

Data Survival on a Seemingly Deserted Island?

A. Michelle Edwards, Cornell University

Berenica Vejvoda, McGill University







“Survival, to continue to live or exist, especially in spite of danger or hardship”


Canadian Historical Census of Agriculture

- Started as a personal interest project
- Transcribing 1871, 1881, 1891, 1901, 1911, and 1921 Census of Agriculture for Ontario, and the Maritime provinces

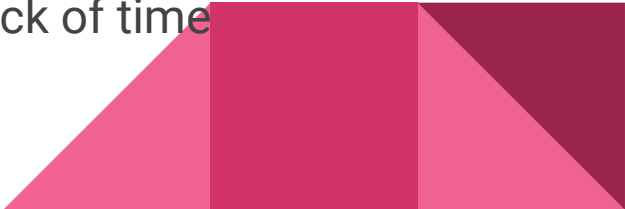
- Challenges:
 - Geography changes - how do you match to Canada today??
 - Data collected changed over time
 - 1901 - only county level data available



Historical Aggregate Census Project (OCUL)

- Born digital aggregated census data - 1961-- 1981
 - Look at who has what; what documentation is available; what format the files are in
 - What is unavailable?
 - Inventory phase
 - 395 files identified
 - 325 at UofT;
 - 171 at StatCan;
 - 105 at Western;
 - 20 have no known location
 - Goal to make these open data
- 

Canadian First Nations

- Census of Canada, 1981: Native People's Files
 - UofT: <http://data.library.utoronto.ca/content/census-canada-1981-native-peoples-Files>
 - Aggregate basic summary tabulations
 - Syntax and raw text files
 - Not processed
 - Some documentation
 - Requested recently by Epidemiologist studying the history of tuberculosis among First Nations in Canada
 - Dilemma: Couldn't utilize unprocessed data due to lack of time
- 


Roper Center Public Opinion Poll Archive

- Many data rescue opportunities available
 - International AIDS survey - 1980s - all on tape
 - Ronald Reagan polls - rescued from tape - next step

- Challenges:
 - Data formatting
 - Resources
 - Deciding which should be rescued



How do we decide what to rescue?

- Will it be of use to more than one researcher?
 - Will data (e.g. models) help predict future changes and detect trends?
 - Time and resources
 - Availability of technologies to convert older file formats
 - Is documentation available?
 - Quality of original data collection and source
 - What if data does not conform to today's standards of protecting against redisclosure risk?
- 

Funding?

- Are you able to fund data rescue projects?
- What grants would support these initiatives?
- Do you have home institution support for spending time on “legacy projects”
- Should funding efforts involve multiple stakeholders and institutions?



Moving forward as a Community

- Members of the RDA Data Rescue Interest Group
- Are there any other groups?
- Interest Group in IASSIST?
 - Identify funding
 - Inventory of potential data rescue projects
 - Collaborative processing & access





Thank-you!!