RECORDLINK, a workshop at the U of Guelph
May 6-7 2009

Monday April 6

0830-0930
Household and Farm Transitions in Environmental Context
Susan Hautaniemi Leonard, ICPSR, University of Michigan

Farm operators' management of resources is dynamic, and offers a meaningful way to evaluate important theory about how households develop and how population and environment are related. The role of household and family in this process has received attention but remains difficult to understand without long series of data that link family, farm, and environment. Taking a household lifecycle approach, we model change in land-use practice as a function of household composition in historical and environmental context. We use data from rich, linked individual-level population and farm census records for 1860-1940 from 25 representative Kansas townships. Using multi-level mixed growth models, we find that environmental conditions are closely allied to the size of farms and acres of crops. Investment in increasing farm land is better explained by head's life course, while investment in cropping is better explained by labor availability and household lifecycle. We then refine our use of environmental data to ask whether climatic shocks – in particular the droughts for which Kansas is famous – uncoupled lifecycle from land-use cycle. We hypothesize that during droughts farmers were not able to use their available household labor effectively, and this would have interfered with growth of farm and cropped acres. Intergenerational transfers may have been jeopardized, accelerating out-migration of young adults.

0930-1030
New Estimates of Migration for the United States, 1850 – 1900
Ron Goeken, Tom Lenius and Becky Vick, Minnesota Population Centre, University of Minnesota

Using recently released linked census data, we will provide new migration rate estimates for late-nineteenth century United States. Our results suggest that previous studies generally overestimate migration rates for this period. We will also examine issues related to places of origin and destination and the impact of household structure and lifetime migration on subsequent migration decisions. The role of wealth (personal and real estate) on migration behavior will also be examined.

Early Life Conditions and Longevity: Linking Data from the 1901 and 1911 Canadian Censuses to Quebec's Civil Registration

Laurence Pilon-Marie (1), Alain Gagnon (2), Bertrand Desjardins (1), and Robert Bourbeau (1)
1 Département de démographie, Université de Montréal
2 Department of Sociology and Department of Epidemiology & Biostatistics, The University of Western Ontario

It has been suspected for a long time that early life circumstances could affect adult mortality, and scholars endorsing a life course perspective have recently reported compelling evidence for the "long arm of childhood." As this niche has not yet been developed in the Canadian context, we have established a record-linkage project tracing the early life conditions of individuals who were listed as children (less than 15 years old) in the 1901 Canadian Census for the Province of Quebec, which contains a wealth of information on socioeconomic and environmental conditions.

Taking advantage of data from the Canadian Family Project, which established a 5% fully computerized sample of the 1901 census for all provinces, we have traced the marriage and the death records of approximately 4,000 French Canadian born between 1885 and 1901 and who died after the age of 40. Marriage and death records were found, respectively, in the BALSAC database (Université de Québec à Chicoutimi) and in the Index des décès of the Institut de la Statistique du Québec. Marriage records are particularly useful in furnishing information on residence and socio-professional statuses, allowing for an examination of residential and social mobility. In parallel, since no study has yet examined whether early life conditions set the stage for survival to very old ages, we are also developing a database on a group of centenarians and their siblings.

So far, approximately 900 families including at least one centenarian listed in the 1901 Census were treated. The 1911 Census (microfilms available online) is also used to supplement the information on early life conditions and to help performing record linkages.

This presentation focuses on the difficulties involved in the linkage process and on how selection biases may affect findings. First, we compare mortality patterns from our samples to the mortality tables derived from the Canadian Human Mortality Database at the Université de Montréal. Second, using logistic regression, we build a sample selection model for which the outcome is whether a death record is found. Many of the predictors included in this model (sex, urban-rural status, father's occupation, literacy, etc.) will also be used in our models predicting mortality at older ages, helping us to assess potential biases and their impacts on findings.

1145-1230

Shifting Sands – Keeping Track of Dynamic Linked Data

John Bass (1), Sandra Silcott (2) and Len Smith (3)
1 Menzies Research Institute, University of Tasmania
2 University of Melbourne
3 Australian National University

Many large-scale historical research projects involve linking multiple data sources into groups of unit records representing individuals. Genealogical information adds a further dimension, grouping the individuals into families. The raw data are often gathered together over several years, creating a dynamic system where the structure may change as groupings are added to, aggregated or split apart and regrouped. Keeping track of links is difficult, particularly where numbers of separate analyses are carried out on data extracted at different times. Many projects will require updates as new data sources are added; this is no easy task where large populations are concerned, and where individuals may have many unit records. Privacy may also become a problem when records from recent times are included. A system developed in Western Australia for linkage of health-related data and associated genealogies is equally pertinent to historical data such as those being collected by the Tasmanian Founders and Survivors project. This project also has a strong emphasis on health information and, because it is planned to follow genealogies up to modern times, a need for privacy-preserving protocols. This presentation will describe how the "linkit" system works, and give examples showing how dynamic changes can be tracked, as well as the way in which genealogical information can be merged with unit record data.

1330-1430

LINKS, LINKing System for Historical Family Reconstruction

Kees Mandemakers, Historical Sample of the Netherlands

The LINKS project which begins June 1 aims to reconstruct all nineteenth and early twentieth century families in the Netherlands. This reconstruction will be based on GENLIAS, which is a digitized index of all civil certificates from this period. For fifteen years numerous volunteers have been working to build the index, which contains not only the names of born, deceased and married persons, but also the names of their parents, places of birth, ages and partly their occupational titles. LINKS will enlarge this system with additional sources like church registers, address books, tax registers and other large nominal administrative sources. As a consequence the system will be extended with functionality for linking addresses and patronyms. LINKS has formulated three requirements for successful reconstruction and dissemination: a) a dynamic parser which converts the input from the sources

into a standardized data structure, b) nominal record linkage procedures with self learning capacities and c) a retrieval system including GIS-references and visualization procedures. Results of a test of reconstructing families for the province of Zeeland will be presented.

1430-1530
Linking Historical Records – Are Both Computerized Record Linkage and Manual Record Linkage Considered Necessary?
Maria Wisselgren and Maria Larsson, Swedish Demographic Database and University of Umea

The Demographic Data Base started as a temporary employment project in 1973. The aim was to computerize parish registers to make them available for research. Today the DDB is a national research source and responsible for ensuring that historical data from parish registers and parish statistics are easily available for researchers from both Sweden and abroad. The number of parish records in our national archives is enormous. DDB have computerized parts of the 18th and 19th parish registers such as: catechetical registers, birth and baptism registers, banns and marriage registers, migrations registers, and death registers. In order to construct both individual biographies and relations between the individuals in all these registers it is necessary to link the individual records from the parish registers. Most of the linkage is performed by a computer program developed at the Demographic Data Base. Some records are linked manually. The question is: are both computerized record linkage and manual record linkage considered necessary? And if so, why?

1545-1645
Different Record Linkage Techniques for the Early and Late 19th Century
Gunnar Thorvaldson, Norwegian Historical Data Centre and University of Tromso

Norway has a wealth of nominative source material computerized for the period 1800 to 1910. In addition to national coverage with censuses, the church books have been transcribed for some regions and localities. When building a longitudinal database on this basis we must adjust our methods according to the variations in quality and contents of the sources over time. With examples from especially the periods around 1800 and 1900 I shall illustrate how existing registers, automatic and manual record linkage techniques can be employed to construct the longitudinal database.

Tuesday April 7

0830-0900
CCRI and LINC
Peter Baskerville, History, University of Alberta

The completion of census samples from 1911, 1921, 1931, 1941 and 1951 marks a landmark advance in digital infrastructure for Canada. This presentation will provide updated information on the data and plans for their distribution/dissemination. A proposal for another large Canadian project, Linked Infrastructure for National Censuses (LINC), is now under adjudication. An overview of LINC and preparations to date will be provided.

0900-0930
Matching Matters: A Comparison of Linkage Algorithms Using Historical Microdata
Chris Minns, Economic History, London School of Economics

This paper compares linkage outcomes under alternative phonetic algorithms in North American census data between 1880 and 1901. Soundex and Double Metaphone algorithms are used to link records from the 1881 Census of Canada to the 1901 Census of Canada, and the 1900 United States Census. This experiment will provide a comparison of the mechanical linkage rate generated by these two methods. Hand inspection of the linkages generated will allow for a rough comparison of the rate of spurious linkage under the two methods. The study will also report to what extent the two algorithms generate a common set of linked records, and whether the methods generate a set of linked individuals with broadly the same economic and demographic.

0930-1000
When Linkage Programs Fail: Assessing False Negative and False Positive Matches
Kris Inwood and Jill Leslie, Economics and History, University of Guelph

Intensive hand-linking of 2049 men and women known to operate industrial establishment establishments in 1871 allows us to track about 1416 of them from the 1871 Canadian census to the 1880 US census and the 1881 Canadian census. The use of family context provides a basis for confirming with a high degree of confidence that these are the correct matches. Taking these 1400 verified links as 'true' we assess the efficacy of the kind of decision rule embedded in most automated matching algorithms. We identify why the automated algorithm will fail to identify some matches, the circumstances in

which an incorrect match (false positive) is identified and the nature of selection bias distinguishing auto-linked records from the full 'population' of 1416 hand-linked and verified records. We argue that large-scale linkage algorithms must be sensitive to the imprecision of databases created from *de jure* census enumeration in nineteenth century North America.

1000-1030
Filling the Gap: The Use of Marriage Records To Help with Inter-census: Linkage of Young Adults in Quebec City (1851-1911)
Marc St-Hilaire, Geography/CIEQ, U. Laval

One of the most hazardous linkages is the one between a single individual living with his family at one census and the young recently married one living with his own family in the next census. This issue is critical as to inter-generational studies as to individual biographies. There are not many ways to overcome the problem: Whether the linkage relies on the individual name and surname only (and marginally – and in a risky way – on the age, the occupation, and the place of abode); whether other sources are used, which give additional clues to strengthen the link. This proposal aims to present the result of the use of marriage records to help with that kind of problematic linkages. The case population is that of Quebec City, from 1851 to 1911, for which the nominal census data has been entered by the Population et histoire sociale de la ville de Québec (PHSVQ) project. The linkage involves one cohort by census (10-year old boys), which is linked to subsequent censuses. The paper presents the results on the linkage using 1) only census data and 2) marriage records (both Catholic and non-Catholic), showing how the use of a second source enhances the overall outcome.

1045-1230
Different Record Linkage Techniques for the Same US Census Data
Ron Goeken, Tom Lenius & Becky Vick, Minnesota Population Centre
Sue Dintelman and Tim Manness, Pleaides Software, Salt Lake City

Two methods of linking late nineteenth century US census data are contrasted. One method uses time-invariant information on individuals in an attempt to minimize selection bias. The second method uses all available information for individuals in their family context.

1300-1345
Census Data to Family Data
David Barss, Jennifer Kerns and Ray Madsen
Historical Family Reconstitution, FamilySearch, Salt Lake City

This presentation will demonstrate how we are converting electronic data files

containing extracted census data into lineage linked family files for family reconstitution. It will show how we are capturing and preserving all of the family linkage data that is present in the data both stated and implied. The results are multiple linked families, and in some cases several generations of ancestry all from a single census entry. We have applied these processes to census data from the United States and Norway, and even to a set of probate data from Canada.

1345-1445
Linking Records for Transported British Convicts
Hamish Maxwell-Stewart, History and Classics, University of Tasmania

Between 1803 and 1853 69,000 male and 13,500 female convicts were transported to the British penal colony of Van Diemen's Land, later renamed Tasmania. By linking information about these individuals it is possible to determine the impact that penal servitude had on their social and physical well-being. This presentation will outline the procedures that have been used to calculate death rates for prisoners while under sentence and to examine colonial marriage and reproduction rates for those who survived penal servitude to become free. It will look at a number of problems that have been encountered during the course of this work and discuss ways in which these have either been overcome, or could be negotiated. Suffice it to say, it has proved much easier to link information about prisoners pertaining to their time under sentence than it has been for the period after their freedom was restored.

1500-1645
Collective Discussion
The concluding discussion will consider issues of software development, quality standards for longitudinal data construction and the potential for sharing of name standardization and other files that may be of broad interest.