

***Forum sur les <métadonnées> au Canada  
Canadian <Metadata> Forum  
19-20 septembre 2003-September 19-20, 2003***

***L'utilisation des métadonnées pour décrire  
les collections d'archives***

***Using metadata to describe  
archival holdings***

***Sarah Klotz & Lorraine Gadoury,  
Bibliothèque et Archives Canada  
Library and Archives of Canada***

L'expérience pratique dont nous allons vous parler ici est celle de la description de documents historiques très anciens, datant des 17<sup>e</sup> et 18<sup>e</sup> siècles. Ces documents ne sont pas conservés par les Archives nationales du Canada mais par les Archives de France. Cependant, comme ils sont au cœur des débuts de l'histoire canadienne, les Archives du Canada ont envoyé en France, dès la fin du 19<sup>e</sup> siècle, des copistes pour décrire et transcrire les documents importants conservés dans les dépôts d'archives de ce pays.

Les documents furent d'abord transcrits puis, avec la mise au point de la technique du microfilmage dans les années 1950, les documents originaux furent reproduits

et mis à la disposition des chercheurs canadiens. Aujourd'hui, avec la nouvelle technique de numérisation de masse des documents à partir des microfilms, c'est une nouvelle étape qui s'amorce qui rendra l'accès aux documents d'archives concernant la Nouvelle-France encore plus facile.

Les documents d'archives de la Nouvelle-France sont de plusieurs types. On trouve par exemple le plus ancien recensement canadien effectué en 1666 par l'intendant Jean Talon et qui énumérait chaque personne habitant la colonie (image 1); le traité de paix intervenu à Montréal en 1701 entre les principales nations amérindiennes et le gouverneur général de la colonie Louis-Hector de Callières, où les chefs ont signé de leurs totems (image 2); des listes de militaires comme celle des officiers de Carignan devenus habitants en 1668 (image 3); de la correspondance entre les administrateurs coloniaux et le Secrétaire d'État à la Marine attestant de décisions importantes concernant l'avenir de la colonie (image 4) ou des ordonnances diverses, par exemple celle-ci qui défend aux habitants de la ville de Québec de laisser vaquer leurs cochons dans les rues (image 5)!

Afin que les documents d'archives soient accessibles, ils doivent être décrits et disponibles par le biais d'une base de données que les chercheurs peuvent interroger

par divers critères de recherche, comme la date des documents, le nom de leurs auteurs et destinataires ou encore les sujets qui sont traités dans ceux-ci.

Or, une grande partie des descriptions des documents d'archives concernant la Nouvelle-France ont été réalisées au moment de leur transcription, c'est-à-dire au tournant du 20<sup>e</sup> siècle. Elles ont été alors publiées dans les *Rapport des Archives publiques du Canada* et se présentent sous la forme suivante (image 6).

Heureusement, en 1986, la Section des archives française de la Division des manuscrits, sous la direction de Victorin Chabot, a mis en place une base de données pour servir de cadre à l'automatisation des instruments de recherche de la section et particulièrement ceux qui concernent les anciennes archives françaises. Cette base est connue sous le nom de PIAF, pour «Programme d'inventaire des archives françaises». Elle fonctionne à l'aide du logiciel Minisis, et son contenu est versé ensuite dans l'outil de recherche de l'institution ArchiviaNet, accessible par l'Internet.

Concrètement, les anciennes descriptions sont saisies dans Word Perfect, puis elles sont vérifiées et complétées par des archivistes ou des assistants-archivistes pour ajouter les éléments qui ne figuraient pas dans les descriptions originales comme

les cotes, les numéros de microfilms, les folios extrêmes, ainsi de suite. On insère enfin les codes pour permettre le transfert dans la base de données PIAF (image 7).

Ces fichiers Word Perfect sont ensuite transférés dans la base de données PIAF sur Minisis. Les archivistes peuvent consulter ces descriptions afin de faire de la recherche et des corrections. Le public, quant à lui, n'a pas accès directement à Minisis qui, comme vous pouvez le constater, n'est pas d'un abord particulièrement facile (image 8).

Les données sont ensuite migrées à intervalle régulier dans ArchiviaNet où elles sont consultées par le public sur le site des Archives nationales du Canada. Les internautes peuvent, sur ce site, faire de la recherche par dates et mots-clés afin de retrouver les documents qui les intéressent. Un écran de recherche permet d'abord d'entrer les éléments désirés (image 9). Le résultat est la liste de tous les documents qui comprennent le ou les éléments demandés (image 10); en sélectionnant un document en particulier, on donne ensuite accès à sa description complète, incluant la référence aux Archives nationales du Canada, ainsi que dans le dépôt où est conservé le document original (image 11).

À l'automne 2001, un projet conjoint de numérisation de documents d'archives a été mis sur pied entre les Archives nationales du Canada, les Archives de France et le Ministère canadien des Affaires étrangères, dans le cadre du programme *Canada-France 2004*. L'objectif du projet est double. L'équipe doit réaliser une exposition virtuelle sur la Nouvelle-France, qui permettra d'attirer les visiteurs et internautes en présentant des documents d'archives se démarquant par leur grande qualité visuelle et en les replaçant dans le contexte historique par des textes d'une grande pertinence.

Le projet vise aussi à décrire, à numériser et à mettre en ligne environ 600 000 pages de documents d'archives relatifs à la Nouvelle-France. Ces documents seront accessibles par le biais d'un portail international qui contiendra les descriptions des documents ainsi que les images et qui, comme ArchiviaNet pour les Archives nationales du Canada, permettra d'y avoir accès. Les chercheurs pourront faire des recherches par fonds d'archives, institutions, mots-clés et par dates. Le tout sera lancé en novembre 2003 en France et au Canada. Ma collègue, Sarah Klotz, va maintenant vous présenter comment les métadonnées ont été au cœur de la mise sur pied de ce portail qui va réunir des descriptions provenant de la France et du Canada.

Although these documents are all related to New France, the French and Canadian descriptions of these documents vary quite significantly. To incorporate both types of descriptions into a single search tool, one metadata language was selected for the mapping. We decided to use Encoded Archival Description (EAD) to transfer information from our National Archives database into a central extended markup language (XML) repository. We selected EAD because it is specific to archives and as it can be tagged in XML, is well suited for an online search tool. Another advantage of EAD is that it does not regulate the content of the descriptions making it easier to tag and reuse existing content.

As Lorraine has mentioned, these records were first described at the turn of the 20<sup>th</sup> century before archival descriptive standards were fully developed in Canada. The purpose of these original descriptions was to explain the content of the documents and to provide users with the information required to consult them. It should be no surprise then that there was always a consistent approach to these descriptions, even if the methodology had not been formalized. In their current state in the Colonial Archives database on the National Archives website, the descriptions from the *Public Archives Report* have been re-keyed and re-purposed for use on the website with a selection of these descriptions now being linked to their digitized copy of the record.

This search tool proved that it was possible to repurpose these descriptions. And so following this rather than creating new descriptions that may be more closely aligned to modern descriptive practices and to the descriptive practices of France, we opted to once again reuse our current descriptions.

Let's take a look at how the database currently appears on the NA website. In this example you have 9 fields. Archives describe information in hierarchical levels. This begins with the fonds, collection or series and works its way down to the file and item level. The majority of the records in this database are items and files, however, some documents have been described as volume. These correspond to a very granular level of description.

The first field indicates the provenance of the item and provides context for the record. This item is part of the Fonds des Colonies/Colonial Archives and of the General correspondence Series, known as Series C11A . The next field relates to the type of document. The records of this database are all textual records. The date of creation of the document is listed as between August and September 1701. This corresponds to the date indicated on the document and if no date is indicated then an approximate date is supplied by the archivist – depending on what is

known about a particular document this can be very general like, century, decade, range of years or more specific like a single year, month and day. The scope and content describes the content of the item. This field can sometimes be quite detailed. Since these descriptions predate databases and the Internet, they were originally created to provide as much detail as possible so that researchers could ascertain the usefulness of the documents without needing to come to Ottawa to refer to the finding aid. The archival reference number, original volume and folio number are used for citation and consultation if a researcher was interested in viewing the copies on microfilm.

Since EAD is an encoding standard, it does not instruct what or how to describe. It provides common elements and tags and thus the means to transfer data. As a first step before mapping to EAD, it was necessary to map the content to the International Standard of Archival Description (ISAD G). The French use ISAD G as their main guide in description and thus the mapping exercise would facilitate the comparison of our descriptions.

Through this comparison, we were able to determine which elements were essential for resource discovery. There were a number of fields in common, but I would like to point out those that were different:

1. When the French archives include descriptions of cartographic material, they require additional elements related to the scale, dimensions, support and form of the maps. These elements will not be included in the descriptions of textual records.
2. All of the institutions have additional descriptive elements that are not considered to be part of the scope and content. This information will be grouped together in a Note field.
3. Levels of descriptions of the National Archives include item, file and volume. According to current standards, a volume is not a true level of description. Rather than beginning the process of re-describing this material, this level of aggregation is accepted.
4. For the National Archives, only file and volume level descriptions have a title provided in a separate field. At the time these descriptions were originally created the title of the item was considered to be part of the general description. As a result the title is included in the scope and content.

The search is performed on a limited number of fields as not all fields exist for every institution. Individual searches will be conducted on the title of the fonds and series, the location of original, and date. One will also be able to search the ti-

tle and scope and content field together. The results screen will display the date, title/description, reference number and location of the original. In the item/file display all fields will be visible with the exception of level of description and language of documents. These descriptions are then linked to their digital copies.

One might think that it would be impossible for descriptions created in the early 1900s and a database created in the 1980s to be compatible with a relatively new encoding system. But if a description is to be useful it must contain a certain number and quality of elements so that a user can determine what the item is and if it is of use to them. To fully exploit all the available information contained in our descriptions, a standard that promotes the use of common as well as divergent elements was essential. EAD is very flexible and recognizes that to facilitate access it is not necessary to employ one single set of descriptive elements. Depending on the content, level of description and media, different elements are required to best describe an item. This has allowed us to make use of not only the fields that we share in common but also those that are unique to individual institutions, descriptions or media.

By using a more stringent standard, we may not have been able to make as much content, or as much detail about the content available to the public. This would

have significantly diminished the usefulness of the descriptions by only making available those fields that were held in common. In effect, we would have lost valuable information simply because it would not have fit with the standard. As well, if we had to significantly re-describe some of our holdings to comply with a standard this would have significantly increased our costs, limited the descriptions and material available, and delayed the time in which this material would have been made available online.

The creation of a shared search tool linked to digitized images was achieved through the use of a common metadata standard that allowed us to find both the similarities and differences between our descriptions. The key to interoperability is not simply the use of a common standard but the use of a standard, like EAD, that is flexible and recognizes the uniqueness of descriptions. Although this project relates only to interoperability within the international archival community – the key to this project was finding a common language. We hope to build upon this project, to provide integrated access to the holdings of different communities or for different audiences. The power of metadata standards then rests with its flexibility and ability to uphold, sustain and promote all of the intricacies that exist in descriptions.