Global Content Management – Challenges and Opportunities for Creating and Using Digital Translation Resources

Gerhard Budin
University of Vienna

Abstract

In this paper the concepts of content management and cross-cultural communication are combined under the perspective of translation resources. Global content management becomes an integrative paradigm in which specialised translation is taking place. In a case study we discuss the Global Content Management strategy of the Center for Translation Studies at the University of Vienna.

Convergence of content management and cross-cultural communication

Two different paradigms that have previously developed independently of each other have converged into a complex area of practical activities: cross-cultural communication has become an integral part of technical communication and business communication, and content management has become a process that is complementary to communication by focusing on its semantic level, i.e. its content. Specialised translation as a form of cross-cultural communication is a content-driven process, thus digital translation resources become a crucial element in content management that takes places in a globalised marketplace.

Content management has emerged as a concept that builds upon information management and knowledge management with an additional focus on content products, such as databases, electronic encyclopedias, learning systems, etc. Due to globalised commerce and trade, such products are increasingly offered on multiple markets, therefore they have to be adapted from a cultural perspective, which also includes the linguistic viewpoint. We will have a closer look at the concept of content, its transcultural dimension, and the role translation resource management plays in this area.

Reflections on concurrent trends

Economic globalisation had been a re-current development during several phases in modern history and several industrial revolutions and has been one of the crucial driving forces in the development of modern engineering, in particular computer technology. Together with rapid advances in telecommunications it was the basis for building databases and global information access networks such as the Internet. Visualisation techniques and constantly increasing storage capacities led to multimedia applications.
This increasingly powerful technology base has then been combined with terminology management practices in the form of termbases, with multilingual communication and translation requirements as well as with cultural adaptation strategies in the form of localisation methods. Language engineering applied to translation in the form of computer-assisted translation, translation memory systems, and machine translation, have recently been combined with localisation methods and terminology management for creating integrated workbenches.

On the economic level, international trade and commerce have increasingly required cross-cultural management and international marketing strategies tailored towards cultural conventions in local markets. This trend towards customisation of products has generated personalised products and services that are based on specific user profiles, customer satisfaction and quality management schemes. The emergence of information and knowledge management systems has been another key development in recent years. Computerisation and economic globalisation are the key drivers in a complex context of the information society, leading to interactive processes between linguistic and cultural diversity, professional communication needs in economic and industrial processes and technological developments. As a result, cross-cultural specialised communication and content management have emerged, both complex processes themselves, as dynamic and integrative action spaces in society.

What is Content?

While terms such as data, information, knowledge, have been defined many times so that we can compare and ideally synthesize these definitions, the term content has not been defined so often. But since this term is essential for our discussion here, and since it is used so often in terms such as content management, eContent, content industry, etc., we have to take a closer look at what this term actually means.

In a modest attempt at distinguishing the different conceptual levels, an iterative and recursive value-adding chain emerges:

\[
\text{data} + \text{interpretation} = \text{information} + \text{cognitive appropriation} = \text{knowledge} + \text{collective representation (in potentially multi-media and multi-modal forms) for specific ways of utilization} = \text{content}
\]

Each higher level of complexity integrates diverse elements of the lower level. Usability aspects are most important on the content level. All lower levels remain crucial on the higher levels, e.g. data management is still an important part of content management.

Looking at the generic concept behind the word content, we would say: Content is what is contained in a written document or an electronic medium (or other containers of such types). We would expect that any content has been created by humans with certain intentions, with goals or interests in their minds. So we can confirm that content is usually created for specific purposes (such as information, instruction, education, entertainment, arts, etc.).

Content is often created in specific domains (arts, sciences, business/industry, government, social area, education, etc.). When specific content that was originally created in a science context, for
instance, it will have to be adapted and re-organised, in order to be able to re-use this content in other contexts, e.g. in secondary education or in industry.

Discussing the term content, we cannot avoid dealing with related terms such as data, information, and knowledge. As we have seen above, it is essential to have a clear distinction between the meanings of (the concepts behind) these terms. From an economic or business perspective, ‘data is a set of particular and objective facts about an event or simply the structured record of a transaction’ (Tiwana 2000: 59f). We derive information by condensing (summarising, eliminating noise), calculating (analysing), contextualising (relating data to concrete environments, adding historical contexts), correcting (revision of data collections on the basis of experience) and categorising data (Davenport/Prusak 1998).

Data management has always been a fundamental activity that is as important as ever. Data repositories and data sharing networks are the basic infrastructure above the technical level in order to facilitate any activity on the levels above, i.e. information management and knowledge management. The transition from information to knowledge can also be described from a systems theory point of view: a certain level of activities has to be reached, so that knowledge ‘emerges’ from information flows. Many knowledge management specialists warn companies not to erroneously equate information flows to knowledge flows.

In order to legitimately talk about knowledge, a number of conditions have to be met:

- **Cognitive appropriation:** knowledge is always the result of cognitive operations, of thinking processes. Yet knowledge is not limited to the personal, individual, subjective level. When people consciously share knowledge on the basis of directed communication processes, it is still knowledge, either referred to as collective or shared knowledge, or as interpersonal, inter-subjective, or objective knowledge. In theories of scientific knowledge, the term ‘objective knowledge’ was mainly explicated by Karl Popper (1972) and is the result of regulated research processes such as hypothesis testing, verification, proof, etc., and that is written down in science communication processes. This is the justification for libraries to talk about their knowledge repositories in the form of books that contain this type of knowledge, i.e. objective knowledge. But as mentioned above, this knowledge is also subjective knowledge in researchers when they created it and when they communicate about it or when they disseminate it to others (e.g. in teaching).

- **Complexity:** the level of complexity is another factor in the transition from information to knowledge. The same processes as on the previous emergence level, from data to information, are relevant: condensation of information (summarising), analysis and interpretation of information gathered, contextualisation (relating information to concrete problem solving situations, embedding and situating information in historical contexts and drawing conclusions from that, correcting (revision of data collections on the basis of experience) and categorising knowledge accordingly.

- **Life span:** the validity of knowledge has to be checked all the time. Again we are reminded by Karl Popper that all knowledge is unavoidably hypothetical in nature and that no knowledge is certain for eternity. Therefore we constantly have to redefine the criteria by which we evaluate our current knowledge for its validity. Another metaphor
from nuclear physics is used for knowledge, especially in scientometrics: the ‘half life’ of knowledge is constantly decreasing, due to the increase in knowledge dynamics, not only in science and technology, also in industry, commerce and trade, even in culture, the arts, government and public sectors, the social sector, etc.

In knowledge management, three basic steps in dealing with knowledge are distinguished (Nonaka/Takeuchi 1998, Tiwana 2000: 71ff, etc.):

• Knowledge acquisition: learning is the key for any knowledge management activity
• Knowledge sharing: the collaborative nature of knowledge is the focus
• Knowledge utilization: knowledge management systems have to allow also informal knowledge to be dealt with, not only formalized knowledge (this is a crucial factor in evaluating knowledge technologies for their suitability in knowledge management environments.

The focus and the real goal of knowledge management are actually directed towards content, i.e. not on the formal aspects of computing, but on what is behind the strings and codes, i.e. the concepts and the messages. When knowledge is then packaged as a product for a certain audience, presented in certain media presentation forms, then we can speak about content, which also has to be managed in specific repositories and to be processed for publishing purposes, for instance.

As soon as we introduce another dimension, that of culture and cultures, communicating content across cultural boundaries becomes a crucial issue. Since we talk about localization as the process of culturally adapting any product to a market belonging to another culture than that of the original market of a product, content also needs to be localized when it should be presented to other cultures. Translation, as a part of the complex process of localization, is one crucial step in this process, but not the only one. Content localization may very well involve more than translation in the traditional sense, i.e. we might have to re-create part of that content for another culture, or at least change fundamentally the way this content is presented to a certain culture.

Since ‘content’ is a relational concept, we have to ask ourselves, what contains something, i.e. what is the container, and what is in this container. A book (with its table of contents), for instance, is such a container, or a database with the information entered in the records as the content. A text or a term can also be containers, with the semantics of sentences and the meaning of the term as the content. But this distinction between container and content cannot be made in a very clear-cut way. We are faced with a semiotic dilemma. Form and content always interact. The medium we choose to present certain information will have some impact on this information; the structure of the information will also lead us in the choice of an adequate medium. Usually we cannot completely separate the container from the content, the form from the content, the term from the concept, the semantics from the text, the medium from the message, etc. Despite the heuristic validity and necessity of an analytical separation, we need a synthesis in the sense of a dynamic interaction, an interactive complementarity. At the same time we also might want to transform one form of knowledge representation into another one, for certain purposes and tasks, and then have to be sure that the content of each knowledge representation does not change – a difficult task.

Similar to typologies of data, information, and knowledge, we also need a content typology.
There are different criteria for distinguishing types of content:

- the domain where specific content is created in: any field of scientific knowledge, a business branch, a profession, a form of art, a type of social activity, etc. For this type of distinction, we may also differentiate different degrees of specialisation (highly technical and scientific, mono-disciplinary or multi-disciplinary, popularised, etc., depending on the audience targeted);
- the form of representation: text, picture, personal action, etc. or the medial manifestation: web site content, the ’story’ of a film, of a video, a piece of music recorded, a digitized scroll, etc.

Here we see again that the form of representing content and the medium chosen to do this is constitutive for distinguishing types of content.

First of all, the purpose of the content: instruction, education, research, aesthetic and artistic purposes, etc. Secondly, the kind of content product that is designed for a particular target audience (e.g. a multimedia CD-ROM for 6-year old children to learn a foreign language, e.g. English). In addition to a content typology, we also have to look at the structures of content. In this respect, and regardless of the content type, we can make use of terminology engineering and ontology engineering. Terminologies and ontologies are the intellectual (conceptual) infrastructures of content, both implicitly (in the form of personal or subjective knowledge of the content generator), or explicitly (as objective knowledge laid down in a specific presentation form).

So we can conclude that concepts are content units (conceptual chunks) and that conceptual structures (the links among concepts) are the structures of content. Again we have to remember that the multi-dimensional content typology will determine the concrete structures of content that users will encounter in specific products.

Global Content Management

After having investigated a little bit into the concept of content, we can now look at content management and how cultural diversity determines this practice. Since the target audience of any content product is always culture-bound, i.e. belonging to one or more cultures, with we can simply state that content management always has to take into account cultural factors in content design and all other processes and tasks of content management. The language(s) spoken by the target audience, social and historical factors, among many others, are examples of criteria for concrete manifestations of content management. Also the meta-level of content management, i.e. those who are content managers, are also culture-bound. Those who have designed and created content products, such as multimedia encyclopedias on CD-ROM, have to be aware that they themselves are belonging to at least one culture (in most cases, there will be one pre-dominant culture in such content management teams), and that this very fact will unavoidably determine the way the content of the product is designed.
In addition to the phases of creation and design of content, there are other key processes of content management at the processing stage:

- Analysis of existing content structures, segmentation of content into units, aggregation of content units into structures, condensation of content (summarization, abstracting, etc.), expansion of content into more detailed forms, transformation of content, etc.
- Presentation of content in different media and knowledge representation forms (see above)
- Dissemination of content on intranets or other web structures, on CD-ROMs, but also more traditionally in the form of books, etc.
- Sharing content in collaborative workspaces
- Using content for various purposes

Taking into consideration the differentiation between data, information, knowledge, and content (see above), we can make a parallel distinction between data management, information management, knowledge management, and content management. It is important to note that each management level is based on the one underneath, i.e. information management is impossible without data management, knowledge management needs both, data management and information management, and content management relies on all three levels below.

Now we should return to the aspect of cultural diversity and the way it determines content management. Global content design, accordingly, is an activity of designing content for different cultures as target groups and is cognizant of the fact that content design itself is a culture-bound process, as shown above.

From the field of cultural studies we can benefit when looking at definitions of what culture is: a specific mind set, collective thinking and discourse patterns, assumptions, world models, etc. Examples for types of culture are corporate cultures, professional, scientific cultures, notably going well beyond the national level of distinguishing cultures. Cultural diversity is both a barrier and at the same time an asset and certainly the raison d’être for translation, localization, etc.

Global Content Management is a complex concept with a specific structure: The term element ‘global’ stands for all the cross-cultural activities such as translation, localization, but also customization, etc. ‘Content’ includes terminologies and ontologies as its infrastructures, products and their design, user documentation, but also pieces of art, etc. And the management component includes all the processes such as markup and modelling, processing, but also quality management, communication at the meta level, etc. Usability engineering is crucial for all these components.

Content management processes cannot do without appropriate knowledge organization and content organization. Terminological concept systems are organized into Knowledge Organization Systems (KOS) that can be used for this purpose of content organization: Thesauri, Classification Systems, and other KOSs, also conceptualized and formalized as ontologies. Such ontologies may be language-related (e.g. WordNet), domain-specific (medicine, etc.), or task-oriented (operating workflows such as in robotics). In order to establish and maintain the interoperability among heterogeneous content management systems, federation and networking
of different content organization systems are necessary in order to facilitate topic-based content retrieval and exchange of content in B2B interactions.

Global Content Management may have very different manifestations. In the area of Cultural Content Management, for instance, cultural heritage technologies have developed in order to build up digital libraries, digital archives and digital museums.

Other applications of Global Content Management systems are:

- ePublishing (single source methodologies)
- eLearning (managing teaching content)
- Cyber Science (Collaborative Content Creation)
- Digital Cities and other Virtual Communities projects.

On the pragmatic level of maintaining content management systems we observe similar problems as on the level of knowledge management, that a corporate culture of knowledge sharing has to be developed and nurtured, that special communicative and informational skills are needed to share knowledge across cultures and that the dynamic changes in content require a management philosophy that is fully cognizant of the daily implications of these constant changes.

Translation resources such as translation memories and other aligned corpora, multilingual terminological resources, reference resources, etc. are typical examples of content that needs to be managed in such global action spaces.

Pragmatic Issues in Global Content Management: a Case Study on the Center for Translation Studies at the University of Vienna

Study programs offered at the Center for Translation Studies include a Bachelor in Transcultural Communication and two Master programs in Translation (specialising into technical (or LSP) translation and literary translation) and Interpretation (specialising into conference interpreting and dialogue interpreting (incl. community interpreting) and cover 14 languages.

50% of all courses held are implementing a blended learning strategy that we have developed and that covers all phases of content creation and content utilization. This strategy explicitly addresses the global aspect of learning content (cross-cultural learning), the technological dimension of using language technologies and knowledge engineering methods for creating and using digital learning objects, and the social dimension of the interaction between teachers and learners with the additional role of tutors who are trained to offer eLearning support to both, teachers and learners.

The Media Lab of the Center operates several servers with content repositories including audio content (recorded speeches that are used for interpreting classes and radio broadcasts in different languages, TV broadcasts and video recordings, web content, full text corpora (written texts), lexical resources (glossaries, terminological data bases), as well as digital learning objects (content units specifically designed for use in an eLearning environment). Using a single
sourcing approach, these central repositories are used by many different people for different purposes (research, publishing, teaching) and in different learning contexts (such as courses).

In the context of a university-wide project called DAMS (Digital Asset Management System), all content objects (that are indeed digital assets) are systematized in a taxonomy of content object types and annotated with metadata that are essential for enhanced search and retrieval functionalities. The collaborative aspect of content management is of particular importance in this context: starting at the personal, individual level, all researchers, teachers and students are able to manage their personal digital objects (their research papers, their annotations on learning objects and on published content, etc.). In the DAMS context any kind of group can be established when their members decide to work collaboratively on a certain content object, such as a joint research paper or a learning object that is collaboratively created and updated. The next level of collaborative work is a stable organisational (sub-)unit or a specific strand of work, such as all teachers in the Spanish program, all teachers offering interpreting courses, the whole faculty staff, etc. Beyond the faculty level the DAMS approach also allows to provide access to local content to colleagues at other faculties as well as to other research groups in other universities.

Our approach to Global Content Management geared towards the many different use contexts in translation studies is based on a highly granular, i.e. conceptual approach. Terminological entries are the most granular units. They are organized in a terminological data base that can be hyper-linked to all text resources that contain the terms documented in the termbase.

The following figure shows an entry of our terminological data base on risk management. In English, German and French basic terms of risk management are documented with definitions, referenced to multiple sources and indexed according to a subject-specific classification system.

The multiple purposes of this database are:

- project communication: the data base is a deliverable in a European project on risk management (WIN – Wide Area Information Network on risk management in Europe, coordinated by Alcatel, with a work package “human language interoperability” coordinated by the University of Strasbourg (Gertrud Greciano). It is being extended to more languages and the underlying data model is also being enhanced. The database will further be used as terminological input into a project-wide ontology for risk management
- web-based access to multilingual risk terminology for various target groups (such as risk experts in different domains such as geology, environmental protection, biology, civil engineering, remote sensing, etc., translators and other language professionals, information managers, students, etc.) – there is a web interface for this database
- learners’ resource: the database serves as a learners’ resource in a number of courses (e.g. on terminology management in order to study the aspects of terminological data modelling, or for translation courses where the database serves as a look-up resource for translation work in different language combinations and translation directions.
- referencing framework: a large multilingual text repository on the topic of risk management and natural hazards has been created, with these texts being analyzed with a term extraction tool in order to collect data on the use of these terms in real-life contexts.
The database is currently being integrated into the eLearning environment at the Center for Translation Studies so that it can be used in the specific learning contexts in many different courses offered in the study programs.

Fig. 1: An entry in the terminological database on risk management

Another example is given for complex learning objects that have been designed and created for use in various translation studies courses. Within the Mellange project (Multilingual eLearning for Language Engineering) (a Leonardo da Vinci II project coordinated by University of Paris Denis Diderot) a whole course program for a European Master on Translation Technologies with the relevant learning content organized in learning objects have been created. The following figure is an example for one of these learning objects that are being created for eLearning in the field of translation technologies. The example shows a course unit on information management for translators in the form of a structured hypertext-based sequenced learning unit. All these SCORM objects have been created in a collaborative way by expert teachers from the members of the Mellange consortium. The eLearning environment used for the project is Moodle. The following figure shows the use of this learning object on Moodle in the Mellange project:
The integration of eLearning, knowledge management, content management, and communication management is a crucial aspect of our blended learning strategy. For us eLearning is a multilingual, cross-cultural process. Members of learning communities, teachers, tutors, etc. increasingly have different cultural and linguistic backgrounds. This aggravates communication problems caused by the specific teacher-learner situation by adding another dimension of cross-cultural communication and its countless pitfalls that most communication partners are not even aware of. Analyzing previous and ongoing eLearning projects (e.g. Budin, 2004) as well as by looking at numerous other concrete projects it seems necessary to generalize from individual activities for formulating eLearning strategies for content development.

In the process of further expanding and developing the eLearning environment at the Center for Translation Studies, the following processes are identified requiring specific support from the DAMS initiative:

- Multilingual co-operative work for collaborative student work (including collaborative annotation of shared content such as learning objects, collaborative writing translations as well as research papers, collaborative glossary preparation, etc.)
• Cross-cultural and cross-disciplinary collaborative work (group work with students from
different countries) with specific support requirements such as meta-communicative
mediation and annotation functions to mediate between diverse cultural groups, to explain
in the required degree of explicitness specialized content to members of other disciplines
or professions
• The multiple (re-)use of multilingual language resource corpora: the re-use of corpora is
an essential element of the content strategy. At the same time there is also a dilemma that
requires specific solutions: on the one hand language resources should be highly re-
usable, which essentially means that they should be as “neutral” as possible in relation to
any specific learning context or learning goal, on the other hand learning objects should
be customised and adapted as much as possible to clearly defined learning contexts and
learning goals. The role of a separate annotation layer as well as a communication layer
becomes obvious – the learning object stays unchanged, but the annotation layer pre-
customises it to a specific eLearning context and a specific learning goal and the
communication layer enables the users (teachers, students, tutors) to post-customise it to
their learning activities.
• Modelling different competence levels for curriculum design for Bologna-type study
programmes such as bachelor and master degrees: competence modelling is an important
perspective for curriculum design as well as for the design of learning objects governed
by specified learning goals. For the translation professions and other language professions
this has become an important research area
• Search and navigation support across various content repositories
• Resource assessment for quality assurance
• Enhancing terminological coherence and consistency across all content
• Text mining with ontology systems, terminology extraction from language resource
corpora, metadata harvesting from learning content, multimedia content management and
content repository management.

The following process model has been proposed (Budin 2005) as the starting point for strategy
development for eLearning environments. The aspect of interactivity is seen as crucial for E-
Learning in the future. Interaction design has become an important principle for learning design
(interactive learning) and in fact for all modules of an eLearning environment. Interactivity is
also crucial for linking the four different dimensions of the model to each other. All four
dimensions only make sense in an interactive model as part of the whole. Therefore it is
mandatory that eLearning environments show all these dimensions, none of them can be
eliminated or simply “forgotten” as it frequently happens these days. The steering dimension is
the left one, i.e. didactic design and learning management, this includes workflows of learning
processes that are monitored, managed, and supported by teachers, also in exploratory
autonomous learning situations. Knowledge management is an important aspect for eLearning
environments, but at the moment only few academic organizations have explicit knowledge
management strategies. In that respect corporate eLearning traditions are far more advanced by
integrating knowledge management and eLearning processes. Designing and using tools for
hypermedia communication and for collaborative learning have become very important processes
for supporting the social dimension of learning. The fourth dimension is obviously another
crucial one, i.e. multilingual content development and content repository management. All four
dimensions are linked to each other in dynamic ways, as figure 3 shows:
Outlook

On the technological level a number of enabling technologies for global content management have emerged that are converging into Semantic Web technologies. Intelligent information agents are integrated into such systems. They are combined with knowledge organization systems (in particular multilingual ontologies). Semantic interoperability has also become a major field of research and development in this respect.

In the field of the so-called content industry different business models have developed that could not be more diverse: on the one hand open source and open content approaches are rapidly gaining momentum, also facilitated by maturing Linux-based applications. On the other hand national, regional and international legislation concerning intellectual property rights is becoming more and more strict and global players are buying substantial portions of cultural heritage for digitisation and commercial exploitation that might eventually endanger the public nature of cultural heritage.

Epistemological issues of global content management will have to be addressed, as well as best practices to be studied in detail in order to develop advanced methods for these complex management tasks. Managing cultural diversity in a dynamic market with rapidly changing consumer interests and preferences, with new technologies to be integrated, also requires a
strategy for sustainable teaching and training initiatives (based on knowledge management teaching and training initiatives) in this fascinating field.

In concluding it seems that strategies for multilingual learning content development for E-Learning environments require a complex approach to modelling learning processes, didactic knowledge organization, ontology creation and multilingual resource support.

References


