



## Standards and Guidelines for the Language Industry

## Table of Contents

About this document .....	5
Summary .....	7
Why are standards important to the language industry .....	8
Standards organizations .....	9
International Organization for Standardization .....	9
ISO/IEC JTC 1 – Information technology .....	9
ISO TC 37 – Terminology and other Language and Content Resources .....	10
ISO TC 171 – Document management applications .....	10
ISO TC46 – Information and Documentation .....	10
ISO Registration Authorities .....	10
Organization for the Advancement of Structured Information Standards .....	11
Localization Industry Standards Association .....	11
World Wide Web Consortium .....	12
Unicode Consortium .....	12
AIIM – Enterprise Content Management Association .....	13
Dublin Core Metadata Initiative (DCMI) .....	13
Internet Engineering Task Force (IETF) .....	13
Object Management Group (OMG) .....	13
Phases in the content creation workflow .....	14
Use of standards in the content creation work flow .....	15
Core standards .....	17
Markup languages and related technologies .....	17
Standard Generalized Markup Language (SGML) .....	17
Extensible Markup Language (XML) .....	17
Hypertext Markup Language (HTML) .....	18
Extensible Hypertext Markup Language (XHTML) .....	18
Extensible Stylesheet Language (XSL) .....	19
XML Schema .....	19
Cascading Style Sheets (CSS) .....	19
Locale Data Markup Language .....	20
Standards relating to character encoding and representation .....	20
UNICODE .....	20
Universal Character Set (UCS) .....	21
RUBY annotation .....	21
Text boundaries .....	22
Standards relating to metadata .....	22
Metadata Registries .....	22
Dublin Core .....	23
Internationalization Tag Set (ITS) .....	23
Standards relating to inventories of real-world entities .....	23
Common Locale Data Repository (CLDR) .....	23

Names of scripts .....	24
Names of languages (and countries).....	24
Quality standards.....	28
Quality management systems .....	29
Web Services.....	30
Content authoring and development .....	31
Software documentation.....	31
Software documentation process.....	32
Design and preparation of software documentation.....	32
Documentation guidelines for computer-based application systems .....	33
Management of software documentation .....	33
Darwin Information Typing Architecture (DITA) .....	34
DocBook.....	34
OpenDocument.....	34
Document language .....	35
Bidirectional text .....	35
XHTML and HTML internationalization.....	36
Business documents.....	36
Internationalization and localization markup requirements .....	37
Translation and localization.....	38
Project management .....	38
Translation contracts.....	38
Global Information Metrics eXchange (GMX).....	40
XML Localization Interchange File Format (XLIFF).....	40
OpenTag .....	43
Translation Web Services.....	43
Quality control.....	44
Quality metrics .....	44
Quality assurance.....	45
Guide for interpretation services .....	46
Computer assisted translation.....	46
Translation Memory eXchange (TMX).....	46
Segmentation Rules eXchange (SRX).....	47
XML:TM – Text Memory .....	48
Multilingual Information Framework.....	48
Terminology and lexicology .....	49
Terminology principles and methods .....	49
Harmonization of concepts and terms .....	50
Terminology of terminology .....	50
Terminography .....	51
Dictionary entries .....	51
Preparation of terminology standards.....	52
Alphabetical ordering .....	53
SGML encoding .....	53
Bibliographic references.....	54

Translation-oriented terminography .....	54
Terminology management systems .....	54
Data categories .....	55
Terminology project management.....	56
Terminology modeling .....	56
Quality assurance.....	57
Assessment of terminology resources .....	57
Concept modeling.....	58
Markup of lexical entries.....	58
XML representation and exchange of terminology data .....	59
XML representation of lexical and terminological data.....	60
Interchange of machine-readable lexicons .....	61
Terminology applied to business modelling.....	61
Content management.....	62
Resource Description Framework (RDF).....	63
Distributed authoring.....	63
Interoperable Enterprise Content Management (iECM) .....	63
Corpus Encoding Standard (CES) .....	64
Web Ontology Language (OWL).....	64
Java Specification Request (JSR).....	65
Generic Electronic Document Interchange (GEDI) .....	66
Records management.....	66

# About this document

Author: Kara Warburton

Dates:

- Version 1: March 27, 2006
- Version 2: February 18, 2007

This document describes international standards, guidelines, and resources that have some relevance to the language industry, specifically, to content authoring, content translation, terminology management, and content management. It was commissioned by the Language Technologies Research Center (LTRC) (<http://www.crtl.ca>) as part of the LTRC's mandate to act upon the recommendations of the Language Industry Association (AILIA) (<http://www.ailia.ca>) which were documented in its Language Technology Roadmaps. These roadmaps identified critical requirements for the language industry in Canada to achieve a competitive advantage and maximize global market share, and defined a technology development strategy to meet these requirements.

The first version of this report was produced in March 2006. It was updated in February 2007 after the author attended the Language Standards for Global Business summit in Vienna, Austria. The main purpose of that summit was to develop an inventory of standards as a source of material for the Wiki for Standards<sup>1</sup>. Most of the standards discussed at that summit were already described in this report, and in fact, summit organizers expressed an interest in the report as a source of material for the Wiki. Several additional standards were also discussed, and these have been added to the report.

It is clear that supporting international standards and recognized best practices is a key requirement for technology development and global competitiveness; this is further explained in the Introduction. This report is a preliminary study of the standards that might be important to the language industry in Canada. Since standards projects are constantly taking shape, this report does not claim to be exhaustive in its coverage; it is possible that some standards relevant to the language industry may not be described in this report at this stage of the project.

To address this and other issues, a more thorough study of the standards may be produced as a subsequent stage of this project. The next phase may include:

- a more focused assessment of the value of each standard to particular language sectors and stakeholders in Canada
- an assessment of which standards are likely to be adopted, in areas where there are competing initiatives
- positioning of specific standards in relation to the critical market-driving factors and situation of the language industry in Canada

---

<sup>1</sup> <http://www.wikiforstandards.org>

- a verification that no key standards are missing
- identification of standards development projects that Canada should participate in
- expansion of the scope to include the other areas within the mandate of AILIA and the LTRC: language training and speech technologies
- expansion of the scope to address knowledge management, content modelling, ontologies, and the semantic Web, all areas that have connections language industries, if the LTRC expands its activities into these areas identification of any language industry sectors where new standards are required.

## Summary

International standards are critical to global economic development in all industry sectors. The language industry is by no means an exception particularly given the massive changes that have taken place in information technologies in recent years. The language industry faces formidable challenges of increasing demands for larger and larger quantities of information, in more and more languages, all electronically interchangeable and easily retrievable. Technology is widely regarded as key to meeting these challenges. The technology must be based on standards in order to be adopted on a global scale. And national players such as Canada's language services, educational institutions, and technology suppliers need to support international standards for process and quality to compete in global markets.

This document describes some standards that could be important to the language industry, specifically in the areas of content development (authoring), content management, translation and terminology. It focuses on the key relevant standards in these areas and does not claim to be an exhaustive study of all language-related standards. With a few exceptions, standards dating prior to 2000 were not included because their currency was considered questionable, particularly for technical standards. In addition, some of the resources described are not actually formal standards, but may refer to internationally-recognized guidelines and best practices, or resources that are required to implement standards. Whenever possible, it is indicated whether the reference refers to a standard, a guideline, or another type of resource. A section describing the key standards organizations for the language industry is also provided. These organizations need to be monitored for new standards projects.

Several standards in the areas of content modelling, business modelling, and Web services have been added as there are relationships between these standards and other fields of the language industry such as terminology. However, these areas have not been covered completely as they are outside of the scope of this project.

The following information is provided for each resource described.

<b>Document title:</b>	The published or official title of the resource
<b>Type of document:</b>	The type of document (standard, guideline, or resource) and in the case of a standard, possibly the stage of standardization (proposal, draft, published standard, under review, etc.)
<b>Organization:</b>	The sponsoring organization
<b>Year:</b>	The year of publication or last update.
<b>Description:</b>	A short description of the standard: its scope, primary purpose, general benefits.
<b>Work flow stage:</b>	The work flow stage in the content development process to which the resource mostly applies: authoring, project management, translation, archival. This information is used as input to a language industry work flow flow chart.
<b>Business value:</b>	The value of the standard to the language industries.

**Web site:**

The primary Web page for further information, current as of the date of this publication.

## Why are standards important to the language industry

The rapid development of information technology in recent years has led to revolutionary changes in the way we create, translate, store, and distribute information. The expansion of global markets and the wide adoption of information technology has led to escalating demands for more and more information in more and more languages, on demand. The Technology Roadmaps produced by the Canadian Language Industry Association (AILIA) present market research that indicates that in several years the translation industry will not be able to meet the market demands for translations. The content authoring industry faces similar challenges and difficulties in finding qualified personnel. Many organizations are turning to technology to find ways to automate processes and increase the efficiency of employees.

In today's global economy, we cannot afford to operate in silos. Technology that works in Canada but nowhere else because we have not implemented international standards will put us at a competitive disadvantage. Standards protect manufacturers and technology developers by ensuring that their products and research will be usable worldwide. Furthermore, standards save organizations time and money by avoiding duplication of work and preventing redesign efforts.

Standards are not just important for manufacturing and technology however. Without business quality controls that are objective and trusted, there would be no means to measure competition. Standards and guidelines can also apply to non-tangibles such as processes, information content, and workflows. These types of standards help to guarantee quality of services for consumers.

International standards such as those produced by International Organization for Standardization (ISO) are very useful. “They are useful to industrial and business organizations of all types, to governments and other regulatory bodies, to trade officials, to conformity assessment professionals, to suppliers and customers of products and services in both public and private sectors, and, ultimately, to people in general in their roles as consumers and end users.”<sup>2</sup>

Today standards are produced by a wide range of organizations in addition to ISO. Organizations such as the World Wide Web Consortium and the UNICODE Consortium play a key role in standards for the World Wide Web. As business processes are increasingly moving to a Web model, these standards are becoming essential for any business wishing to compete in tomorrow's markets.

In addition to official standards, there are a number of internationally-recognized unofficial resources such as guidelines and recommendations that can really benefit Canada's language industry.

The key standards, guidelines, and other global resources can assist Canada's language industry to grow

---

<sup>2</sup> [www.iso.org](http://www.iso.org)

and remain competitive are described in this document.

## Standards organizations

The standards organizations that produce the language-related guidelines and standards described in this report are presented in this section. For information about additional standards organizations, consult the World Standards Services Network (WSSN) at the following Web site:

<http://www.wssn.net/WSSN/index.html>

### **International Organization for Standardization**<http://www.iso.org/>

<http://www.iso.org>

The International Organization for Standardization (ISO) is the world's largest producer of standards. ISO is an association of national standards institutes from over 150 countries. ISO started first as the International Electrotechnical Commission (IEC) in 1906, and its early focus was in the areas of electrotechnology and engineering. In 1947 its name changed to ISO and its scope expanded to include all industrial standards. Since 1947 it has published more than 15,000 international standards. One of the most widely known standard is ISO 9000 which establishes a framework for quality management of business processes.

The ISO committees that produce standards in areas that have impacts on the language industry are described in the following sections.

### **ISO/IEC JTC 1 – Information technology**

This joint technical committee of ISO and the International Electrotechnical Commission develops, maintains, promotes and facilitates Information and Communication Technologies (ICT) standards required by global markets meeting business and user requirements concerning the following areas<sup>3</sup>:

- design and development of IT systems and tools
- performance and quality of IT products and systems
- security of IT systems and information
- portability of application programs
- interoperability of IT products and systems
- unified tools and environments
- harmonized IT vocabulary
- user friendly and ergonomically designed user interfaces

In addition to the nearly 70 national body members, approximately 2100 experts from around the world contribute to the work of the committee. JTC 1 is composed of 18 subcommittees, the following three being the principal ones for language technologies:

- SC2 – Coded Character Sets - <http://std.dkuug.dk/jtc1/sc2/>
- SC7 – Software and system engineering - <http://www.jtc1-sc7.org/>

---

<sup>3</sup> Committee mission statement

- SC32 – Data management and interchange - <http://jtc1sc32.org/> and <http://metadata-standards.org/>
- SC34 – Document Description and Processing Languages - <http://www.jtc1sc34.org/>

## **ISO TC 37 – Terminology and other Language and Content Resources**

<http://www.iso.org/iso/en/stdsdevelopment/tc/tclist/TechnicalCommitteeDetailPage.TechnicalCommitteeDetail?COMMID=1459>

The scope of this ISO technical committee is the “standardization of principles, methods and applications relating to terminology and other language and content resources in the contexts of multilingual communication and cultural diversity”<sup>4</sup>. It consists of four subcommittees:

1. Principles and methods
2. Terminographical and lexicographical working methods
3. Systems to manage terminology, knowledge and content
4. Language resource management - <http://www.tc37sc4.org/>

For information about ISO publications, visit the ISO Catalog: <http://www.iso.org/iso/en/prods-services/ISOstore/store.html>

## **ISO TC 171 – Document management applications**

This ISO technical committee is responsible for the standardization of quality control and integrity maintenance in the field of document management.<sup>5</sup> Among the many more traditional aspects of document management such as indexing, storage, retrieval, and so forth, the committee is also involved in defining standards around document management work flow, including Web-based work flow.

## **ISO TC46 – Information and Documentation**

Standards relating to libraries, documentation and information centers, publishing, archives, records management, museum documentation, indexing and abstracting services, and information science. Working group 4 is responsible for terminology of information and documentation.

## **ISO Registration Authorities**

Some of the ISO technical committees have appointed a respected organization to act as a registration authority, to ensure that the standards they produce are implemented effectively and correctly. The following is a partial list of ISO registration authorities for standards in the language industry.

---

<sup>4</sup> Committee scope statement

<sup>5</sup> Scope statement of the committee.

- ISO Central Secretariat
- International Information Center for Terminology (INFOTERM)
- The Library of Congress
- British Standards Institution
- American National Standards Institute
- Unicode Consortium

For more information about ISO registration authorities visit the following Web site:

<http://www.iso.org/iso/en/catsupport/maralist.html>

## ***Organization for the Advancement of Structured Information Standards***

<http://www.oasis-open.org>

The Organization for the Advancement of Structured Information Standards (OASIS) is a non-profit international member-based organization that focuses on the development, harmonization, and promotion of e-business standards. It produces standards for Web services in areas such as security, e-business, public sector management, and specific application markets.

The history of OASIS goes back to 1993 when it started as a consortium of vendors and users dedicated to supporting the adoption of SGML. Today, OASIS has more than 5,000 participants representing more than 600 organizations and individual members in approximately 100 countries.

OASIS plays a key role in the development of standards and guidelines in the following areas related to language technologies:

- Topic-based authoring
- Exchange of files for localization
- Computer documentation
- Business file formats
- Translation Web services

OASIS also hosts two major information portals on XML and Web services standards:

- Cover Pages: <http://xml.coverpages.org/>
- XML: <http://www.xml.org/>

## ***Localization Industry Standards Association***

<http://www.lisa.org>

The Localization Industry Standards Association (LISA) is an international membership-based

association for organizations that conduct business globally. It produces best practices and guidelines in international business and language-related procedures, and hosts frequent conferences all over the world. It has approximately 500 members in the public and private sectors, of which the largest industry represented is the translation and localization industry. LISA plays a key role in the development of guidelines and standards in translation, localization, and terminology.

## **World Wide Web Consortium**

<http://www.w3.org/>

The World Wide Web Consortium (W3C) is an international consortium of individuals and businesses that develops Web standards. Its mission is “to lead the World Wide Web to its full potential by developing protocols and guidelines that ensure long-term growth for the Web.”<sup>6</sup> In order for the Web to reach its full potential, Web technologies must be compatible with one another. In addition to its membership, the W3C has a full-time staff and collaborates with researchers, allowing it to develop standards-compliant software and deliver education programs.

The W3C has produced more than ninety standards and guidelines since its creation in 1994. W3C standards are called “W3C Recommendations.” It also hosts a number of international e-mail forums for implementers and users of standards around the world to exchange ideas, contribute to standards development, and provide peer support.

The W3C produces recommendations in the following language-related areas:

- markup languages
- Web accessibility
- Web services
- Web search

## **Unicode Consortium**

<http://www.unicode.org/>

The Unicode Consortium is a non-profit organization that developed the Unicode character set standard. Membership comprises a wide range of stakeholders from software companies to research institutions.

The Unicode Consortium is responsible for extending this standard to meet the constantly changing needs of the information technology industry, and for promoting its use world-wide. It also publishes guidelines for Unicode implementation, and an international reference for names of languages, countries, currencies, and other locale information.

---

<sup>6</sup> <http://www.w3.org/Consortium/>



## ***AIIM – Enterprise Content Management Association***

<http://www.aiim.org/>

AIIM is the global association for users and suppliers of Enterprise Content Management (ECM) solutions. It is the recognized international authority on ECM. ECM comprises the strategies, services and technologies that enable organizations to capture, manage, store, maintain, and deliver information to support business processes.

AIIM is also a standards development organization. It acts as the secretariat for ISO TC171 which is focused on Information Management Compliance issues. Over 80 of AIIM's standards, recommended practices, and technical reports, have been approved by ANSI.

## ***Dublin Core Metadata Initiative (DCMI)***

<http://dublincore.org/>

The Dublin Core Metadata Initiative is an organization that encourages the adoption of interoperable metadata standards and develops metadata vocabularies for describing specialized resources. Metadata standards are important to facilitate development in the areas of knowledge discovery and other intelligence systems. It also maintains a registry of existing metadata definitions for specific industries.

## ***Internet Engineering Task Force (IETF)***

<http://www.ietf.org/>

The Internet Engineering Task Force ([IETF](#)) is a large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. The following is the IETF mission statement.<sup>7</sup>

“The mission of the IETF is to produce high quality, relevant technical and engineering documents that influence the way people design, use, and manage the Internet in such a way as to make the Internet work better. These documents include protocol standards, best current practices, and informational documents of various kinds.”

## ***Object Management Group (OMG)***

<http://www.omg.org/>

OMG is an international computer industry consortium. OMG Task Forces develop enterprise integration standards for a wide range of technologies and industries. OMG's modeling standards,

---

<sup>7</sup> <http://www.ietf.org/rfc/rfc3935.txt>

including the Unified Modelling Language (UML), enable design, execution and maintenance of software and other processes. OMG's middleware standards and profiles are based on the Common Object Request Broker Architecture (CORBA®) and support a wide variety of industries.

## Phases in the content creation workflow

The following phases have been identified in the end-to-end content creation work flow. For each standard, the phases to which the standard applies is indicated by the numbers in the Work flow stage row of the tables.

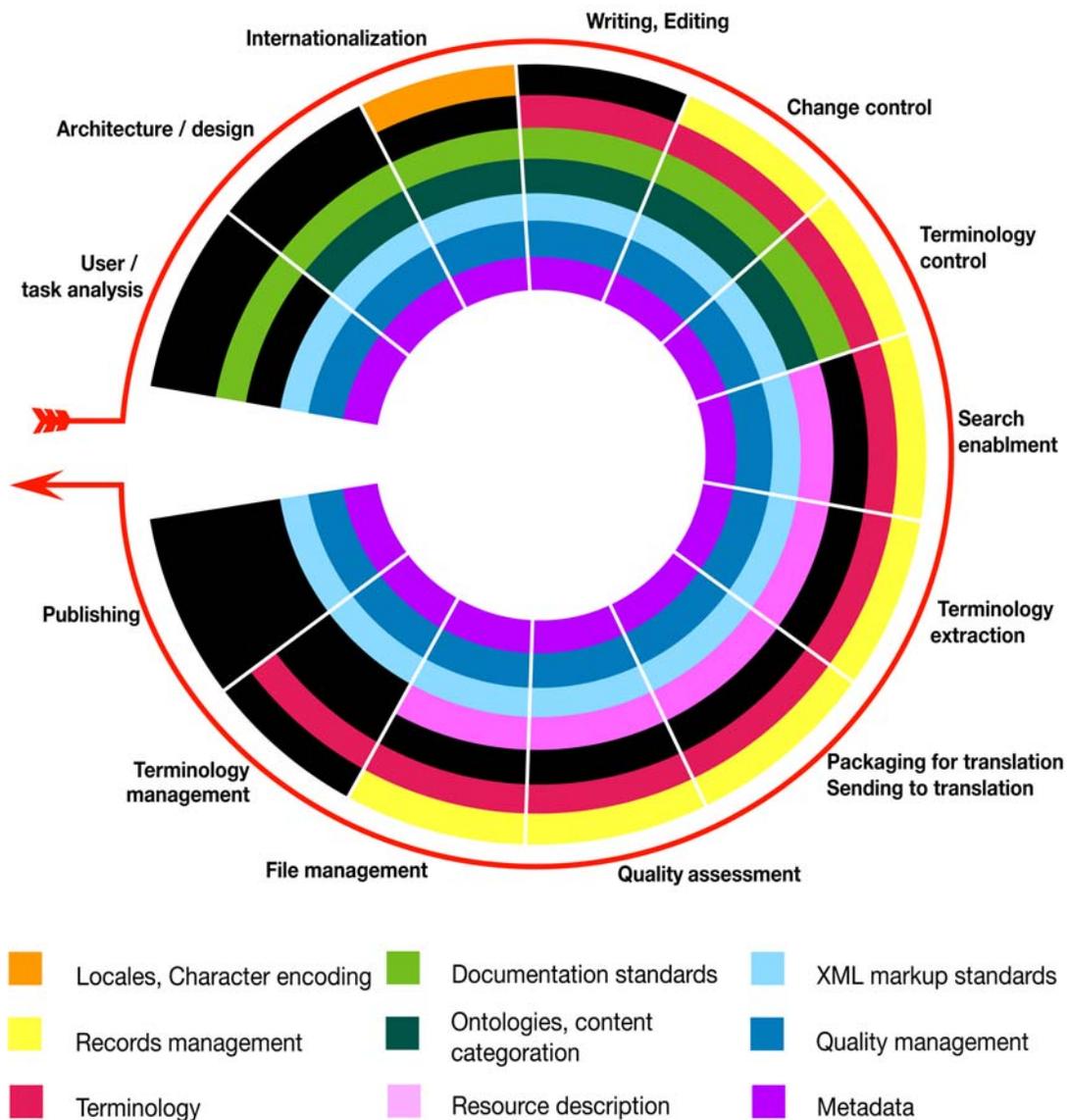
1. Content creation
  1. User analysis / task analysis
  2. Architecture / design
  3. Internationalization
  4. Writing
  5. Editing
  6. Version control / change detection
  7. Terminology control
  8. Enablement for search
  9. Terminology extraction
  10. Packaging for translation
  11. Sending to translation
  12. Quality assessment
  13. File management and archiving
  14. Terminology archiving
  15. Publishing
2. Translation
  1. File preparation
  2. Costing
  3. Text extraction
  4. Terminology preparation
  5. Project distribution
  6. Translation
  7. Quality management
  8. Verification / testing
  9. Delivery
  10. Terminology archiving
  11. Memory archiving
  12. Billing

## **Content management**

### **Use of standards in the content creation work flow**

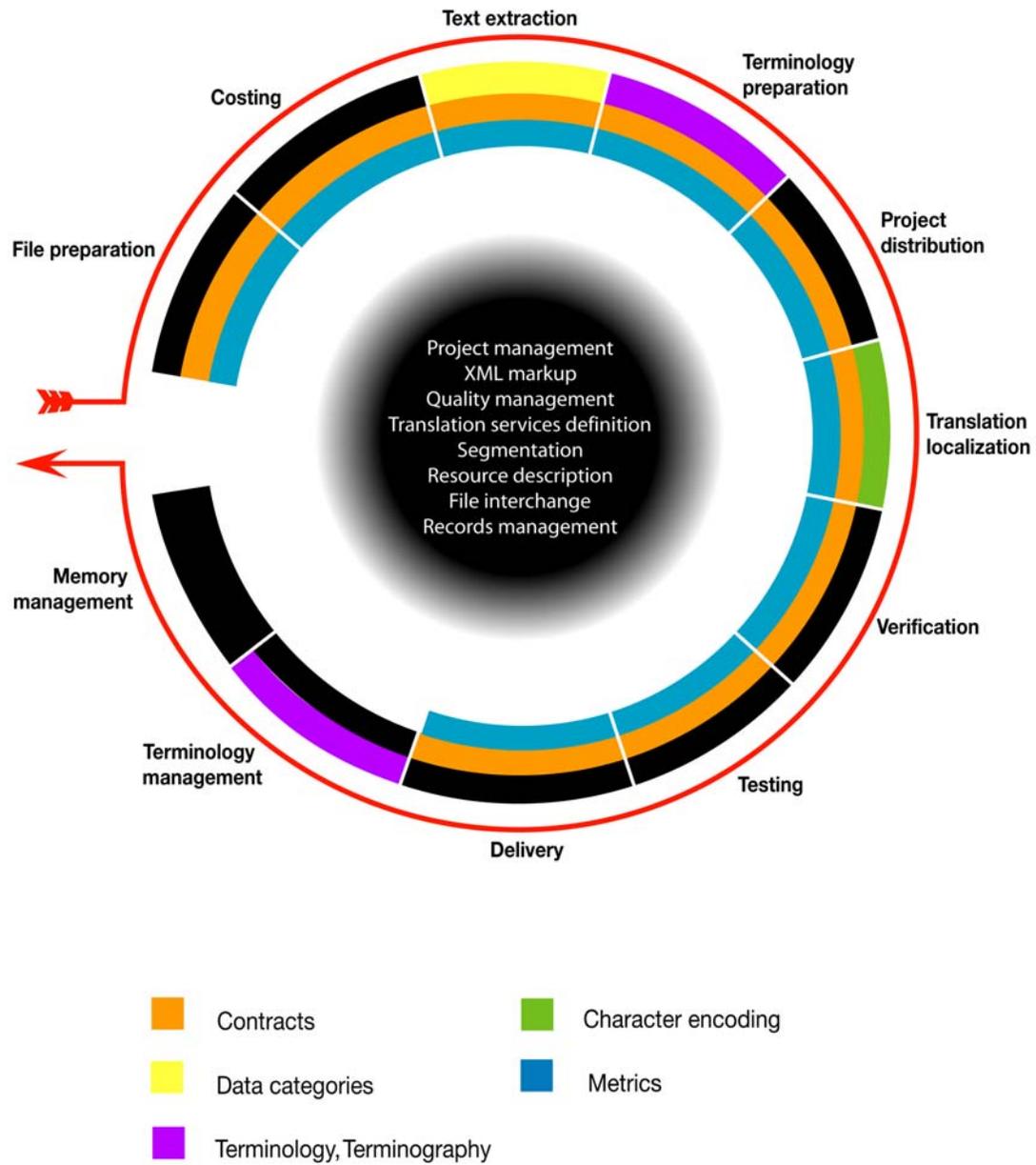
The standards described in this document were then categorized into broad categories represented by the colored bands in the following two flow charts to visually represent which types of standards apply to each stage in the content creation process. For example, standards relating to quality obviously apply to all stages of content creation, whereas resource description standards primarily apply to phases after content is initially created. This categorization and phase assignment is preliminary and requires validation through a more detailed study of the standards and their application.

## CONTENT AUTHORING



In the next flow chart, depicting the application of standards to the translation process, it was found that many standards apply to the entire process rather than to specific stages. To simplify the chart, these standards were simply listed in the black central area.

# TRANSLATION



## Core standards

Standards in this section apply to many stages and activities in content development, and can be considered applicable to the language industry in general.

### *Markup languages and related technologies*

#### Standard Generalized Markup Language (SGML)

<b>Document title:</b>	<b>ISO 8879:1986 – Information processing – Text and office systems - Standard Generalized Markup Language (SGML)</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO/IEC JTC1
<b>Year:</b>	1986
<b>Description:</b>	A standard format for embedding descriptive markup within a document, and a standard method for describing the structure of a document. SGML is a metalanguage for defining markup languages for documents.
<b>Work flow stage:</b>	1.2, 1.3, 1.8, 1.9
<b>Business value:</b>	By putting documents into SGML, organizations protect their information investment from being locked into a single vendor's technology or methodology. By tagging data with its role and any other useful identifiers, SGML allows information to be readily located and re-used. Being supported by many open source applications, businesses are not locked into a specific vendor's format or tool and are free to choose from a wider range of tools to create, manage or retrieve documents and the data in documents.
<b>Web site:</b>	<a href="http://xml.coverpages.org/sgml.html">http://xml.coverpages.org/sgml.html</a>

#### Extensible Markup Language (XML)

<b>Document title:</b>	<b>XML - Extensible Markup Language</b>
<b>Type of document:</b>	Recommendation
<b>Organization:</b>	W3C
<b>Year:</b>	2006
<b>Description:</b>	Extensible Markup Language (XML) is a simple, flexible text format derived from SGML (ISO 8879). Originally designed to meet the needs of large-scale electronic publishing, XML is now being increasingly used to exchange a wide variety of data on the Web and elsewhere.
<b>Work flow stage:</b>	1.2, 1.3, 1.8, 1.9

<b>Business value:</b>	The aim of XML was to achieve the benefits of SGML in a Web environment, to eliminate some of the inherent problems of HTML and extend the powers of the Internet. The use of XML is critical for any organization wishing to operate over the Internet.
<b>Web site:</b>	<a href="http://www.w3.org/XML/">http://www.w3.org/XML/</a>

### Hypertext Markup Language (HTML)

<b>Document title:</b>	<b>HTML – Hypertext Markup Language</b>
<b>Type of document:</b>	Recommendation
<b>Organization:</b>	W3C
<b>Year:</b>	1999
<b>Description:</b>	HTML is the standard markup language for publishing content on the World Wide Web. Based on SGML, HTML can be processed by a wide range of tools. The HTML Working Group has been established to modify HTML into an XML-based language, make it easier to modularize with other markup languages, and correct some problems in areas such as internationalization, accessibility, device independence and forms processing.
<b>Work flow stage:</b>	1.2
<b>Business value:</b>	HTML is a necessary standard for any organization that has a Web site.
<b>Web site:</b>	<a href="http://www.w3.org/MarkUp/">http://www.w3.org/MarkUp/</a>

### Extensible Hypertext Markup Language (XHTML)

<b>Document title:</b>	<b>XHTML – Extensible Hypertext Markup Language</b>
<b>Type of document:</b>	Recommendation
<b>Organization:</b>	W3C
<b>Year:</b>	2002
<b>Description:</b>	XHTML is a family of document types and modules that reproduce, subset, and extend HTML. XHTML document types are based on XML.
<b>Work flow stage:</b>	1.2
<b>Business value:</b>	Converting HTML documents to XHTML allows organizations to view, edit, and validate the documents with standard XML tools. XHTML documents can utilize applications that are designed for either HTML or XML. XHTML documents are more likely to interoperate with existing and future XHTML browsers and environments.
<b>Web site:</b>	<a href="http://www.w3.org/TR/xhtml1/">http://www.w3.org/TR/xhtml1/</a>

## Extensible Stylesheet Language (XSL)

<b>Document title:</b>	<b>XSL - Extensible Stylesheet Language</b>
<b>Type of document:</b>	Recommendation
<b>Organization:</b>	W3C
<b>Year:</b>	2001
<b>Description:</b>	A language for representing stylesheets. It consists of two parts: a language for transforming XML documents (XSLT) and an XML vocabulary for specifying stylesheet formatting elements.
<b>Work flow stage:</b>	1.2
<b>Business value:</b>	XSL enables organizations to manipulate arbitrarily structured XML documents for any type of layout, and for any type of output medium, from browsers and hand-held devices to catalogs and books.
<b>Web site:</b>	<a href="http://www.w3.org/TR/xsl/">http://www.w3.org/TR/xsl/</a> <a href="http://www.w3.org/TR/xslt">http://www.w3.org/TR/xslt</a>

## XML Schema

<b>Document title:</b>	<b>XML Schema</b>
<b>Type of document:</b>	Recommendation
<b>Organization:</b>	W3C
<b>Year:</b>	2004
<b>Description:</b>	XML schemas provide a standard way to define the structure, content, and semantics of XML documents, thereby making it easier to share content and enable machine interpretation of content. XML Schema definition language provides the means to describe the structure and add constraints to XML documents, using namespaces.
<b>Work flow stage:</b>	1.2, 1.3, 1.8, 1.9
<b>Business value:</b>	XML schemas increase capabilities of XML document type definitions (DTDs) to be used for Internet-based business processes such as for e-commerce. The standard includes schemas for some common e-commerce transactions such as invoicing and purchase orders.
<b>Web site:</b>	<a href="http://www.w3.org/XML/Schema">http://www.w3.org/XML/Schema</a>

## Cascading Style Sheets (CSS)

<b>Document title:</b>	<b>CSS - Cascading Style Sheets 2.1</b>
------------------------	---

<b>Type of document:</b>	Working Draft
<b>Organization:</b>	W3C
<b>Year:</b>	2005
<b>Description:</b>	A simple language for adding style (fonts, colours, spacing, and so forth) to HTML and XML documents.
<b>Work flow stage:</b>	1.2
<b>Business value:</b>	Simplifies Web authoring and site maintenance by separating presentation style from content.
<b>Web site:</b>	<a href="http://www.w3.org/Style/CSS/">http://www.w3.org/Style/CSS/</a>

### Locale Data Markup Language

<b>Document title:</b>	<b>LDML: Locale Data Markup Language (V1.4)</b>
<b>Type of document:</b>	Technical standard
<b>Organization:</b>	Unicode
<b>Year:</b>	2006
<b>Description:</b>	<p>A locale is an identifier (id) that refers to a set of user preferences that tend to be shared amongst a specific set of users who share socio-cultural, linguistic, or geographic properties. The locale identifier provides access to a set of definitions for formatting and parsing of dates, times, numbers, and currencies; for measurement units, for sort-order (collation), plus translated names for timezones, languages, countries, and scripts. The locale definitions can also include text boundaries (character, word, line, and sentence), text transformations (including transliterations), and support for other services.</p> <p>This document describes an XML format for the exchange of structured locale data. This format is used in the Common Locale Data Repository maintained by the Unicode Consortium. LDML can also be used to create a custom locale for a minority language.</p>
<b>Work flow stage:</b>	1.3, 1.8, 2.8
<b>Business value:</b>	LDML facilitates the specification of computer environments for all the language communities of the world. It therefore supports the language industry as a base technology to enable the provision of a wide range of services and data.
<b>Web site:</b>	<a href="http://unicode.org/reports/tr35/">http://unicode.org/reports/tr35/</a>

## ***Standards relating to character encoding and representation***

### **UNICODE**

<b>Document title:</b>	<b>UNICODE</b>
<b>Type of document:</b>	Standard
<b>Organization:</b>	Unicode Consortium
<b>Year:</b>	2003
<b>Description:</b>	Unicode is the world standard for character data. It provides a unique machine-readable code identifier for all characters in all the languages of the world.
<b>Work flow stage:</b>	1.2, 1.3
<b>Business value:</b>	Applications must support Unicode to ensure interoperability between systems, environments, software, and languages. Incorporating Unicode offers significant cost savings over the use of legacy character sets. It enables a single software product or a single Web site to be used for multiple platforms, languages and countries without re-engineering. It allows data to be transported through many different systems without corruption or loss.
<b>Web site:</b>	<a href="http://www.unicode.org/">http://www.unicode.org/</a>

### **Universal Character Set (UCS)**

<b>Document title:</b>	<b>ISO/IEC 10646:2003 – Information technology – Universal Multiple-Octet Coded Character Set (UCS)</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO/IEC JTC 1
<b>Year:</b>	2003
<b>Description:</b>	Defines the Universal Character Set (UCS) as a character encoding. ISO 10646 and Unicode define the same characters with the same numbers, however, Unicode adds additional constraints and rules. ISO 10646 is a simple character map, an extension of previous standards like ISO 8859.
<b>Work flow stage:</b>	1.2, 1.3
<b>Business value:</b>	The official way to implement this standard is to use Unicode, since Unicode imposes additional constraints that are not part of ISO 10646.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### **RUBY annotation**

<b>Document title:</b>	<b>RUBY</b>
------------------------	-------------

<b>Type of document:</b>	W3C Recommendation
<b>Organization:</b>	W3C
<b>Year:</b>	2001
<b>Description:</b>	RUBY annotation is a markup format for inline annotation of text.  <i>Ruby</i> is the term used for a run of text that is associated with another run of text, referred to as the <i>base text</i> . Ruby text is used to provide a short annotation of the associated base text. It is most often used to provide a <i>reading</i> (pronunciation guide). Ruby annotations are used frequently in Japan and in China to indicate pronunciation and sometimes to indicate the meaning of a word.
<b>Work flow stage:</b>	1.3, 1.4
<b>Business value:</b>	Ruby text, encoded through RUBY annotation, is important for the language publishing industry for Chinese and Japanese. It is also used to provide interlinear translations of obscure languages, such as English translations of the original Bible. RUBY annotation is part of XHTML 1.1.
<b>Web site:</b>	<a href="http://www.w3.org/TR/ruby/">http://www.w3.org/TR/ruby/</a>

## Text boundaries

<b>Document title:</b>	<b>Unicode Standard Annex #29: Text boundaries (5.0.0)</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	Unicode Consortium
<b>Year:</b>	2006
<b>Description:</b>	This standard describes guidelines for determining default boundaries between certain significant text elements: grapheme clusters (user-perceived characters), words, and sentences.
<b>Work flow stage:</b>	1.9, 2.3.
<b>Business value:</b>	This standard could be relevant to a word segmentation standard which does not yet exist.
<b>Web site:</b>	<a href="http://unicode.org/reports/tr29/">http://unicode.org/reports/tr29/</a>

## Standards relating to metadata

### Metadata Registries

<b>Document title:</b>	<b>ISO 11179 - Metadata Registries (MDR)</b>
------------------------	--

<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO/IEC JTC 1
<b>Year:</b>	2005
<b>Description:</b>	A multi-part family of standards that specifies basic aspects of metadata element composition and definition.
<b>Work flow stage:</b>	1.2
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Dublin Core

<b>Document title:</b>	<b>ISO 15836 - Information and documentation - The Dublin Core metadata element set</b>
<b>Type of document:</b>	
<b>Organization:</b>	<b>Dublin Core Metadata Initiative (DCMI) and ISO TC46</b>
<b>Year:</b>	2003
<b>Description:</b>	A standard for cross-domain information resource description.
<b>Work flow stage:</b>	1.2, 1.10, 1.13
<b>Business value:</b>	Useful for describing the digital resources of an information system in a standard, machine-readable way. This standard has been used by organizations with large information systems such as educational organizations and governments as a basis for electronic description of information resources. The Dublin Core metadata element set has been formally endorsed by other standards organizations such as the WWW Consortium and ANSI as well as numerous governments.
<b>Web site:</b>	<a href="http://dublincore.org/documents/dces/">http://dublincore.org/documents/dces/</a> <a href="http://www.niso.org/international/SC4/n515.pdf">http://www.niso.org/international/SC4/n515.pdf</a>

### Internationalization Tag Set (ITS)

<b>Document title:</b>	<b>Internationalization Tag Set (ITS)</b>
<b>Type of document:</b>	Working draft
<b>Organization:</b>	W3C
<b>Year:</b>	2006
<b>Description:</b>	This draft standard defines an XML vocabulary and set of rules that can be used to specify whether specific content elements of XML documents are to be translated or not. ITS is designed to be used with XML schemas to support the internationalization and localization of schemas and documents.

<b>Work flow stage:</b>	1.2, 1.3, 1.9, 1.10
<b>Business value:</b>	The ITS standard is essential to enable the error-free translation of XML documents.
<b>Web site:</b>	<a href="http://www.w3.org/TR/its/">http://www.w3.org/TR/its/</a> <a href="http://www.w3.org/TR/2005/WD-itsreq-20050805/">http://www.w3.org/TR/2005/WD-itsreq-20050805/</a>

## ***Standards relating to inventories of real-world entities***

### **Common Locale Data Repository (CLDR)**

<b>Document title:</b>	<b>Common Locale Data Repository (CLDR)</b>
<b>Type of document:</b>	n/a – Repository of locale information
<b>Organization:</b>	Unicode Consortium
<b>Year:</b>	2005
<b>Description:</b>	A standard, official resource of names of regional entities such as languages, currencies, locales, and countries, which is available in multiple languages and in a standard machine-readable XML format. The purpose of the Common Locale Data Repository project is to provide a general XML format for the exchange of locale information for use in application and system development, and to gather, store, and make available a common set of locale data generated in that format.
<b>Work flow stage:</b>	1.3, 1.10
<b>Business value:</b>	Useful reference for locale information such as names of countries, applicable time zones, currencies, and so forth, provided in multiple languages.
<b>Web site:</b>	<a href="http://www.unicode.org/cldr/">http://www.unicode.org/cldr/</a>

### **Names of scripts**

<b>Document title:</b>	<b>ISO 15924 - Information and documentation - Codes for the representation of names of scripts</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO TC 46/ SC 2
<b>Year:</b>	2004
<b>Description:</b>	Provides a code for the presentation of names of scripts which were devised for use in terminology, lexicography, and linguistics, but which may be used for any application requiring the expression of scripts in coded form.
<b>Work flow stage:</b>	1.3, 1.10
<b>Business value:</b>	

<b>Web site:</b>	<a href="http://www.unicode.org/iso15924/">http://www.unicode.org/iso15924/</a>
------------------	---

## Names of languages (and countries)

A user's language preferences need to be identified so that appropriate processing can be applied. For example, language preferences in a Web browser can be used to select Web pages in the preferred language of the user. Language preferences can also be used to select among tools, such as dictionaries, to assist in the processing or understanding of content in different languages.

In addition, knowledge about the particular language used in content is required by some types of processing; for example, spell-checking, computer-synthesized speech, Braille transcription, or high-quality print renderings.

Language can be identified by labelling the information content with an identifier or "tag". These tags can also indicate additional language attributes, such as the dialect, writing system, or orthography. This information can be required to process or render the content into an appropriate form or style.

Machine-readable language identifiers are required for all software programs used in the language industry.

The following standards relate to the identification of languages.

<b>Document title:</b>	<b>ISO 639 series – Codes for the representation of names of languages</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO TC 37
<b>Year:</b>	2005
<b>Description:</b>	<p>A six part series providing machine readable codes for the representation of names of languages. Part 1: Alpha-2 code. Part 2: Alpha-3 code. Part 3: Alpha-3 code for comprehensive coverage of languages. Part 4: Implementation guidelines and general principles for language coding. Part 5: Alpha 3 code for language families and groups. Part 6: Alpha-4 representation for comprehensive coverage of language variation.</p> <p>Part 6 is the most developed of the parts. It provides the identification and documentation of language varieties using 4 letter codes and unique linguistic identifiers. It establishes a hierarchical framework of the relationships between language varieties and language families, language groups and individual languages. Part 6 also includes a roadmap for developing a metadata registry of language-related data categories, with an accompanying list of data categories describing the properties of languages, such as the writing system, geography, population, and history.</p>
<b>Work flow stage:</b>	1.3, 1.8, 1.10
<b>Business value:</b>	Applicable to a broad range of applications requiring language identification, particularly computer systems: terminology, lexicography, information

	management, documentation archival and retrieval, information services, information technology, search engines, and multilinguistic cultural and economic exchange.
<b>Web site:</b>	<a href="http://www.loc.gov/standards/iso639-2/langcodes.html">http://www.loc.gov/standards/iso639-2/langcodes.html</a>

<b>Document title:</b>	<b>ISO 3166 – Codes for the representation of names of countries and their subdivisions</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO TC 46
<b>Year:</b>	2006
<b>Description:</b>	ISO 3166 is a three part standard that specifies codes for countries and their subdivisions, and includes guidelines for the implementation and maintenance of the codes. It is intended for use in any application requiring the expression of current country names in coded form.
<b>Work flow stage:</b>	1.3, 1.8, 1.10
<b>Business value:</b>	Applicable to a broad range of applications requiring language identification, particularly computer systems: terminology, lexicography, information management, documentation archival and retrieval, information services, information technology, search engines, and multilinguistic cultural and economic exchange.
<b>Web site:</b>	<a href="http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html">http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html</a>

<b>Document title:</b>	<b>UN M.49 – Standard country or area codes for statistical use</b>
<b>Type of document:</b>	Standard
<b>Organization:</b>	United Nations, Statistical Services Branch
<b>Year:</b>	2006
<b>Description:</b>	This is a list of numerical codes used by the United Nations for statistical purposes. It maps the numerical codes to the ISO 3-digit alphabetical codes.
<b>Work flow stage:</b>	1.3, 1.8, 1.10
<b>Business value:</b>	Applicable to a broad range of applications requiring language identification, particularly computer systems: terminology, lexicography, information management, documentation archival and retrieval, information services, information technology, search engines, and multilinguistic cultural and economic exchange.
<b>Web site:</b>	<a href="http://unstats.un.org/unsd/methods/m49/m49.htm">http://unstats.un.org/unsd/methods/m49/m49.htm</a>

<b>Document title:</b>	<b>IETF RFC 3066 – Tags for the identification of languages</b>
<b>Type of document:</b>	Published best practice
<b>Organization:</b>	IETF
<b>Year:</b>	2001
<b>Description:</b>	<p>This document describes a language tag for use in cases where it is desired to indicate the language used in an information object, how to register values for use in this language tag, and a construct for matching such language tags. It is a best practices document for using the tags specified in ISO 639.</p> <p>This document has been replaced by RFC 4646 and RFC 4647 (see below)</p>
<b>Work flow stage:</b>	1.3, 1.8, 1.10
<b>Business value:</b>	Applicable to a broad range of applications requiring language identification, particularly computer systems: terminology, lexicography, information management, documentation archival and retrieval, information services, information technology, search engines, and multilinguistic cultural and economic exchange.
<b>Web site:</b>	<a href="http://www.ietf.org/rfc/rfc3066.txt?number=3066">http://www.ietf.org/rfc/rfc3066.txt?number=3066</a>

<b>Document title:</b>	<b>IETF RFC 4646 – Tags for identifying languages</b>
<b>Type of document:</b>	Published best practice
<b>Organization:</b>	IETF
<b>Year:</b>	2005
<b>Description:</b>	<p>This document describes the structure, content, construction, and semantics of language tags for use in cases where it is desirable to indicate the language used in an information object. It also describes how to register values for use in language tags and the creation of user-defined extensions for private interchange. It is a best practices guideline for using the language codes in ISO 639. This document, in combination with RFC 4647, replaces RFC 3066, which replaced RFC 1766.</p> <p>Note: RFC 4646 and RFC 4647 were also published together under the number IETF BPC 47.</p>
<b>Work flow stage:</b>	1.3, 1.8, 1.10
<b>Business value:</b>	Applicable to a broad range of applications requiring language identification, particularly computer systems: terminology, lexicography, information management, documentation archival and retrieval, information services, information technology, search engines, and multilinguistic cultural and

	economic exchange.
<b>Web site:</b>	<a href="http://www.ietf.org/rfc/rfc4646.txt?number=4646">http://www.ietf.org/rfc/rfc4646.txt?number=4646</a>

<b>Document title:</b>	<b>IETF RFC 4645 – Initial language subtag registry</b>
<b>Type of document:</b>	Published best practice
<b>Organization:</b>	IETF
<b>Year:</b>	2006
<b>Description:</b>	This document defines the initial contents of the IANA Language Subtag Registry that was defined in IETF RFC 4646.
<b>Work flow stage:</b>	1.3, 1.8, 1.10
<b>Business value:</b>	Applicable to a broad range of applications requiring language identification, particularly computer systems: terminology, lexicography, information management, documentation archival and retrieval, information services, information technology, search engines, and multilinguistic cultural and economic exchange.
<b>Web site:</b>	<a href="http://www.ietf.org/rfc/rfc4645.txt?number=4645">http://www.ietf.org/rfc/rfc4645.txt?number=4645</a>

<b>Document title:</b>	<b>IETF RFC 4647 – Matching of language tags</b>
<b>Type of document:</b>	Published best practice
<b>Organization:</b>	IETF
<b>Year:</b>	2006
<b>Description:</b>	<p>This document describes a syntax, called a "language-range", for specifying items in a user's list of language preferences. It also describes mechanisms for comparing and matching these to language tags. This document, in combination with RFC 4646, replaces RFC 3066.</p> <p>Note: RFC 4646 and RFC 4647 were also published together under the number IETF BPC (Best Current Practice) 47.</p>
<b>Work flow stage:</b>	1.3, 1.8, 1.10
<b>Business value:</b>	<p>Applicable to a broad range of applications requiring language identification, particularly computer systems: terminology, lexicography, information management, documentation archival and retrieval, information services, information technology, search engines, and multilinguistic cultural and economic exchange.</p> <p>Possible applications include programmed selection of a language among multiple valid values, or content selection.</p>
<b>Web site:</b>	<a href="http://www.ietf.org/rfc/rfc4647.txt?number=4647">http://www.ietf.org/rfc/rfc4647.txt?number=4647</a>

## **Quality standards**

The ISO 9000 standards are a set of guidelines that assure that businesses meet certain quality control and management standards. They are designed to define the requirements of good quality management system (QMS), irregardless of the specific industry sector or the size of the business.

The ISO 9000 standards are particularly useful for industries that have many competitors, since the standards help the business focus on customer requirements and customer satisfaction. The ISO 9000 standards establish a generic framework and principles for quality assessment. A key principle is the audit process.

The ISO 9000 standards have been implemented by over 760,000 organizations in over 150 countries world-wide.

There are over 30 ISO standards related to quality management. Many are specific to particular industries. Two of the general ISO 9000 standards that could serve as a foundation for establishing quality processes in the language industry are described below. For the complete list of ISO quality related standards, visit the following Web site:

Quality management and quality assurance standards:

<http://www.iso.org/iso/en/CatalogueListPage.CatalogueList?ICS1=3&ICS2=120&ICS3=10&scopelist>

### Quality management systems

<b>Document title:</b>	<b>ISO 9000 – Quality management systems. Fundamentals and vocabulary.</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO TC 176
<b>Year:</b>	2005
<b>Description:</b>	ISO 9000 describes the fundamentals of quality management systems, and forms the basis for the ISO 9000 family of standards. It also defines quality management terms. It is recommended for organizations that intend to implement a quality management system to establish confidence in their customers and their suppliers about the quality of their products. It is also recommended for individuals who work with quality management systems such as auditors and trainers.
<b>Work flow stage:</b>	1.12, 1.13, 2.7, 2.8, 3
<b>Business value:</b>	This is a fundamental standard that all businesses should adopt as an objective means of establishing quality.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

<b>Document title:</b>	<b>ISO 9001 – Quality management systems. Requirements.</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO TC 176
<b>Year:</b>	2004
<b>Description:</b>	This standard specifies the actual requirements for a quality management system. It is generic so that it can be applied to any organization irregardless of its type, size, or products produced.
<b>Work flow stage:</b>	1.12, 1.13, 2.7, 2.8, 3
<b>Business value:</b>	This is a fundamental standard that all businesses should adopt as an objective means of establishing quality.
<b>Web site:</b>	<a href="http://www.iso.org/">http://www.iso.org/</a>

<b>Document title:</b>	<b>ISO/IEC 90003 – Software engineering – Guidelines for the application of ISO 9001:2000 to computer software</b>
<b>Type of document:</b>	Published standard

<b>Organization:</b>	ISO JTC1
<b>Year:</b>	2004
<b>Description:</b>	This standard provides guidelines on how to apply ISO 9001 to the acquisition, supply, development, operation and maintenance of computer software and related support services.
<b>Work flow stage:</b>	1.1, 1.12, 1.13, 2.7, 2.8, 3
<b>Business value:</b>	This is a fundamental standard that software companies should adopt as an objective means of establishing quality.
<b>Web site:</b>	<a href="http://www.iso.org/">http://www.iso.org/</a>

## **Web Services**

A Web service is a software system that can be accessed and operated over a network such as the Internet. In the translation industry, a Web service could, for example, be used to provide translation technology similar to the computer-assisted translation tools that we have today, only over the Internet, thereby eliminating the need for desktop applications.

<b>Document title:</b>	<b>Web Services Description Language (WSDL)</b>
<b>Type of document:</b>	W3C Candidate Recommendation
<b>Organization:</b>	W3C
<b>Year:</b>	2006
<b>Description:</b>	<p>Web services technology is increasingly important for automated business interaction and application integration between enterprises. Web services allow software components to interoperate using the Internet as a transport mechanism.</p> <p>WSDL is an XML format for describing network services as message carriers of document or procedural information. It provides a standard way of describing Internet-based services as communication endpoints capable of exchanging messages. WSDL is thus a means of automating communication between applications over the Internet.</p>
<b>Work flow stage:</b>	
<b>Business value:</b>	<p>WSDL allows the development of Web services in a standardized way, thereby extending the Internet from a source of information to a full-service platform.</p> <p>Recent research in this area at the W3C includes the development of semantic annotations for WSDL aimed towards integrating semantic Web technology and concepts into Web services.</p>

<b>Web site:</b>	<a href="http://www.w3.org/TR/wsdl20/">http://www.w3.org/TR/wsdl20/</a>
------------------	---

<b>Document title:</b>	<b>Simple Object Access Protocol (SOAP)</b>
<b>Type of document:</b>	Recommendation
<b>Organization:</b>	W3C
<b>Year:</b>	2003
<b>Description:</b>	SOAP is a protocol for exchanging XML-based messages over computer networks. SOAP defines the XML-based information which can be used for exchanging structured and typed information between systems in a decentralized, distributed environment.
<b>Work flow stage:</b>	
<b>Business value:</b>	SOAP is an enabling protocol for Web services.
<b>Web site:</b>	<a href="http://www.w3.org/TR/soap/">http://www.w3.org/TR/soap/</a>

## Content authoring and development

### Software documentation

<b>Document title:</b>	<b>IEEE 1063:2001 – Standard for Software User Documentation</b>
<b>Type of document:</b>	Standard
<b>Organization:</b>	IEEE
<b>Year:</b>	2001
<b>Description:</b>	Presents the minimum requirements for the structure, information content, and format of user documentation, including both printed and electronic documents used in the work environment by users of systems containing software.
<b>Work flow stage:</b>	1.1, 1.2, 1.4, 1.5, 1.6, 1.7, 1.8, 1.12, 1.13
<b>Business value:</b>	This standard helps software developers ensure that software information and documentation is consistent, complete, accurate, and usable.
<b>Web site:</b>	<a href="http://standards.ieee.org/">http://standards.ieee.org/</a>

<b>Document title:</b>	<b>IEC 61506 - Industrial Process Measurement and Control - Documentation of Application Software</b>
<b>Type of document:</b>	International standard
<b>Organization:</b>	IEC

<b>Year:</b>	1997
<b>Description:</b>	Defines requirements for the documentation of software in industrial process measurement and control systems.
<b>Work flow stage:</b>	1.1, 1.2, 1.4, 1.5, 1.6, 1.7, 1.8, 1.12, 1.13
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://domino.iec.ch/webstore/webstore.nsf/artnum/021929">http://domino.iec.ch/webstore/webstore.nsf/artnum/021929</a>

### Software documentation process

<b>Document title:</b>	<b>ISO IEC 15910 – Software User Documentation Process</b>
<b>Type of document:</b>	Standard under revision
<b>Organization:</b>	ISO/IEC JTC1/SC7
<b>Year:</b>	1999
<b>Description:</b>	Provides guidelines for specifying the process used in developing documentation, thereby ensuring the quality of the delivered materials.
<b>Work flow stage:</b>	1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.12, 1.13
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www.hci.com.au/iso/">http://www.hci.com.au/iso/</a>

### Design and preparation of software documentation

<b>Document title:</b>	<b>ISO 18019 – Software and system engineering – Guidelines for the design and preparation of user documentation for application software</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO/IEC JTC1/SC7
<b>Year:</b>	2004
<b>Description:</b>	This standard provides guidelines on how to determine what information users need, how that information should be presented, and how to prepare and present the information.
<b>Work flow stage:</b>	1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.12, 1.13
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www.hci.com.au/iso/">http://www.hci.com.au/iso/</a>

<b>Document title:</b>	<b>BS 7649 – Guide to the design and preparation of documentation for users of application software</b>
------------------------	---

<b>Type of document:</b>	Standard, confirmed
<b>Organization:</b>	British Standards Institution
<b>Year:</b>	1993
<b>Description:</b>	
<b>Work flow stage:</b>	
<b>Business value:</b>	(Replaced by ISO 18019)
<b>Web site:</b>	

<b>Document title:</b>	<b>BS 7830 – Guide to the design and preparation of on-line documentation for users of applications software</b>
<b>Type of document:</b>	Standard, confirmed
<b>Organization:</b>	British Standards Institution
<b>Year:</b>	
<b>Description:</b>	
<b>Work flow stage:</b>	
<b>Business value:</b>	(Replaced by ISO 18019)
<b>Web site:</b>	

### Documentation guidelines for computer-based application systems

<b>Document title:</b>	<b>ISO/IEC 6592 – Information technology. Guidelines for the documentation of computer-based application systems</b>
<b>Type of document:</b>	
<b>Organization:</b>	ISO
<b>Year:</b>	2000
<b>Description:</b>	Not available.
<b>Work flow stage:</b>	1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.12, 1.13
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Management of software documentation

<b>Document title:</b>	<b>ISO 9294 – Guidelines for the management of software documentation</b>
<b>Type of document:</b>	Technical report

<b>Organization:</b>	ISO/IEC JTC1/SC7
<b>Year:</b>	2005
<b>Description:</b>	Provides guidelines for the management of software documentation to managers responsible for the production of software products.
<b>Work flow stage:</b>	1.13
<b>Business value:</b>	Standardizing software documentation will reduce production and translation costs and will make the information more usable and more reusable.
<b>Web site:</b>	<a href="http://www.hci.com.au/iso/">http://www.hci.com.au/iso/</a>

## Darwin Information Typing Architecture (DITA)

<b>Document title:</b>	<b>DITA - Darwin Information Typing Architecture</b>
<b>Type of document:</b>	Committee draft
<b>Organization:</b>	OASIS
<b>Year:</b>	2005
<b>Description:</b>	An XML-based document creation and management specification that builds content reuse into the authoring process. It organizes information into modular and extensible topics. New information types can be defined as specializations of existing types.
<b>Work flow stage:</b>	1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 1.13
<b>Business value:</b>	Optimizes information assets for reuse and modular assembly of output deliverables. Improves search and navigation of information. Prepares the information for content management applications. Overall increases productivity of information development, effectiveness of information, and customer satisfaction
<b>Web site:</b>	<a href="http://xml.coverpages.org/dita.html">http://xml.coverpages.org/dita.html</a>

## DocBook

<b>Document title:</b>	<b>DocBook 4.5</b>
<b>Type of document:</b>	Working draft
<b>Organization:</b>	OASIS
<b>Year:</b>	2006
<b>Description:</b>	A general purpose XML schema for writing books and papers. The scope of DocBook is computer hardware and software documentation. It provides a common structure and vocabulary for developing computer documentation including print and online formats as well as a range of information types including reference guides and tutorials.
<b>Work flow stage:</b>	1.1, 1.2, 1.3, 1.4, 1.5
<b>Business value:</b>	Facilitates the development and reuse of computer documentation. Increases productivity of employees who can use standard formats and templates, and makes information easier to use due to the common look and feel.
<b>Web site:</b>	<a href="http://www.docbook.org/specs">http://www.docbook.org/specs</a>

## OpenDocument

<b>Document title:</b>	<b>OpenDocument 1.0 (ODF)</b>
------------------------	-------------------------------

<b>Type of document:</b>	Standard
<b>Organization:</b>	OASIS
<b>Year:</b>	2005
<b>Description:</b>	An XML-based file format for office files such as spreadsheets, text, charts, and graphics. This standard is supported by a range of office applications and companies including Arbortext, OpenOffice, IBM, and Sun Microsystems. It is currently being submitted for approval as an ISO standard. It has generated interest by governments operating large scale information networks. It is not supported by MSOffice.
<b>Work flow stage:</b>	1.1, 1.2, 1.3, 1.4, 1.5
<b>Business value:</b>	This type of standard protects content from being locked into an application or vendor-specific file format. It also allows users to realize the benefits of having their content in a standard XML format without changing their authoring tools.
<b>Web site:</b>	<a href="http://www.oasis-open.org">http://www.oasis-open.org</a>

### Document language

<b>Document title:</b>	<b>Authoring Techniques for XHTML &amp; HTML Internationalization: Specifying the language of content 1.0</b>
<b>Type of document:</b>	Working draft
<b>Organization:</b>	W3C
<b>Year:</b>	2005
<b>Description:</b>	A standard specification for indicating the language of HTML and XHTML content.
<b>Work flow stage:</b>	1.3, 1.8
<b>Business value:</b>	Required for any global deployment of Web sites. Useful for implementing a wide range of language dependent functions from language-sensitive searching to language-specific display settings.
<b>Web site:</b>	<a href="http://www.w3.org/TR/2005/WD-i18n-html-tech-lang-20050224/">http://www.w3.org/TR/2005/WD-i18n-html-tech-lang-20050224/</a>

### Bidirectional text

<b>Document title:</b>	<b>Authoring Techniques for XHTML &amp; HTML Internationalization: Handling Bidirectional Text 1.0</b>
<b>Type of document:</b>	Working draft
<b>Organization:</b>	W3C
<b>Year:</b>	2004
<b>Description:</b>	Provides guidelines for using markup and CSS to create Web pages for

	languages that use bidirectional text, such as Arabic and Hebrew.
<b>Work flow stage:</b>	1.3, 1.8
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www.w3.org/TR/2004/WD-i18n-html-tech-bidi-20040509/">http://www.w3.org/TR/2004/WD-i18n-html-tech-bidi-20040509/</a>

## XHTML and HTML internationalization

<b>Document title:</b>	<b>Authoring Techniques for XHTML &amp; HTML Internationalization: Characters and Encodings 1.0</b>
<b>Type of document:</b>	Working draft
<b>Organization:</b>	W3C
<b>Year:</b>	2004
<b>Description:</b>	Provides guidelines about characters, character sets, and encoding matters for producing internationalized HTML and XHTML content.
<b>Work flow stage:</b>	1.3, 1.8
<b>Business value:</b>	Important for any organization that produces Web content for international audiences.
<b>Web site:</b>	<a href="http://www.w3.org/TR/2004/WD-i18n-html-tech-char-20040509/">http://www.w3.org/TR/2004/WD-i18n-html-tech-char-20040509/</a>

## Business documents

<b>Document title:</b>	<b>Universal Business Language (UBL)</b>
<b>Type of document:</b>	Standard
<b>Organization:</b>	OASIS
<b>Year:</b>	2006
<b>Description:</b>	UBL defines a generic XML interchange format for business documents. UBL also provides: <ol style="list-style-type: none"> <li>1. A library of XML schemas for reusable data components such as “Address,” “Item,” and “Payment” — the common data elements of everyday business documents.</li> <li>2. A set of XML schemas for common business documents such as “Order,” “Despatch Advice,” and “Invoice” that are constructed from the UBL library components and can be used in generic procurement and transportation contexts.</li> </ol>
<b>Work flow stage:</b>	1.2, 1.4, 1.5
<b>Business value:</b>	The widespread use of XML has led to the development of multiple industry-specific XML versions of such basic documents as purchase orders, shipping notices, and invoices. UBL is designed to provide a standard syntax for

	<p>business documents.</p> <p>UBL has the potential to deliver many benefits, including lowering the actual costs of business software by making them easier to develop, and lowering the cost of integrating the software into business environments.</p> <p>UBL is designed to plug directly into existing business, legal, auditing, and records management practices, eliminating the re-keying of data in existing fax- and paper-based supply chains and providing an entry point into electronic commerce for small and medium-sized businesses.</p>
<b>Web site:</b>	<a href="http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=ubl">http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=ubl</a>

### Internationalization and localization markup requirements

<b>Document title:</b>	<b>Internationalization and localization markup requirements</b>
<b>Type of document:</b>	Working draft
<b>Organization:</b>	W3C
<b>Year:</b>	2005
<b>Description:</b>	A description of the key requirements for schemas (XML schemas, DTDs, etc.) to ensure that they meet the needs of authors of content intended for international audiences, and also the needs of content localizers. It defines requirements for solving the main challenges and issues of internationalizing and localizing XML documents. It will be used as a framework for developing further standards to address these challenges. It is developed by the same workgroup that is creating the ITS (Internationalization Tag Set).
<b>Work flow stage:</b>	1.2, 1.3, 1.8, 1.9
<b>Business value:</b>	Of interest to any Web developer who creates Web content for international audiences.
<b>Web site:</b>	<a href="http://www.w3.org/TR/2005/WD-itsreq-20051122/">http://www.w3.org/TR/2005/WD-itsreq-20051122/</a>

## Translation and localization

Standards for the translation and localization industry facilitate interoperability of tools and file formats, and compatibility and harmonization of the industry in diverse global markets. The standards in this section are divided into three categories: project management, quality control, and computer assisted translation.

Localization refers to the process of translating and adapting content for a specific target market. As translation work became more and more computerized through the development of computer assisted translation software, standards had to be defined to address more technical aspects such as file formats and data interchange.

The key localization formats -- XLIFF, TMX, TBX, and SRX -- are fairly complementary in their purpose and scope. The focus of XLIFF is to simplify the handling of file formats and facilitate localization work flow. The purpose of TMX is to increase portability of translation memories among tools. SRX addresses a different aspect of translation memory portability: segmentation. And TBX is specifically designed for terminology databases.

Developers of these standards collaborate to avoid the risk of duplication and overlap. TBX and XLIFF have adopted TMX styles for enclosing markup tags. TBX uses data categories and XML styles proposed by ISO TC 37. TMX will be improved once SRX and a standard set of segmentation rules are defined.

Nevertheless, there is much that still needs to be done to achieve full integration of these standards in a one-stop localization solution. This level of integration might not occur until each of these individual standards has matured to a stable level and gained wider adoption by the localization industry.

### ***Project management***

Translation project management standards enable translation and localization service providers to apply objective measurements and criteria in areas such as job sizing and billing, quality assessment, and contracts. Furthermore, file interchange standards ensure that content subject to translation and localization can be easily exchanged from client to service provider, and from authoring software to translation software, without corruption or loss of data.

## Translation contracts

<b>Document title:</b>	<b>Translation Contracts: A Standards-Based Model Solution</b>
<b>Type of document:</b>	Published book
<b>Organization:</b>	Author: Uwe Muegge
<b>Year:</b>	2005
<b>Description:</b>	Guidelines for establishing translation contracts between clients and service providers, based on DIN 2345, ONORM 1201, and ASTM 515.48. Provides a collection of customizable forms.
<b>Work flow stage:</b>	2.2, 2.7, 2.9, 2.12
<b>Business value:</b>	Useful for translators, project managers, and anyone involved in planning, specifying, managing, and evaluating translation and localization projects.
<b>Web site:</b>	Available from <a href="http://www.amazon.com">http://www.amazon.com</a>

<b>Document title:</b>	<b>DIN 2345 – Translation Contracts</b>
<b>Type of document:</b>	Standard
<b>Organization:</b>	Deutsches Institut für Normung (German Standardization Institute)
<b>Year:</b>	1998
<b>Description:</b>	Describes the various conditions for concluding contracts between translators and clients. Addresses the selection of a translator, cooperation between all parties involved, documentation, source text, target text, checking and revision. Available in English, French and German.
<b>Work flow stage:</b>	2.2, 2.7, 2.9, 2.12
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www2.din.de">http://www2.din.de</a>

<b>Document title:</b>	<b>ÖNORM D 1200 – Translation and interpretation services – Translation services – Requirements for the service and the provision of the service</b>
<b>Type of document:</b>	
<b>Organization:</b>	ÖNORM – Österreichisches Normungsinstitut (Austrian Standards Institute)
<b>Year:</b>	2000
<b>Description:</b>	Includes recommendations regarding the qualifications of translators.
<b>Work flow stage:</b>	2.2, 2.7, 2.9, 2.12
<b>Business value:</b>	

<b>Web site:</b>	<a href="http://www.on-norm.at">http://www.on-norm.at</a>
	<a href="http://www.on-norm.at/index_e.html">http://www.on-norm.at/index_e.html</a>
	<a href="http://www.on-norm.at/index_e.html">http://www.on-norm.at/index_e.html</a>
<b>Document title:</b>	<b>ÖNORM D 1201 – Translation and interpretation services – Translation services – Translation contracts</b>
<b>Type of document:</b>	
<b>Organization:</b>	ÖNORM – Österreichisches Normungsinstitut (Austrian Standards Institute)
<b>Year:</b>	2000
<b>Description:</b>	Includes recommendations regarding the qualifications of translators.
<b>Work flow stage:</b>	2.2, 2.7, 2.9, 2.12
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www.on-norm.at">http://www.on-norm.at</a>

<b>Document title:</b>	<b>UNI 10574 – Definition of services and activities of translation and interpreting enterprises</b>
<b>Type of document:</b>	
<b>Organization:</b>	UNI – Ente Nazionale Italiano di Unificazione
<b>Year:</b>	1996
<b>Description:</b>	
<b>Work flow stage:</b>	2.2, 2.7, 2.9, 2.12
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www.uni.com">http://www.uni.com</a>

### Global Information Metrics eXchange (GMX)

<b>Document title:</b>	<b>GMX - Global Information Management Metrics eXchange</b>
<b>Type of document:</b>	Standard
<b>Organization:</b>	LISA
<b>Year:</b>	2006
<b>Description:</b>	A 3-part family of proposed standards concerning translation-related metrics: (1) volume, (2) complexity and (3) quality. A review draft of the volume component - GMX-V has been made available for public comment. This volume component establishes guidelines for quantifying the workload for a given localization or translation task. Going beyond simple word counts, it provides a more precise definition of the statistics necessary for translation billing and sizing purposes.

<b>Work flow stage:</b>	2.2, 2.7, 2.12
<b>Business value:</b>	Provides a neutral benchmark for estimating and billing purposes for the translation industry. This makes the translation industry less vulnerable to fluctuations in billing practices.
<b>Web site:</b>	<a href="http://www.lisa.org/standards/gmx/">http://www.lisa.org/standards/gmx/</a>

## XML Localization Interchange File Format (XLIFF)

<b>Document title:</b>	<b>XLIFF - XML Localization Interchange File Format</b>
<b>Type of document:</b>	Committee specification
<b>Organization:</b>	OASIS
<b>Year:</b>	2004
<b>Description:</b>	A specification for the loss-less interchange of localizable data and its related information. It is tool-neutral, has been formalized as an XML vocabulary (through XML schema), and features an extensibility mechanism.
<b>Work flow stage:</b>	1.3, 1.6, 1.10, 1.13, 2.1, 2.3, 2.5, 2.9, 2.11
<b>Business value:</b>	Streamline the translation process by reducing file management and file formatting tasks.
<b>Web site:</b>	<a href="http://www.oasis-open.org/committees/xliff/documents/xliff-specification.htm">http://www.oasis-open.org/committees/xliff/documents/xliff-specification.htm</a>

XML Localization Interchange File Format (XLIFF) is an XML vocabulary designed for the specific purpose of interchanging translatable content among members of the localization industry. It aims to enable a consistent and regular format for presenting content for translation to human translators as well as to translation automation while supporting the whole translation work flow process.

One of the greatest challenges that translation project managers have to overcome is the sheer number of files that may have to be handled - a modern project may comprise hundreds, if not thousands, of files. In addition, files themselves may exist in many different file formats, from plain text to multiple different proprietary formats, each of which requires different processes and tools. And multiple versions of these files might even exist because localization has likely taken place in conjunction with product development. The following list would not be an uncommon subset of the file types making up the smallest of projects:

- Microsoft® Windows® resource files (RC, DLL, EXE)
- Plain Text
- HTML
- Java™ Resource Bundle
- Java
- XML
- XSL
- Rich Text

- Graphics (BMP, GIF, JPEG)
- Script files (JavaScript, VBScript)
- Documentation (Windows help, printed documentation)

Most translation vendors will not have any problems dealing with the file types from the list above, but most projects also contain many other proprietary file types. Each proprietary file type may require its own process, or necessitate a dedicated tool, or necessitate some special translator expertise or experience. Such overhead adds both a financial and a time cost to a localization project.

XLIFF does not try to address all of the challenges encountered in a localization project but it does aim to simplify file management and interchange between a localization team and the translation vendor. In particular, XLIFF addresses the following issues:

- *The number of different types of files in a translation package*  
An XLIFF document may contain the translatable text or items from all of the types of files in a project. It is up to the localization team to filter the file types to generate the XLIFF document, but only one type of file must be delivered to the translation vendor and, in turn only one type of file is returned.
- *The number of files in a translation package*  
A single XLIFF document may contain all of the translatable content for a project; that is, every translatable file in the project may be encapsulated in a single XLIFF document. An XLIFF document may also include binary files for translation, such as graphics files (GIF, JPEG, etc.), allowing all of the files for a given version or revision of a project to be kept together.
- *Managing, storing, and using project metadata*  
Since XLIFF contains either whole binary files or extracted data from a file, there are placeholders for associated metadata that accompanies a file during localization. Information such as the date the XLIFF was generated, the build number and product revision for a file, and the translatable word count may be conveniently provided.
- *Ease of translation*  
XLIFF was designed so that a translator could work directly within the XLIFF document should the need arise. Provisions are also made to accommodate complex tool combinations.
- *Historical changes to translations and review process*  
XLIFF provides a convenient way to track historical changes to a translation using ‘alternative’ translations, stored with the actual current translation for an item in an XLIFF document. That way, a final reviewer may be offered a choice of the correct translation to use. This feature provides for a method of correctly updating translation memories on completion of a project. XLIFF also introduces ‘phase’ information. This information can be used to indicate at what stage a translation was changed and by whom.
- *Resource Data*  
Apart from translatable content, XLIFF also facilitates storing modifiable resource data such as coordinate or font information.
- *Interoperability*  
XLIFF was designed with interoperability in mind, offering advantages to both translation

vendors and software companies. Interoperability offers ease of transition among vendors should the need arise, negating the need for expertise in handling particular formats or processes. Translation vendors, on the other hand, may now become more competitive, knowing they can quote on and immediately handle a given job.

- *Enforces a consistent tools support approach*

The need to produce only one file format for translation enforces a consistent tools support strategy for all file types.

The necessity for XLIFF arose from the limitations of the existing globalization XML formats for use in the translation stages of localization. TMX was defined specifically for the purpose of moving existing translations from one application or store to another and were not suited for the actual translation process. TBX (TermBase Exchange) is an XML exchange format specifically for terminological data. OpenTag, while originally designed as a translatable format, was incomplete for use in localization projects for three reasons:

1. It was intended more for documentation-type content
2. It was too open and generic
3. It didn't sufficiently address the need for project data or metadata.

XLIFF is specifically intended to handle bilingual data in the translation process (which is typically bilingual in nature). XLIFF also allows for the inclusion of specific translation instructions on a per-language basis. Finally, the amount of collaboration from the contributing parties and their respective companies (representing an extremely wide cross-section of the localization industry) to define XLIFF, produced a complete and thoroughly interoperable format.

Apart from handling the normal translation cycle movement of files, XLIFF is also ideal for use in Web services. Since it is XML based, containment within a SOAP envelope is a trivial matter, and XLIFF accommodates all other necessary content for job tracking and processing.

## OpenTag

<b>Document title:</b>	<b>OpenTag</b>
<b>Type of document:</b>	Specification
<b>Organization:</b>	OpenTag.com
<b>Year:</b>	2001
<b>Description:</b>	A format used for the extraction of translatable text from a source document, and the reinsertion of the translated text back into the original document format.
<b>Work flow stage:</b>	2.3
<b>Business value:</b>	OpenTag has been replaced by XLIFF.
<b>Web site:</b>	<a href="http://www.opentag.com/opentag.htm">http://www.opentag.com/opentag.htm</a>

OpenTag is the predecessor to XLIFF. It was developed in 1997 by International Language Engineering Corporation (ILE) and then made public. However, this format was not widely adopted within the software industry despite its usefulness and ingenuity. Many of the concepts defined in OpenTag form integral constituents of XLIFF.

OpenTag is a very generic format, and designed more for the extraction, translation and re-insertion of documentation-type content than for handling other software formats. It was first published nearly four years prior to the development of XLIFF and in the early days of XML. It is now no longer maintained and XLIFF should be used in its place.

### Translation Web Services

<b>Document title:</b>	<b>Translation Web Services technical committee</b>
<b>Type of document:</b>	n/a (Technical Committee)
<b>Organization:</b>	OASIS
<b>Year:</b>	n/a
<b>Description:</b>	The Translation Web Services technical committee is developing a standard that provides an encapsulation of all the information required to support the following objective: "Any publisher of content to be translated should be able to automatically connect to and use the services of any translation vendor, over the Internet, without any previous direct communication between the two" <sup>8</sup> . This technical committee is investigating the use of Web services to automate many of the administrative and process flow tasks of localization projects. The TC is working in collaboration with the XLIFF TC to ensure that XLIFF is supported in translation Web services.
<b>Work flow stage:</b>	2.1, 2.2, 2.5, 2.6, 2.9, 2.12
<b>Business value:</b>	A standard for Web services for the translation business has the potential to make translation project management much more automated than it is today, thereby saving translation companies much cost and effort.
<b>Web site:</b>	<a href="http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=trans-ws">http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=trans-ws</a>

### Quality control

#### Quality metrics

<b>Document title:</b>	<b>SAE J2450 - Translation Quality Metric</b>
------------------------	---

<sup>8</sup> From the committee's charter statement.

<b>Type of document:</b>	Recommended practice
<b>Organization:</b>	Society of Automotive Engineers (SAE)
<b>Year:</b>	2001
<b>Description:</b>	SAE standard designed to develop an industry-wide metric for the evaluation of translation quality, the objective of which is to establish a consistent standard against which the translation quality of automotive service information can be objectively measured. Measurable aspects that are included in this standard include completeness, correctness of numbers, and spelling and grammar mistakes.
<b>Work flow stage:</b>	2.6, 2.7, 2.8
<b>Business value:</b>	Improves translation quality in the automotive industry. Principles can be applied to other industries.
<b>Web site:</b>	<a href="http://www.sae.org/technicalcommittees/j2450p1.htm">http://www.sae.org/technicalcommittees/j2450p1.htm</a>

### Quality assurance

<b>Document title:</b>	<b>WK2953 - Consumer-Oriented Guide to Quality Assurance in Translation and Localization</b>
<b>Type of document:</b>	New Work Item
<b>Organization:</b>	<b>ASTM TC F15.48</b>
<b>Year:</b>	
<b>Description:</b>	A guide designed to identify factors relevant to the quality of language translation and localization services for each phase of a project and to provide a framework within which the participants in a services agreement can define the specifications necessary to arrive at a product that satisfies defined customer needs.
<b>Work flow stage:</b>	2.2, 2.7, 2.8
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www.astm.org/">http://www.astm.org/</a>

<b>Document title:</b>	<b>LISA Quality Assurance Model</b>
<b>Type of document:</b>	Guideline
<b>Organization:</b>	<b>LISA</b>
<b>Year:</b>	2004
<b>Description:</b>	Provides guidelines for quality assurance for all stages of a localization project, from source content preparation to translation and final testing.

<b>Work flow stage:</b>	1.2, 1.3, 1.4, 1.7, 1.12, 2.2, 2.4, 2.5, 2.7, 2.8
<b>Business value:</b>	Valuable reading for managers to get a general understanding of the requirements for a localization project.
<b>Web site:</b>	<a href="http://www.lisa.org/products/qamodel/">http://www.lisa.org/products/qamodel/</a>

<b>Document title:</b>	<b>EN-15038:2006 – Translation services – Service quality</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	<b>EUATC – European Union of Associations of Translation Companies and CEN – Comité européen de normalisation</b>
<b>Year:</b>	2006
<b>Description:</b>	<p>A standard for translation services recognized in Europe. It includes quality criteria based on human and technical resources, quality management, project management, quotations, contracts, rights and obligations of the supplier and customer, translation service procedures (administrative, technical and linguistic work, review and revision), localization, desktop publishing, translation memory management, glossary compilation, and so forth. It is accompanied by a series of checklists and compliance procedures.</p> <p>It will replace local quality standards such as Germany's DIN 2345, Italy's UNI10574, Önorm D 1200.2000 Service Requirements, and Önorm D 1201.2000 Translation Contracts.</p>
<b>Work flow stage:</b>	2.4, 2.6, 2.7, 2.8
<b>Business value:</b>	This standard will help language service providers differentiate themselves on the basis of formal process and quality -- and thus help buyers of their services make better purchase decisions.
<b>Web site:</b>	<a href="http://www.cen.eu/catweb/03.080.20.htm">http://www.cen.eu/catweb/03.080.20.htm</a>

### Guide for interpretation services

<b>Document title:</b>	<b>ASTM F2089-1 – Standard Guide for Language Interpretation Services</b>
<b>Type of document:</b>	
<b>Organization:</b>	<b>ASTM</b>
<b>Year:</b>	
<b>Description:</b>	Identifies the components of quality language interpretation services and defines criteria for each component including quality assessment. Identifies a procedure for selecting an interpretation service.
<b>Work flow stage:</b>	2.2, 2.7, 2.9, 2.12
<b>Business value:</b>	

<b>Web site:</b>	<a href="http://www.astm.org/">http://www.astm.org/</a>
------------------	---

## ***Computer assisted translation***

### **Translation Memory eXchange (TMX)**

<b>Document title:</b>	<b>TMX - Translation Memory eXchange</b>
<b>Type of document:</b>	Standard
<b>Organization:</b>	LISA
<b>Year:</b>	2004
<b>Description:</b>	An open XML standard for the exchange of translation memory data created by computer aided translation and localization tools. It allows easier exchange of translation memory data between tools and/or translation vendors that support TMX, with little or no loss of critical data. Tools vendors can be certified TMX-compliant by the Localization Industry Standards Association.
<b>Work flow stage:</b>	2.1, 2.3, 2.6, 2.11
<b>Business value:</b>	Allows translation and localization services to have flexibility of choice between CAT tools. Provides greater reuse of translation memory assets.
<b>Web site:</b>	<a href="http://www.lisa.org/standards/tmx/">http://www.lisa.org/standards/tmx/</a>

TMX was designed by the OSCAR special interest group of the Localization Industry Standards Association. It is a globalization-related XML vocabulary designed to provide a standard mechanism for moving translation memory information among different translation tools.

The obvious aims of TMX are to maximize translation reuse, increase translation consistency, and reduce costs by facilitating the movement of translation memories among different processes and stages within a translation work flow. TMX might be used, for example, when a translation vendor supplies a TMX document, exported from the translation vendors' possibly proprietary translation memory tool, of all translations used in a project, on delivery of that translated project. This would allow the customer to import the translation memory into their own translation memory tool for use during pre-translation of the next version of the project.

Encapsulating markup in the manner prescribed by TMX allows the translation memory application to either ignore the markup altogether, or protect the markup from corruption or loss during import or export of the memory. The markup techniques defined in TMX for handling inline codes have been incorporated into XLIFF.

## Segmentation Rules eXchange (SRX)

<b>Document title:</b>	<b>SRX - Segmentation Rules eXchange</b>
<b>Type of document:</b>	Standard
<b>Organization:</b>	LISA
<b>Year:</b>	2004
<b>Description:</b>	An XML-based standard for the description of the ways in which translation and other language-processing tools segment text for processing. Designed to enhance the TMX standard to optimize the exchangeability of translation memories.
<b>Work flow stage:</b>	2.1, 2.3, 2.11
<b>Business value:</b>	Differences in segmentation are the main causes of incompatibilities in translation memories between CAT tools. This standard can help to increase compatibility between CAT tools, thereby increasing options and facilitating tools use by translation companies. It may also help in developing term extraction tools.
<b>Web site:</b>	<a href="http://www.lisa.org/standards/srx/">http://www.lisa.org/standards/srx/</a>

One of the challenges preventing 100 percent portability of translation memories among tools is differences in segmentation: the ways that the tools divide text into translation segments. This issue is addressed by the Segmentation Rule Exchange Format. This is not actually a segmentation standard, but a standard XML language for documenting segmentation rules. The Segmentation Workgroup of LISA felt that documenting existing segmentation rules in a common language to facilitate comparison was a necessary preliminary step to defining an actual standard set of segmentation rules.

An SRX document identifies the characters that are used for white spaces, words, non-words and digits. It also defines the segmentation rules of the application and how the rules are applied. For example, it identifies the characters that start or end a segment, and any exceptions that apply.

## XML:TM – Text Memory

<b>Document title:</b>	<b>XML:TM</b>
<b>Type of document:</b>	Specification (draft)
<b>Organization:</b>	LISA
<b>Year:</b>	2006
<b>Description:</b>	An XML-based text memory specification. The specification defines a vocabulary for storing text memory information within an XML document using the XML namespace syntax.
<b>Work flow stage:</b>	1.4, 1.13, 2.3, 2.6, 2.11
<b>Business value:</b>	XML:TM integrates the concept of author memory with translation memory.

	This is a radical new approach that has the potential to facilitate the linking of authoring and translation processes within one integrated XML environment. Such a seamless environment has the potential to reduce process and data overhead in the authoring and translation processes.
<b>Web site:</b>	<a href="http://www.lisa.org/standards/xmltm/">http://www.lisa.org/standards/xmltm/</a>

## Multilingual Information Framework

<b>Document title:</b>	<b>Multilingual Information Framework (MLIF)</b>
<b>Type of document:</b>	New Work Item Proposal
<b>Organization:</b>	ISO TC37 SC4
<b>Year:</b>	2006
<b>Description:</b>	This proposed standard aims to define a specification for a computer-oriented representation of multilingual data. It could be applied to a large variety of applications such as translation memories, localization, computer-aided translation, multimedia, or electronic document management. The MLIF will introduce a metamodel and data categories that will be integrated within the TC37 data category registry.
<b>Work flow stage:</b>	1.2, 1.3, 2.3, 2.11
<b>Business value:</b>	This standard has the potential to improve interoperability of authoring, translation, and localization applications.
<b>Web site:</b>	<a href="http://www.tc37sc4.org/">http://www.tc37sc4.org/</a>

## Terminology and lexicology

Managing terms in the source language and during translation is an essential part of the localization process. As opposed to words that have a common meaning, and words that have a purely syntactic purpose (articles, prepositions, pronouns, and so forth), which are intuitively handled by translators, terms frequently have a narrow meaning in a given context (technical, scientific, etc.). Therefore, they must be used in a consistent, standard fashion both in the source language and in translation. This is why translators frequently compile dictionaries to assist them during translation, and why many products are accompanied by a glossary.

Terms are smaller units than full translation segments, which are normally the length of one or more sentences. In addition, terms have properties that larger text segments do not, such as a part of speech value and a definition. Therefore, translation memory applications that focus on translation segments are not suitable for tracking terminology use and consistency throughout the document authoring and

translation process. Thus, standards such as XLIFF and TMX are not appropriate for managing terminology.

Through its Technical Committee 37, ISO has been developing terminology standards for decades. They include an SGML markup standard (MARTIF), an extensive catalog of data categories (ISO 12620), and an architectural model for terminology collections, called Terminological Markup Framework (TMF - ISO 16642). These lay the groundwork for further XML-based markup standards specific to the localization industry

ISO also publishes hundreds of standards on the vocabulary of specific fields of industry. ISO standards can be purchased at [www.iso.org](http://www.iso.org).

### Terminology principles and methods

<b>Document title:</b>	<b>ISO 704 - Terminology work - Principles and methods</b>
<b>Type of document:</b>	Draft international standard
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2005
<b>Description:</b>	Describes the general principles and methods for conducting terminology work: selecting terms, identifying concepts, writing definitions, and so forth.
<b>Work flow stage:</b>	1.7, 1.8, 1.9, 2.4, 2.10
<b>Business value:</b>	Very useful guideline teaching the fundamentals of terminology work for anyone who creates terminology resources, from database entries to simple glossaries.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Harmonization of concepts and terms

<b>Document title:</b>	<b>ISO 860.2 - Terminology work - Harmonization of concepts and terms</b>
<b>Type of document:</b>	Draft international standard
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2006
<b>Description:</b>	This standard specifies a methodological approach to the harmonization of concepts, concept systems, definitions and terms. It applies to the development of national and international terminologies in both a monolingual and a multilingual context. Revision of 2nd edition produced in 1996. Describes steps in the process such as needs assessment, analysis of concepts, identification of characteristics, comparison of concept systems, and so forth. Includes examples of equivalent concepts and non-equivalent concepts and concept systems. Provides guidelines on term harmonization. Addresses

	situation where concepts cannot be harmonized. Includes a helpful flowchart of the concept harmonization process.
<b>Work flow stage:</b>	1.7, 1.8, 2.4, 2.10
<b>Business value:</b>	Eliminate barriers to communication through clarification of differences between concepts in different socio-economic communities.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Terminology of terminology

<b>Document title:</b>	<b>1087-1 - Terminology work - Vocabulary – Theory and application</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2000
<b>Description:</b>	Establishes a basic vocabulary for the theory and application of terminology work. It provides a systematic description of the concepts in the field of terminology and clarifies how the terms are used in the field.
<b>Work flow stage:</b>	1.7, 2.10
<b>Business value:</b>	Valuable for anyone involved in the field of terminology.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

<b>Document title:</b>	<b>1087-2 - Terminology work - Vocabulary - Computer applications</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2000
<b>Description:</b>	Provides a systematic description of the concepts and terms for computer applications in terminology work.
<b>Work flow stage:</b>	1.7, 2.10
<b>Business value:</b>	Valuable for anyone involved in the field of terminology.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

<b>Document title:</b>	<b>ISO 21829 – Terminology for language resources</b>
<b>Type of document:</b>	Working draft
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2005

<b>Description:</b>	A preliminary study of terms used in ISO TC37 documents, with proposed definitions.
<b>Work flow stage:</b>	1.9, 1.7, 2.3, 2.10
<b>Business value:</b>	None at this stage.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Terminography

<b>Document title:</b>	<b>1951:1997 - Lexicographical symbols and typographical conventions for use in terminography</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO TC37
<b>Year:</b>	1997 (to be revised)
<b>Description:</b>	
<b>Work flow stage:</b>	1.14, 2.4, 2.10
<b>Business value:</b>	(Replaced by ISO 1951, below)
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Dictionary entries

<b>Document title:</b>	<b>ISO 1951 - Presentation/representation of entries in dictionaries</b>
<b>Type of document:</b>	Draft international standard
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2005
<b>Description:</b>	This is an update of the above standard 1951:1997. This version extends the scope beyond lexicographical symbols and typographical conventions to provide a much wider set of standards and guidelines covering the entire content and structure of a modern dictionary entry.
<b>Work flow stage:</b>	1.14, 2.4, 2.10
<b>Business value:</b>	Essential for dictionary publishers.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Preparation of terminology standards

<b>Document title:</b>	<b>10241 - International terminology standards - Preparation and layout</b>
<b>Type of document:</b>	Published standard (under revision)

<b>Organization:</b>	ISO TC37
<b>Year:</b>	1992
<b>Description:</b>	Establishes rules for preparation, working procedures, entry structure and other considerations for developing standard terminologies. Currently being revised and will be reissued in two parts (see below).
<b>Work flow stage:</b>	1.14, 2.4, 2.10
<b>Business value:</b>	Essential for any group involved in establishing a standard terminology.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

<b>Document title:</b>	<b>10241-1 - Terminological entries in standards - Part 1: General requirements</b>
<b>Type of document:</b>	Standard under development
<b>Organization:</b>	ISO TC37
<b>Year:</b>	(target date) 2007
<b>Description:</b>	A specification designed to standardize the way terminology is included in ISO standards. Currently it is found in the “Terms and Definitions” sections of standards but there is a great deal of variation in the completeness, structure, and content of this section. Many ISO committees are not aware of the standards and guidelines that should apply for this section.
<b>Work flow stage:</b>	1.14, 2.4, 2.10
<b>Business value:</b>	Standardizing the terminology in ISO standards will facilitate the migration of this information into the future ISO TermBank.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

<b>Document title:</b>	<b>ISO 10241-2 Terminological entries in standards - Part 2: Localization of international terminology standards</b>
<b>Type of document:</b>	New work item proposal
<b>Organization:</b>	ISO TC37
<b>Year:</b>	n/a
<b>Description:</b>	This proposed standard will establish requirements and guidelines for the development of national equivalents of international terminology standards in languages other than the languages of the original standard, as well as multilingual versions. It will provide practical approaches for handling concept inequivalency in different languages. It will include preparation steps and layout guidelines. There are plenty of examples, based mostly on adaptations between English and Japanese.

<b>Work flow stage:</b>	1.3, 1.14, 2.4, 2.10
<b>Business value:</b>	International cooperation and trade need harmonized multilingual terminologies for efficient communication.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Alphabetical ordering

<b>Document title:</b>	<b>ISO 12199 - Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2000 (reconfirmed in 2005)
<b>Description:</b>	
<b>Work flow stage:</b>	1.14, 2.4, 2.10
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### SGML encoding

<b>Document title:</b>	<b>12200 - Computer applications in terminology - Machine-readable terminology interchange format (MARTIF) - Negotiated interchange</b>
<b>Type of document:</b>	Standard
<b>Organization:</b>	ISO TC37
<b>Year:</b>	1999
<b>Description:</b>	SGML terminology interchange format derived originally and significantly modified from TEI-Term, which later evolved into LISA TBX.
<b>Work flow stage:</b>	1.14, 2.4, 2.10
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Bibliographic references

<b>Document title:</b>	<b>12615 - Bibliographic references and source identifiers for terminology</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2004

<b>Description:</b>	Specifies requirements for recording bibliographic information in terminology resources, including data elements.
<b>Work flow stage:</b>	1.14, 2.4, 2.10
<b>Business value:</b>	Facilitates the validation of information in terminology records by providing a standard way of representing the source of the information.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Translation-oriented terminography

<b>Document title:</b>	<b>12616 - Translation-oriented terminography</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2002
<b>Description:</b>	Describes the application of terminology theory and practices for projects whose goal is to provide terminology resources to assist in the translation process.
<b>Work flow stage:</b>	1.9, 1.14, 2.4, 2.10
<b>Business value:</b>	Valuable for the translation industry, providing guidelines on how to apply terminology methodologies specifically to benefit translators and the translation process. It describes methods to enable translators to record, maintain and retrieve terminological information.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Terminology management systems

<b>Document title:</b>	<b>26162 - Computational aids in terminology - Design, implementation and use of terminology management systems</b>
<b>Type of document:</b>	Committee draft
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2006
<b>Description:</b>	This is an update of an older ISO standard (12628). This upcoming standard provides guidelines about designing, implementing, and using terminology databases, including, user-centered design, user interfaces, data modeling, data interchange formats, software development process, and more.
<b>Work flow stage:</b>	1.7, 1.9, 1.14, 2.4, 2.10
<b>Business value:</b>	Essential for anyone involved in establishing a terminology database, whether developing an entirely new database or customizing an off-the-shelf product.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

## Data categories

<b>Document title:</b>	<b>12620:1999 - Computer applications in terminology - Data categories</b>
<b>Type of document:</b>	Published standard (under revision)
<b>Organization:</b>	ISO TC37
<b>Year:</b>	1999
<b>Description:</b>	A catalog of data categories for terminology collections.
<b>Work flow stage:</b>	1.7, 1.14, 2.4, 2.10
<b>Business value:</b>	Useful basis for defining data categories for a customized terminology database. Even though this standard is under revision, it remains the only official standard on terminological data categories. It is due to be replaced by the Data Category Registry described below.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

<b>Document title:</b>	<b>12620-1 - Computer applications in terminology - Data categories - Part 1: Model for description and procedures for maintenance of data category registries for language resources</b>
<b>Type of document:</b>	Committee draft
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2006
<b>Description:</b>	A formalized framework for defining data categories for all language resources, and their registration in an official registry. Subsequent parts 2, 3, and so forth, will specify specific sets of data categories for specific types of language resources (lexical resources, terminological resources, and so forth).
<b>Work flow stage:</b>	1.2, 1.3, 1.7, 1.14, 2.3, 2.4, 2.10, 2.11
<b>Business value:</b>	Any industry involved in a language discipline will need data categories. These data categories will eventually be recorded in the registry defined in this standard. Canadian language industries can contribute to the registry to help shape the language industry, and can use the registry as a reference for obtaining information about data categories used by other language industries around the world.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

<b>Document title:</b>	<b>12620-2 - Computer applications in terminology - Data categories - Part 2: Terminological data categories</b>
<b>Type of document:</b>	Withdrawn

<b>Organization:</b>	ISO TC37
<b>Year:</b>	n/a
<b>Description:</b>	A committee draft of this standard was prepared in 2003 and 2004 with the intention to produce an updated version of 12620:1999. Committee members decided to withdraw the project after it was agreed that data categories for terminology shall be registered in the Data Category Registry defined in ISO 12620-1 and therefore need not be published as a formal standard..
<b>Work flow stage:</b>	1.7, 1.14, 2.4, 2.10
<b>Business value:</b>	None.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Terminology project management

<b>Document title:</b>	<b>ISO 15188 - Project management guidelines for terminology standardization</b>
<b>Type of document:</b>	Under review
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2001
<b>Description:</b>	Guidelines for the phases, procedures, and overall management of terminology standardization projects. Covers general standardization projects and projects for international standardization bodies. It does not cover the principles of terminology work (see ISO 704), methodological approaches (see ISO 860), the preparation and layout of standards (see ISO 10241), or the development of terminological databases (see ISO 12618). Covers the various project management phases of a terminology standardization project (preparation, design, implementation, review). Includes various organizational models. Helpful flowcharts. In spite of its date, the principles are still valid today. Tools section is outdated.
<b>Work flow stage:</b>	1.7, 2.4, 2.7, 2.10
<b>Business value:</b>	Economic benefits due to more efficient communication, facilitates trade and improves productivity, facilitates access to international markets.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Terminology modeling

<b>Document title:</b>	<b>ISO 16642 - Computer applications in terminology - Terminology markup framework (TMF)</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	ISO TC37

<b>Year:</b>	2003
<b>Description:</b>	A metamodel that defines underlying structures and mechanisms needed for the computer representation of terminological data that is designed to express interchange information between different TMLs (Terminology Markup Languages).
<b>Work flow stage:</b>	1.7, 1.14, 2.4, 2.10
<b>Business value:</b>	Of particular value to anyone designing a terminology database.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Quality assurance

<b>Document title:</b>	<b>ISO 22128 - Terminology work – Effectiveness and quality of terminology products</b>
<b>Type of document:</b>	Working draft
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2006
<b>Description:</b>	This future guideline will provide a typology of the whole range of terminology products and services, with criteria for quality assessment and effectiveness of purpose. It is intended to be complementary with ISO 23185 which focuses on conformity assessment criteria of terminology resources.
<b>Work flow stage:</b>	1.4, 1.5, 1.7, 1.9, 1.14, 2.4, 2.6, 2.7, 2.10
<b>Business value:</b>	This future standard has the potential to assist the terminology industry with establishing objective criteria for measuring quality.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Assessment of terminology resources

<b>Document title:</b>	<b>ISO 23185 - Assessment and benchmarking of Terminological data holdings -- General concepts and principles</b>
<b>Type of document:</b>	Working draft
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2005
<b>Description:</b>	Provides reference criteria for evaluating terminological data holdings. Establishes a generalized model of characteristics for the usability of a terminological data holding. It covers desirable features of a terminological data collection itself, as well as aspects of maintainability and usability. Includes a typology of the kinds of human socio-economic activities that rely on terminology data. Includes a case study of evaluating terminological data

	holdings for an ISO terminology standardization project.
<b>Work flow stage:</b>	1.4, 1.5, 1.7, 1.9, 1.14, 2.4, 2.6, 2.7, 2.10
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Concept modeling

<b>Document title:</b>	<b>ISO 24156 - Guidelines for applying concept modeling in terminology work</b>
<b>Type of document:</b>	Working draft
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2005
<b>Description:</b>	This standard describes how to apply object modeling techniques to analyze terminological concepts. It proposes the use of graphic techniques developed for object oriented programming to represent the results of terminological concept analysis. It explains how to design model graphs to represent concepts and their relationships. Some instructions on how to use the concept model as an aid to writing definitions is provided.
<b>Work flow stage:</b>	1.7, 1.8, 1.14, 2.4, 2.10
<b>Business value:</b>	The methodology presented aids in the systematic study of concept-oriented terminology. It can be useful for developing object models for the organization of information, processes and data.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

### Markup of lexical entries

<b>Document title:</b>	<b>ISO 24613 - Language Resource Management - Lexical Markup Framework (LMF)</b>
<b>Type of document:</b>	Committee draft
<b>Organization:</b>	ISO TC37
<b>Year:</b>	2006
<b>Description:</b>	An abstract metamodel that provides a standardized framework for the development of computational lexicons. It provides a standard representation of lexical resources, including morphological, syntactic, and semantic information.
<b>Work flow stage:</b>	1.7, 1.9, 1.14, 2.4, 2.10
<b>Business value:</b>	The business purpose of LMF is to provide a common model for the creation and use of large scale lexical resources, to facilitate interchange and reuse.

	This standard would be beneficial to any large organization that develops lexical resources.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

## XML representation and exchange of terminology data

<b>Document title:</b>	<b>TermBase eXchange (TBX)</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	LISA
<b>Year:</b>	2002
<b>Description:</b>	Open XML-based standard format for terminological data. TBX is compliant with ISO 16642. This format is particularly suited to terminology databases that are part of computer aided translation software. It was modeled on the ISO MARTIF, with specific data categories selected for the translation and localization industry. As a kind of "lite" MARTIF, it may not be robust enough for large-scale enterprise terminological databases. Vendors of CAT tools have not yet implemented the format but are being lobbied by their clients to do so.
<b>Work flow stage:</b>	1.7, 1.14, 2.4, 2.10
<b>Business value:</b>	Would greatly facilitate the exchange of terminology assets between CAT tools thereby increasing leverage of terminology data by the translation industry.
<b>Web site:</b>	<a href="http://www.lisa.org/standards/tbx/">http://www.lisa.org/standards/tbx/</a>

TermBase Exchange is a standard format for the exchange of terminology for the localization industry. TBX is largely based on DXLT. TBX is designed to support the analysis, representation, dissemination, and exchange of information from human-oriented terminological databases (termbases). TBX complies with the relevant ISO terminology standards.

TBX essentially provides a standard XML structure for terminological entries that might be recorded in a database or other terminology management system. Terminology collections that wish to meet this standard, such as for the purpose of exchange among applications or among organizations, could either adopt TBX as a native format, or map to TBX from a proprietary format. TBX provides some flexibility to account for some unique features of proprietary terminology collections. TBX uses the features for encapsulating inline markup that are provided in TMX.

TBX is not yet widely adopted by the localization industry, primarily because the localization industry is only just beginning to recognize the importance of terminology management and implement appropriate tools and processes. TBX is expected to gain importance as terminology requirements increase for localization.

<b>Document title:</b>	<b>TBX Lite</b>
<b>Type of document:</b>	Proposed standard
<b>Organization:</b>	LISA
<b>Year:</b>	n/a
<b>Description:</b>	A standard under development which would be a lighter version of TBX, comprising less data categories and a simpler entry structure. No draft is currently available.
<b>Work flow stage:</b>	1.7, 1.14, 2.4, 2.10
<b>Business value:</b>	Facilitate standardization and exchange of terminology data for terminology repositories of limited complexity, which includes the vast majority of data repositories. TBX-lite is expected to increase the compliance rate of smaller terminology collections to TBX, thereby making it easier to merge and integrate terminology collections into different repositories.
<b>Web site:</b>	<a href="http://www.lisa.org/term">http://www.lisa.org/term</a>

<b>Document title:</b>	<b>TBX Link</b>
<b>Type of document:</b>	Standard under development
<b>Organization:</b>	LISA
<b>Year:</b>	2005
<b>Description:</b>	An XML namespace that enables specific identified terms within an XML document to be linked to a specific TBX XML document. Provides a rigorous notation for linking embedded terms in an XML document to their entries in a TBX document or a TBX database repository.
<b>Work flow stage:</b>	1.4, 1.7, 1.9, 1.14, 2.4, 2.10
<b>Business value:</b>	Useful for organizations that author documents in XML and maintain a terminology database as it allows terms in documents to be directly linked to the database. This can be used for example to retrieve the definition of a term for end-user applications such as hover-help.
<b>Web site:</b>	<a href="http://www.lisa.org/standards/tbxlink/">http://www.lisa.org/standards/tbxlink/</a>

### XML representation of lexical and terminological data

<b>Document title:</b>	<b>DXLT – (Default) XML representation of Lexicons and Terminologies</b>
<b>Type of document:</b>	SALT working draft
<b>Organization:</b>	SALT – Standards-based Access to Lexicons and Terminologies
<b>Year:</b>	2000

<b>Description:</b>	A framework for XML formats designed to facilitate the exchange of lexicons and terminologies, and in particular, between those used for human translation and those used for machine translation. Much of it is based on the ISO TC/37 specification MSC – MARTIF with Specified Constraints, but other standards were considered such as OLIF. It is the basis from which TBX was developed by OSCAR standards committee of LISA. DXLT is also compliant with ISO 16642 Terminology Markup Framework (TMF).
<b>Work flow stage:</b>	1.7, 1.14, 2.4, 2.10
<b>Business value:</b>	This standard provides a solid theoretical and XML compliant foundation for developing a terminology collection.
<b>Web site:</b>	<a href="http://www.ttt.org/oscar/xlt/dxlt.html">http://www.ttt.org/oscar/xlt/dxlt.html</a>

### Interchange of machine-readable lexicons

<b>Document title:</b>	<b>OLIF - Open Lexicon Interchange Format</b>
<b>Type of document:</b>	Open standard
<b>Organization:</b>	OLIF 2 Consortium
<b>Year:</b>	2002
<b>Description:</b>	A markup format for lexical data that focuses on the interchange of data among lexbase resources from various machine translation systems.
<b>Work flow stage:</b>	2.4, 2.6, 2.10
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www.olif.net">http://www.olif.net</a>

Open Lexicon Interchange Format was originally developed in the context of the European Community OTELO project. OLIF was the result of a collaboration to define a common format for representation of lexical information for Natural Language Processing (NLP) tools, in particular for use with machine translation. OLIF2 is a reformulation in XML of OLIF.

### Terminology applied to business modelling

<b>Document title:</b>	<b>Semantics of Business Vocabulary and Business Rules (SBVR)</b>
<b>Type of document:</b>	Specification
<b>Organization:</b>	Object Management Group (OMG)
<b>Year:</b>	2006
<b>Description:</b>	This specification defines the vocabulary and rules for documenting the semantics of business vocabulary, business facts, and business rules; as well as an XMI schema for the interchange of business vocabularies and business rules among organizations and between software tools.
<b>Work flow stage:</b>	1.1, 1.2, 1.4, 1.7, 1.8

<b>Business value:</b>	<p>A business vocabulary contains terms and definitions of concepts that an organization or community uses in the course of doing business.</p> <p>SBVR is designed to support interchange of business vocabularies and rules among organizations by enabling linguistic analysis of text. It is applicable to all kinds of business activities. It is designed to be used for business purposes, independent of information systems designs.</p> <p>SBVR adopts many of the theories and principles of terminology and applies them to business modelling. As such it has a potential to promote the use of terminology methods and services in business applications.</p>
<b>Web site:</b>	<p><a href="http://www.businessrulesgroup.org/sbvr.shtml">http://www.businessrulesgroup.org/sbvr.shtml</a>  <a href="http://www.omg.org/docs/dtc/06-03-02.pdf">http://www.omg.org/docs/dtc/06-03-02.pdf</a></p>

## Content management

The concept of “content management” can be considered with a large or a narrow scope depending on the primary industry focus. The broad scope would be content management as it applies to any business process. This definition is the one adopted by ALIA's Canadian Content Management Industry Technical Roadmap (2004):

Content Management research centers and firms develop technologies and applications that are used to organize, categorize, and structure information resources so that they can be stored, published, and reused in multiple ways; to automatically create, interpret and analyze unstructured information (e.g. Word documents or emails) and semi-structured information (e.g. forms or Web pages with metadata); and to extract knowledge from information.

This interpretation includes in content management emerging fields such as text analysis and knowledge discovery, and state-of-the art technologies such as natural language processing and automatic summarization. The fields and technologies are wide-ranging and the description of all the related standards is beyond the scope of this report.

For people directly involved in the language industry such as writers and translators, content management refers more specifically to the technologies and applications that are used to organize, categorize, and structure information so that it can be stored, published, and reused in direct relation to their work. One of the greatest challenges to content developers and translators today is how to manage the huge and continuously increasing volumes of information in any typical language-based enterprise. A “content management system” in an enterprise could refer to the centralized libraries of electronic documents provided with features and interfaces to enable collaborative authoring, version control, and archiving. A “content management system” for a translator could refer to a translation work flow

management system that incorporates file management capabilities.

As organizations increasingly adopt content management systems and integrate them into business processes, there is an urgent need for common standards for accessing, exchanging, managing, and integrating content. Standards are needed to provide guidance to clients, vendors, and system integrators on how to design their content-centric applications and business processes. These standards would facilitate the seamless exchange of information and files between the authoring stage and the translation stage of content development.

This section describes some of the standards that can further the development of content management applications for content developers and translators. Aside from some of the key standards such as RDF, standards for the broader scope of content management have not been included. It should also be noted that some of the standards for content management-related tasks are described in previous sections of this document, such as standards for markup and terminology management

### Resource Description Framework (RDF)

<b>Document title:</b>	<b>RDF – Resource Description Framework</b>
<b>Type of document:</b>	W3C recommendation
<b>Organization:</b>	W3C
<b>Year:</b>	2004
<b>Description:</b>	A set of semantic Web specifications defining a framework for representing information on the Web. It allows structured and semi-structured data to be exchanged across different applications. It provides facilities to enable automated processing of Web resources, thus providing the basic features to support the Semantic Web. RDF can be used for a variety of content management purposes from search to cataloging and knowledge discovery.
<b>Work flow stage:</b>	1.2, 1.3, 1.6, 1.8, 1.10, 1.13, 2.1, 2.11, 3
<b>Business value:</b>	RDF will make information and data more interoperable through the Web and other applications. It will also enable better machine interpretation of the “meaning” of content. This will enable enterprises to much more effectively manage content and greatly increase the effectiveness of search applications.
<b>Web site:</b>	<a href="http://www.w3.org/RDF/">http://www.w3.org/RDF/</a>

### Distributed authoring

<b>Document title:</b>	<b>RFC 2518 - WebDAV - World Wide Web Distributed Authoring and Distribution</b>
<b>Type of document:</b>	Standard protocol
<b>Organization:</b>	IETF Delta-V Working Group

<b>Year:</b>	2002
<b>Description:</b>	A protocol for a standard infrastructure for asynchronous collaborative authoring across the Internet. It will provide a standard interface between authoring tools and Web content, with version management and locking and management of metadata such as author and last modification date. Essentially, it is a set of extensions to the HTTP protocol which allows users to collaboratively edit and manage files on remote Web servers. The first standard in the series, Versioning Extensions to WebDAV, has been published.
<b>Work flow stage:</b>	1.2, 1.4, 1.6, 1.11, 1.13, 2.1, 2.5, 2.8, 2.9, 3
<b>Business value:</b>	WebDAV would enable organizations to more easily collaborate by using the Internet.
<b>Web site:</b>	<a href="http://www.webdav.org/deltav/">http://www.webdav.org/deltav/</a> <a href="http://www.webdav.org/">http://www.webdav.org/</a>

### Interoperable Enterprise Content Management (iECM)

<b>Document title:</b>	<b>iECM - Interoperable Enterprise Content Management</b>
<b>Type of document:</b>	n/a
<b>Organization:</b>	AIIM
<b>Year:</b>	n/a
<b>Description:</b>	A new initiative that aims to create an international standard comprised of SOA and Web services to facilitate the interoperability of enterprise content management related systems, portals and enterprise applications. It will define functional requirements for process-oriented Web services to enable disparate content management systems and environments to interoperate, thereby facilitating the exchange, integration, and management of structured and unstructured content.
<b>Work flow stage:</b>	3
<b>Business value:</b>	This standards activity has the potential to enable the automation of many content management services over the Internet, and to enable pervasive access to directly relevant content on the Internet.
<b>Web site:</b>	<a href="http://www.aiim.org/standards.asp?ID=29284">http://www.aiim.org/standards.asp?ID=29284</a>

### Corpus Encoding Standard (CES)

<b>Document title:</b>	<b>CES - Corpus Encoding Standard</b>
<b>Type of document:</b>	Published standard
<b>Organization:</b>	EAGLES
<b>Year:</b>	2002

<b>Description:</b>	Specifies a minimal encoding level that corpora must achieve to be considered standardized in terms of descriptive representation (marking of structural and typographic information) as well as general architecture, so as to be maximally suited for use in a text database. It also provides encoding specifications for linguistic annotation, together with a data architecture for linguistic corpora.
<b>Work flow stage:</b>	1.2, 1.3, 1.6, 1.8, 1.10, 1.13, 2.1, 2.5, 2.11
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www.cs.vassar.edu/CES/">http://www.cs.vassar.edu/CES/</a>

### Web Ontology Language (OWL)

<b>Document title:</b>	<b>OWL – Web Ontology Language</b>
<b>Type of document:</b>	W3C Recommendation
<b>Organization:</b>	W3C
<b>Year:</b>	2004
<b>Description:</b>	A set of specifications that define a language that is intended for describing ontologies in a manner that makes them compatible with the architecture of the Web. It is intended for use by applications that need to process the content of information rather than just presenting it. It provides additional vocabulary to represent semantics that extend the capabilities of XML and RDF to interpret content. OWL makes use of the URI and RDF standards. OWL is a revision of the DAML and OIL web ontology language.
<b>Work flow stage:</b>	1.8, 1.13, 1.14, 2.10, 3
<b>Business value:</b>	Together with RDF and XML, OWL is one of the standards that will enable intelligent processing of information over the Web. Any organization that wants to make its data optimally usable and retrievable on the Web should investigate implementing an ontology using OWL to enhance its XML program.
<b>Web site:</b>	<a href="http://www.w3.org/2004/OWL/">http://www.w3.org/2004/OWL/</a>

### Java Specification Request (JSR)

<b>Document title:</b>	<b>JSR 283 – Content Repository for Java Technology API Version 2.0</b>
<b>Type of document:</b>	Specification (in progress)
<b>Organization:</b>	JCP – Java Community Process
<b>Year:</b>	2006
<b>Description:</b>	A project whose aim is to produce a content repository API that provides a standard way to access content bi-directionally and granularly. The API would enable such tasks as author based viewing, full text searching, access control,

	content categorization and event monitoring. JSR 283 is a further enhancement of JSR 170.
<b>Work flow stage:</b>	1.4, 1.6, 1.8, 1.13, 1.15, 2.5, 2.9, 3
<b>Business value:</b>	The databases underlying many applications aren't particularly suited for content management, due to requirements specific to content management for handling objects such as documents and images. Content repositories add functionality to the underlying databases to make them more effective for content management. An API is required to allow applications to interact with the content repository. JSR 170 defines a standard Java-based content repository API. JSR 170 facilitates the development of connections to content repositories. It also enables the same content repository to be accessed by different applications and portals, and the same applications to connect to different content repositories. It benefits enterprises by opening access to content repositories, eases development, and streamlines repository management.
<b>Web site:</b>	<a href="http://www.jcp.org/en/jsr/detail?id=283">http://www.jcp.org/en/jsr/detail?id=283</a>

<b>Document title:</b>	<b>JSR 170 – Content Repository for Java Technology API</b>
<b>Type of document:</b>	Final specification
<b>Organization:</b>	JCP – Java Community Process
<b>Year:</b>	2006
<b>Description:</b>	The goal of a content repository API is to abstract the details of application data storage and retrieval such that many different applications can use the same interface, for multiple purposes, without significant performance degradation. Content services can then be layered on top of that abstraction to enable software reuse and reduce application development time.
<b>Work flow stage:</b>	1.4, 1.6, 1.8, 1.13, 1.15, 2.5, 2.9, 3
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://jcp.org/aboutJava/communityprocess/final/jsr170/index.html">http://jcp.org/aboutJava/communityprocess/final/jsr170/index.html</a>

### Generic Electronic Document Interchange (GEDI)

<b>Document title:</b>	<b>ISO 17933 - Generic Electronic Document Interchange (GEDI)</b>
<b>Type of document:</b>	Published standard (in review)
<b>Organization:</b>	ISO TC46
<b>Year:</b>	2000

<b>Description:</b>	A format for the exchange of electronic documents between computer systems. The format includes the definition of a header containing information about the requester, supplier, and format of the document and relevant bibliographic information.
<b>Work flow stage:</b>	1.6, 1.10, 1.11, 1.13, 2.1, 2.5, 2.9, 3
<b>Business value:</b>	
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

## Records management

<b>Document title:</b>	<b>ISO 23081-1 - Information and documentation - Records management processes - Metadata for records – Part 1: Principles</b>
<b>Type of document:</b>	International standard
<b>Organization:</b>	ISO TC46
<b>Year:</b>	2006
<b>Description:</b>	This standard covers the principles that govern records management metadata. It establishes the importance of records management metadata in business processes and identifies the different roles and types of metadata that support business and records management processes, and provides a framework for managing metadata. It will later be complemented by two additional parts covering implementation and evaluation.
<b>Work flow stage:</b>	1.2, 1.6, 1.10, 1.11, 1.13, 2.1, 2.5, 2.8, 2.9, 2.11, 3
<b>Business value:</b>	This standard supports business by enabling records to be used in any application or information system. Note that the standard addresses records management metadata from a general perspective, and not metadata specific to the language industry. Nevertheless the general principles apply irregardless of the business industry.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

<b>Document title:</b>	<b>ISO 15489-1 - Information and documentation - Records management – Part 1: General</b>
<b>Type of document:</b>	Standard under revision
<b>Organization:</b>	ISO
<b>Year:</b>	2001
<b>Description:</b>	Provides a high-level framework for record keeping and outlines the benefits of records management, and defines responsibilities of the actors involved. It also addresses records management requirements, the design of recordkeeping systems, and the various processes involved such as record capture, retention,

	storage, and access. It also covers audit operations and staff training.
<b>Work flow stage:</b>	1.2, 1.6, 1.10, 1.11, 1.13, 2.1, 2.5, 2.8, 2.9, 2.11, 3
<b>Business value:</b>	This standard assists businesses to apply the general quality control principles of ISO 9001 with respect to managing document and information records.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>

<b>Document title:</b>	<b>ISO 15489-2 - Information and documentation - Records management – Part 2: Guidelines</b>
<b>Type of document:</b>	Standard under revision
<b>Organization:</b>	ISO
<b>Year:</b>	2001
<b>Description:</b>	Provides practical and detailed guidelines about how to implement part 1.
<b>Work flow stage:</b>	1.2, 1.6, 1.10, 1.11, 1.13, 2.1, 2.5, 2.8, 2.9, 2.11, 3
<b>Business value:</b>	This standard assists businesses to apply the general quality control principles of ISO 9001 with respect to managing document and information records.
<b>Web site:</b>	<a href="http://www.iso.org">http://www.iso.org</a>