The Perfect Outfit

Microsoft tailors Encarta for international markets.

by Bernhard Kohlmeier

The Challenge: Take a Successful Product to Other Markets

Since 1993 Microsoft has distributed the US version of *Encarta Encyclopedia,* a digital reference title based on a US print encyclopedia. In order to broaden its user base, Microsoft attempted to take this successful product to other markets. The idea of creating the first electronic encyclopedia in multiple languages targeted for numerous markets was born.

In 1995 Microsoft decided to introduce multiple international versions in key markets, including German and Spanish versions. Bowne Global Solutions was tasked with carrying out the project and with its ongoing extension and maintenance.

It was obvious that a plain translation of the US version would have no chance in the international market, because all content (articles, style, media, and features) was targeted to US users.

From a German or Spanish point of view local content (e.g., locale-specific biographies, historic events, and geographic entities) was missing, while content developed for the US audience was confusing to local customers. Due to the target audience, examples, comparisons, institutions, laws, etc. were US-related. Further, the way of presenting and judging facts was

problematic for a non-US audience, particularly in articles concerning historical, legal, or political matters.

The challenges of executing the project were concentrated in the following areas:

- Content adaptation and creation
- Project size
- Project management
- Budget tracking
- Media selection and production
- Updating management throughout the project
- Paper management

Meeting the challenges was only possible with the close cooperation of all parties involved. Microsoft was consulted on all issues involving budget and schedule changes.

Added to the German version of the Timeline

Setting Standards for Correctness vs. Quality

Projects with a high degree of creative content that require the contribution of many individuals risk inconsistencies in style and detail.

Each team member has his own definition of what is correct and what is not. This relates not merely to "the facts," but also to their presentation style and level of detail.

Although some areas are covered in the headword review process (see below), much more detail must be considered: spelling of proper names, date format, format of the first paragraph of an article, standard for what information needs to be included, how media are referenced in articles, and what media entities need to be chosen.

To remove potential ambiguities, a style guide was created outlining the content and media requirements for the target locale. Although it is easy to create style guides of several hundred pages, the information needs to be concise enough to be usable for translators, editors, and copy-editors.

Continuous updates of the style guides and requirements documents allow for corrective action throughout the project, constantly improving the quality. This process/requirements, measurement, comparison to standards, and corrective-action loop allows project quality to be controlled in close to real time.

A Long and Winding Road

Production time of 18 (1st version of *Encarta*) to 10 months (following versions) seems practically an eternity at the start of the project. But as soon as the project manager starts to break down the total time available into time available per entity, the schedule turns out to be very tight. Each entity must be tracked according to a strict throughput model. Targets are defined and review periods broken into small and manageable increments. Intranet-based systems are applied to track individual throughput and each content entity (video, map, chart, article, etc.). Obviously, the throughput had to be monitored in conjunction with the quality of the content.

Of course it would be naive to think that every editor could accomplish the same throughput every day and every week. Content editing and creation is not comparable to conveyor-belt production. It does however, have similar requirements: a certain throughput needs to be achieved in

Scope of Encarta Localization

	Version 1	Version 2
Articles	27,000	33,000
Words	7,200,000/9,000,000	8,200,000/10,000,000
Media elements	10,000	11,000
Photos and illustrations	7,000	7,600
Audio	2,000	2,000
Maps/Charts	1,000	1,250
Videos and animations	100	110
Web links	0	1,500
Bibliographical entries	0	3,500
Yearbook builder	25 articles monthly	25 articles monthly

a given timeframe. A further obstacle is that the complexity of each article can differ significantly. Some need extensive research. Others can be edited without additional investigation. An article with only a few words could therefore require a trip to a university library, whereas an article with multiple pages could be considered straightforward. Project management must react accordingly: a simulation was created to calculate projected throughput based on actual accumulated numbers. This model allows for real-time simulation of time remaining on the project and empowers the project and content-management teams to adjust.
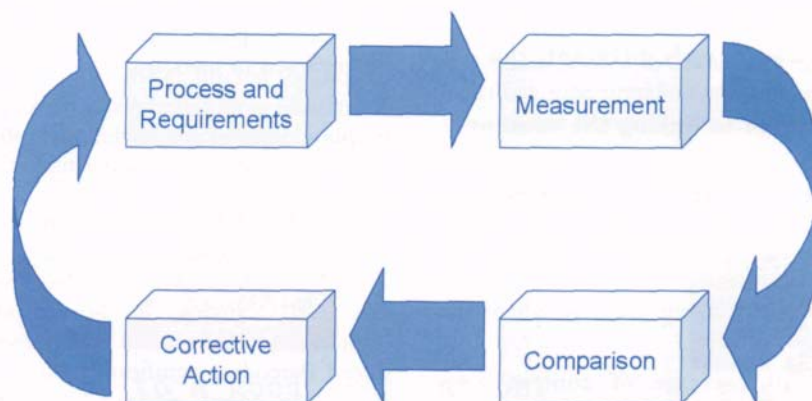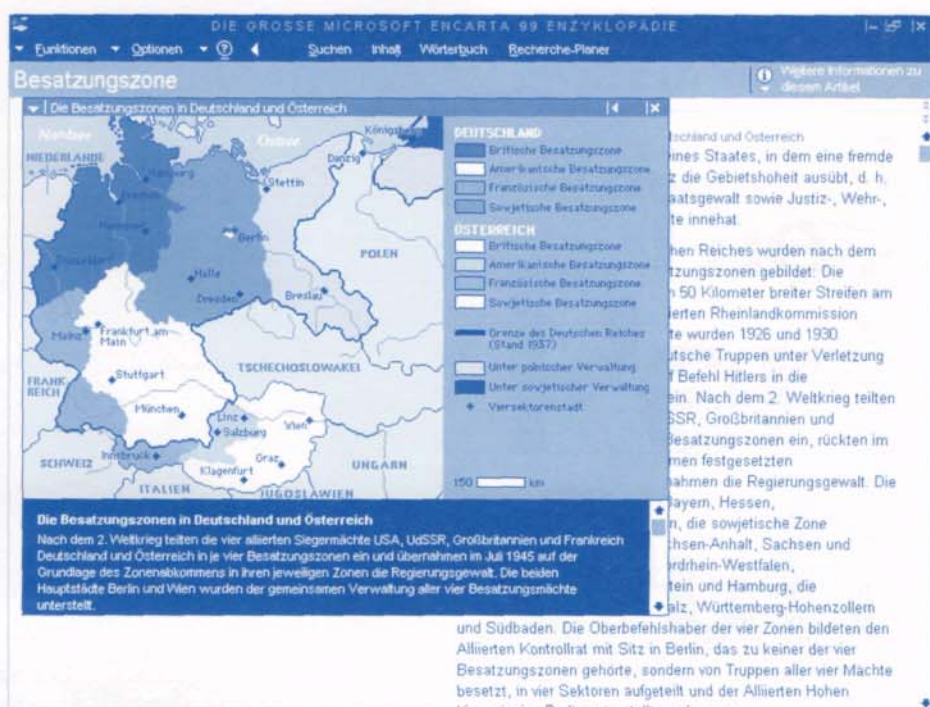
Change is the Only Constant

An encyclopedia is a living product. Our world changes every second and many of those changes affect a product like *Encarta Encyclopedia*. New explorations, changes in geopolitics, and biographical data are but a few examples. Electronic publications live from their accuracy and timeliness. This influences not only online updates but also the revision cycle from the first version on. Events must be tracked and associated to the appropriate content entities. In case of an update, the content entity needs to be assigned to the appropriate resource and QA, test, and build steps need to be carried out. For example, when Yitzhak Rabin was assassinated at a peace rally on November 4, 1995 in Tel Aviv's Kings Square, the news affected not only the statesman's biography but also many articles about Israel, the Near East, and international politics. All related articles needed to be updated with the latest information. The workflow-management system allowed for an efficient means of locating related entities and reassigning them to the proper editors.

This allows editors to finish an article while leaving open the possibility that the information could change before product release.

Process Flow for Encarta Localization

Besatzungszone

Die Besatzungszonen in Deutschland und Österreich

DEUTSCHLAND
Britische Besatzungszone
Amerikanische Besatzungszone
Französische Besatzungszone
Sowjetische Besatzungszone

ÖSTERREICH
Britische Besatzungszone
Amerikanische Besatzungszone
Französische Besatzungszone
Sowjetische Besatzungszone

Grenze des Deutschen Reiches (Stand 1937)
Unter polnischer Verwaltung
Unter sowjetischer Verwaltung
Viersektorenstadt

Die Besatzungszonen in Deutschland und Österreich

Nach dem 2. Weltkrieg teilten die vier alliierten Siegermächte USA, UdSSR, Großbritannien und Frankreich Deutschland und Österreich in je vier Besatzungszonen ein und übernahmen im Juli 1945 auf der Grundlage des Zonenabkommens in ihren jeweiligen Zonen die Regierungsgewalt. Die beiden Hauptstädte Berlin und Wien wurden der gemeinsamen Verwaltung aller vier Besatzungsmächte unterstellt.

und Südbaden. Die Oberbefehlshaber der vier Zonen bildeten den Alliierten Kontrollrat mit Sitz in Berlin, das zu keiner der vier Besatzungszonen gehörte, sondern von Truppen aller vier Mächte besetzt, in vier Sektoren aufgeteilt und der Alliierten Hohen

Customizing Media

Working on media presented similar challenges. "Entities" were tracked in much the same way as the articles. New media had to be requested in close cooperation with the Microsoft Media Acquisition group, though not everything was possible, due to copyright restrictions. New solutions had to be found in these cases and articles had to be modified accordingly.

Map production and editorial work on those media elements had to respect the

> A true cultural adaptation of the work effectively leads to a new "hyperspace" of content relationships specific to the locale. Not only is the content the user reads different, the approach to linking the content entities together is redefined.

geopolitical framework. These issues can be extremely complex and controversial in the various different markets where the product is sold.

Additional coverage of content areas sometimes requires additional media work.

The introduction of the euro is such an important event in Europe that a link in the Timeline was added in the German product (see illustration).

Managing Paper— Lots of Paper

At some stages printouts were necessary for both articles and media. Given the sheer amount of content entities, the number of printouts quickly reached 100,000 pages. Not only did the paper have to be tracked and the electronic version "locked," the hard copies had to be archived in a manner that allowed quick retrieval and a complete paper trail of work performed. Revisions marked on paper had to be proofed against the corrections entered by text-prepping staff. Another important aspect is the differentiation between a "live" hard copy currently being processed and a "dead" one used for personal reference by team members. The need to mark copies as "live" or "dead" becomes clear by imagining the scenario of two paper copies of the same article with different sets of markups.

A solution was found in a place where keeping track of many different physical entities is critical: the supermarket. Barcoding all printouts and using a print routine that differed for current versions was part of the answer. Barcoding not only simplified data entry, it also eased inventory and archiving. It also provides a "watermark" that cannot be easily reproduced. Barcoding significantly cut the time needed to access electronic versions of

paper copies, and also minimized the potential of data-entry mistakes.

Headword List Review

Content balance is key in producing an encyclopedia. Although the original version provided a great level of detail in some areas, time had to be dedicated to ensure that the level of detail applied to headwords of higher relevance to the target locale was in balance with headwords of less relevance. This also meant that numerous topics and categories needed expansion. A Spanish or German user, for example, expects more topics and a greater level of detail on European and national history than are provided in the US version.

Editorial management and the subject editors reviewed the headword list. During this phase, decisions were made whether to keep or cancel certain articles. The next step was to determine the process for

> A Spanish or German user expects more topics and a greater level of detail on European and national history than are provided in the US version.

adapting the article belonging to a headword to the target market, and which articles should be written from scratch and which translated in whole or in part.

Fact-Checking and Research

The choice of where to turn in verifying facts as well as terminology turned out to be critical. Further complicating the matter were the many potential sources of "facts" and terms—each with its own bias based on politics, geopolitical perspectives, etc. This influences large parts of a product like *Encarta Encyclopedia,* necessitating extensive fact-checking throughout the editorial effort. The product needs to represent the generally accepted view of the locale that can be found in the most important reference material. In numerous instances, the editors performed additional research to provide information that could not be translated in a straightforward manner. A good example is the way titles

of movies and books are referred to in a German encyclopedia. The title of the original work is given along with the local title. As the local title is only rarely a literal translation of the original, research had to identify the correct translation: the movie *Saving Private Ryan* has the German title *Der Soldat James Ryan* (the soldier James Ryan).

Culturally Sensitive Topics

Areas of high cultural sensitivity require special attention from the editorial team. A certain fact or name may be used in one language and locale and at the same time be inappropriate or even insulting in a different locale. Geopolitical issues are particularly sensitive due to their close relation to a nation's or people's history and sense of self.

One example from the Spanish *Encarta Encyclopedia* is the headword "Falkland Islands." Although this headword is used in English-speaking locales and is totally acceptable in the US it cannot be used in the Spanish product due to its unacceptability in the Argentine market where the product is sold. The Spanish product refers to the island group as "Islas Malvinas."

The same principle applies to disputed areas like borders, where numerous subject-matter experts and even governmental institutions in the target countries must be consulted.

Importance of Local Content

In culturally adapting the encyclopedia, it is not sufficient to merely tailor the content and style of existing articles. Completely new articles must also be created, resulting in new content relationships. This approach effectively leads to a new "hyperspace" of content relationships specific to the locale. Not only is the content the user reads different, the approach to linking the content entities together is redefined.

This challenge is illustrated in the case of the German reunification. Although the US version of the product covers the topic in several articles, the topic does not exist as a separate headword. Due to the issue's obvious importance to the German locale, the headword was created, giving substantial in-

> Although "Falkland Islands" is used in English-speaking locales and is acceptable in the US, it cannot be used in Spanish due to its unacceptability in the Argentine market. The Spanish product refers to the island group as "Islas Malvinas."

formation about the political situation that led to reunification. These included videos and a new map of the zones of occupation after the Second World War as a result of the Potsdam conference.

Another example is Bovine Spongiform Encephalopathy (BSE), better known as mad-cow disease. The scientists who discovered the disease, Hans-Gerhard Creutzfeldt and Alfons Jakob, are German and their biographies appear as separate headwords in the German version. Creutzfeldt and Jakob were sufficiently covered in the BSE article of the US product.

About the Case Study

The above is an introductory excerpt from an extensive case study on the Encarta localization, and will appear in full in a book to be published in 1999 by John Benjamins Publishing Co. in the ATA monograph series. The book, *Case Studies in Language Management*, will feature real-life examples of language-technology and management techniques at global companies, large and small.

About the Author

Bernhard Kohlmeier studied technology and cartography, earning a degree in 1985 from the Fachoberschule für Technik in Munich, Germany. He em-

Here are further changes made in the German version due to locale differences:

- German sign-language conventions differ from US standards. The respective media element required a complete adaptation.

- Due to the extremely sensitive nature of articles about National Socialism and Adolf Hitler, related content had to be heavily adapted to the German market. Special attention was given to the latest research and current discussion in Europe and Germany. In this area of history, articles were added due to their significance to the locale. This included these German keywords: Arbeitsdienst, Arier, Bücherverbrennung, Dolch-stoßlegende, Entnazifizierung, Ermäch-tigungsgesetz, Roland Freisler, Gleichschaltung, Hitler-Stalin-Pakt.

For the German user, the depth of the original US articles concerning National Socialism was insufficient (background, connections, developments, ideology, etc.). In contrast, articles concerning American history (biographies, historical incidents like the American Civil War, etc.) were too detailed for the locale. The editorial team spent many long and tense hours puzzling over such complex issues as the presentation of historical events and their relative importance.

barked on a career in data-processing by starting his own company when IBM introduced its first personal computers.

In 1992, Kohlmeier joined Gesellschaft Gesellschaft für Technische Information mbH (Gecap), a pioneer in Germany in computer-aided word and graphics processing. As vice president of strategy and planning, Kohlmeier guided localization strategies for Gecap's subsidiaries and customers, including the first German version of the Microsoft Encarta Encyclopedia, as well as other Microsoft multimedia titles and Office 97. Kohlmeier moved to Seattle and became executive vice president, strategic development, for Bowne Global Solutions.