

## Abstract

This research project will study and archive the distinctive heritage of Belgian and French immigrant communities in the Western Canadian Prairies, by gathering images of geographic landmarks associated with French and Belgian immigrants in Western Canada and relevant biographical, historical, and cultural information tied to specific locations. The objective of this project is to ensure that the French and Belgian immigrant historical, literary, and cultural memory relating to geographical locations is preserved and is publicly accessible. A metadata scheme describing personal names, biographical information, geographical locations, and archival collections has been developed to describe the unique data to be created, preserved, and made public.

This data management plan was created using the University of Alberta DMP Template in the <u>DMP Assistant platform</u>. This template has additional questions which are not present in the standard Portage DMP template. If you are using this template as a model for work in the Portage template, you may need to consider what the appropriate location for different types of information would be.

## Administrative Details

#### **Project Name:**

Belgians and French in the Prairies (v. 1.0)

#### Principal Investigator / Researcher:

Dr. Rao

#### **Project Data Contact:**

Denis Lacroix

#### **Description:**

This research project will study and archive the distinctive heritage of Belgian and French immigrant communities in the Prairies, because we want to understand the impact that these two minority groups have had on the Canadian Prairies' landscape and collective memory, in order to facilitate access to and analysis of Belgian and French Western Canadian historic and



literary documents. Similar projects have recorded and archived the cultural history of various ethnic groups, like Western Canadian Ukrainians, often focusing on a cultural aspect, like sacred objects. This project will focus on geographic landmarks associated with French and Belgian immigrants in Western Canada and relevant biographical, historical, and cultural information tied to specific locations. Covering Western Canadian Prairies is the final outcome of this project; however, the project will first focus on Alberta, then Saskatchewan, and finally Manitoba. Some of the questions to be considered include: How can we preserve the memory of Belgian and French landmarks that may soon physically disappear? How can we bring the landmarks alive by associating biographical, historical, and literary information to them? How can we preserve biographical information of prominent community members in a dynamic, updated, and freely accessible database? What links can be made between people, landmarks and memory institutions, like archives, to help better understand and access historical information?

#### Institution:

University of Alberta



## **Data Collection**

## What types of data will you collect, create, link to, acquire and/or record?

Data/assets will consist of the following document types:

- text files containing bibliographic entries, archival fonds descriptions including geographic locations, biographical entries of prominent members of the Canadian Prairie Belgian and French immigrant communities, list of migratory waves with dates and geographic locations
- images of prominent members of the Canadian Prairie Belgian and French communities
- images of the Canadian Prairie Belgian and French heritage
- A map with geographic data linking architectural and geographical landmark images to their spatial location using a Geographic Information System (GIS)

Prominent members of the two French and Belgian immigrant communities will be identified using a number of reference resources, e.g. the *Canadian Encyclopedia*, *Canadiana*, and Gamila Morcos' *Dictionnaire des artistes et des auteurs francophones de l'Ouest canadien*. French geographical location data will be identified using toponymic and geographic research, e.g. Carol Léonard's *Noms De Lieux Français En Saskatchewan*. Metadata for each document type will be collected in a spreadsheet and standardized to ensure consistency and findability.

Identifying the reference resources helps to explain what the contents of the text files are likely to be.

## What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

The data will be collected in file formats suitable for access and preservation based on industry best practices. The following file formats are widely used by the community and allow for data re-use, sharing, and long-term access. Some of the files are better suited for preservation, e.g. Extensible Markup Language (XML) or Rich Text Format (RTF), while others are better suited for sharing and viewing, e.g. Portable Document Format (PDF).



- **Textual files**: eXtensible Mark-up Language (XML) text according to an appropriate Document Type Definition (DTD) or schema (.xml) OR Rich Text Format (.rtf) AND Portable Document Format (.pdf) for public downloading of files
- **Textual files** containing bibliographic entries: Bibtex and RIS (Research Information Systems bibliographic tag format) and PDF. The first two formats will be produced using RefWorks (an online bibliographic reference manager).
- **Textual files** containing alphabetical list of literary works with summaries produced within a spreadsheet: Comma-Separated Values (CSV) and XML and PDF
- Quantitative data: CSV file produced from a spreadsheet
- **Maps**: geo-referenced Tagged Image File Format (TIFF) (.tif) or a TIFF image stored in a GIS application (.tfw) and/or Keyhole Markup Language (KML) (.kml) files for representing maps in an XML based file format
- **Images**: TIFF uncompressed (.tif) for downloading and preservation purposes and Joint Photographic Experts Group format (JPG) for public display on website

#### How much data do you anticipate collecting? Include an estimate of how much storage space you will require (in megabytes, gigabytes, terabytes). This estimate should also take into account storage space required for file versioning, backups, and the growth rate over time.

Textual and quantitative files will require minimal storage. Data will grow during the active life of the project (i.e., next five years), but will likely not exceed the amounts below. Images: 7GB Maps: 3GB



#### Are there any existing data that you can re-use? If so, please explain how you will obtain that data and integrate it into your research project.

Yes, textual and image files currently exist on the researchers' computer hard-drives. More photos will be taken and text created by the researchers. They will all be included in the database (Dataverse, Omeka and/or another file management platform) with support from the <u>UA Arts Research Centre (ARC)</u>. The database, the files, and the metadata will be preserved by the University of Alberta Library. Some images also exist in archives, both physical and virtual.

Geographic locations of French and Belgian settlements will be identified for Saskatchewan using Carol Léonard's <u>Mémoire des noms de lieux d'origine de d'influence françaises en</u> <u>Saskatchewan</u>, as well as the online <u>Musée virtuel de la Saskatchewan</u>, and for Alberta using <u>Geographical Names of Alberta gazetteer</u> and Carol Leonard's research on Alberta Francophone toponymic data. Similar reference sources will be used for Manitoba. Archival finding aids and databases will also be consulted and various encyclopedic sources will be used to identify relevant fonds and prominent immigrants. File metadata will be entered into a spreadsheet.

Embedding links in your DMP will direct readers to external sources and resources that can provide more detailed information about discipline- and institution-related topics.

## What conventions and procedures will you use to structure, name and version-control your files to ensure that your data are well-organized?

Data files will follow the naming conventions below (see Appendix for descriptions and examples):

- RIS, Bibtex, and PDF files containing bibliographic entries: BIBLIO and Cultural group name and upload date (e.g. BIBLIO-Belgian-oct2020). Annotations with summaries of literary works will be included.
- archival fonds descriptions including geographic locations will consist of a spreadsheet entitled ARCHIVAL FONDS with the following fields: fonds title, Dataverse Digital Object Identifier (DOI), fonds location, fonds longitude/latitude, Canadian geographic descriptors, file number, name of immigrant or institution, nationality, dates, subjects, and description.



- biographical entries of prominent members of the Canadian Prairie Belgian and French communities will consist of a spreadsheet entitled Biographies with the following fields: name of immigrant or institution, Dataverse DOI, locations inhabited, Canadian geographic descriptors, nationality, birth date, death date, description, and subjects.
- images of prominent members of the Canadian Prairie Belgian and French communities will consist of a spreadsheet entitled Person-images with the following fields: file names, Dataverse DOI, name of immigrant or institution, Image longitude and latitude, Canadian geographic descriptors, date when image was taken, description. and subjects. Each image will receive a name based on the model PERSONIMAG-Nationality(ie. French or Belgian)-LastName-FirstName-Optional Number for multiple images (e.g. PERSONIMAG-French-Michelet-Alex-1)
- images of landmarks from the Canadian Prairie Belgian and French heritage will consist of a spreadsheet entitled Landmark-images with the following fields: file names, Dataverse DOI, name of immigrant or institution, image longitude and latitude, Canadian geographic descriptors, date when image was taken, description, and subjects. Each image will receive a name based on the model LANDIMAG-Nationality(ie. French or Belgian)-Location Name (e.g. Edmonton)-Landmark Name (e.g. Michelet House)-Optional number for multiple images
- map with geographic data linking architectural and geographical landmark images to their spatial location using GIS will consist of a spreadsheet entitled Land-images and consist of a separate kml file.
- Dataverse provides automatic version control, along with individual DOIs, for each uploaded file.





## **Documentation and Metadata**

What documentation will be needed for the data to be read and interpreted correctly in the future? This includes study-level documentation, data-level description, and any other contextual information required to make the data usable by other researchers.

- A file naming convention has been established and will be followed. It will be documented in a helpfile that will accompany the data (see Appendix for a list of metadata categories complete with descriptions, syntax and examples).
- All files containing bio-bibliographic entries and other content will be in non-proprietary, preservation ready formats such as csv, txt, or pdf.
- Images will be named according to the established file naming conventions and will be described according to project standards and guidelines. Accompanying metadata will be in non-proprietary, preservation ready formats such as csv.
- Geographical information will follow standard GIS conventions and will be in nonproprietary, preservation ready formats, such as CSV or a geospatial vector data shapefile format (SHP).
- The helpfile will also include study-level information: who created/collected the data, when it was created, location of file preservation copies, and conditions of use.
- Training of research team members (e.g. students) will go over the documentation/helpfile, the project data, and the protocols in place to clarify how to create and manage the data appropriately.

#### Please list the metadata standard and tools you will use to document and describe your data. If there is not an appropriate standard, please explain how you will ensure consistency in your documentation.

- Project specific conventions and standards will be used to describe the prominent immigrants, images, and locations. These conventions will be described in the helpfile. When possible and appropriate, standard vocabularies will be used for names of people, places, and organizations, as well as subject terms.
  - Standard spelling of given names as they appear in an authoritative source, e.g. AURORA (<u>https://bac-lac.on.worldcat.org/discovery</u>), *The Canadian encyclopedia* (<u>https://www.thecanadianencyclopedia.ca/en</u>), or the *Dictionnaire*



des artistes et des auteurs francophones de l'Ouest canadien (<u>https://www.library.ualberta.ca/catalog/2072670</u>).

- Profession titles and other subject terms will be standardized using the Canadian Subject Headings dictionary to identify the appropriate terms in English/French (<u>http://www.collectionscanada.gc.ca/csh-bin/search/I=0</u>) or AURORA Subject headings (<u>https://bac-lac.on.worldcat.org/discovery</u>) and scan the Répertoire de vedettes matières (RVM) for French terms (<u>https://rvmweb.bibl.ulaval.ca/en/recherche/</u>).
- Image files will have embedded GIS (i.e., georeferencing layers) from the camera used in taking them. This information will also be included in the accompanying metadata.
- Bibliographical entries will follow conventions listed in the Chicago Manual of Style 17th edition. Bibliographical data will be stored in RefWorks and exported in RIS/Bibtex format on a regular basis for data preservation.
- Geographical entries will use longitude and latitude conventions. Google Maps will be used to determine, track, and visualize geographical data. In conjunction with Google maps, <u>Natural Resources Canada's geographical place names database</u> and Carol Léonard's *Mémoire des noms de lieux d'origine et d'influence françaises en Saskatchewan* will be consulted.
  - Natural Resources Canada's geographical place names database and Carol Léonard's Mémoire des noms de lieux d'origine et d'influence françaises en Saskatchewan will be consulted. Carol Léonard also has data regarding Alberta francophone toponyms. Convert Legal Sub Divisions (LSD), quarter section, section to longitude and latitude using <u>http://legallandconverter.com/</u>.
  - Geonames <u>http://www.geonames.org/</u> will be used to determine the accepted form of a place name.

Note that both project-specific conventions and more widely recognized conventions are described in some detail.



## How will you make sure that documentation is created or captured consistently throughout your project?

- There will be ongoing collaboration between the UofA Library's metadata team, website/database developers, and researchers around metadata creation and capture.
- Student assistants who enter metadata will receive training from the UofA Library's metadata librarian and researchers. Regular quality control will also be part of the workflow.
- When an image file is uploaded to the collection, some initial metadata describing it will be given to it. Further metadata will be entered subsequently in the spreadsheet used for describing images.
- Project documentation will be in non-proprietary, preservation ready formats such as TXT or PDF and will accompany the data.

Note the identification of the different team members involved in the project, and their roles in ensuring consistency. Particularly on larger projects, different team members may not interact regularly, so it is important to establish clear and explicit procedures that everyone must follow.

## Storage and Backup

## How will your data be stored and backed up during your research project?

Data will be stored on the principal researcher's hard-drive and shared institutional Google drive with regularly scheduled uploads to <u>Dataverse</u>, the University of Alberta's data repository. Data will also be uploaded using Web FTP or file transfer protocol to to a Web publishing platform, such as Omeka or WordPress, maintained by <u>UA Arts Resource Centre (ARC)</u> where additional metadata will be incorporated. Images and geographic data will also be uploaded to Google maps, which in turn will provide a .kml/.kmz file for the preservation of the maps created for this project. Bibliographic data will be stored in RefWorks and exported in RIS/Bibtex format for preservation. Data stored in DataVerse and the public website created for the project will be permanently preserved by the UA Library and <u>Archivelt</u> respectively.



## How will the research team and other collaborators access, modify, and contribute data throughout the project?

All data will be fully shareable and open access through Google Drive and Dataverse.

## How will you ensure that sensitive data is stored securely and only accessible to the research team during the research project?

All data will be fully shareable and open access. There is no intention to capture sensitive data. The research data will receive a Creative Commons attribution share-alike license (CC-BY-SA).

## Preservation

Which data are selected for preservation and access will depend on potential reuse value, whether there are obligations to either retain or destroy data, and the resources required to properly curate the data and ensure that it remains usable in the future. In some circumstances, it may be feasible to preserve all versions of the data (e.g., raw, processed, analyzed, final), but in others, it may be preferable to keep only selected data (e.g., transcripts instead of audio interviews).

This project will produce data in various formats. The research team will follow digital preservation industry best practices in the selection of file formats. All types and versions of data will be preserved at the UofA Library. This includes digital objects such as images, documents, geographical data, etc., and metadata created for these objects. The digital objects will be saved in file formats suitable for long term preservation and the metadata will be exported to XML format and preserved. The website created for this project will be archived using a web archiving tool, such as Archive-It, and preserved along with the data.



## At the end of your research project, where will you deposit your data for long-term preservation and access?

The UofA Library will be responsible for the long-term preservation of the digital objects and associated metadata. All the data objects, associated metadata, and websites created for this project will be preserved in the UofA Library digital preservation repository. The UofA Library uses appropriate strategies and industry best practices to preserve content of enduring value. Digital preservation requires continual development and migration of the content and infrastructure and the UofA Library employs strategies to cope with this challenge. Working data will be saved onto a computer and/or external hard-drive, as well as on a shared Google drive. Data will be deposited to the UofA Library's Dataverse repository regularly.

# Please describe how you will prepare the data for preservation and access, including any necessary procedures for data cleaning, normalization or de-identification. Explain how you will prevent data from being lost while processing and converting files.

During the data collection phase, researchers will be guided to produce data in the recommended file formats mentioned above with a minimal risk of data loss. Data produced in a non-standard file format will be normalized to formats recommended for long term preservation.

## Sharing and Reuse

What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final). Consider which data may need to be shared in order to meet institutional or funding requirements, and which data may be restricted because of confidentiality/privacy issues.

All data, including metadata, will be made open access. Images will be shared in both raw and reduced sizes. The project's website will provide users with the main access point; however, users may also access data files from the Dataverse repository.



# How will you be sharing your data (e.g., institutional repository, a specialized data archive, project website, informal/on-request sharing)? Include a brief description of any resources needed to share your data (equipment, systems, expertise, etc.).

Data will be shared from the University of Alberta's Google Drive and Dataverse institutional repository, as well as the project's website and the Archive-It collection. An expert in database/website creation and management will be key in ensuring this project is accessible and visible. Furthermore, data stored in Dataverse will receive a Digital Object Identifier (DOI), a permanent identifier for digital objects (documents, datasets, etc.), which will be used for citation purposes in publications and conference presentations.

#### Please describe whether there will be any restrictions place on your data when they are made available and who may access them. If data are not openly available, describe the process for gaining access.

No restrictions will be placed on the data.

## What type of end-use license will you include with your data? Please include a copy of this license with your data management plan.

We will use a Creative Commons Attribution Share-alike 4.0 International license, see <u>https://creativecommons.org/licenses/by-sa/4.0/</u>.



## **Responsibilities and Resources**

Who will be responsible for data management during the project (i.e., during collection, processing, analysis, documentation)? Identify staff and organizational roles and their responsibilities for carrying out the data management plan (DMP), including time allocations and training requirements.

Dr. Rao at the University of Alberta will be the main project coordinator or principal researcher. Other researchers and graduate students will participate as the project develops; however, anyone working on the project will need to back up their data to Google Drive and Dataverse at the end of each work period. The principal researcher will also back up the uploaded data to his hard-drive on a weekly basis.

## What will happen if personnel changes occur or if the Principal Investigator leaves the institution before the project has concluded?

If the project comes to an end, the UofA Library will assume control of archiving and access. Regular data deposits will ensure that preservation and sharing occurs early in the project's lifecycle and that the UA Library's long-term stewardship will be a continuous process built into the project from the start, thanks to Dataverse and Archive-It. The collection will be ingested into the UAL digital preservation infrastructure and maintained there according to the terms of deposit.

If the principal investigator (PI) leaves the institution, he will continue using Dataverse to preserve the data. The use of Archive-It will continue as long as the UA continues to pay into the service or the PI's new institution provides for this service. If the PI decides not to continue working on the project, his department will be asked to find a replacement, e.g. a colleague or a graduate student.

## What resources will you require to implement your plan? Will extra people, time, hardware, or storage be required? How much will this cost (estimation)?

Student employees, paid from existing and anticipated grant funds, will be required. The project will use the infrastructure at the University of Alberta. The size of the textual and image data can be absorbed into the existing infrastructure at the UofA Library.



## **Ethics and Legal Compliance**

If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?

No sensitive data will be collected.

## If applicable, what strategies will you undertake to address secondary uses of sensitive data?

No sensitive data will be collected.

It is preferable to explicitly say that no sensitive data will be collected, rather than leaving the question blank or putting 'N/A', as it shows that you have considered the question in relation to your specific project.

#### How will you manage legal, ethical, and intellectual property issues?

- Only public domain information will be used in the metadata. Textual descriptions will not be copied and pasted from published sources that are not in the public domain.
- Images created for this project will receive a CC-BY-SA license
- Images copied from existing archives will be in the public domain or the permission of the rights-holder will be sought so as to allow a CC license.



## Appendix: Metadata Categories

Categories	Description and Syntax	Examples
Name of Immigrant or Institution Place Name	Last Name; First Name or Name of institution. Use the standard spelling and given names as they appear in an authoritative source, e.g. AURORA ( <u>https://bac-lac.on.worldcat.org/discovery</u> ), <i>The Canadian encyclopedia</i> ( <u>https://www.thecanadianencyclopedia.ca/en</u> ), or the <i>Dictionnaire des artistes et des auteurs francophones de l'Ouest canadien</i> ( <u>https://www.library.ualberta.ca/catalog/2072670</u> )	Michelet; Charles Alexandre Alliance française
Locations Inhabited	Enter multiple Longitudes and Latitudes separated by semi- colons in the following order and format: latitude -longitude. Use photo's embedded geolocation for greater accuracy when appropriate. Google Maps will be used to determine, track, and	54.01791 - 113.71282 ; 53.547637 - 113.492580
OR	visualize geographical data. In conjunction with Google maps, Natural Resources Canada's geographical place names database and Carol Léonard's <i>Mémoire des noms de lieux</i>	
Image Longitude and Latitude	<i>d'origine et d'influence françaises en Saskatchewan</i> will be consulted. Carol Léonard also has data regarding Alberta francophone toponyms. Convert Legal Sub Divisions (LSD), quarter section, section to longitude and latitude using <u>http://legallandconverter.com/</u> or the <u>ATS converter for Alberta</u> .	
OR		
Fonds Longitude and Latitude		



Categories	Description and Syntax	Examples
Canadian Geographic Descriptors	Textual description of the locations inhabited or the location's longitude and latitude. They will consist of the name of the Canadian province and the city or town name. Separate each descriptor by a vertical bar, and each location group by a semicolon.	Alberta   Edmonton ; Alberta   La Calmette
Nationality	Use either French or Belgian to describe the immigrant person's origin.	French
Birth Date	The date takes the form yyyy-mm-dd. Enter 00 for unknown data.	1911-12-05
Death Date	The date takes the form yyyy-mm-dd. Enter 00 for unknown data.	2010-12-05
Image Date	When image was taken.The date takes the form yyyy-mm-dd. If only the year is known, replace the unknown numbers with a 0. If it is a date range, separate starting and end date with a colon.	2010-12-05 1912-0-0 : 1918-0-0
Landmark Date	Date when landmark was built.The date takes the form yyyy- mm-dd. If only the year is known, replace the unknown numbers with a 0.	2010-12-05 1912-0-0
Description	Enter biographical, image, or archival fonds description here. Explain what is the relationship of the image with and how does it represent the French or Belgian identity in Western Canada. Include name of Province inhabited and any other defining locations relevant to the person or institution described, e.g. street address. Data regarding Saskatchewan French geographical names, see names located in Léonard's	Image of Alex Michelet when he was editor of the newspaper Le Courrier de l'Ouest,



Categories	Description and Syntax	Examples
	Mémoire des noms de lieux d'origine et d'influence française en Saskatchewan (call number FC 3506 L46 2010). Do not include commas.	whose offices were located at 654 2nd Street in Edmonton, Alberta.
Subject: Geographic Location	This is a location that is an addition to the geographic descriptors used to describe the Canadian province and town/city. Separate multiple terms using a semicolon . Use Geonames <u>http://www.geonames.org/</u> to determine the accepted form of a place name.	France; Bourg-en- Bresse; United- States; Washington;
Subject: Profession or Domain of Activity -	This is a subject that describes the profession of the person described or the domain in which the institution performed its activities. Use English and French terms to describe this concept separated by a vertical bar. Separate multiple concepts using a semicolon. Search the Canadian Subject Headings dictionary to identify the appropriate terms in English ( <u>http://www.collectionscanada.gc.ca/csh-bin/search/l=0</u> ) or AURORA Subject headings ( <u>https://bac-lac.on.worldcat.org/discovery</u> ) and scan the RVM for French terms ( <u>https://rvmweb.bibl.ulaval.ca/en/recherche/</u> ).	Writers   Écrivains ; Journalists   Journalistes
Subject: Name of structure or object	Name of structure or object represented	Michelet House
Subject: Topical Subjects	Use standard subject headings relating to events.	World War, 1914-1918



Categories	Description and Syntax	Examples
File Names	PERSONIMAG-Nationality(ie. French or Belgian)-LastName- Optional Number for multiple images	PERSONIMA G-French- Michelet- Alex-1
	LANDIMAG-Nationality(ie. French or Belgian)-Location Name (e.g. Edmonton)-Landmark Name (e. g. Michelet House)- Optional number for multiple images	LANDIMAG- French- Edmonton- Michelet House-1
File Number: Use file number associated with the fonds as attributed by the archive.	Use file number associated with the fonds as attributed by the archive.	PR1997.0117
Fonds Title	Use title indicated in the archival database, e.g.	Fonds Alexandre Michelet
Fonds Location	Use the name of the archival institution and/or the digital repository name separated by a semi-colon, if appearing in combination.	Provincial Archives of Alberta ; DataVerse
Dataverse DOI	DOI permanent URL provided by DataVerse upon uploading of the file.	https://doi.org /10.7939/DV N/10160





Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) Portage Network | portage@carl-abrc.ca | portagenetwork.ca