

Use of historical documents in a digital world: comparisons with original materials and microfiche

[Wendy M. Duff](#) and [Joan M. Cherry](#)

Faculty of Information Studies
University of Toronto
Toronto, Ontario, Canada

Abstract

The paper reports on a user study of a digital library collection of Early Canadiana material, with comparisons to the use of the material in original paper and microfiche formats. The study included a survey of individuals who had used Early Canadiana in original paper, microfiche or WWW format, focus group sessions, and server log analysis. The purpose of the study was to compare use and user satisfaction across the three formats to identify ways to improve the WWW format. Although, as expected, many people liked the paper format the best, over half of those who had experience with all three formats thought that the WWW format would be most useful in their work. However, some users expressed concerns about the authenticity of the WWW format. This raises questions for digital libraries to make explicit the relationship between the original paper and digital formats. The research led to 26 recommendations. To date, over half have been implemented or are in the process of being implemented. The paper concludes with suggestions for future research.

Introduction

Early Canadiana Online/Notre Memoire En Lingne (ECO) is a digital collection of Canadiana¹ published before 1900 and includes both individual GIF images for each page of an item and the full text of all material in a searchable ASCII file. This material is also available in the original paper format and on microfiche. The digital collection consists of over 3,000 English and French language books and pamphlets and is particularly strong in literature, women's history, native studies, travel and exploration, and the history of French Canada. ECO was developed by the Canadian Institute for Historical Microreproductions in partnership with Laval University Library, the National Library of Canada and the University of Toronto Library. The ECO project began in April 1997 with a grant from the Andrew W. Mellon Foundation and funds from numerous other agencies. The overall goal of the project was to investigate the cost, usage patterns, and long term feasibility of a digital collection.

This paper reports on a usage study that was conducted in the spring of 1999. The study included a survey of people who use Early Canadiana in original paper, microfiche, or WWW format. The purpose of the survey was to compare use and user satisfaction across the three formats and to identify ways to improve the WWW format. By studying the use of the original paper and microfiche formats we wished to inform the future development of the digital format ([O'Hara & Sellen, 1997](#); [Sellen & Harper, 1997](#)).

Literature Review

The first generation of digital library research projects focused primarily on technology and content and, to a lesser extent, on user aspects ([Marchionini, 1999](#)). There is a general acknowledgement that incorporating user input into the design and development of digital libraries will result in the construction of better systems ([Dillon, 1999](#); [Jones, Gay, & Rieger, 1999](#); [Kilker & Gay, 1998](#); [Rieger & Gay, 1999](#)). Unfortunately, accepted benchmarks and goals for digital library research are not yet available, and thus evaluation methods must be iterative ([Kilker & Gay, 1998](#)).

Previous user studies of digital libraries have employed online questionnaires ([Bishop et al., 2000](#); [Hill et al., 1997](#); [Mandel, Summerfield, & Kantor, 1997](#); [Spink, Bateman, & Jansen, 1999](#)), participant observation ([Bishop et al., 2000](#); [Kengeri, Seals, Harley, Reddy, & Fox, 1999](#)), interviews ([Bishop et al., 2000](#)), focus group sessions ([Bishop et al., 2000](#); [Hill et al., 1997](#); [Kilker & Gay, 1998](#)), and transaction logs ([Bishop et al., 2000](#); [Hill et al., 1997](#)). Most of these studies depend upon self-reporting and are based on fairly small samples ([Mandel et al., 1997](#); [Spink et al., 1999](#)).

Many users want digital libraries to remove traditional barriers to information (i.e., poor access or condition of documents), while providing all of the functionality of paper documents (copy, file, annotate, edit), and to be more than "online photocopies" ([Kilker & Gay, 1998](#), p. 64). Speed in searching, retrieving, and downloading documents are also considered very important by many users ([Jones et al., 1999](#); [Kengeri et al., 1999](#)). Users want the ability to perform both simple and advanced searches, to filter their search results, and to save queries ([Kengeri et al., 1999](#)). Since subject specialists can be computer novices, systems should be responsive to the users' skill levels ([Jones et al., 1999](#)). Often small, technical barriers can create significant deterrents to access ([Bishop et al., 2000](#)). Finally, one study found that users need access to an overview of a digital library's layout ([Kengeri et al., 1999](#)).

Methods

This study employed a user survey, an analysis of server logs, and focus group sessions. By gathering data using three methods, we hoped to gain a more complete picture of the use of the material that would help in improving the design of the WWW format.

User survey

The survey utilised six questionnaires: an English and French questionnaire for each format (original paper, microfiche, and the WWW). A copy of the WWW format of the questionnaire is available at http://vax.library.utoronto.ca/htbin/eco_survey01/HTTP/1.0.

The questions on the questionnaires were identical, except in those cases where specific wording reflected differences in various formats, e.g., "printing" for microfiche and WWW, rather than "photocopying". In the WWW survey some questions contained one more or one less option than the same question on the Microfiche or Paper surveys, and the WWW questionnaire had one additional question which asked about the location of the computer they were using. The questions were grouped into four sections: (1) questions about the person's use of Early Canadiana in general; (2) detailed questions about the person's use of the Early Canadiana item just used; (3) questions about the person's use of computers and the Internet; and (4) demographic information. The survey was designed to capture information about how people generally use Early Canadiana, as well as to gather data about the use of an item in a particular format immediately after that use. The questionnaires included both closed-ended (multiple choice) questions, and open-ended questions. In designing the instrument, we consulted questionnaires used in other digital library surveys and tried to benefit from the lessons learned in the administration of these earlier surveys. For example, previous digital library projects found that response rates were extremely low for WWW-based surveys on topics of this nature ([Mandel et al., 1997](#); [Spink et al., 1999](#)). Therefore, we offered the participants in our survey an entry for a draw with a chance to win \$500.

The WWW questionnaire was available for three weeks in April, 1999 on the ECO site. The initial page of the site provided a link to a notice that gave background information about the study including details regarding the confidentiality of the data. Users could access the questionnaire from any page by clicking on an image in the upper right hand corner of the page. At the end of the questionnaire, the user could click on a button to receive an entry form for the draw. The data from the questionnaires were written to a file that was subsequently transferred to an Access database. The entry form was kept in a separate file and entered in the draw.

The questionnaires in the Paper and Microfiche surveys were colour coded (for format and language) and numbered to aid in identification. We originally planned to distribute these surveys at six sites for a time period similar to that for the WWW questionnaire. However, we had to add more sites and continue distribution of the Paper and Microfiche questionnaires for four months, from March 22 to July 16, 1999, due to the small number of questionnaires completed in the first few weeks. Three libraries distributed the Microfiche questionnaire while the Paper questionnaire was given out at six libraries. Not all libraries distributed the questionnaires for the full four months, but at any given time during this period at least one library was distributing the questionnaires. Each site

received two deposit boxes (one for completed questionnaires and one for entry forms for the draw) and a set of instructions for administering the questionnaire. The data from the questionnaires were entered into an Access database.

Two hundred and sixty-five people answered the questionnaires: 167 for the WWW, 37 for the Paper, and 61 for the Microfiche. Respondents in the survey ranged in age from under 26 to over 55. The age groups were evenly distributed with the largest group aged 36-45 (22.6%) and the smallest group aged 26-35 years (18.3%). In the case of occupations, next to "Other", students formed the largest group of respondents: 17.1% of all respondents were graduate students, 11.6% were undergraduate students, and 2.3% were high school students. The large student representation is not surprising since five of the nine libraries that distributed the Paper and Microfiche questionnaires were university libraries. Overall, most respondents identified history (32.8%) as their discipline, with the second largest group being genealogy (30.8%). Once again this is not surprising considering the content of the collection. Most of the respondents from all three surveys were users of the Internet. Over half of the respondents had been using the Internet for more than two years, and 32.6 % said they used it for more than 15 hours a week. In rating their knowledge of the Internet, 38.5% of the respondents reported "excellent".

World Wide Web server logs

The WWW server logs recorded site activity for the period April 1, 1999 until May 2, 1999, a total of 32 days. During that period, the ECO collection consisted of approximately 1,400 documents and 240,000 page images. A notice on the first page of the site notified users that their activity was being logged and would be analyzed for research purposes. The system assigned a session ID number when a user connected to the site and included this number in all communication between the server and the user's workstation. We retained the session ID when we extracted the data from the logs rather than the user's IP address for reasons of privacy. We were able to track the activity carried out in a particular session with the session ID and a time stamp, but we were not able to track a user across visits. The server logs captured 8,226 user sessions.

Focus group sessions

The focus group methodology followed the guidelines suggested by Morgan and Krueger ([1998](#)). Participants were drawn primarily from among those who indicated on the entry form for the draw in the user survey that they were interested in the focus group sessions, and whose addresses were in the Toronto area. We also recruited two groups of participants who had not participated in the user survey - students from the History, Anthropology, and English departments at the University of Toronto, and scholars who have a special interest in Early Canadiana. We identified five Early Canadiana scholars and asked them to participate in the study. All but one accepted the invitation. All participants had some experience with the ECO site prior to the focus group sessions.

The script for the focus group sessions was developed using the preliminary data gathered by the survey and the WWW server logs. The sessions were organized to proceed from general topics (to put participants at ease and establish backgrounds) to more specific topics. After a general discussion, participants watched a demonstration and then commented on the Search function and then the Browse function. This was followed by a discussion of the general attitudes and opinions toward the ECO site, and suggestions for improving the site. Each participant was asked to rank the importance of a list of new features that respondents in the surveys had identified as desirable. They rated each of the 11 features on a three-point scale (not important, somewhat important, very important). Participants then chose and ranked the three features they felt were the most important. The session concluded with a discussion of the features, a general discussion of the pros and cons of the site, and other suggestions for ECO.

Audio tapes of the focus group sessions were transcribed verbatim and analyzed using NVivo software for qualitative analysis. Transcripts were coded independently for themes by two researchers. Coding differences were resolved through discussion.

In total, the fourteen participants in the four focus groups included five scholars, two graduate students, one undergraduate student, a journalist, two business people, and three people who work in museums or libraries. Four of the five scholars participated in the same focus group session. The interests of the participants included history, genealogy, Canadian studies, anthropology/archeology, English literature and linguistics. Participants represented all age groups. All participants had used the Internet. Nine had used the Internet for over 2 years and 6 said they used it for 11- 15 hours per week. Half (7 participants) rated their knowledge of the Internet as "excellent".

Findings

We present our findings in the context of the three research questions that guided the collection and analysis of the data. The research questions were:

1. Does use of Early Canadiana differ across the three formats?
2. Does user satisfaction with Early Canadiana differ across formats?
3. How can the World Wide Web format of Early Canadiana be improved?

Does use of Early Canadiana differ across the three formats?

One section of the questionnaires gathered data about the item the respondents had just used. It asked people how long they spent using the item today, their reason for using it, how they used it, what features they used, and how satisfied they were with the format of the item they had just used. In the WWW survey, if a person had not used an item, they skipped to the next section. Some questions asked users to check more than one choice therefore in some cases the data adds to more than 100%.

The reasons for using an item differed significantly across the three survey groups ($\chi^2 = 83.929$, $df = 12$, $p = .000$) with the majority of respondents in the WWW survey (57.4%) choosing "Personal interest/hobby" or "Curiosity" as their reason. Far fewer of the respondents of the Paper (18.8%) and the Microfiche surveys (12.3%) provided these reasons. In contrast, the majority of users in the Paper (50.1%) and the Microfiche (68.5%) survey gave "Research project" or "Student assignment" as their reason for using the item while only 13.3% of respondents to the WWW survey chose this category. The results are shown in Figure 1.

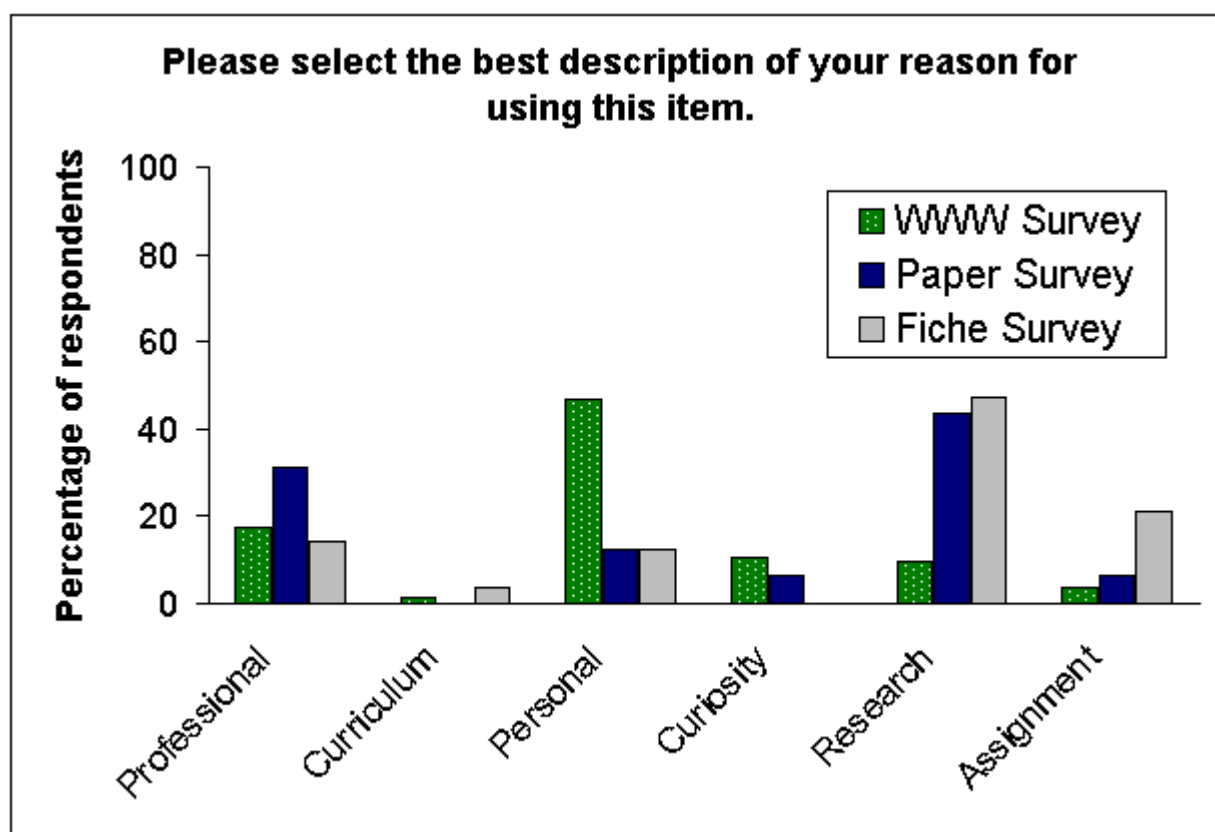


Figure 1: Reason for using item

Figure 2 shows how the respondents in the three surveys found out about the Early Canadiana item they were using. Most users in the Paper survey (75.6%) and Microfiche survey (90.1%) had a reference/citation, searched a library catalogue, or used an index or bibliography. In the WWW survey only 9.6% of the respondents reported these methods. Moreover, many respondents (35.3%) in the WWW survey found out about the item from another person. At first we wondered if the respondents who said they found out about the item from another person had really been told about the *site*, rather than the specific *item* used. However, the data from the WWW server logs, discussed in the next paragraph, seem to support the data from the WWW survey. Surfing the WWW or using a search engine

was also a frequently reported method for finding out about an item with 27.6% of the respondents of WWW survey, 22.9% users of the Microfiche survey, and even 16.2% users of the Paper survey finding out about the item this way.

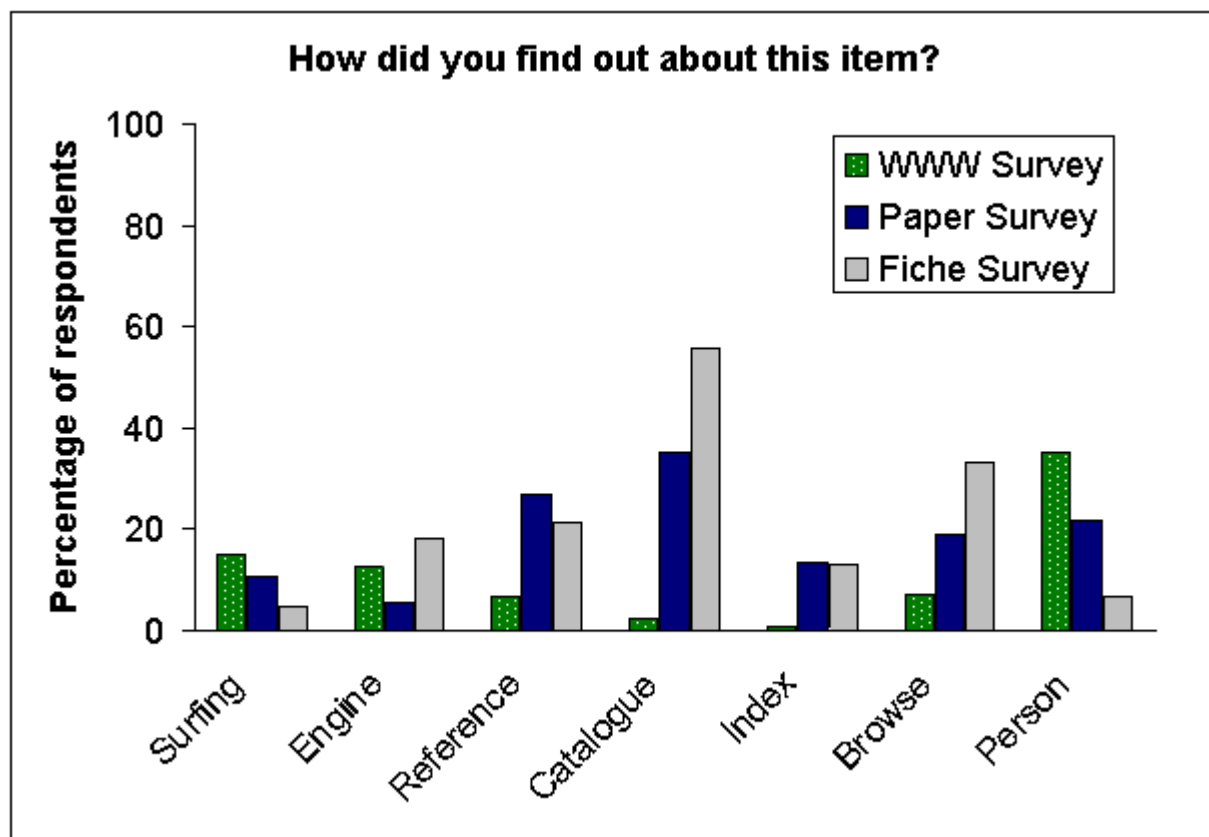


Figure 2: How users found out about the Early Canadiana item

The server logs also provided insights into how people in the WWW survey found out about the item they had just used. An examination of the first action of the 8,226 sessions logged showed that queries are the initial action for most users (55.9%). However, viewing pages was the first action in 35.0% of the sessions. This could only occur if the user had the URL for the page (passed on by a friend or colleague or in a bookmark). This supports the data from the WWW survey where 35.3% reported that they found out about the item from another person. We investigated this further by examining the logs in more detail. To better understand the sessions that started with a page view, we identified items that were frequently retrieved this way. The logs revealed two items that were involved in numerous sessions that started with a page view. Both items were biographical dictionaries. Perhaps colleagues forwarded URLs to each other for these items.

The server logs showed that only 7.7% of users began their session with browsing. Again this supports the data from the WWW survey where 7.2% of users reported that they had found out about their item by browsing the collection. The WWW server logs also showed that full-text searching was the most popular type of search with 79.3% of searches being of this type.

We asked respondents how often they had used this item in any format in the last four weeks. There were no significant differences in the responses across the survey groups (chi-sq = 6.478, df = 8, p = .594). Over 70% of respondents in each of the survey groups were using *this* item for the first time. However, we note that some of these users had used *other* items of Early Canadiana in the last four weeks. (This will be seen later in Table 4.)

The results in Table 1 show how people used the Early Canadiana item. Few people (10.4%) read the entire item. The Microfiche group had the largest percentage of respondents reading the entire item (24.6%), compared to 6.5% for the WWW and 2.9% for the Paper.

	WWW	Paper	Fiche	Total
Looked up something	34.8%	28.6%	19.3%	30.0%

Browsed through the item	24.6%	20.0%	19.3%	22.6%
Read parts of the item	32.6%	40.0%	26.3%	32.2%
Read the entire item	6.5%	2.9%	24.6%	10.4%
Other	1.4%	8.6%	10.5%	4.8%

Table 1: How people used the Early Canadiana item

Overall people did not spend much time using the Early Canadiana item. As shown in Table 2, less than one-third of the respondents (27.7%) used the item for more than 30 minutes and one-fifth (20.9%) used it for ten minutes or less. There were no significant differences across the three survey groups (chi-sq = 10.826, df = 8, p = .212).

Table 2: Time spent using the Early Canadiana item

Time spent	WWW	Paper	Fiche	Total
0 - 10 minutes	23.4%	18.9%	16.4%	20.9%
11 – 20 minutes	31.4%	29.7%	32.8%	31.5%
21 – 30 minutes	20.4%	13.5%	23.0%	20.0%
31 – 60 minutes	14.6%	8.1%	11.5%	12.8%
More than an hour	10.2%	29.7%	16.4%	14.9%

We asked people to indicate which features they had used in the Early Canadiana item. Figure 3 shows the data gathered by this question in the three survey groups. Overall, only 7.5% of people did not use any of the features listed, suggesting that users find these features useful. The respondents in the WWW survey used all traditional features less. For example a smaller percentage of respondents in the WWW survey (25.7%) used the table of contents than in the Paper survey (62.2%) and the Microfiche survey (59.0%). Similarly, a smaller percentage of users in the WWW survey (6.0%) used footnotes/endnotes, compared to the Paper survey (27.0%) and the Microfiche survey (24.6%). Also, a smaller percentage of users in the WWW survey used an index (17.4%), compared to the Paper survey (37.8%) and the Microfiche survey (32.8%). Almost 60% of the users in the WWW survey used the full-text search capabilities.

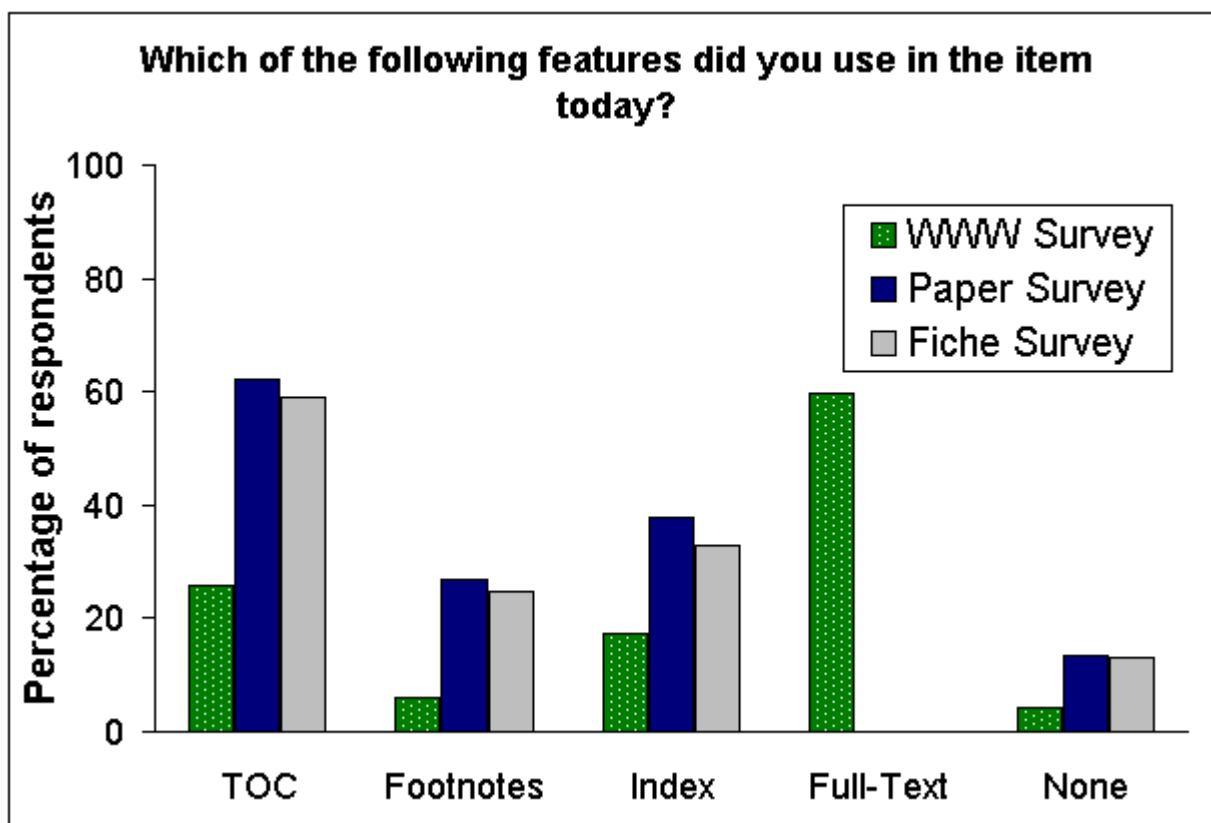


Figure 3: Features which people used
 (Percentage represents the percentage of respondents in the survey group that reported using the feature.
 Respondents could check more than one feature)

The differences in the features used may indicate that users of the WWW format do not need these features as much as users of the original paper or microfiche formats. However, there may be other explanations. For example, some of the items the respondents of the WWW survey used may not have had these features. We do not know what items the respondents had just used and it is unlikely that they used the same ones. Many items in the ECO collection are fiction and therefore are less likely to have indices or footnotes. If a respondent in the WWW survey group had just used an item that did not have an index or footnotes, he/she could not have used these features. On the other hand, access to full-text search capabilities may eliminate the need for an index.

Participants in the focus group sessions discussed these features and their value in the WWW format. Most participants stressed their importance and suggested that information about the table of contents, the table of illustrations, and any indices should be included in the bibliographic record that describes an item. They stated that the table of contents would indicate the existence of more extensive discussions of the topics of their research, e.g., indicating entire chapters rather than just the page where the search term occurred. One participant suggested that the software should create a table of illustrations when the original paper item did not have one because of its importance. The importance of an index was also mentioned several times with one participant stating that an index speeded up the research process. No one mentioned the connection between the full text search function and an index.

Another question asked respondents how often they needed to consult the original document rather than a facsimile when using Early Canadiana. There were significant differences across the survey groups ($\chi^2 = 18.804$, $df = 4$, $p = .001$). The data are shown in Table 3. Approximately half of the respondents indicated that they needed to consult the original document occasionally or frequently in the WWW group (49.7%) and Microfiche group (53.5%). Not surprisingly, the percentage was much higher in the Paper group (86.2%).

	WWW	Paper	Fiche	Total
Never	50.3%	13.9%	46.6%	44.4%
Occasionally	38.7%	55.6%	39.7%	41.2%

Frequently	11.0%	30.6%	13.8%	14.4%
------------	-------	-------	-------	-------

Table 3: How often users need to consult the original document when using Early Canadiana

We recognize that differences in use presented in this section may be due to differences in the users themselves across the three survey groups. In the remainder of this section we present data on the users. This data relates to their previous use of Early Canadiana, their experience with the Internet, and demographics.

One section of the questionnaire focused on the respondent's use of Early Canadiana *in general*. The first question asked how often they had used Early Canadiana (in any format) during the last four weeks. Users differed across the three survey groups (chi-sq = 36.090, df = 8, p = .000). The percentage who were using Early Canadiana for the first time in four weeks was largest in the WWW survey (66.9%), followed by 37.8% in the Paper survey, and 30.0% in the Microfiche survey as shown in Table 4. On the other hand, the Microfiche group had the largest percentage (18.3%) of users who had used Early Canadiana in some format more than 20 times in the last four weeks.

Table 4: Number of times people used Early Canadiana in any format in the last four weeks

Times used	WWW	Paper	Fiche	Total
This is my first time	66.9%	37.8%	30.0%	54.4%
2-5 times	18.7%	35.1%	36.7%	25.1%
6 – 10 times	7.2%	16.2%	8.3%	8.7%
11 – 20 times	3.0%	2.7%	6.7%	3.8%
More than 20 times	4.2%	8.1%	18.3%	8.0%

One question asked people to indicate which formats of Early Canadiana they had ever used. Of interest here is whether respondents in each of the surveys had used the other formats. In the WWW survey, 31 % of the respondents had used one of the other two formats. In the Paper survey, 43% of the respondents had used one of the other two formats. In the Microfiche survey, 39% had used one of the other two formats. Forty-seven respondents had used all three formats.

We collected demographic data on gender, age, occupation, and discipline. There was no significant difference in the gender distribution across the three survey groups (chi-sq = 3.322, df = 2, p = .190). The percentage of females was 46.6% overall, with 42.9% in the WWW survey, 45.7% in the Paper survey, and 56.7% in the Microfiche survey. In terms of age, the distribution across the age groups differed in the three surveys (chi-sq = 60.197, df = 8, p = .000) as shown in Table 5. In the WWW survey, there was a larger percentage than expected in the higher age categories: 76.5% of the respondents were in the three categories representing age over 35. In the Paper survey 50.0% of the respondents were over 35, and in the Microfiche survey 30.5% were over 35.

Age	WWW	Paper	Fiche	Total
<Under 26	7.0%	30.6%	47.5%	19.8%
26 – 35	16.6%	19.4%	22.0%	18.3%
36 – 45	26.1%	16.7%	16.9%	22.6%
46 – 55	28.7%	8.3%	5.1%	20.2%

Over 55	21.7%	25.0%	8.5%	19.0%
---------	-------	-------	------	-------

Table 5: Percentage of respondents in each age category

One question asked respondents about their occupations, but over 42.6% chose "Other" indicating that our list of occupations was not a good match to the actual occupations of the respondents. However, we note that the WWW group had a smaller percentage of respondents (27.5%) from the university community (undergraduates, graduate students, and university faculty/instructors) than did the Paper group (32.4%) or the Microfiche group (64.9%).

Table 6 shows the distribution of disciplines across the three survey groups. Overall, genealogists and historians are the largest categories, representing 63.6% of respondents. Genealogists were the largest group in the WWW survey, and historians were the largest group in the Paper and Microfiche surveys.

Discipline	WWW	Paper	Fiche	Total
Anthropology/Archeology	3.1%	16.1%	5.5%	5.3%
Canadian Studies	3.1%	6.5%	10.9%	5.3%
English Literature	2.5%	0.0%	0.0%	1.6%
French Literature	2.5%	0.0%	0.0%	1.6%
Genealogy	41.6%	9.7%	10.9%	30.8%
History	25.5%	51.6%	43.6%	32.8%
Law	1.9%	0.0%	3.6%	2.0%
Sociology	1.9%	0.0%	0.0%	1.2%
Other	18.0%	16.1%	25.5%	19.4%

Table 6: Disciplines of respondents

While the WWW survey was administered online, one-half of the sites that distributed the Paper and Microfiche surveys were academic libraries. This difference may have affected the type of user who answered the three surveys and resulted in a higher percentage of respondents from the university community answering the Paper and Microfiche surveys than the WWW survey. The members of this community are more likely to be younger and involved in more academic pursuits than the general population.

Discussion of use

The respondents in the three survey groups used the items differently. The respondents in the WWW survey group were more likely to be using the item for personal interest/hobby or curiosity while the users in the Paper or Microfiche groups were more likely to be involved in student assignments or research projects. When using the item, the respondents in the WWW group were more likely to have looked up something or browsed through the item than the respondents in the Original Paper or Microfiche groups. Users of the original paper and microfiche formats were more likely to have used traditional methods when locating an item such as searching a library catalogue or using a bibliography, while users of the WWW format were more likely to have found out about the item from another person. These differences may reflect differences in the types of respondents who were using the item and who answered the different surveys. Respondents to the WWW survey were more likely to be infrequent users of Early Canadiana, less likely to be part of the university community, and more likely to be involved in genealogical research.

Just before the WWW survey began, a publicity campaign was conducted, and the ECO site was chosen as Pick of the Week by Yahoo! Canada. Furthermore, the day after the survey began, the site was named CTV's Webmania Pick of the Week. This publicity may have attracted more casual or curious users, which may in turn have affected the type of respondents answering the WWW survey. Five of the ten libraries that distributed the Paper and Microfiche surveys were university libraries, the type of institution in which most large collections of the original paper and microfiche formats of Early Canadiana are held. This resulted in a more "academic" set of respondents in these survey groups.

Does user satisfaction with Early Canadiana differ across formats?

Overall, respondents were satisfied with the format of the Early Canadiana item they had just used. Across the three survey groups, 47.3% were very satisfied, another 39.4% were satisfied. There were no significant differences across the survey groups ($\chi^2 = 6.093$, $df = 8$, $p = .637$).

In addition to satisfaction ratings that were given by all respondents for the item they had just used, we looked at the preferences expressed by those who reported that they had used all three formats. Forty-seven respondents across the survey groups had experience with all three formats (original paper, microfiche, and WWW). Forty-six of these respondents (thirty-four in the WWW group, six in the Paper group, and seven in the Microfiche group) answered the three questions related to format preferences. The results for these questions are shown in Figure 3. Overall, the people who had used all three formats liked the WWW format best, the microfiche format least, and said that the WWW format would be most useful in their work. It is noteworthy that although 41.3% *liked* the original paper format most, only 17.4% said it would be *most useful* in their work.

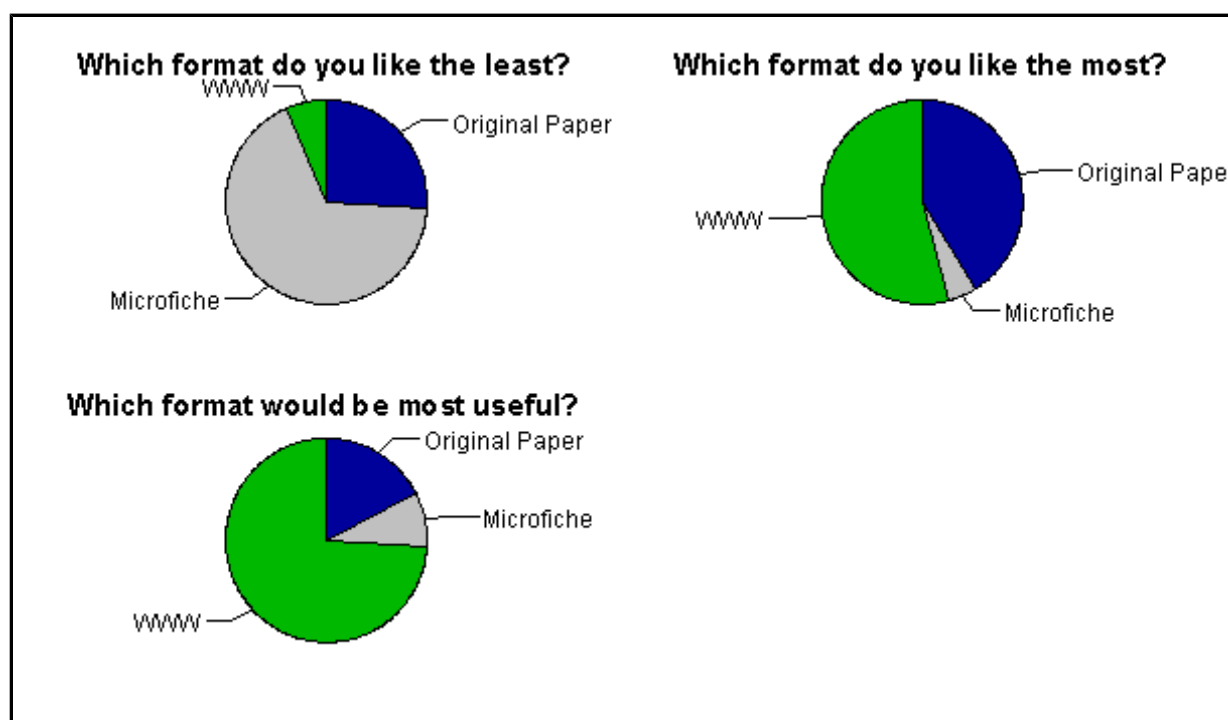


Figure 4: Preferred format of Early Canadiana

For those who had used all three formats of Early Canadiana (original paper, microfiche, and World Wide Web)

We also asked respondents to write comments about *why* they liked, disliked or found a format most useful. This information reveals important attributes of the three formats and some barriers to the use of the formats.

Reasons for liking a format most or least

Many users stated that they liked the original paper format most because it was the easiest to read and to navigate. Furthermore, it provided a sense of the whole document. They highlighted the importance of the physical attributes of the original paper and commented on its authenticity, accuracy, trustworthiness, and completeness.

I feel I've seen the real thing.

Closer to the source; certainly there is no risk of error in transcription; the further one is from original

source the more likelihood of problems - also just the physical feel of handling old docs is a pleasure.

Changes in form modify the content.

Because I like to look through the whole volume and take notice of the entire text, the binding, other illustrations and publication.

The original medium is the most authoritative and accurate.

Certain to be as complete as possible.

However, not everyone liked the original paper best, and some ranked it as the format they liked least. Their reasons included the restriction that libraries and archives placed on accessing it and copying it, the time it takes to locate information especially if no index exists, allergies, and their fear of damaging it.

Guarded by over cautious archivists who reluctantly allow you access.

I would not want to damage it and that in itself slows down the search, plus I don't want others handling it for the same reason.

I think it [original paper] is slower than microfiche. It doesn't have the convenience of the World Wide Web.

Unless there is an index, it can take a lot of time.

A few respondents liked the Microfiche format most. Respondents liked microfiche because it was easy to print and navigate, and it provided a reliable facsimile of the original. Many respondents felt it protected the original.

Easy to use, fast way to get resources.

Ability to photocopy - speeds research, allows easy future reference.

Ease of moving through document. Ease of photocopying. No conservation issues vs. with paper. Not requiring extensive technology - not just text but a facsimile of an original copy. Therefore, minimal editorial interference/margin for error.

Although some respondents liked the Microfiche format best, the majority liked it the least. Respondents suggested that the Microfiche format was inaccessible, difficult to read, resulted in poor reproductions, and required machines that were difficult to use, and often dirty. The condition of the microfiche equipment available in some libraries seems to present barriers to the use of this material.

Microfiche is a pain - difficult to read, limited type size.

Difficulty of reading handwritten stuff, finding your way around the document. Often poor machines.

As the quantitative data showed, the majority (54.3%) of the people who had used all three formats liked the WWW format the most. The prime reason was its accessibility. Many commented that this format allowed them to access material from their homes at a time that was convenient to them, a feature they applauded.

Although nothing can replace the feel of an original document, it is the ease and access of the Internet which is special: available conveniently on our computer at home, a home which is 200 km removed from the significant archives.

Convenience of being able to do research at home at the time you are free to use the Internet. Copies are obtainable just by pressing the print button and are available 24 hours a day.

Some of the respondents noted that the WWW format increased their ability to reproduce or copy the item while others appreciated the fact that they could gain access to historical information without handling fragile documents.

Ease and speed of access. Eliminates handling fragile originals. Provides access for many people --

Canadians and others, nationally and internationally -- to materials which may otherwise be totally inaccessible.

However, not everyone liked the WWW format. Some suggested that the text is difficult to read on the screen and that WWW documents are more likely to contain errors. Some users simply did not trust documents in this format.

Don't trust input - I always verify info.

Web still too difficult to search and too artificial to read.

Benefits of searching/home access but drawbacks of not being a facsimile, greater margins for copy/editorial error, selection process would not prioritize items I'm most likely to use, temptation to print (download) entire document too great - rather than do research/notes, too much time on Internet would be required etc.

The WWW format of ECO documents is a facsimile of the microfiche and it has not been altered. Even blank pages were scanned and are available for viewing. This respondent, however, must not have realized that a GIF image is a facsimile.

Reasons for choosing a format as most useful

Almost three-quarters (73.9%) of those who had used all three formats identified the WWW as the most useful format. Many users who stated they liked the original paper or microfiche format best also pointed out that the WWW format was often the most *useful* because it provided them with the ability to search quickly, to make copies and to access the material where and when they wanted.

The internet is the most useful because anyone with a computer and access to the web has the ability to do research from their home any time they wish to do research. Most web sites are very informative and list all the important details in one location.

Though the original paper is nice, the web enables excellent search capabilities downloading, printing etc. All without damage to the original.

Accessibility. I am a Ph.D. student with an 18 month old son. Need I say more?

Two respondents with disabilities commented on the convenience of the WWW format. One stated, *I think the web stuff would be easiest. I am permanently disabled and I do not travel well. I must do most of my research at home or spend as little time in the libraries or archives as I can.*

Some users rated the original paper format as the most useful because it was easy to use and navigate, it provided a sense of the whole document and one could use it without a machine. As one user put it, *I want to see the entire package and look through the whole book, [and] may possibly see other interesting articles.*

Discussion of user satisfaction

Early Canadiana in the original paper and microfiche formats present many barriers to use. Many items lack indices; therefore, locating required information can be tedious and time-consuming. Moreover, many libraries restrict the use and copying of Early Canadiana to protect the original material. Respondents appear to understand the reasons for these restrictions and depend upon microfiche to overcome these barriers. Microfiche readers, however, seem to present different obstacles. The majority of respondents did not like microfiche and rated it as the least useful format because of poor resolution, limited navigability, and faulty equipment. The WWW format presents an opportunity to overcome many of the restrictions placed on the use of the original paper format. However, Web users have high expectations. Users wanted the system to respond quickly whether they were locating relevant material, navigating through an item, using it, or printing it. Although many of the users were somewhat accepting of the time consuming nature of research using the original paper format, when they use the WWW they are less tolerant of a system's poor response time or activities that are cumbersome or overly time consuming.

The Early Canadiana Online site provides a separate GIF image of each page of an item and even includes images of

all blank unnumbered pages to protect the integrity of the original item. However, users still worried that the image might have been changed and said they would continue to rely on the original material to ensure an item's authenticity. Acceptance of the WWW format by users will depend upon the degree to which they trust it, believe it is authentic, and find it easy to use. Therefore, research on how different document attributes affect evaluations of the trustworthiness or authenticity of documents is needed.

How can the WWW format be improved?

The questionnaires solicited ideas about features that would make the WWW format more useful. Some respondents requested new functionality and discussed existing features that they found confusing or frustrating. Their comments provided insights into the characteristics and functions that helped them in their research, the features that limited their use of the WWW format, and why. The focus group sessions also provided numerous ideas for improving the site.

Respondents to the survey provided a number of suggestions that related to the search engine, the help features and the content of the site. Replies included requests to increase the comprehensiveness of the database, to provide hit-lists ranked in order of relevance, to identify the original source of the digitized image, to enable browsing by date, to supply more instructions about the search function, and to provide a cut and paste feature.

The participants in the focus groups sessions were asked to select from a list of eleven features, the three features that they would most like the system to have. The top choice was the ability to print an entire document, followed by fast response time, and then the ability to download an entire document.

In the following sections we elaborate on some of the improvements that users requested in the survey and the focus group sessions.

Downloading

ECO has treated each page of an item as a separate GIF file. Therefore, a user must view or print each page individually. Depending on the connection and the computer a person is using, the rendering of a page can be slow. Some respondents in the WWW survey criticized this feature.

It would be enormously useful to have the option to download the entire document as a PDF file, rather than having to visit each page individually. This is a time-consuming process, especially when a text is 2[00] to 300 pages long.

Is it possible to view two pages at once? The time spent in recalling each page seems long.

After the study, ECO's management decided to increase the number of pages downloaded at one time. They are presently trying to determine the optimum number of pages to download at one time. At the time of writing this number had not yet been determined.

However, participants in the focus group sessions, particularly the scholars, were pleased that the original image was reproduced as a GIF image. One participant emphasized that "*if it's scanned in, I don't want the image cleaned. I want it in the original state that the book is in*".

Highlighting search terms

Respondents to the WWW survey wanted the system to identify or highlight the section or phrase that contained the subject or term they had searched for. One respondent complained, *I couldn't immediately see how to pinpoint within the full text the phrase I had initially searched*. Participants in the focus group sessions also wanted search terms highlighted in the displayed text to shorten the time needed to identify the relevant text on a page. According to one of the scholars, this would help make searches "*as powerful and usable as possible*". Participants liked the scanned image, but they also wanted the search term highlighted in the text. However, they did not comment on which they wanted more. One participant wanted the search term that retrieved the hit list or the item displayed in the corner of the screen.

Terminology

Users had problems understanding some of the terminology used on the site. As a result, some survey respondents requested features that the system already has.

I didn't notice a "previous page/next page" choice on the screen. This would be useful when you find that the info you want continues to the next, or from the previous, page. If that capability is there, then I need another cup of coffee this morning!

Each screen contains a "Previous" and "Next" option, but this respondent must not have seen the feature, or may have misunderstood the meaning of the terms. Participants in the focus group sessions pointed out that the term "next" could mean the next hit in the hit list rather than the next page in the document. They suggested changing the terms to "Previous Page" and "Next Page". Some participants wanted to be able to go to the next item in the hit list as well as the next page of the document.

Many participants in the focus group sessions commented on other terminological problems. For example, participants did not understand "non-blank", the word used to designate pages that had text but no page number. Someone suggested that the term "unpaginated" be used instead. Most participants (both scholars and others) agreed with this idea. However, one scholar commented that the term "unpaginated" could be interpreted in more than one way. The scholars felt that the information about blank and nonblank pages should remain, but distinctions between non-blank pages and illustrations should be noted.

Some computer terms, e.g., PDF, were foreign to many users. The system offers the option of invoking a PDF copy of each GIF image for printing by clicking on the term "PDF". The majority of the participants in the focus group sessions thought that the term PDF was confusing and suggested that the phrase "Print This Page" be used instead. Someone also commented on the confusion that might be created by the appearance of the Acrobat Reader when one clicks on the term PDF.

Authenticity

Many respondents to the survey said they liked the original paper format because it was more authentic or trustworthy. The scholars in the focus group sessions were also concerned about the authenticity of the WWW format. In their opinion problems over authenticity might be mitigated if more information about the original item was provided. Other participants suggested that information about the original source of the image could help overcome questions concerning the authenticity of the material. In their opinion, the more information the system provides about an item, the more they will trust it. One participant wanted the source of the item added to the top of the page when the image was displayed. Another participant expressed concern about the authenticity for legal reasons. The Microfiche format contains information about each original document in a target page. This target page is included in the WWW format of each item, but is not obvious to users. It is placed before the title page and the system opens a document at the title page. During the discussion, participants were asked about the importance of this information. Many participants commented that the target page was very important, particularly two of the scholars, two of the researchers, and a library worker.

Formatting information

Two of the scholars mentioned including formatting information about the original item, e.g. the size of the book and the pagination. One participant commented that *"if you're a historian, you want to be able to see that it's in its original, and that's why the microfiche was so important - because it is in its original form. It hasn't been doctored at all."* The integrity of the original has been maintained in ECO by providing an "undoctored" GIF image of each page, including all blank pages. The participants also discussed the presence of blank pages and their inclusion in the bibliographic display. The history student and genealogist suggested that blank pages should be removed. The library worker disagreed stating that this information was important to bibliographers. The scholars did not comment on this issue directly, but they commented generally on the importance of replicating the original.

Searching

While everyone in one focus group session liked that the full text search function was the default, participants expressed a need to narrow searches. Among the suggestions included were the ability to limit the search by date, by illustration, by publication location, by latest update, and by language. The system allows the user to limit a search by French and English, but the participants also wanted to limit a search by other languages such as German

and Dutch. The existing choice of language (English and/or French) confused some participants. It was not clear to them what they would retrieve, even after they had searched for material. Some participants wanted to retrieve both French and English documents when they searched on an English term, while others did not.

Benefits

Many of the participants in the focus group sessions discussed the benefits of using the site and corroborated what was found in the survey comments. Participants pointed out that the system would reduce the use of original documents. Other benefits mentioned included easier access, convenience and speed of access. Three of the scholars mentioned that this site would help them conduct their research faster. A scholar and a genealogist mentioned that they would probably start reading outside their research areas and felt that this site might help them make serendipitous links in their work. One scholar noted the potential of the site as a teaching tool. Another scholar stated that it would allow greater depth in research and "*... It allows me to look at documents ... to give an anthropological gaze . . . without doing the stuff I don't like doing, which is going to the archives.*"

Discussion of improvements

The survey and focus group sessions gathered many ideas for improving the site. Many of these comments were specific to the ECO site but others provide insights into the use of digital library collections in general. Many were concerned about the trustworthiness or authenticity of the WWW format. Providing an exact facsimile of the original material augmented by information about the original material could reduce some of these concerns. However, participants in the focus groups did not totally agree. For example some wanted the blank pages removed while others wanted the WWW format to be a faithful copy of the original paper format. This supports the view that different interfaces may need be required for different users ([Jones et al., 1999](#)). Further research in this area is needed.

Some users do not understand the difference between machine-readable text and a GIF image. These users will need information about how a digital copy is created. Providing information about the source of the document, and providing a cross reference to the microfiche format may also reduce some of these problems.

Speed in retrieving, downloading, and printing documents and locating relevant text on a page was particularly important. This finding is consistent with the findings of previous studies ([Jones et al., 1999](#); [Kengeri et al., 1999](#)). People have come to expect quick retrieval of numerous sources that they can manipulate when they use the Web and they will not tolerate restrictive, cumbersome or slow systems. They want the material freely available and they want to cut and paste text into other documents, download, and print at will. They also want search terms highlighted to save time. Searching and browsing by date is highly desirable in a historical collection and users want the computer to enhance the original material by creating hyperlinked indexes and tables of contents to improve access to the material. As in previous studies, they wanted the WWW format to have greater functionality than the Paper or Microfiche formats ([Kilker & Gay, 1998](#)). Many respondents commented that the WWW format would remove many barriers presented by the original paper or microfiche formats. Moreover, a clear overview of an item was viewed as desirable which supports ([Kengeri et al., 1999](#)) finding that users want an overview of the layout of the site.

The majority of respondents in our study felt that the WWW format would be the most useful in their work, and many participants in the focus groups thought it would help them conduct their research faster. We contrast this with the findings of the Columbia Online Project ([Mandel et al., 1997](#)). In that study, respondents were asked how the use of online books would affect the productivity and quality of their work. Findings showed that the majority of respondents did not think that using online books would improve the productivity or the quality of their work. However, the respondents in the Columbia study had primarily used the online format of the Oxford English Dictionary and that survey gathered information about productivity and quality rather than usefulness. Moreover, the material in the two digital collections is very different. Locating and using Early Canadiana in original paper or microfiche can be extremely time consuming due to the barriers libraries put upon their use. The WWW format removes many of these barriers and improves its usefulness.

Based on our findings, we made 26 recommendations for improvements to the ECO site. At the time of writing, 7 of the 26 recommendations have been implemented and 8 more are in the process of being implemented. Some of the remaining recommendations are still being discussed and may be implemented in the future.

Future Research

In spring 2000 we will conduct a follow-up WWW user survey. The first survey covered a time period when publicity for the ECO site was heavy. This publicity may have affected the volume of use, the nature of use, and type of users. The follow-up survey will provide data for a more typical period and will assess the impact of improvements to the site, including the increase in the number of titles available in ECO collection.

The ECO collection has the potential to have a significant effect on how school children learn about Canadian history. Marketing to school teachers and students could be followed with a user study of that group at some point following the marketing effort.

The focus groups suggested that information about the source of the digitized image might help them evaluate the authenticity of the item. Future research should investigate methods and techniques for judging the authenticity of a digitized image. We need to know what factors increase the trust people place in these facsimiles.

The findings of this study suggest that both GIF images and formatted text should be available for all items. We need to study the types of users who need a GIF image and the reasons why they need it. The findings also show that the original paper format is still needed at times. Future research should investigate the reasons why this format is still needed and determine which, if any, enhancements to the WWW format might reduce this need.

The findings of the study raise questions about the effect that digital libraries will have on scholarship. Scholars suggested that the ECO collection would allow them to make more links, consult new sources, and delve deeper into these sources. Future research should explore how scholars use this collection and the effect that this use has on their research. One-on-one observation of scholars using Early Canadiana materials would be an appropriate way to do this. This type of observation would also be useful with genealogists. In both cases, this observation should include use of Early Canadiana materials in the original paper and microfiche formats. This would increase our understanding of how people use Early Canadiana materials, help us recognize uses that could be supported in the digital format, and identify functionality desired by researchers that is not possible with the original paper and microfiche formats but may be viable in a digital environment.

Acknowledgements

We would like to thank Gerry Oxford for conducting the the WWW server log analysis and Cheryl Buchwald for conducting the focus group study. We would also like to thank the following individuals for their contributions to this project: Steve Szigeti, Guy Teasdale, Clement Arsenault, Zoran Piljevic, Rick Kopak, Joan Bartlett, and Joseph Desjardin. Finally, we express our appreciation to Karen Turko, ECO Project Manager, and Pam Bjornson, CIHM Executive Director for their support and cooperation throughout the project.

References

- Bishop, A.P., Neumann, L.J., Star, S.L., Merkel, C., Ignacio, E., and Sandusky, R.J. (2000). "Digital libraries: situating use in changing information infrastructure." *Journal of the American Society for Information Science*, **51**(4), 394-413.
- Dillon, A. (1999). "TIME - a multi-leveled framework for evaluating and designing digital libraries." *International Journal on Digital Librarianship*, **2**, 170-7.
- Hill, L.L., Dolin, R., Frew, J., Kemp, R.B., Larsgaard, M., Montello, D.R., Rae, M.-A., and Simpson, J. (1997). "User evaluation: summary of the methodologies and results for the Alexandria Digital Library, University of California at Santa Barbara." *ASIS '97: Proceedings of the 60th ASIS Annual Meeting, Washington, DC*. Medford, NJ: American Society for Information Science. pp. 225-43, 369.
- Jones, M.L. W., Gay, G.K., and Rieger, R.H. (1999). "Project Soup: comparing evaluations of digital collection efforts." *D-Lib*, **5**(11), Retrieved December 10, 1999 from the World Wide Web: <http://www.dlib.org/dlib/november99/11jones.html>.

- Kengeri, R., Seals, C.D., Harley, H.D., Reddy, H.P., and Fox, E.A. (1999). "Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, NDLTD." *International Journal on Digital Libraries*, **2**, 157-69.
- Kilker, J., and Gay, G. (1998). "The social construction of a digital library: a case study examining implications for evaluation." *Information Technology and Libraries*, **17**(2), 60-70.
- Mandel, C.A., Summerfield, M.C., and Kantor, P. (1997). "Online Books at Columbia: measurement and early results on use, satisfaction, and effect." Paper presented at the *Scholarly Communication and Technology Conference, Emory University, Atlanta, GA*. Retrieved January 6, 2000 from the World Wide Web: <http://arl.cni.org/scomm/scat/summerfield.toc.html>.
- Marchionini, G. (1999). "Digital library research and development challenges circa 2000." *Enhancing Canada's Digital Information Resources: Report of the HCI and the Digital Library Research Institute*. Toronto, ON. pp. 7-13
- Morgan, D.L., and Krueger, R.A. (1998). *Focus Group Kit*. Thousand Oaks, CA: Sage.
- O'Hara, K., and Sellen, A. (1997). "A comparison of reading paper and on-line documents." *CHI '97: Human Factors in Computing Systems Conference Proceedings, Atlanta, GA*. New York: Association for Computing Machinery, Inc. pp. 335-342.
- Rieger, R., and Gay, G. (1999). "Tools and techniques in evaluating digital imaging projects." *RLG DigiNews*, **3**(3), Retrieved December 17, 1999 from the World Wide Web: <http://www.rlg.org/preserv/diginews/diginews3-3.html#technical1>.
- Sellen, A., and Harper, R. (1997). "Paper as an analytic resource for the design of new technologies." *CHI '97: Human Factors in Computing Systems Conference Proceedings, Atlanta, GA*. New York: Association for Computing Machinery, Inc. pp. 319-26.
- Spink, A., Bateman, J., and Jansen, B.J. (1999). "Searching the WEB: a survey of EXCITE users." *Internet Research*, **9**(2), 117-28.

¹ Canadiana includes documents (books, periodicals, brochures, etc.) published in Canada as well as documents published in other countries but written by Canadians or about Canada. [return to text](#)

How to cite this paper:

Duff, Wendy M. and Cherry, Joan M. (2001) "Use of historical documents in a digital world: comparisons with original materials and microfiche". *Information Research*, **6**(1) Available at: <http://InformationR.net/ir/6-1/paper86.html>

© the authors, 2000. Updated: 27th August 2000

Check for citations, [using Google Scholar](#)

[Contents](#)

7 2 1 3
[Web Counter](#)

[Home](#)
