integrated with SDL Tridion R5, the Web Content Management system of SDL Tridion.

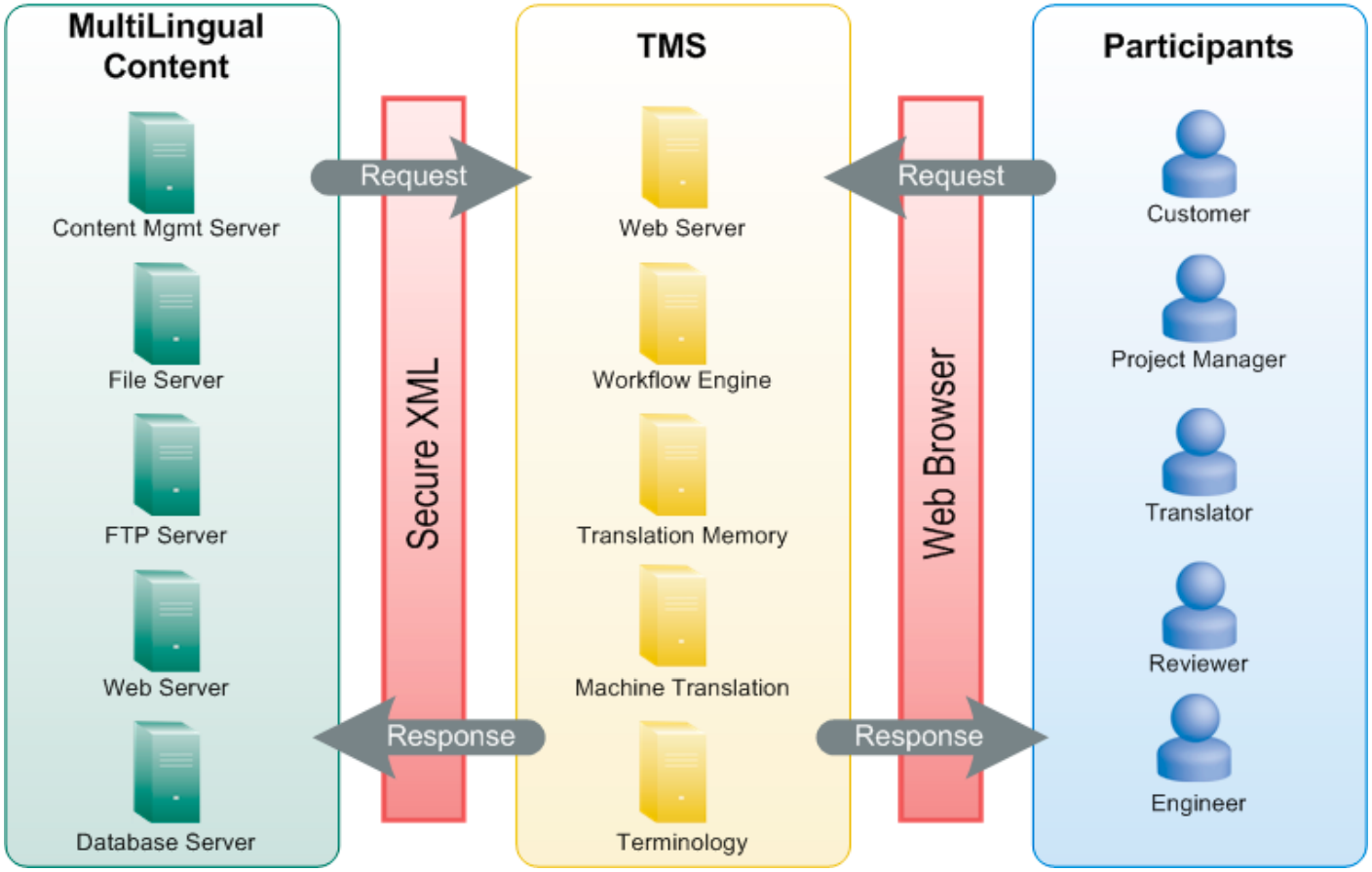

Figure 3 Managing complex localization processes with a Translation Management System

SDL Translation Management System ties all the human and machine components together through predefined workflows that automate key processes: extract changed content, prepare content, apply translation memory, update translation memory, include approved terminology and manage post-translation processing. Companies using these systems can implement consistent processes and platforms worldwide, thereby eliminating manual, redundant and time-consuming efforts. One of the most beneficial features is the centralized overview, a powerful management tool, which provides a snapshot of the current status of localization projects throughout the entire workflow process.

SDL Translation Management System, for example, unifies the translation and localization supply chain, providing the collaboration, control, integration and process flow required to prepare Web and other content for a global audience. The system is built on a modern, XML-based architecture that ensures content retains structure and style throughout the translation process. Data is transferred in and out of the translation management system via an advanced and secure XML-based messaging system. Enterprise content reuse strategies can also take advantage of the latest support for processing of DITA -compliant XML ('Darwin Information Typing Architecture' – a comprehensive framework for authoring, managing, and distributing topic-oriented information in XML).

TIP: Use the centralized overview to stay on top of all your localization projects.

7.2 Localizing graphics

Static Graphics

Many graphics used in website design are static; they are neither animated, nor layered. The most common formats currently in use are GIF and JPEG files. Some sites use graphics made up of many smaller images, alterations to which can adversely impact the overall appearance of the whole page. The most common problem associated with graphics is the use of text overlays, which are especially prevalent in banner advertisements and site navigation bars.

> **Tip: Always give the localization vendor original source graphics files that have the text in a separate, editable layer.**

Moving graphics

Moving graphics fall into two main categories: animated GIF files and files created using tools, such as Macromedia Flash or Java animations. Animated GIF files contain multiple frames, much like a movie reel, which are played in quick succession to give the appearance of animation. The localizer opens the files in an appropriate tool and determines the number of frames that need to be localized. This process is similar for Java animations and Macromedia Flash. To recreate animated GIF files, it is usually sufficient to have blank original images along with the text and effects used to create the original. The tools will duplicate the effect with the localized text.

GIF animation provides less control and flexibility than other animation formats, but is very popular because it is supported by nearly all Web browsers and the files are comparatively small. Macromedia Flash employs vector technology to created animated graphics. The resultant graphic is bandwidth friendly and browser independent. As long as the user's browser is equipped with the necessary plug-ins, Flash animations will run properly.

> **Tip: To localize Flash graphics you will need the source files(.fla), rather than the executable Flash file (.exe or .swf).**

7.3 Localizing active content

Content may be interactive, such as a form, or dynamic, such as Java applets, JavaScripts, streaming audio and video, and ActiveX applications. Where the translatable components of active content are embedded in program code, they must be extracted from the application, translated and reinserted. Customized tools are required to perform the extraction and reinsertion.

> **Tip: Move translatable content elements into external files wherever possible.**

An alternative is to create custom tags in the application that denotes translatable text. Extraction and reinsertion can then be automated. Following are descriptions of the most common types of active content.

Interactive, server-side content

A form is a formatted document containing editable fields that users can fill in with data. Electronic forms are especially common on the Web because the coding language has built-in codes for displaying form elements, such as text fields and check boxes. The data entered into a Web-based form is usually processed by response handlers particular to the application server technology being used by the server.

CGI is a specification for transferring information between a Web server and a CGI program. A CGI program runs on the Web server and is designed to accept and return data that conforms to the CGI specification. Programming languages, such as Visual Basic, Visual Studio.NET and C# (see www.microsoft.com), Java (see www.sun.com/java/) and Perl (www.perl.org), are used to create CGI applications.

An Active Server Page (ASP) is a dynamically created Web page that uses server-side scripting in the form of a Visual Studio.NET script or JavaScript code. When a browser requests an ASP, the Web server generates a page with code and sends it back to the browser.

> **Tip: When you localize forms, ensure your servers and any databases or applications that receive the form data can cope with the special language characters you might receive.**

Dynamic, client-side content

A more powerful way to provide a dynamic experience for Web users however, is to include scripts or programs that run on the user's computer, rather than on the Web server. These programs can be Java applets, JavaScripts, ActiveX controls, Python, Dynamic HTML (DHTML) pages or multimedia content.

An applet is a program designed to be executed from within another application. Web browsers can interpret applets from Web servers. Because applets are small in file size, cross-platform compatible and highly secure, they are ideal for small Internet applications accessible from a browser.

JavaScript helps Web authors design interactive sites. It shares many features and structures of the full Java language, but was developed independently. JavaScript interacts with HTML source code, allowing Web authors to spice up their sites with dynamic content. For example: you can use dynamic overlays created in JavaScript to enhance graphics. Text-based strings, embedded in JavaScript, are displayed over a background image. These images are usually localized as part of JavaScript identification and extraction. However, if the translated text expands too much, the underlying graphics may need to be resized.

An ActiveX control can be downloaded and executed by Internet Explorer running on Windows. An ActiveX control is similar to a Java applet except that ActiveX controls have full access to the Windows operating system, and are neither cross-browser nor cross-platform compliant. This introduces a certain risk that the applet may damage software or data on the user's computer. To control this risk, Microsoft developed a registration system so that browsers can identify and authenticate an ActiveX control before downloading it.

7.5 Localizing search engine keywords

When users enter keywords in search engine home pages, the keywords are compared to indexes and the appropriate links to Web pages are supplied as search results. Search engines use two different methods to create an index of the information contained in Web pages. One indexing method occurs automatically, the other requires human intervention.

In the automatic method, used by sites such as Google and Yahoo, a small software application called a spider or crawler accesses the Web page, indexes all the words on the page, then goes to any linked pages and indexes them. Since your localized pages will be linked to your home page, the crawlers will eventually find and index them as well. International users will find your localized pages when they search in a specific language without any extra intervention on your part. Many of the larger search engine companies, such as Google and Yahoo, also have country-specific sites.

> **Tip: Terminology management will help increase your search engine optimization (SEO) as you will be ensuring that the correct term is used more frequently throughout your site. So ensure your SEO terms are localized and used by translators or in-country authors.**

Human-powered directories, such as Open Directory and LookSmart are also available. You submit a description of the site to the particular directory or a review is submitted by an editor. Search results are taken from the submitted descriptions.

> **Tip: If you use a directory service, you will want to localize your page titles and site description into the appropriate languages and re-submit them to the directory.**

7.6 Localizing multimedia

Multimedia assets, such as movie and sound files, are localized using media-specific tools in exactly the same manner as if they were standalone assets. For example, an asset containing spoken-word elements would need to have the script translated, the audio portion re-recorded and the sound track replaced. Text elements would need to be translated and reinserted into the presentation. The localized assets are stored on the Web server, referenced on the web page and pulled in by the user's browser. Following are descriptions of the most common types of multimedia content.

Audio

MP3 is a popular format for storing compressed sound in files. RealAudio is a popular standard for streaming audio data over the Web that supports FM-stereo-quality sound. MIDI (Musical Instrument Digital Interface) is an older standard adopted by the electronic music industry for controlling devices that emit music, such as synthesizers and PC sound cards.

Video and Audio Combined

QuickTime is a video encoding algorithm for use on PC and Mac. Microsoft's Video for Windows stores video and audio information in files that use the .AVI extension.

RealVideo is a technology developed by RealNetworks for streaming video over the Internet.

MPEG stands for Moving Picture Experts Group, and refers to the family of digital video compression standards and file formats developed by the group.

8.0 Managing change

Once the initial localization is complete, a new set of challenges arises. One of the biggest advantages that websites bring to international business is the ease with which content can be changed and updated. However, this advantage can become a costly disadvantage if it is not managed properly. The number of languages you maintain on your website acts as a cost multiplier for any changes you make to the original site. Large companies today may have upwards of twenty other-language websites that have been localized to different depths. Here are some best practices to help manage these complex structures.

Content Management

Website content has various sources throughout the enterprise. Translations may already exist for some of this content, but may be scattered throughout the company in different output formats. Single-sourcing approaches based on a global XML strategy help create and maintain one set of source information from which each of the output formats is derived. The benefit of these approaches is that they can greatly reduce the number of new words that need to be translated for the different presentation media used by the organization. Implementing a Content Management System based on global XML may produce many cost and time savings that directly and indirectly reduce the cost of maintaining and updating a global website.

Translation Management

Websites that have been integrated into a Translation Management System, such as SDL Translation Management System, benefit from automated transaction processing. This virtually eliminates project management costs for incremental updates and is the most cost-effective approach when you need to update the sites on an ongoing basis. Translation Management Systems interact directly with Content Management Systems to reduce localization costs even more. For example, whenever changed content is checked back into the Content Management System, the Translation Management System is triggered; it extracts the content and submits it for translation according to predefined workflow rules. Project personnel can concentrate on managing the business side of localization rather than being swamped by all the overhead tasks required to push a myriad of small projects through the system.

Frequency of Updates

Website updates are more likely to be numerous small changes, rather than a single large project. When managed manually, project management costs may quickly exceed the actual translation costs.

> **Tip: Batch smaller updates into a single update project at regular intervals (weekly, for example), rather than manage each update individually.**

Content developers are often unaware of the cost implications of their changes; even a relatively 'minor' change can become costly when it needs to be cascaded through thirty languages.

> **Tip: Create a spreadsheet formula to approximate change costs and use with content developers to help combat gratuitous change requests.**

9.0 Conclusion

Global Information Management (GIM) unifies and automates fragmented information lifecycle processes across your global operations. The net result is an enhanced customer experience, improved brand consistency, faster time-to-market for products and services, and reduced costs. GIM applies as much to the creation, development and maintenance of an organization's Web domain as it does to any other channel – including customer support, point-of-sale, collateral, and packaging.

Clearly, the corporate website can be a cost-effective vehicle to reach a large, global audience. However, not all content needs to be available in all languages. Striking the right balance between global and local content shows your customers that you not only care about their market, but that you also know how to manage resources effectively.

There is a horde of considerations to evaluate as part of the corporate website development. Including global XML within an overall Global Information Management strategy enables an organization to store information in an XML format – and then replicate it in multiple languages across the global network of web sites. Starting with a properly internationalized website means that no matter how many localized versions you subsequently create, you can derive them all from one source. And then there are the inevitable localization issues to consider, such as the need to respect different cultures, data protection issues, centralized versus decentralized localization workflow and the depth of localization.

Best practices for managing global websites can be summed up in a few themes: centralize and standardize technology platforms; centralize, standardize and automate processes.

To find out more about how you can maximize the global effectiveness of your Web presence, please contact your local SDL office.

About SDL

SDL is the leader in Global Information Management (GIM) solutions that empower organizations to accelerate the delivery of high-quality multilingual content to global markets. Its enterprise software and services integrate with existing business systems to manage the delivery of global information from authoring to publication and throughout the distributed translation supply chain.

Global industry leaders rely on SDL to provide enterprise software or hosted services for their GIM processes, including ABN-Amro, Best Western, Bosch, Canon, Chrysler, CNH, Hewlett-Packard, Microsoft, Philips, SAP, Sony, SUN Microsystems and Virgin Atlantic.

SDL has implemented more than 400 enterprise GIM solutions, has deployed over 150,000 software licenses across the GIM ecosystem and provides access to on-demand translation portals for 10 million customers per month. Over 1,000 service professionals deliver consulting, implementation and language services through its global infrastructure of more than 50 offices in 30 countries.
For more information, visit www.sdl.com.