WHAT WE LEARNED FROM FRANKLIN'S FIRST EXPEDITION ON CD

Bosko. D. Loncarevic
Pinetree Information Technology
302 - 780 Windmill Road
DARTMOUTH, NS, B3B 1T3, Canada

ABSTRACT: Saving printed material as electronic files appeared to answer all our prayers: lower production costs, saving of shelf space, easy transmission and sharing of material and, with the evolution of Internet, global access to source material. That was the theory but in practice three issues emerged as stumbling blocks: The cost of conversion, the authenticity of electronic files, and the permanence of the electronic media.

There are two approaches to electronic document storage.

Early efforts of electronic storage resembled medieval scriptoria. Modern day scribes set down at their word processor and copied (retyped) manuscripts. The oldest of these projects (Project Gutenberg) is still going strong. The spin-offs include a major effort of various Electronic Text Centres (ETC) to use special coding (SGML) to render electronic text as a faithful copy of the original. While many librarians seem to be committed to this approach, we believe that the progress in information technology offers a better approach.

The second approach substitutes electronic imaging for photo reproduction (microfilm). We explore this second approach through a case study of converting Sir John Franklin's "Journey to the Shore of the Polar Sea" (London, 1823) to a CD-ROM.

Professional scanning equipment makes it possible to produce high quality electronic images with relative ease, except that for the best results the pages should be loose so that bound books cannot be used. The electronic images can be greatly enhanced by adjusting brightness and contrast, by cropping and sizing, de-speckling and de-skewing. The purpose of this electronic processing is to get an image that can be successfully recognized by the Optical Character Recognition (OCR) software, and requires a minimum of editing to remove "suspect" words.

Converting the image to recognized text has several advantages and at least one disadvantage. The main advantages are the ease of reading the text on the computer monitor, the word search capability, and the ability to extract portions of text for pasting in other documents. The

disadvantage is that the processed text may have diverged from the original and is no longer its true representation. It is easy, however, to provide links to images of original pages for verification.

Reading an electronic book should be an experience similar to reading a paper copy. Thus one should be able to go to the Table of Contents to select a desired chapter, browse through pages or go to an index, and examine the illustrations. This can be accomplished by using "navigation buttons", and considering carefully their design and typical usage.

Finally, one of the reasons for using Franklin's "Journey ..." for this pilot project was the desire to evaluate the quality of reproduction of the engraved illustrations which represent an important feature of this work. By experimentation we have developed techniques to render these illustrations as faithful reproductions of the original, as will be demonstrated at the Conference.