DATA LIBRARY PRE-CONSULTATION WORKSHEET

Finding data is challenging. Use this outline to conceptualize your data requirements.

SECONDARY DATA ANALYSIS

Secondary data analysis is re-using data for analysis other than its original collection purpose. The University of Alberta Data Library has a large collection of valuable microdata and aggregate data, available to all students, faculty, and staff.

BEFORE YOU BEGIN

When looking for secondary economic or social data, there are three basic starting concepts:

1) Identify the **substantive social or economic content** of your research topic or research question. (Some examples: the links between educational attainment and earnings; the purchasing choices made by retired people; stock and bond price movements during wars). Then, find an available data source containing variables which measure your chosen substantive content.

   Q1) What is the substantive content of your topic?

   Q2) List some potential relevant variables.

2) Identify the **geographic context** -- all social content occurs within a unit of space or geography. Do you want national level (e.g., Canada), provincial (e.g., Alberta), or local (e.g., Edmonton neighbourhoods) data? Do you want to compare jurisdictions (e.g., OECD countries)? Have geographic boundaries or terminologies changed over time (e.g., "former Yugoslavia")?

   Q3) What is the geographic context of your topic?

   Q4) Have there been any geographical boundary or terminology changes over time?

3) Identify the **time frame** for your research. In fact, time itself may be the important organizing factor for your analysis (e.g, trend analysis). Consider the frequency of observations (e.g. annual, monthly) and how many observations you require. (See "time series data" next page)

   Q5) What is the desired time span of your topic?

   Q6) What is the frequency of observations?

   Q7) What is the minimum number of observations you need for valid analysis?

DATA SOURCES

Perhaps the most useful question to ask when searching for data is: who would be interested in this data enough to undertake the effort of collecting and organizing it? It could be government, financial organizations (e.g. banks, investment advisors), marketing firms, trade associations, etc. Previously-published academic articles are a valuable source for locating data source leads and techniques.

Q8) What sort of organization might have collected the data you're interested in?

Q9) List 1-3 academic articles you've found and read relating to your topic.

UNITS OF OBSERVATION & ANALYSIS

The unit of observation is the entity (e.g., a survey respondent) about which measurements have been made and subsequently stored as variables. The unit of analysis is the entity about which research generalizations are being made. (Often, the unit of observation and the unit of analysis are the same entity; if not, additional processing is involved to produce the desired unit of analysis.) A data file is usually organized with units of observation/analysis in the rows and the variables in the columns.

Q10) What is your unit of observation/analysis?

Q11) What variables have you chosen to describe your unit of observation/analysis?

DATA TYPES

Microdata: files contain the characteristics of individuals, institutions, or events at the level that the information was collected (e.g., all actual responses from each respondent in a census)
- **Public use microdata** are anonymized to protect the individual identities of entities or persons. They may be available through licenses (e.g. Statistics Canada's DLI).
- **Confidential microdata** have _not_ been anonymized. They are accessible only in a secure area and require prior permission via formal application. The **Research Data Centre** is one such source located at the UofA.

Sources of public-use microdata sets available through the Data Library include:
- The DLI (including such datasets as the Survey of Household Spending and the Labour Force Survey), ICPSR Direct, or Roper Center datasets.

Aggregate data: files contain statistical summaries organized in a data structure often organized by time or geography; e.g. time series data, geo-referenced tabular data, and census tables.

Some of the sources of aggregate data available through the Data Library are:
- CANSIM
- the Canadian Census of Population
- Labour Force Survey (CANSIM tables and microdata)

Q12) Does your research topic require microdata or aggregate data?