eMigrate: Aggregating Government Open Data for Enhanced Job Category Selection in Support of Immigration Applications

Eva Hourihan Jansen, University of Toronto
Jenna Jacobson, University of Toronto
Gabby Resch, University of Toronto

Abstract
With the growth of open data, much of the information required when applying to immigrate to Canada is available directly from government websites; however, an information problem persists, as making sense of the necessary information remains challenging. In this project, we describe a web-based application combining three sources of open data to communicate with potential newcomers during the pre-arrival stage. *eMigrate* combines policy documents and occupational descriptions and employment demographics in one system to aid newcomers in identifying appropriate occupational categories for their application. In the case study, we specifically identify how *eMigrate* can provide social media professionals strategic information to strengthen their immigration application.

1   Introduction

The immigration experience is understood as a complex process (Caidi et al., 2010) and may involve navigating multiple sources of information provided by the Government of Canada. With the growth of open data, much of this information is now available directly and transparently from government websites. The Canadian federal government increasingly makes data it collects on immigration and occupational trends available, and recently launched an Open Government website (http://open.gc.ca) which provides online access to a variety of federal datasets. Access is a key concept in exploring everyday life information seeking practices (ELIS) of immigration. Barriers have been identified in relation to language, information and communication technologies (ICTs), cultural relevance, status, and trust among others (ibid). Is the increased availability of government open data relevant to addressing some of these barriers? We draw on preliminary findings from an ongoing ethnographic study of standard occupational classification. Our project responds to the need for employment information by creating an information environment for sharing open data about occupational categories, policies, and demographic data.

2   Purpose of this Study

As information user groups, immigrants also called *newcomers,* are extremely diverse, with information practices and needs that vary substantially and are situated among complex policy issues (Caidi et al., 2010). In this project, we combine immigration policy information, standard occupational classification, and job posting information to backend a web-based application intended support pre-arrival information seeking for newcomers to Canada. This service, *eMigrate*, aggregates occupational information from these disparate sources in order to assist potential newcomers to Canada as they select relevant occupational categories in their application process.

When choosing to identify oneself through an occupational category applicants must use the robust categories of the NOC, Canada's *National Occupational Classification* (HRSDC, 2013). The NOC is a standard classification that organizes Canadian occupational information. "Many of Canada's immigration programs use it to decide if a job, or type of work experience, is valid for that program's criteria" (Government of Canada, 2013). The process of selecting a NOC code and selecting an immigration program under which to apply is not seamless, and requires consulting multiple data sources. How can we help potential newcomers identify appropriate occupational categories as they prepare their application for immigration to Canada? *eMigrate* augments the current information flow by providing an

extended, more holistic view that aggregates occupation data mined from immigration policy documents, occupational codes directly from the NOC, and emergent category data from Government of Canada job banks. The rich aggregate view enabled by *eMigrate* affords people the opportunity to make informed decisions about which categories to select and how to reflect their current occupation or past experience to support their application to immigrate.

In the current regime, many applications are assessed for eligibility based on a smaller subset of NOC codes that are also subject to eligibility caps. By aggregating three sources of relevant open data and providing recommendations based on the aggregate output, *eMigrate* aims to provide a more comprehensive picture of the policy environment, the occupational landscape, and the employment opportunities said to exist. The following Canadian open data sources are combined to provide strategic information to potential applicants:

Data Source	Motivating Questions
1. Immigration Policy Documents	Which occupational categories actually qualify for which immigration programs?
2. National Occupational Classification (2011)	Which NOC codes describe the occupations in an applicant's employment history? Which of these occupations will strengthen an application according to current immigration policy and projections?
3. Government of Canada Job Bank	What is the employment outlook for the selected occupational category?

3   Framework

Library and Information Science literature describes a range of ICTs (including computers and the internet, mobile phones, radio, television, and satellite systems) that act as significant information resources for immigrants (Caidi et al., 2010). While ICTs are shown to foster social inclusion in general, employment-related activities in particular are documented among the main reasons immigrants use ICTs at institutions such as community networks, settlement agencies, and public libraries (Caidi, Allard, Dechief, & Longford, 2008 in ibid).

In this project, we embrace "presentation of self" as a lens to understand occupational categories. Goffman (1959) outlines how people use the presentation of self to foster specific impressions in other people's minds through strategic performance. An individual, as a social actor, engages in a performance where identity is negotiated and developed as part of one's impression management: "Thus when an individual appears in the presence of others, there will usually be some reason for him to mobilize his activity so that it will convey an impression to others which it is in his interests to convey" (p. 4).

A potential immigrant may be required to list their current occupation and up to five others held within 10 years preceding the date of application. For each of these choices, the applicant must identify the occupation with the appropriate four-digit NOC code, along with the dates of employment, number of years of continuous full-time or equivalent part time experience, and a description of main duties. From this list, the applicant must also select the primary occupation and the corresponding four-digit NOC code that will be used to assess whether the application meets the minimum program requirements to be eligible for processing.

By embracing Goffman's dramaturgical model, we posit that this process offers more than one way of representing job history in order to fit Canadian entry requirements. This presentation process is based on the notion of communicating a positive impression on the immigration application in order to enhance the likelihood of subsequent acceptance and arrival into the country. We attempt to make the process more transparent by combining policy and occupational choices in one interface to help applicants select the most appropriate and advantageous four-digit NOC codes in support of this presentation process.

4   Methodology and Design

eMigrate is being developed through an iterative prototyping process to identify useful sources of open government data that will create a more extensive picture of occupational category requirements. The project is informed by an ongoing ethnographic study of standard occupational classification in the context of employment services for immigrant professionals by co-author Hourihan Jansen. We assert that an interpretive understanding of occupational categories may inform users in this context. The specific case study is further informed by qualitative interviews, undertaken by co-author Jacobson, with social media professionals in Canada.

4.1   Data Sources

4.1.1   Immigration Policy

http://www.cic.gc.ca/english/immigrate/apply.asp
The Federal Minister of Citizenship and Immigration issues instructions regarding eligibility and processing of applications to the Federal Skilled Worker Program (FSWP), Federal Skilled Trades Program (FSTP), and Canadian Experience Class (CEC). *eMigrate* mines these instructional documents for topical language related to occupational categories.

4.1.2   National Occupational Classification

http://open.canada.ca/data/en/dataset/bc4d5e01-cc19-45f3-bebc-72969b6d4b42
Immigration policy refers applicants to consult "the NOC". The National Occupational Classification (NOC) 2011 is described as the authoritative resource on occupational information in Canada, used by people to understand the jobs found throughout Canada's labour market. In the U.S. labour market, the comparable standard for classifying occupations is O*Net.

4.1.3   Job Bank

http://www.jobbank.gc.ca/occupation_search-eng.do
Canada's Job Bank is a job listing service delivered by Employment and Social Development Canada (ESDC) that lists information about jobs provided by employers from across Canada. Jobs are listed by occupational code, education program, wage, outlook, as well as by skills and knowledge expected for career entry.

4.2   *eMigrate* Description

Combining information from these open data sources provides potential immigration applicants an indication of their odds for gaining successful entry into Canada, as well as finding employment in a role commensurate with their current pre-immigration job.

1.   Users begin by selecting an occupational category from a list of all 500 occupational groups, each with a unique code, in Canada. The code selected will be indexed against 50 currently eligible codes according to Citizenship and Immigration Canada in 2014. If a user selects a code that is not currently eligible, *eMigrate* notifies the user to try other immigration program options.
2.   If a user selects an eligible occupational code, the application indicates the odds that an application may be accepted, and provides pathways for users to find out more information. If the eligible occupational group has reached its capped number of applicants, *eMigrate* will recommend that the user explore job titles and descriptions listed in another occupational code. [return to step 1.]
3.   If the user elects to find out more, *eMigrate* will reveal the status that the occupation is in relation to the cap. (i.e. how many other applicants a user is competing with in this category). They may then go back to step 1 or find out more information.
4.   If the user elects to find out more, *eMigrate* will reveal the relationship between an applicant's occupational selection and the overall immigration cap.
5.   Users may also find out the number of current job postings available in specific occupational codes. With this, *eMigrate* assesses the potential applicant's competitive odds for those jobs.
6.   Finally, users can find out where those jobs are, which is visualized on a map identifying the number of jobs in a specific province.

Additionally, *eMigrate* will display a user's odds of finding timely employment in a chosen NOC category in a specific geographic location. If odds are poor, it will recommend that the user select another location or go back and select another occupational area. The built in logic of the system allowing for applicants to self-identify under multiple categories runs counter to the NOC's assumption of specificity and mutual exclusion of each occupation.

5   Case Study

eMigrate is particularly significant for immigrants whose training or recent work experience relates to what might fall under the broad categories of social, new, or digital media work. Because the NOC is only updated every 5 years - with the current NOC last updated in 2011 - the NOC categories could often be considered outdated. Social media, for example, is currently not represented in the larger occupational categories, which poses a significant problem for social media professionals seeking to immigrate.

As the social media landscape changes rapidly, businesses have adopted and developed positions to encompass social media management. According to the U.S. Department of Labor, there are currently over 80 job titles that identify the use of social media (Kwon, Min, Geringer, & Lim, 2013). However, social media job titles have been poorly described and defined. A problem exists whereby social media professionals often hold job titles that do not reflect their actual work experience, and tend to adopt "traditional" job titles (e.g. brand management; marketing; design). This is compounded by the fact that at this time the NOC completely omits any recognition of social media labour as an emergent category. Interviewees often spoke of "working in social" to describe their work as a social media professional, which highlights the ambiguity of the recognition and classification of their labour. This case study illuminates the precarity of social media workers and seeks to contribute to new media, labour, and classification scholarship.

eMigrate provides foreign social media professionals information to position themselves strategically in their immigration application. This is pertinent to social media professionals who work in hybrid positions that may cross various occupational classifications.

6   Contribution

Overall, the promise of open data needs to be met with tangible results. Our cross disciplinary perspectives from knowledge organization, social media studies, and STS combine to produce an fresh perspective on a complex information environment. *eMigrate* provides algorithmic transparency to the immigration process by automating the combination of policy and occupational choices in one system to aid potential immigrants in identifying the most appropriate and advantageous employment classification. *eMigrate* anticipates the information needs of immigrants by locating and synthesizing occupational information in relation to immigration policy. The web application is situated in the pre-migration stage, when information is mainly gathered from formal sources, such as government agencies' publications and websites (Caidi, Allard, & Dechief, 2008 in Caidi et al., 2010).

This prototype web application presents opportunities for exploring the myriad ways immigrant communities locate and access information. It is intended to help researchers learn about how potential immigrants become aware of immigration and related occupational information. Long-term deployment will aid researchers exploring the skills potential immigrants utilize as they navigate various foreign information institutions and related technologies (ibid). Ultimately, *eMigrate* foregrounds some of the consequences that standard occupational classification brings to bear on the localized practice of sharing information for immigration.

References

Caidi, N., Allard, D., & Quirke, L. (2010). Information practices of immigrants. *Annual Review of Information Science, 44*(1), 491-531.

Goffman, E. (1959). *The presentation of self in everyday life.* Garden City, NY: Doubleday.

Government of Canada. (2013). *What is NOC?* Retrieved from http://www.cic.gc.ca/english/immigrate/skilled/noc.asp

Human Resources and Development Canada. (2013). *Welcome to the National Occupational Classification 2011.* Retrieved from http://www5.hrsdc.gc.ca/NOC/English/NOC/2011/Welcome.aspx

Kwon, O., Min, D., Geringer, S., & Lim, S.-K. (2013). Students perception of qualifications for successful social media coordinator. *Academy of Marketing Studies Journal, 17*(1), 109-128. http://www.cic.gc.ca/english/immigrate/skilled/noc.asp