



## Social Science Data Services: Data Research Process

<http://libraries.mit.edu/guides/subjects/data/>

<u>Steps of the Process</u>	<u>Resources:</u>
<b>1. Define the Research Question</b>	<ul style="list-style-type: none"><li>▪ Colleagues, faculty or TAs associated with your course.</li></ul>
<b>2. Define the Type of Data Needed</b>  Consider: <ul style="list-style-type: none"><li>▪ What variables/measures are needed?</li><li>▪ Should the study take place over time and/or be repeated with the same subjects?</li><li>▪ At what level of geography do you need the data (e.g., city, state, country, etc.)?</li><li>▪ Do you want to conduct geospatial analysis?</li><li>▪ In what form do you need the information? Data can be categorized along the spectrum from data for information (e.g., reference book such as Statistical Abstract of the United States) to data for research (e.g., large survey data set such as the General Social Survey)</li></ul>	<ul style="list-style-type: none"><li>▪ Understanding Data Files: <a href="http://libraries.mit.edu/guides/subjects/data/training/tutorials/understanding.html">http://libraries.mit.edu/guides/subjects/data/training/tutorials/understanding.html</a></li><li>▪ GIS Services at MIT: <a href="http://libraries.mit.edu/gis">http://libraries.mit.edu/gis</a>.</li></ul>
<b>3. Identify Potential Data Sources</b>	<ul style="list-style-type: none"><li>▪ Data Access: <a href="http://libraries.mit.edu/guides/subjects/data/access/">http://libraries.mit.edu/guides/subjects/data/access/</a></li><li>▪ ICPSR: <a href="http://libraries.mit.edu/get/icpsr">http://libraries.mit.edu/get/icpsr</a></li></ul>
<b>4. Determine Usefulness of the Data</b>  Consider: <ul style="list-style-type: none"><li>▪ Assess whether or not potential data sources will help to answer your research question and fit within your project.</li><li>▪ Consider: sample design, method of collection, quality of the data, measures and units of analysis, variables, file structure, and validity of the source.</li><li>▪ For raw data files, you need the codebook, the documentation of the contents and structure of the data file. It contains a description of the research project and methods, the data collection instrument, and a data dictionary listing the variables, their locations, and the coded values.</li></ul>	<ul style="list-style-type: none"><li>▪ Understanding Data Files.: <a href="http://libraries.mit.edu/guides/subjects/data/training/tutorials/understanding.html">http://libraries.mit.edu/guides/subjects/data/training/tutorials/understanding.html</a></li><li>▪ Data Access: <a href="http://libraries.mit.edu/guides/subjects/data/access/">http://libraries.mit.edu/guides/subjects/data/access/</a> Also provides access to data documentation.</li></ul>

## Social Science Data Services: Data Research Process (cont'd)

---

<b>5. Access the Data</b>	<ul style="list-style-type: none"><li>▪ Data Access: <a href="http://libraries.mit.edu/guides/subjects/data/access/">http://libraries.mit.edu/guides/subjects/data/access/</a></li><li>▪ Locating Data: <a href="http://libraries.mit.edu/guides/subjects/data/access/locating.html">http://libraries.mit.edu/guides/subjects/data/access/locating.html</a></li><li>▪ ICPSR: <a href="http://libraries.mit.edu/get/icpsr">http://libraries.mit.edu/get/icpsr</a></li><li>▪ HMDC: <a href="http://libraries.mit.edu/get/hmdc">http://libraries.mit.edu/get/hmdc</a></li><li>▪ Keep in mind terms for responsible use, namely to maintain confidentiality of survey respondents.</li></ul>
<b>6. Format the Data</b>  Consider: <ul style="list-style-type: none"><li>▪ If the data is in the form of easily-retrievable statistics, no further analysis may be needed. If you have a raw data file, then you need to do some formatting and analysis.</li></ul>	<ul style="list-style-type: none"><li>▪ Import the data into the preferred software package (e.g. Stata, R, SAS, SPSS)</li><li>▪ If the data is available in a format compatible with another software package, you can use transfer software (e.g., Stat/Transfer) to convert it to your preferred package; see software availability: <a href="http://libraries.mit.edu/guides/subjects/data/software/">http://libraries.mit.edu/guides/subjects/data/software/</a>.</li><li>▪ Extract variables of interest (e.g., if you are interested in economic but not lifestyle variables).</li><li>▪ Subset observations of interest (e.g., if you are interested in only respondents from a certain age group).</li><li>▪ Note: some older data sets may be in print only; decide whether it's worth it to manually enter the data.</li><li>▪ Working with Data Files.: <a href="http://libraries.mit.edu/guides/subjects/data/training/tutorials/working.html">http://libraries.mit.edu/guides/subjects/data/training/tutorials/working.html</a></li></ul>
<b>7. Analyze the Data</b>	<ul style="list-style-type: none"><li>▪ Software and Hardware: <a href="http://libraries.mit.edu/guides/subjects/data/software/">http://libraries.mit.edu/guides/subjects/data/software/</a></li><li>▪ Tutorials on Working with Data Files: <a href="http://libraries.mit.edu/guides/subjects/data/training/tutorials/working.html">http://libraries.mit.edu/guides/subjects/data/training/tutorials/working.html</a></li><li>▪ You may also consider geospatial methods of analyzing the data. See GIS Services at MIT: <a href="http://libraries.mit.edu/gis">http://libraries.mit.edu/gis</a>.</li><li>▪ Citation Software (for writing your paper): <a href="http://libguides.mit.edu/references">http://libguides.mit.edu/references</a></li></ul>
<b>For More Information</b>	<ul style="list-style-type: none"><li>▪ Social Science Data Services Web Site: <a href="http://libraries.mit.edu/guides/subjects/data/">http://libraries.mit.edu/guides/subjects/data/</a></li><li>▪ Katherine McNeill, Social Science Data Services Librarian, <a href="mailto:mcneillh@mit.edu">mcneillh@mit.edu</a>, 617-253-0787</li><li>▪ Need help? Ask Us!: <a href="http://libraries.mit.edu/ask-us/">http://libraries.mit.edu/ask-us/</a></li></ul>