

# Evidence and Accountability: Digital Archives and Public Policy

by Mark Vajčner

PhD student, Edith Cowan University

Paper presented to the Association of Canadian Archivists

Calgary, 16 May 2009

## **Introduction**

Today archival collections increasingly provide raw material for public policy. They serve up documents that record past decision and action. They exhibit the raw statistics, data, and facts on which those decisions and actions were based. While archival institutions were once seen as serving a small and select population of academics, genealogists, or community historians, today the traditional perception of archives as the exclusive domain of history is being challenged. More than ever, archives are viewed by the public as sources for evidence of rights and as instruments in the search for accountability.

This is our conference theme, and the sessions that we will have over this weekend attest to this fundamental shift. Our program examines the impact of archives on issues as far ranging as residential schools, environmental change, workplace and gender, the protection of personal data, and the Canadian blood system. From outside of Canada we will

hear about the records of Saddam Hussein's campaign against the Kurds and of various reconciliation efforts in Africa, the Pacific, and Australia.

Tandem to this shift in the place of archives there is also change in how citizens interact with their governments. Not only do we seek evidence and accountability, but we also demand convenience. There is a growing comfort level in using the internet to receive government service. Electronic means of communication play an ever greater role in the development, implementation, and questioning of public policy.

A decade ago, before the 9/11 tragedy, talk of media convergence and the future shape of e-government was all the rage in Washington. Al Gore after all, according to the more inane media reports of the day, "invented" the internet. That and his "lock box". Three years ago Senator Ted Stevens was much maligned for referring to the internet as a "series of tubes". But the march toward e-government service in Canada, the United States, and other western countries has proceeded apace. A simplified process that makes access to government information convenient and cost-effective is an enticing and powerful idea. Witness the pre-clearance of Customs as businesses electronically submit their waybills before goods arrive at the boarder. Witness also, as tax season has just ended, the convenience of the *NetFile* service in separating you from your money. There is, in fact, a whole host of electronic services that the government provides. A quick visit to the Government of Canada website and you can

watch streaming video on how to prepare for a natural disaster, see job advertisements and weather conditions, download a myriad of forms publications and reports, read the latest on the H1N1 flu virus, and, for those of you with masochist tendencies, read in detail the 2009 federal budget.

More and more our public policy discourse is being affected by the internet. In his 2001 book, *The Future of Ideas*, Lawrence Lessig described the internet as a commons where, in its ideal incarnation, everyone is free to share and contribute. This equality of participation has had marked impact already. More and more spectators want to be players on the internet. And more and more these informal participants are leading innovation in the fields to which they contribute. Technology allows groups of amateurs to do what professionals exclusively did before. Witness the re-emergence of amateur participation in astronomy. Here networks of people, connected via the internet and with access to increasingly sophisticated amateur equipment, are making important discoveries mapping the night skies. Technology has had a profound impact on the practice of journalism. The rise of blogs has given voice to a much wider range of views than that covered by the established media. Think also of how many times you have seen video shot by amateurs on their camera-phones featured on the evening news. Cell phones, text messaging, and social-networking sites such as *Facebook*, *YouTube* and

*Twitter* are revolutionizing the way we organize political movements, protest marches, and other forms of public policy debate.

We now interact with government and shape public policy in a myriad of electronic ways. So the question now becomes: Will archives be able to serve their new public policy roles electronically? To answer this question we must focus on two aspects of the problem. First, are archival digital resources meeting expectations of evidence? The evidential role of archives is paramount in a world where archival documents increasingly inform public decisions and actions. Evidential documents are also critical in enforcing accountability. Non-evidential archival documents, or records that would not stand the rigor of a courtroom, are of little value in land claims, the residential schools question, or any other policy and accountability debate. The second aspect of the problem is an accountability problem itself. We must ask: Are digital archives being constructed in accordance with the information accountability standards that public policy increasingly dictates?

### **Evidence and Digital Archives**

But first things first. Lets examine the evidential value of digitized records. When the World Wide Web and the webpage made its debut in the early 1990s archives were among the early heritage institutions to take advantage of this new means of promotion. These early homepages

advertised location, hours, services, and content. They were followed shortly by collection-level descriptions as provincial councils and various institutions began to create the equivalent of on-line union catalogues for their archival records. These catalogues soon covered most institutions in their provinces and were networked into a national portal first known as the Canadian Archival Information Network and later renamed Archives Canada. It became possible to do one search and receive hits on resources across the country. These systems allowed researchers to easily ascertain where relevant collections of records could be found.

Samplings of the records themselves quickly followed. Federal funding to increase Canadian content on the internet brought digitized archival texts and images to the World Wide Web. These virtual exhibits have tended to be selective of the materials they capture. This is contrary to long established principles and practices that find archival meaning in the interconnections of the whole *fonds*. Since the time of the French Revolution archival practice has been based on two key principles – *provenance* and original order. These principles demand that archival materials be organized in such a way that their creator is known, that the records of different creators form distinct collections, and that any organizational structure within these collections be preserved. This allows each document to remain within its context giving it greater evidential weight as it draws credibility, or discredit, from the other documents with which it is associated.

This contextual organization is fundamental as archives increasingly become sources of the type of evidential documents required in public policy formulation and debate. As the archival user/informed citizen moves to the internet to conduct research it is critical that archives present their digitized collections in a contextual and thus an evidential way. If the only archives that our user/citizen visits is online, if the only archival documents that he/she wants are digital, then archives must present their digital materials in a contextual way making their online collections meaningful and as evidential as the boxed collections in their vaults. *Provenance* and original order, so critical in the way we organize archives must also be central to the way we present archives. Some digitization projects attempt that. Rather than presenting a small sampling of records, as virtual exhibits do, these more comprehensive digitization projects aim to digitize the majority of records in a given collection and some are presenting their materials contextually.

An example of a well presented collection is “The Empire That Was Russia: The Prokudin-Gorskii Photographic Record Recreated” (<http://www.loc.gov/exhibits/empire/>) which features photographic surveys of the Russian empire made by photographer Sergei Prokudin-Gorskii. Of the 2,607 distinct images in the collection, the United States Library of Congress digitized over 1,900 glass negatives, over 700 prints for which there are no negatives, and album pages showing all the prints

in the collection. These images, that document daily life across the vast Russian empire from 1905 to 1915, are in brilliant full colour. Prokudin-Gorskii used a photography process where three black-and-white negative exposures were taken at one time. These exposures were made through blue, green, and red filters to produce photographs that could be printed or projected in colour.

This website provides extensive technical details on his photographic process including information on the equipment used by Prokudin-Gorskii. There is also a biography of him and his ambitious plan for a photographic survey of Russia. His project won the support of the Russian Czar and Prokudin-Gorskii spent numerous years completing photographic surveys of eleven regions of the empire, travelling in a specially equipped railroad car provided by the imperial ministry responsible for transportation. There is further information on the site providing historical context on the Russian empire during the time that these impressive photographs were taken. The site also contains information about how the collection came into the possession of the Library of Congress in 1948.

Further technical details on the digitization process employed by the library are provided. Once scanned the images were reproduced through an innovative process known as digichromatography. The website details this process showing examples of an image in its three black and white

stages, how the original colour image would have looked, and the image as it appears cleaned up through the library's digitization process. A detailed database links all images. Links in the database place each photograph in the context of the album from which they originate, as well as links to other contextual and technical information. The user may search the database directly, navigate through the various series of the collection, or view an exhibition of the photographs.

The context that the Library of Congress built into this website augments the research value of this presentation. The colour slides of Russia are interesting in themselves, but once one knows that they were taken at a time when colour photography, as we later came to know it, did not exist, they become utterly fascinating. The technical complexity, and thus the work involved in making these images, becomes clear only in the contextual information that goes along with the images. This was not a simple photographer travelling around the country side taking "snaps"; it was an involved process that required a large investment of time and resources. Although the text on the website doesn't allude to it, such an intensively planned photographic project probably didn't leave its subject matter to chance. It is only logical that the images captured were also carefully planned. You have a sense of this in images of various ethnic minorities dressed up in their finest costume, or the extra-ordinary cleanliness of a factory floor in another set of images. Contrast this to another set of photographs, this time taken in rural Saskatchewan in the



1950s and 1960s (by grain agent Everett Baker), and you see a somewhat more candid representation of rural life (at <http://www.shfs.ca/baker.php>). The ease of the technology and its reduced cost makes the more relaxed images of the 1950s possible. But there surely are other factors to consider, factors such as differences in culture and even politics – remember the Czarist government sponsored Prokudin-Gorskii, we see no colour images of strikes or revolutionaries.

Images taken for the Czar probably won't have much impact on Canadian public policy, but there are other collections of documents that will. The archival records of the Department of Indian and Northern Affairs, for example, have played an extensive role in shaping our modern relations with first nations. These records have been used in court cases and treaty negotiations and play a vivid role in formulation and debate on current policy. The presentation of these materials online, in a contextual and evidential way would go a long way to advancing research, debate, and hopefully policy formulation.

We see at the University of Saskatchewan the beginnings of such an online collection. "Our Legacy" (at <http://www.scaa.sk.ca/ourlegacy>) presents materials relating to aboriginal, Métis, and Inuit peoples found in various Saskatchewan cultural and heritage collections. Over 4,000 descriptions and over 50,000 digitized archival and published items are available. Where appropriate, some finding aids will soon also be

available. The site also includes library and museum material. Materials can be searched by institution, keyword, subject, *provenance*, format of record, date, and various other ways such as treaty area. Various exhibits and slideshows have been created and there is also a detailed contextual essay on aboriginal treaties written by Professor J.R. Miller of the Department of History.

Digitization is thus more than the scanning and mounting of material. It involves the provision of context that makes it possible for the researcher/citizen to ask meaningful questions. The questions that we began to ask of the Prokudin-Gorskii slides, for example, are only possible because they have a context and can be analyzed in comparison to the contexts of other collections, such as that of Everett Baker. This is what we need to strive for as we enter an age of digital archives. The contextual essay is an important tool in this regard. In 2005 Tom Nesmith outlined what a contextual essay system could look like and how it would serve it's propose. "There could be," Nesmith wrote, "a series of essays... these essays would be a guide to thinking about and using" the record. Essays could be available on the history of the archives itself and its collection development policy. Essays could be available on societal contexts, creators, mandates, relevant laws, functions, record-keeping systems and processes, organizational cultures, information technologies, custodial histories, and a whole host of other relevant access points. These essays would prepare the user to go beyond the materials to the less visible and

complex ideas and trends behind them. A user could always go directly to the materials and avoid the essays at will, or choose to read some and ignore others. These essays would be especially valuable in working digitally without direct access to an archivist. An essay could point to related literature as well. A researcher could also contribute to these essays by sharing interesting facts learned during research. In this manner existing essays could be updated or new essays written.

Indeed scholars and other experts could be commissioned as part of the digitization project, as Prof. Miller at the University of Saskatchewan was, to write these contextual essays. The short essays which accompany the digitized diary of Prime Minister William Lyon Mackenzie King (<http://king.collectionscanada.ca/EN/default.asp>) are embryonic examples of the work proposed by professor Nesmith. Nesmith envisions a formally developed stream of contextual essays written and vetted by experts. But that needn't always be the case. These essays could just as easily follow the model of the short user-generated book reviews and book suggestions found on such sites as Amazon.com. They may evolve as a Wiki connected to the digitized resource. Ultimately there may be different levels of participation and different tools to serve different functions for the same digitized resource. We may see a more open tool for general users to post their comments and recommendations, and a more scholarly peer-reviewed one for in-depth expert opinion with the ability of users to move effortlessly between the two.

In our Department of Indian and Northern Affairs example, were a future digital archive to exist a stream of user participation could revolve around the policy debates the archival records inform. Wikis and user-generated essays could comment on what public policy issues referenced the records. Where these uses successful? Did courts accept or reject the arguments constructed with the record? Essays could bring to attention other sets of records, digitized or not, that might be useful in the debate. The possibilities are limitless.

The future digital archive envisioned here has several characteristics. It has the bulk of the records in a collection digitized. Those records are presented in a manner that allows access by *provenance* and original order (but note that this doesn't exclude subject or other access points). Finally the context of the collection is developed through essays and user comment. It is hoped that this will create an evidentially strong digital archives. It addresses the first aspect of our problem today: Evidence and the digital archives. Let us now briefly turn to the second aspect: Accountability.

### **Accountability and Digital Archives**

As we have seen, the digital resources that archives construct will be used to hold institutions to account. Think of the use of church archives in

the residential schools question. But archives themselves will need to be accountable in the records they select and present online. The two central laws that govern archives in this regard are copyright and freedom of information/protection of privacy. There really isn't much that I need to say on this. The laws are in place and it is simply a matter of archivists being aware of these laws and applying them appropriately to their digital records. Archivists must ensure they have donor and transfer agreements in place. Where materials aren't in the public domain, they must also ensure they have the right to use intellectual property or specific permission to place collections online. Digital archivists must learn and remain current with both Copyright and Freedom of Information laws. The "remain current" piece may prove to be the most challenging. Copyright law in particular is changing and evolving and is doing so in ways that are not always in the best interest of archives and archival research. Several attempts have been made in past years to significantly change Canadian copyright law. In the United States copyright law has changed more in the last 60 years than in all the years before. The cynical say that each time Mickey Mouse is due to enter the public domain there is a push to extend the provisions of copyright. But the cynical needn't go to the United States. Here in Canada we have seen the efforts of the estate of Lucy Maud Montgomery to extend the protective provisions of copyright. There is a valid argument for increased copyright protection as well. The ease with which copies can be made and distributed in the digital world has dramatic effects on industries that trade in intellectual property. Think

only of the dramatic changes that we have seen in the music recording industry.

It is in the interest of archives to sit at the vanguard of copyright and intellectual property issues. We should not just know about these issues and explain the rules to researchers, but we should endeavor to effect them on behalf of our users interests. Our professional associations should be more visible in these efforts and develop a core expertise that individual archives haven't the time or money to develop. This is another great collaborative arena for archives and our researchers and would allow us to participate in public policy formulation rather than simply providing the raw documentary materials.

Finally, I'd like to touch on one more point. How we authenticate digital materials. In the absence of comparison between digital and original materials – an absence which may become more common as familiarity of researchers with the digital medium grows. We may see a shift in the authentication process away from a rigorous examination of the item, to an examination of the credentials of the archives that digitizes it. Without access to the original it is difficult to determine if all the pages of an item were scanned. Or if the color reproduction of an art work is close to the original, or if its proportions are correct. Without ready access to the original how can we know that the digital item is an accurate representation of that original? The researcher must trust that it is so. To

trust it, he or she must be confident that the digitizing institution followed a rigorous process in producing the digital item. Contextual essays on how collections were chosen for digitization and how that process proceeded is the way to build that trust. Add to that further essays written by experts in the field and the level of trust grows. It grows to the point that over time, simply citing the name of a rigorous institution is enough to elicit trust. This is exactly what happens in the publishing industry. Which publisher would the classical scholar turn to for the most trustworthy copy of a Latin text – some small under-funded, little-known publisher or Oxford University Press? Selection, review, and editing at an academic publishing house are rigorous. Furthermore selection, review, and editing in the scholarly publishing model are user driven. It is academics who select, review, and edit the works of their peers for publication. It is quite possible that the archive that wishes to position itself as a truly reputable and accountable research institution will, in a sense, have to replicate the instruments created by scholarly presses. Materials cannot simply be digitized and mounted on the internet, they will need to be subject to careful selection, to a quality digital reproduction process, and provide a high level of contextual and metadata information. Like the publishing process, on which it is based, the whole process may involve many professionals – archivists, librarians, computer technicians, scholars, and other users.

This collaboration serves not only to build the accountability of an archive in the digital materials it provides, but to build the authoritativeness of the archives itself. Having an active, involved, and supportive user community and having active and involved ties with other professionals shows the archives as useful, innovative, and effective. All these features are incredibly important for archival administrators when they approach their parent institutions and funding agencies.

## **Conclusion**

I've enjoyed sharing some of my ideas with you today. I'm sure that many of the suggestions I've put forward aren't necessarily new to you either, nor, are they necessarily right. Nevertheless I hope that in raising them today I am providing impetus to think about them anew. I'd enjoy hearing your thoughts and engaging in discussion for the time we have left in this session.

Thank you.