Longitudinal Analysis, Historical Sources and Generational Change

A workshop at the University of Guelph May 24-25 2010
OVC LifeLong Learning Centre Rm 1713
http://www.recordlink.org/

MONDAY

0845
Record linkage at the Minnesota Population Center
Ron Goeken, Lap Huynh, Tom Lenius, and Rebecca Vick (Minnesota Population Center)
Paper   Tables Powerpoint
This paper will present an overview of methods used to link various samples of the United States Population Censuses to a complete-count database of the 1880 United States population. Topics include name standardization, construction of similarity scores and the use of support vector machines to classify linked records (SVMs). We discuss our preliminary data release and subsequent work on our final release. Topics include the construction of name commonness scores and birth density measures, and their impact on the final linked data. We also present a number of indirect measures assessing the accuracy of our linked data. We also discuss the construction of weights to deal with linkage rate differentials.

0945
An Automated Record Linkage System – Linking 1871 Canadian Census to 1881 Canadian Census
Luiza Antonie (U of Guelph), Peter Baskerville (U of Alberta), Kris Inwood (U of Guelph) and Andrew Ross (U of Guelph)
Paper   Powerpoint
This paper describes a recently developed linkage system for historical Canadian censuses and its application for tracking people from 1871 to 1881. The record linkage system incorporates a supervised learning module for classifying pairs of records as matches and non-matches. The classification module was trained using a set of true links that was created by experts. We evaluate the first results and provide a road map for further experimentation.

1045 break

1100
Using family lineage data to improve record linkage success
David Barss (Family Search)

This paper demonstrates how using lineage linked family data from census records expands the data that can be gathered from the census and thereby improves the record

linkage opportunities and success when merging census data with other census years or other record collections. It shows results from using lineage linked data samples and versions of the same data that are not lineage linked. It also shows several family relationships that can be preserved from the census that are lost using the household perspective. As well as extended relationships that can be captured with "derived records" generated as place holders to connect identified family members.

1145
"To Fill Dishonored Graves": Assembling life course data for transported British convicts
Hamish Maxwell-Stewart (University of Tasmania)
Slides (pdf)
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. While the documentation for these individuals is highly detailed, information about death was only occasionally recorded on each convict's file. Our aim is to fill this gap by linking with other classes of records, including the surgeons' reports for the voyage to Australia and the civil death registers for Van Diemen's Land/Tasmania. As a result of this we have been able to build up a detailed picture of death rates during the voyage to the Antipodes and the initial years in the colony while the convicts were still under sentence. We are now attempting to extend this picture by exploring death rates for former convicts. This process has raised a number of interesting issues and I will outline a range of approaches we are exploring in an attempt to address these.

1230 lunch

1315
What accounts for the movement of rural household heads in Logan Township?
Peter Baskerville (University of Alberta)
Paper
This paper provides a first report on a project which has linked/traced residents of Logan township (pop: 3196) in 1871 to an unusually large catchment area: the whole of Canada in 1881 and the United States in 1880. The linkage was done by hand in accordance with rigorous rules to provide a set of true links both for use in testing and establishing a computer generated linkage program and to further a project which focuses on credit and community in Perth County Ontario in the late 19th and early 20th centuries. The paper focuses on household heads in 1871. Of the 521 heads for whom we have no death information, we linked 415 (79.7%) to the Canadian or US census in 1881 and 1880 respectively. We could not link 106 of the 521 heads. Two hundred and ninety-nine (72%) of the linked heads stayed in Logan and 116 (21%) moved. Through a series of logistic regressions this paper seeks to establish the personal, familial, and environmental attributes that most influenced the probability of Loganites persisting or moving in the 1870s and compares the situations of those who moved to a new region with those who persisted in Logan.

1415

To combine Swedish historical data with modern population registers
Elisabeth Engberg (Umeå University), Maria Larsson (Umeå University) and Maria Wisselgren (Umeå University)

The Demographic Data Base (DDB) started out as a temporary employment project in the early 70's. The aim of the organization was to computerize parish registers to make them available for research. Today DDB is a national research resource and responsible for ensuring that historical data from parish registers and parish statistics are easily available for researchers in both Sweden and other counties. Since the 70's DDB has digitized parish registers and constructed one of the largest historical population databases in Europe, based on church records from the 18th and 19th century. The individual historical database currently contains information about more than one million people, has a depth of about ten generations, and includes around eighty parishes. The database is available for research and has been used by researchers both in the social sciences and humanities as well as in medicine and science, both in Sweden and internationally. However, interest in and demand for population data from the 20th century have increased and the question have been raised about the possibility of linking historical data with modern population registers. In Sweden there is a lack of digitized population data on an individual level during the period 1900 to 1950's, from where Statistics Sweden is having digitized data. In order to meet the present needs within several fields of research, a new infrastructure is being developed by the DDB in close cooperation with Statistics Sweden. In this presentation we will talk about the preparatory work behind this new infrastructure. The different stages of linkage will be in focus as well as methods of secure linkage that has been developed by the DDB.

1500 break

1515
Reconstructing the history of morbidity: the Hampshire Friendly Society and its records
Martin Gorsky (London School of Hygiene and Tropical Medicine), Aravinda Guntupalli (University of Southampton), Bernard Harris (University of Southampton), Andrew Hinde (University of Southampton)
Paper   Powerpoint
During the last two decades, economic, social and demographic historians have achieved significant advances in our understanding of the history of health and disease. However, the majority of these studies have been concerned either with the history of 'positive health indicators', such as height, or with the history of mortality. It has proved much harder to reconstruct the history of non-fatal illnesses, despite the valiant efforts of researchers such as James Riley, George Alter, Herb Emery and John Murray.
In addition to this, there has also been a growing interest in the use of historical records to understand what is often termed 'lifecourse epidemiology'. Much of this work was stimulated by David Barker's research into the impact of early-life experiences on adult development and mortality. However, other researchers, such as George Davey Smith, have argued strongly that we also need to take account of the impact of 'insults' across the lifecourse in our efforts to understand mortality at higher ages.

The records maintained by the Hampshire Friendly Society offer an opportunity to address both of these issues. The Society recorded information about the sickness episodes experienced by its members from 1825 onwards. In May 2007 we were awarded a grant by the UK Economic and Social Research Council which enabled us to construct a database containing information about 5552 individuals who joined the Society between 1825 and 1939 and experienced sickness episodes between 1825 and 1981. We are currently in the process of analysing these data, and hope to be able to present new results showing how sickness patterns changed over the course of these men's lives and between different membership cohorts. We also hope to be able to present new results which illustrate the relationship between the sickness episodes experienced in earlier life and later-life mortality.

1615
Parsing data from several sources in the Netherlands with LINKS
Kees Mandemakers (Historical Sample of the Netherlands)
Outline
LINKS stands for LINKing System for historical family reconstruction. In first instance the project aims at reconstructing 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. In second instance the system will be enlarged with other sources like church registers, address books, tax registers and other large nominal administrative sources. With the church registers we will attempt to link 18th century material into families as well and make a connection with the 19th century material, especially the death certificates.
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. For a schematic overview, see the scheme below. Important is also the feedback given to the archives of all kind of errors and inconsistencies that we will find in the sources delivered by the archives themselves.
In my presentation I will concentrate on the parsing part of the project. By parsing we mean converting and standardizing the data from all kinds of sources in a universal format suited for the linking process in a way as efficient as possible. I will elaborate on the different sources and the way we handle/standardize the data before starting the matching process. Because a lot of data are hidden in fields called 'miscellaneous' quite an effort is put in systematic scanning, separating and retrieving data out of these fields. Preparing the data for the matching procedures means that quite a lot of redundancy is stored in the database called LINKS_cleaned. Actually LINKS_cleaned is not one database but a system of five interconnected but separated database-systems. In my presentation I will explain this system, sketch the several parts and go into the operational results.

TUESDAY

0845

Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-60

Jeremy Atack (Vanderbilt University), Fred Bateman (University of Georgia), Michael Haines (Colgate University) and Robert Margo (Boston University)

[Paper](#)

Using a newly developed geographic information system transportation database, we study the impact of gaining access to rail transportation on changes in population density and the rate of urbanization between 1850 and 1860 in the American Midwest. Differences-in-differences and instrumental variable analysis of a balanced panel of 278 counties reveals only a small positive effect of rail access on population density but a large positive impact on urbanization as measured by the fraction of people living in incorporated areas of 2,500 or more. Our estimates imply that one-half or more of the growth in urbanization in the Midwest in the late antebellum period may be attributable to the spread of the rail network.

0945

Community Trees: From concept to publication

Ray Madsen (Family Search)

1030 break

1045

A longitudinal perspectives on the French-English stature divide

John Cranfield (University of Guelph), Kris Inwood (University of Guelph) and Asher Kirk-Elleker (University of Guelph)

[Paper](#)

In this paper we bring together medical examinations during World War One with household information for 9000 soldiers located in the 1901 census. The linked data provide evidence of intergenerational occupational mobility on a scale sufficient to caution against the use of own-occupation as a proxy for socio-economic status during childhood. The two measures of socioeconomic status, own-occupation and that of household head during childhood, lead to analysis of height differentials indicating that Quebec-born and labourers were especially short and that stature for all groups declined more or less continuously during the last third of the 19th century. The 1901 census linkage adds new information that confirms being francophone is the principle source of smaller stature in Quebec. The intergenerational perspective adds useful detail and sharpens our impressions of significant differences physical well-being by region and occupational class and of a decline at the end of the 19th century.

1130

How did teacher recruitment and teacher career paths change as school provision became centralized? The case of Victorian Britain

David Mitch (University of Maryland)

[Powerpoint](#)

On the accession of Victoria to the British throne, much of elementary schooling was operated on a private, adventure basis. By the time of her death in 1901, elementary schooling was largely state funded with an extensive system of rules, inspection and localized educational authorities in place to supervise its operation. One recurring issue in the history of Victorian education has been whether Victorian elementary school teachers were increasingly recruited from those with middle class parentage as standards of teacher qualification became more rigorous. Another important question is what happened to the turnover of teachers and the related point of their duration in the profession as schools were subject to funding according to examination performance. This paper will address these questions by linking those listing teacher related occupations in British censuses between 1841 and 1901.

1215 lunch

1300
Between family and household: the linkage of civil records and census data, a pilot project on Quebec City, 1851-1911
Marc St-Hilaire (University of Laval) and Hélène Vézina (Université du Québec à Chicoutimi)
Paper
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.

1400
Reconstituting the population of Antigonish, Nova Scotia using probabilistic record linking
Sue Dintelman (Pleiades Software), Tim Maness (Pleiades Software) and David Barss (Family Search)

Reconstituting populations is valuable for many reasons. Historians and demographers reconstitute populations to do longitudinal population studies of such things as migration and birth rates. Geneticists use populations to study inheritance mechanisms and to help locate specific genes. Genealogists are interested in tracing ancestry back through time and finding context for their ancestors. Reconstituting populations from vital statistics

records such as births, deaths and marriages and from census records is attractive, but until recently, projects of any size required a large investment. Today with available software and low cost hardware smaller organizations and even individuals can undertake population reconstitution projects. This paper outlines the specific steps used to merge birth and marriage records from Antigonish Nova Scotia into a genealogy.

1445 break

1500
A longitudinal database for Norway 1800-2010
Gunnar Thorvaldsen (University of Tromsø)
Paper   Map   Powerpoint
A population registry is longitudinal database in the sense that the aim is to maintain a continuously updated overview of the population in a geographic area. The registry may be national, regional or local. Longitudinal population registers are topical for two reasons in the historical context. First, as a contemporary methodology, they have for some decades in country after country replaced the traditional population control instruments based on censuses or vital registers. Second, historians in several countries are building historical population registers in order to be able to base their research on continuous collective biographies. This paper illustrates these contemporary and historical developments primarily with examples from Norway, where such registers have a century-long history and where we are currently building a historical population register on the basis of the NAPP censuses. Supplemented with other source material, this aims to include as many of the 9.7 million people that were born or immigrated between 1735 and 1964 as possible. The latter year marks the introduction of the national or Central Population Register (CPR), which superseded contemporary local registers with a long history.