

Databases for Early Cinema Research

Databases are one of the most common ways to link digital tools and cinema. We will show the specificity of two databases designed by GRAFICS, pointing out how they were developed by film scholars involved in various fields of research, how they contributed their own concepts to the database design, what information each database includes and how it is structured, and why those databases already helped deepening the knowledge of film activities during the silent era in Québec. At the same time, we present results of a cooperation between a university and a film archive. Finally, we will discuss new perspectives of linkage between GRAFICS, the *Cinémaèque québécoise* and the *Bibliothèque et Archives nationales du Québec* (BAnQ) in respect to a research project on early film activity in Québec as reflected in the contemporary press. The project aims at a global website with many internal links for research, reference, and documentation.

GRAFICS (*Groupe de recherche sur l'avènement et la formation des institutions cinématographique et scénique*) is a research group working on early cinema in Québec and elsewhere in the world. It was created in 1994 at the Université de Montréal, building on the pre-centennial mood that was prevailing in the western world at the time.¹ Studying the cinematographic and theatrical institutions, it developed many axes of research:

- 1 Reception – audiences / social practices
- 2 Exhibition; Theatres: buildings – programs / intermediatic practices
- 3 Music – Sound effects
- 4 Lecturer
- 5 Production
- 6 Cultural representation / identity
- 7 Travelling exhibition
- 8 Institutional discourse

1 GRAFICS is headed by André Gaudreault, and Pierre Véronneau is one of the researchers affiliated to it. Over the years, it has received financial assistance from Fonds québécois de la recherche sur la société et la culture (www.fqrsq.gouv.qc.ca) and the Social Sciences and Humanities Research Council of Canada (www.sshrc-crsh.gc.ca).

One of the first research projects undertaken by GRAFICS was the Québec newspaper project. Up to that point, newspapers had occupied a paradoxical place in film historiography: though widely available in research libraries and generally acknowledged as an exceptional source of data on film exhibition and audiences, daily newspapers proved to be too intimidating a resource for many film historians. It is easy to understand why: looking for specific information in hundreds of thousands of pages is really like looking for the proverbial needle in a haystack. However, this difficulty was greatly reduced by GRAFICS' collaborative structure, which enabled several researchers sharing an interest in early cinema to pool their resources and hire a team of research assistants. The assistants were asked to systematically collect every document pertaining to moving and projected images, as well as amusement places and popular visual culture, printed between 1895 and 1915 in three Montreal newspapers: *La Presse* (still published today, and still touting itself as "America's largest francophone daily"), *La Patrie*, and *The Montreal Daily Star*— then Canada's largest newspaper. It should be noted that moving pictures had enjoyed an exceptional level of popularity and acceptance in the largely francophone province of Québec in the nickelodeon era.² This was reflected by the coverage given to moving pictures by the selected newspapers, which turned out to be more important and sustained than that of most American and English-Canadian newspapers of the same era.³

The documents collected by GRAFICS in Québec newspapers varied widely in nature and content: there were reports on screenings, exhibitor ads, opinion pieces on moving picture shows, drawings and photographs of filmgoers, news items (for instance on ladies refusing to take off their hats in moving picture theatres), etc. Even the classifieds turned out to be an exceptional source of data. By the early 1910s, not a week went by without at least one small and otherwise undocumented moving picture theater showing up in the "For Sale" section. The wealth of documents generated by the project eventually filled several filing cabinets, and soon became almost as daunting as the newspapers themselves. It was consequently decided to index the collection. This led to the creation of a FileMaker database, which by 1997 held close to 16,000 files indexing as many documents from *La Patrie* and *La Presse*. The database was subsequently copied to CD-ROM and made available, along with bound copies of the collected documents, to visitors of the *Cinémathèque québécoise's médiathèque*. Over the past decade, several Canadian and international research-

2 On the popularity of early cinema in the province of Quebec, see: Gaudreault/Lacasse (1996).

3 In recent years, Richard Abel, in the United States, and Paul S. Moore, in Canada, have also extensively studied the coverage of silent cinema in newspapers.

ers have made the trip to the *Cinémathèque québécoise* to consult this exceptional collection.

An important development pertaining to GRAFICS' newspaper project came about in 2005 when the *Bibliothèque et Archives nationales du Québec* – one of the province's very few cultural organizations not chronically short on funds – launched a major digitization project aiming to make most of Québec's historical periodicals available online by the end of the decade. One of the first titles available on the BAnQ's website was *La Patrie*. This did not go unnoticed by GRAFICS' members, who by this point were aware of the fact that their ten year old FileMaker database was becoming obsolete, and was thus in need of a major overhaul. The situation was eventually solved with the help of the federal government's *Canadian Culture Online* program.⁴ By joining forces with the BAnQ and the *Cinémathèque québécoise*, GRAFICS was able to convince *Canadian Culture Online* to provide funding for a project entitled "Silent Film in Québec, 1896-1930". The project's objectives were a) to enable the BAnQ and the *Cinémathèque québécoise* to digitize part of their collections and make these available on the Internet, and b) to give GRAFICS the means to pursue its research in these newly digitized collections, and subsequently use its discoveries to build an educational website.

During this second phase of its newspaper project, GRAFICS collected several thousands of new documents in about twenty recently digitized historical periodicals: daily big city and weekly small town newspapers, women's and fan magazines, specialized weeklies, etc. These were indexed in a new version of GRAFICS' newspaper database developed with *Ruby on Rails*, an open-source database-backed web software, and are now being made available to anybody with an Internet connection through a search engine featured on the website of the "Silent Film in Québec, 1896-1930" project. This search engine will generate results featuring the retrieved documents' bibliographical references as well as hyperlinks permitting the user to view them on the BAnQ's website.⁵

Before we launch into a detailed description of the types of researches permitted by the project's database and search engine, it is necessary to recapitulate the arguments that convinced GRAFICS to pursue its systematic survey and indexing of Québec's historical newspapers in a technological context very different from that of the project's beginnings. It is safe to say that in the early days of the project nobody at GRAFICS could have predicted – or even

4 *Canadian Culture Online* is part of the *Department of Canadian Heritage's* strategy to encourage a uniquely Canadian presence on the Internet. See www.culture.ca.

5 The website of "Silent Film in Québec, 1896-1930" (www.cinemamuetquebec.ca) was launched in September 2008.

dreamed – that by 2007 an ever-growing number of historical newspapers would be available in their entirety as searchable digital files on the Internet. The phenomenal progress of digital technologies since the mid-1990s consequently forced GRAFICS to reconsider the methods, the objectives and, indeed, the validity of its newspaper project. Given the phenomenal development of digital tools, and more particularly of Optical Character Recognition software (OCR) now routinely used to convert digitized pages into searchable textual documents⁶, was the systematic collecting and indexing of film-related newspaper articles and documents really worth the investment in time and money? After taking into account a variety of issues ranging from the purely technical to the epistemological, GRAFICS' members agreed that it definitely was.

There is no denying the extraordinary possibilities opened by the intersection of computerized search engines and OCR software. For researchers who had to unwind kilometers of worn-out microfilm in remote research libraries in order to locate a few interesting documents, the possibility of instantly gathering hundreds, or even thousands, of relevant documents by simply typing a few words in an on-line search engine is nothing short of revolutionary. Still, there are some tasks that OCR software cannot perform yet, and many others that in all likelihood it never will. Regarding GRAFICS' project, a first limitation of the various OCR software is that, according to tests performed by the BAnQ, they still have a relatively high error rate when dealing with old newspapers reproduced on rather less than perfect microfilms. And even when they do correctly capture the content of the printed page, the use of OCR software is still limited by the staggering amount of misprints and misspellings that can be found in historical newspapers. As opposed to the human reader, OCR software cannot – and indeed, should not be made to – see a “kinetoscope” where a “kenitoscope” has been printed. Apart from that, search engines solely relying on OCR often prove rather ineffectual when asked to perform several types of basic searches. Simple search requests involving personal names, for instance, commonly turn out to be nearly impossible. A case in point is that of Léo-Ernest Ouimet – the single most important individual in the history of silent cinema in the province of Québec. Since “Ouimet” is one of the province's most common surnames, any search using this criterion performed on a Québec newspaper will return literally thousands of irrelevant documents. (First names are not very useful for this kind of request, since – if used at all –

6 See, for instance the Library of Congress's *Chronicling America: Historic American Newspapers* project, part of the *National Digital Newspaper Program* (www.loc.gov/chroniclingamerica); Cornell University's *Making of America* project (cdl.library.cornell.edu/moa); *ProQuest Historical Newspapers* (www.proquest.com/products_pq/descriptions/pq-hist-news.shtml).

they tend to vary widely in form. “Léo-Ernest” can thus appear in historical newspapers as “Léo”, “Ernest”, “L.”, “E.” or “L. E.”.)

Search engines relying on OCR software will also be regularly stalled by the extreme instability of the nomenclature of film-related terms and categories in the pre-institutional period covered by the project. At the same time, this instability constituted the main justification of and a great challenge to GRAFICS’ indexing enterprise. In order to return a comprehensive list of results when interrogated through the website’s search engine, the project’s database had to rely on a strict categorization and normalization of the data entered on each document’s file. Obviously, this does not sit well with silent – and more particularly early – cinema, which by definition is a period largely devoid of standardized practices: Between the 1890s and the late 1920s, films were exhibited in a variety of venues and contexts, genres and formats were ever changing, and labor division in film production had yet to stabilize in clearly denominated functions such as producer, director and cameraman. Most of the categories of information indexed in the project’s database consequently had to rely on standardized lists of predetermined entries. For instance, a single unified category – “cinema” – was inscribed on the file of every document referencing “kinetographic records”, “moving pictures”, “motion pictures”, “animated views”, “photoplays”, “motography”, “films” or “movies” (to quote only some of the most common variations on a single type of attraction). This normalization principle most notably applied to the “function” field associated to the indexed organizations and individuals, to the fields dedicated to the types of venues and attractions mentioned by the document, as well as to another field registering some of the indexed attractions’ particular features, such as the presence of lecturers and the use of sound effects or color processes. A standardized subject headings system inspired by those of the BAnQ and the *Fédération internationale des archives du film* (FIAF) has also been devised and incorporated in the project’s database. Its various entries deal with issues related to the general historical context (“World War I”), with specific debates pertaining to film (“Legislation: Sunday laws”), to film audiences (“Women”, “Ethnic groups: French-Canadians”), as well as to film uses and contents (“Education”).

Names of individuals and organizations have also been strictly normalized according to the *Anglo-American Cataloguing Rules 2* and, as much as possible, covered by a single unified entry. The fact that the *London Bioscope Co.* – a well-known traveling company based in Montreal – is called the *London Film* in one particular article does not mean much. Film titles, however, have been systematically indexed in the exact formulations in which they can be found in the treated documents, since their variations can be very telling. For instance, when the 1902 *Edison Film Manufacturing Co.* production *Jack and the Beanstalk* is

screened in Montreal under the title *Jacques et les tiges de fer* (translates *Jack and the Iron Stalk*), much is revealed about local strategies of appropriation. In order to facilitate search requests, a “normalized title” field has nevertheless been added to the database. (It should however be noted that, even with the help of such exceptional online resources as the *Silent Era Québec Filmography*, which we will cover later, and the American Film Institute’s *Silent Film Catalog*, the identification of the quoted films’ “original” titles has turned out to be a task far too time-consuming to be done systematically.) Finally, once again in order to facilitate search requests, the database allows entering the names of the producer(s) and director(s) of each quoted film, even if they do not appear in the indexed document. However, every piece of information not coming from the indexed document has been segregated and clearly identified as such, in order to retain the possibility of limiting a search request to documents explicitly referring to particular agents.

Apart from these normalization issues, another argument in favor of GRAFICS’ indexing enterprise is that the database and search engine on which it relies can also be made to perform some forms of automated statistical analysis. For example, a researcher interested in the rise of the star system might want to see how many indexed newspaper documents from the year 1909 refer to film actors and actresses (“performers” function) and then, by repeating the same query for the subsequent years and comparing the resulting figures with those obtained for the “production” and “direction” functions, uncover some salient features of the evolution of promotional discourse. It would moreover be possible to generate figures dealing specifically with advertisements or reviews, since the types of documents, along with the types of publications, have also been indexed in the database.

Such combinations of research criteria could simply not be accommodated by search engines relying solely on OCR-generated files. Another fundamental drawback of OCR is that it greatly reduces the likeliness of an unexpected discovery. As anybody who has done historical research can testify, groundbreaking discoveries are often made by accident rather than through targeted searches. OCR only permits one to find what one is looking for. On the other hand, OCR unquestionably remains the best tool for the study of the evolution of the nomenclature of film-related terms. It can easily retrieve the earliest uses of such connoted words as “nickelodeon” and “photoplay”, or identify the periods during which the use of these words peaked. It also remains extremely useful for preliminary samples aiming to gauge a newspaper’s level of coverage of moving pictures. In the end, it really seems that indexing and OCR should be conceived and used as two complementary research tools. There is no arguing that in a perfect world, historical newspaper collections would all be both indexed and treated with OCR software.

We hope to have demonstrated that the phenomenal progress of digital tools did not render obsolete the type of systematic survey and indexing work pursued by GRAFICS since the 1990s. Quite on the contrary, it is our opinion that, by disseminating the rewards of this innovative enterprise outside of a small select circle, digital technologies have made it more relevant than ever. Let's now hope that this exponentially increased ease of access will stimulate researchers, but also film and history devotees, as well as anybody with a passing interest in early and silent cinema, to find new uses for the surprising variety of documents that can be found in old newspapers and magazines. Over its first decade, GRAFICS' newspaper project has generated several major research projects, such as the *Silent Era Québec Filmography* and Germain Lacasse's study of film lecturers. There is no doubt that several other projects still lie dormant in GRAFICS' newspaper database and in the digitized collections of its partners.

The filmography of the films shot in Québec during the silent era clearly illustrates the importance of local newspapers during this period.⁷ Researchers analyzed all available information relating to cinema not only in the newspapers, but also in the trade press, companies catalogues, censorship files, and other archives. This research brought to light, first of all, the existence of hitherto forgotten films. Secondly, it made it possible to determine how films were produced, advertised, and received by the public and assimilated by specific communities. The copious amount of information available on film exhibition made it possible to introduce a considerable amount of documentation that is not provided in film catalogues or the trade press. The way a film's title could vary throughout the province, for example, demonstrates the importance of the exhibitor during this pre-institutional era of film history. Exhibitors re-named films in order to highlight whichever elements they thought more attractive to their own audience. The filmography is thus pre-institutional in that it provides disparate information and documents the heterogeneity of film-related activities. The filmography also embodies the period's intermediality, drawing on sources other than those of the film world and documenting the way cinema interacted with other media. We thus see the special place moving pictures occupied at the time: cinema was a new, unstable and hybrid practice which was appropriated by other practices and institutions, but it also soon freed itself from this appropriation in order to establish its autonomy and legitimacy and constitute its own specific sphere of influence

7 This research was subsidized by the *Social Sciences and Humanities Research Council of Canada*. It is headed by Pierre Véronneau, with André Gaudreault and Germain Lacasse as co-researchers.

within an effervescent, intermedial environment, in which large-circulation newspapers, telephony, radio and other media were also rapidly expanding.

The indexing of the newspapers led to the discovery that many films were shot in Québec and in Canada, both by Canadians and foreigners. So another project grew out of the first: a transnational filmography of films shot in Québec. For this team project, we decided to explore as many sources as possible to find information on the films in question. This first filmographic database was built in FileMaker, and a website was launched at the 2002 Domitor conference held in Montréal. Each entry provides technical information on the film (production company, date, etc.) as well as descriptions from period publications and critical commentary by members of the research team.

The goal of this filmography is two-fold: first, to establish an inventory of films shot in the province of Québec in the early years of film history and to make this information publicly available; and second, to attempt to understand how these films were received at the time they were produced and first viewed. This latter goal, of course, is much more difficult to accomplish than the former, but attempting it will allow us to demonstrate the specificity of early cinema and the transformations cinema has undergone over the past century. A study of the entries in this filmography will not only provide readers with information about the individual films but will help them to understand these films' history and, more specifically, the task of writing film history.

The filmography has been established on the basis of geographical territory. The films found here were not made by a particular production company or filmmaker, or produced in a particular country; instead, they were made within a given territory.⁸ In some ways, it is a filmography that reveals the variety of films made within Québec and the diverse strategies that existed for distributing and exhibiting these films from a sort of pre-institutional and transitional era to the era of institutionalised practices and film-related government regulations. The filmography makes it possible to understand how film production and reception in Québec developed rather than simply highlighting the role of any given community or individual. It is also possible to use the filmography to study how various groups and communities responded to the appearance of a radically new apparatus and its equally novel practices and texts.

Within a comparative filmography, the research must pay attention to the origin of the production agents. During the pre-institutional period, many films were shot by foreigners wishing to show exotic locations and original situations. When the same locations and topics were shot by local agents, the

8 This point was elaborated by Pierre Véronneau at the 2006 Domitor conference held in Ann Arbor, Michigan.

point of view was different, as if those agents wanted to appropriate their own “foreignness”, to show their own people and play on local identification. A good example for this are the films shot for the tercentenary of Québec in 1908: the one from Urban (*Québec: The Tercentenary Celebration*), the one from Gaumont British (*Québec Pageant*), the two from Vitagraph (*Discoverers: A Grand Historical Pageant Picturing the Discovery and Founding of New France, Canada and Québec Tercentenary Celebration*), and the three shot by Léo-Ernest Ouimet (*Fêtes du tricentenaire, première série; Fêtes du tricentenaire, deuxième série: Tableaux historiques des Pageants; Fêtes du tricentenaire, troisième série: Partie indienne et de la cour*).

There were many flaws and failures in the way the database handled the information. So in 2003 we decided to build a new database on a more rigorous and systematic basis, and import the content of the old database into it. The second database was done in Microsoft Access, and it was published on the Internet in fall 2006 at <http://cri.histart.umontreal.ca/Grafics/fr/filmo/default.asp>. In the final part of this article, we would like to point out some important features of this new version.

The original title is a key issue for a film. Generally, this is the title found in a producer’s catalogue or on the film itself. It is important to indicate the source of the title. If the title is found in a source that is less reliable, e.g. a contemporary newspaper, the cataloguer can choose chevrons to emphasize that the title is only presumed to be the original title. In the absence of any contemporary documents, the cataloguer creates hypothetical titles, or draws on titles already attributed to the work by archivists or historians. Square brackets indicate a non-contemporary title, and a box can be checked in the case of an attribution. This small example shows how digital tools oblige us to structure and divide the information in separate and semantically well-defined data to allow the researchers to intersect the information and eventually draw their conclusions.

One of the major issues in building digital tools to handle audiovisual and film-related collections is to start with a strict analysis of the information that needs to be processed, to determine how it can be controlled with authority files, and to write elaborate cataloguing rules. The quality of information retrieval depends largely on the precision of the organization of data entry. Filmographic tools are useful in film studies that focus on production and exhibition. We are able to store information on shooting locations, screenings, persons and companies involved in the production, different kinds of sources, etc. The database also allows us to provide important content information, like the film’s subject and genre. It was decided to use the FIAF subject headings as a standard. There is also a “Film Comment” field where you find comments on the film production and exhibition, and a “Film Context” field that documents the historical context of the film’s subject. Sometimes even the

entering of data might call for some annotation. We decided to complement many fields with a “Note” button which leads to comments on the data, transcriptions of articles and other pertinent information.

In 2004, a partnership agreement was set up between the *Cinéma-thèque québécoise* and the *Université de Montréal* to share some information gathered by GRAFICS. For instance, clippings and film catalogues were made available at the CQ media library. Part of the filmography was included in the CQ database *Ciné-TV*. This SQL database is divided into many modules that are linked together, for their filmographic part, by a core module called *Répertoire*, a “database” that may be a reference by itself (like the *IMDb*), and works as a point of access to all the collections (films, documents, and film-related collections like stills, posters, archives) and all the research tools of the *Cinéma-thèque*. The existence of this module allows the researcher to cross-cut information and to find all that is accessible at the *Cinéma-thèque québécoise* on a specific title, agent or topic. Four collections, *Canadian Cinema and Television*, *Videos*, *Film-Related Collections*, and *Documents*, are already on the Web (collections.cinema-theque.qc.ca) and soon to be completed by the *Répertoire de la production audiovisuelle au Québec* (www.repertoireaudiovisuelquebec.ca).

In conclusion we may say that databases are among those digital tools that can have a great effect on film research and film studies, providing an in-depth indexation and standardization of the information, especially for subjects, names and titles. It is also important to interconnect between databases as much as possible to share their strength and multiply their effects. The example of the two databases of the GRAFICS (filmography and newspapers) and the one of the *Cinéma-thèque québécoise* (*Ciné-TV*) shows that through co-operation between institutions and research projects, it was possible to develop powerful tools to access and document silent cinema in Québec, and make it available on the Internet.

Bibliography

Gaudreault, André, and Germain Lacasse: *Au pays des ennemis du cinéma: pour une nouvelle histoire des débuts du cinéma au Québec*. In collaboration with Jean-Pierre Sirois-Trahan. Québec: Nuit blanche, 1996.