The Catalogue of Canadian Art on CD-ROM
at the National Gallery of Canada

Pierre B. Landry
Associate Curator, Canadian Art, National Gallery of Canada
380 Sussex Drive, P.O. Box 427, Station A, Ottawa, Ontario, Canada, K1N 9N4
tel. (613) 990-0488; fax. (613) 990-8689; e-mail: landryp@fox.nstn.ca

Introduction

As stated in its mission, "the strength of the National Gallery of Canada lies in its collection of art, especially Canadian art, and its accessibility to the public across the country." Founded in 1880, the National Gallery of Canada now includes over 42,000 objects of fine and decorative art, of which only a small proportion is shown either in permanent installations or temporary exhibitions. Following the practice of major international institutions, the National Gallery of Canada recognized the need to make its permanent collection known to scholars, museum curators and the public at large by publishing catalogues. From the first check list printed in 1882, these catalogues have evolved over the years into sets of illustrated volumes as the collection grew and as the practice of art history changed.

By the late 1970s, when the last hard-copy edition was initiated, the catalogue of the Canadian collection alone had grown to be a set of four volumes. Progress was slow and the first volume in the series was published in 1988, while the second was released only in 1994. By then the Canadian collection had already doubled and it had become clear that the project was not only marred by an aging concept but also under serious threat from a severe shortage of resources, a situation that showed no sign of improvement in the near future. The project needed an in-depth restructuring in order to drastically reduce compilation time and production costs. The result was a proposal to produce within 2 years a catalogue covering the 16,000 works in the Canadian collection and to publish it on CD-ROM. The proposal was approved in April 1994. This paper will explore the benefits and implications of abandoning the hard-copy publication in favour of the CD-ROM, as well as the methodology involved in the preparation of the new Catalogue of Canadian Art of the National Gallery of Canada.

Scope

The new electronic edition is an exhaustive catalogue of the Canadian school, including all accessioned paintings, sculptures, prints, drawings, decorative arts, photographs, film, and video. Over 16,000 entries will represent the work of some 2,500 artists who were either active in Canada or had some connection with Canada. The information is provided in a format similar to the summary catalogues published, for instance, by the Tate Gallery, the Louvre, or the National Gallery of Art in Washington, D.C. Each entry includes the artist's name and dates, and, for the work of art, title, date, medium, support, dimensions, and source of acquisition. Over 65% of the entries will be illustrated, most of them in colour.

Text preparation

By the late 1980s, the National Gallery of Canada had achieved a high level of accuracy and consistency in the automated description of its holdings. The information is structured according to the *Humanities Data Dictionary*, designed by the Canadian Heritage Information Network (CHIN) in collaboration with its client museums, and housed on the CHIN mainframe. Although essentially a collection management system, this database includes the information traditionally found in a summary collection catalogue. During more than a decade of usage, the information contained in the database has been used, reviewed, and edited by registration and curatorial staff.

In the first phase of the CD-ROM project, missing information in a predetermined set of fields was identified and entered on the database. The full data set is then extracted and downloaded from the mainframe to a local computer, and reassembled into the CD-ROM format in English and French versions. Character set and code page configurations are strictly respected on all computers as the data is downloaded from the mainframe to a PC as ASCII text files (under DOS) and later converted to ANSI (under Windows). With code page 850 Multilingual (Latin I), data integrity is maintained during the transfer and conversion of all texts, including French.

Designing the computer programme to process the data from the original collection management format to the final CD-ROM text required a full understanding of the data structure and, also, of the sometimes unorthodox compromises reached to satisfy curatorial requirements. The reformatting of the text is accomplished with a programme written by the author in ObjectPAL, the programming language of Paradox for Windows. To illustrate the process, the following lines represent typical acquisition data as downloaded from the mainframe computer:

AQD=19620418
MO=purchase
SR=Dominion Gallery
SRMUN=Montreal
SRPR=Quebec

The programme reads each field mnemonic (AQD, MO, SR, etc), analyses and reassembles the information as:

Purchased from Dominion Gallery, Montreal, 1962

The same programme looks up proper names and place names in translation tables. In this particular case, the source municipality Montreal has a French equivalent "Montréal"; it is also found among a list of major cities for which the mention of a country or province is not essential. The name of the acquisition source is evaluated for gender and elision as the preposition and article ("au", "aux", "à la", "à l'") are inserted. The French acquisition line reads as follows:

Acheté à la Galerie Dominion, Montréal, 1962

This automated process was designed to handle over 90% of all entries without error and, if not foolproof, it drastically reduces the time required for the compilation of the texts. The repurposing of the information already stored in digital format, if not as impressive as the digital image capture, has really made this project possible. The electronic generation of a bilingual manuscript would require only a few days, while the editing may take up to four months.

Image capture

At the onset of this project, background information about scanning was available in a report produced by the Electronic Imaging Committee, a group of National Gallery staff chaired by the author and mandated to explore potential applications of imaging technologies. High resolution scanning, i.e., scanning 4" × 5" or 8" × 10" transparencies at prepress quality, was prohibitively expensive and therefore not an option. Because of their low cost and availability, 35-mm slides were the obvious choice as source material for scanning. However, the digital images would only be used for display on computer monitors since 35-mm slides do not meet the publication standards of the National Gallery of Canada.

In view of the market expectations projected for the date of release (spring 1996), the level of quality was set at 24 bit (16.7 million colours) for colour depth and full screen SVGA for size. After a first attempt at setting a format based on a standard height and width, it was decided instead to use a uniform

file size of approximately 1 megabyte. Given the unpredictable aspect ratio of the original images, the file size, which is proportionate to the surface, proved to be a better indicator of visual impact.

At first, the Kodak Photo-CD scanning process was chosen for quality and low cost:

•Colour space and resolution equal or exceed the requirements for display quality.

•The cost of scanning is minimal (CAN$1.15 per 35-mm slide).

•The Photo-CD is stable, easy to store, and does not require the acquisition of scanner or mass storage devices.

During the summer and fall of 1994, some 7,000 slides were sent out to be scanned on Photo-CD. Unfortunately the scans could not be relied on entirely. Because of the fixed aspect-ratio of the Photo-CD, systematic cropping was required to eliminate extraneous artifacts such as slide frames, easels, picture frames, mounts, etc. Even in the best cases, some adjustments to colour balance, contrast, and sharpness were necessary. Additional retouching was often required to remove dust marks, blemishes, and so on.

It was decided early on to have the final image manipulations and quality control handled by curatorial staff under the supervision of the general editor, rather than left to the colour lab involved in scanning. This, combined to the extensive post-scanning manipulations, implied that the Photo-CDs could only be used as an intermediate format. In fact, all digital images are saved as 24-bit TIFFs (Tagged-Image File Format) and archived without compression on CD-R (Compact Disk Recordable).

Unforeseen during the initial testing of the scanning process, some critical problems arose. Our stock of slides has been constituted over some 25 years and has never been stored under optimal temperature conditions. The deterioration of the colour components of the emulsion resulted in colour shifts which were grossly exaggerated during scanning. The colour lab offered colour and density correction free of charge, but the unpredictability of colours in the original works of art made it next to impossible for the scanner operator to achieve good results. The images remained too dark, often with a strong green cast, or else the highlights were burnt to the point of losing all colour. Even worse, for images on white backgrounds such as drawings and prints, the Photo-CD scans showed a systematic darkening along the sides which could only be corrected partially and with great difficulty. Faulty cropping and exaggerated contrast resulted in some 8% of the slides being returned for second scanning. As a result, post-scanning editing was more labour intensive than expected. Colour corrections and retouching with Adobe Photoshop could take as much as 40 minutes per image in extreme cases.

In December 1994, a slide scanner, the Polaroid SprintScan 35, was acquired. Scans from Photo-CD and from this inexpensive scanner were compared and, given the specific requirements of the project (low resolution, 1 megabyte, RGB colour space), the Polaroid SprintScan 35 proved to be superior on three points:

- Accuracy of colour with minimal adjustments
- Evenness of large white areas
- Speed

In view of our requirements, some of the major features of Photo-CD, such as the YCC colour space (suitable for prepress), high resolution, and stable storage became irrelevant. More importantly, scanning in-house allowed us to merge the processes of scanning and image editing, ensuring better density and colour balance from the start. It also significantly reduced the handling of slides and simplified the tracking of digital files.

At the time of writing, over 6,000 images had been scanned and corrected. Our experience has shown that technology alone fails to provide consistent results and that a high level of human intervention is necessary. Unpredictable inaccuracies arise at all steps: in the slide itself, from the initial photography and from rapid aging. Colour patches and grey scale may tell how distorted colours have become, but absolute RGB values for old Kodak scales are not available and therefore adjustments need to be approximated visually. When colour and grey scales are missing, colour corrections can only be intuited by the operator based on his personal knowledge of the work of art; often, the original work needs to be looked at. All new photography made specifically for this project now includes a three step grey scale with white, grey, and black from which density is adjusted. Hardware calibration, ambient light, operator's mood or fatigue, all these factors may seriously impact the quality and consistency of the results. For these reasons, the digital images produced for this project, as attractive as they may appear, are of little archival value beyond the context for which they were created.

Software

The Catalogue of Canadian Art on CD-ROM was designed as a visual database rather than a multimedia presentation, so the software was selected from image management packages rather than multimedia authoring systems. The software, a generic Windows image management system customized by 3D Integrated Technologies, of Ottawa, combines a simple interface with excellent image handling.

The user's interface relies on the popular metaphor of slides on a light table. Thumbnails can be browsed and a mouse click will call the full size image or the descriptive text. The entire text is searchable, adding to the product a flexibility that is not possible in hard copy. Resulting sets can be further refined and then saved on the computer's hard disk (only pointers to the images and text are saved).

The software was chosen in part for its efficient image handling in 24-bits and its ability to display in 8- or 16-bits, depending on the video display settings. It uses rapid, high quality compression, based on Lead Technologies' compression algorithm. The full version is used for image compression, text importing and indexing, and a pared down browser version will be packaged with the CD-ROM for distribution. As part of their contract, 3D Integrated Technologies implemented support for international characters, a multi-language module and a French interface.

Unlike standard multimedia productions, there is no authoring per se. The data, both images and texts, is prepared in-house and integrated using the full version of the software. As simple as the process may seem, it involves keeping track of large quantities of text and images and maintaining the links between them. The author made extensive use of Paradox for Windows for the management of the data. All documents, whether text or image, include a reference to the accession number of the work of art. As the National Gallery of Canada uses 5-digit accession numbers, these unique identifiers can be used as file names for images as well as keys for text. A strict naming convention for digital images allows database programmes to read directory contents, recognize the component that represents the accession number, and match it to a reference table that records the full name, path, date, and size of the file, as well as the name of the image editor, the source, and the type of scanner used. With these tools in hand, errors of identification are found early on during quality check and duplication of files is avoided.

Copyright

Over 10,000 works of art will be reproduced on the CD-ROM (or set of 2 CD-ROMs). A significant number of them are in the public domain, with the National Gallery of Canada holding the rights to photographic reproductions. In all other cases, permission to reproduce will have to be obtained from the rights holders, either the artists, their descendants, or artist's collectives. Considering the production costs and the limited market for Canadian art in general and for collection catalogues in particular, this product is not expected to generate substantial revenues. Therefore negotiations will rely heavily on the educational nature of the product, a position already adopted by the National Gallery of Canada for the earlier hard-copy edition of its permanent collection catalogues. Some copyright holders may be reluctant to give the permission to reproduce given the impossibility of controlling unauthorized copying of the images in the digital domain. However, it is important to note that because of the lossy

compression and the relatively low resolution, the CD-ROM images do not have any use other than computer display and therefore will be of little commercial value on their own.

Since the National Gallery of Canada retains control over the entire production, no licensing agreement with an exterior producer is required. As specified in the contract between the National Gallery and 3D Integrated Technologies, royalties will be paid for the software distributed on the CD-ROM on the basis of the number of disks sold.

Conclusion

A previously unpopular and under-funded project, the Catalogue of Canadian Art in its new form has attracted outside funding and was widely publicized by the media on the occasion of a fund-raising gala organized by the Friends of the National Gallery of Canada. Its profile was further raised by the hype surrounding the Electronic Highway, as it could be adapted for network distribution; as a matter of fact, images scanned for this project are now shared with the Canadian Heritage Information Network for a prototype version of the Humanities database distributed on Internet.

The Canadian catalogue project demonstrates that the introduction of technology can be cost effective — the overall cost will probably be less than 25% of the hard-copy edition originally planned. It also underlines that a product-oriented methodology based on well defined requirements, as opposed to archiving at optimum levels, is essential to remain within the boundaries of feasibility. Given the nature of museum collections and the large number of items involved, most of the work needs to be done in-house by either regular or contractual staff in order to maintain tight quality control and efficiently manage the movement of source materials and digital files.

The reception of the product, targeted primarily at an academic audience, remains difficult to predict. Initial responses have been enthusiastic but in the end, the sales and use of the CD-ROM for research and enjoyment will determine its validity and future.