Dataset Purchase Guidelines and Best Practices

❖ Dataset requests should go through CMTAG for approval.
❖ Electronic and Continuing Resources will negotiate the license and communicate with relevant subject liaison as necessary
❖ Please contact Dr. Rob O'Reilly (roreill@emory.edu) as well for additional assistance in vetting a dataset request.

Subject Librarian should gather the following information:
❖ Publisher/producer of the data, vendor (foreign or domestic)
❖ Data product name or title
❖ Years available and years requested
❖ If there are different geographies available, what they are and which the researcher wants
❖ Price
❖ Where specifically the researcher found information about the data set (a citation in an article, on a website, etc.)
❖ Justification for the need of the database (what is the research to be supported)
❖ Whether the data can be made available campus-wide for all faculty, staff, and students currently affiliated with the university
❖ When possible, documentation describing the contents of the data – geographic area and level, time period, frequency of data (e.g. annual, quarterly, one-off), and substantive content (e.g. a list of variables in the data)
❖ How the data can be accessed, e.g. via a database interface that is available on-line, via client software that has to be installed on specific workstations, as a file or collection of files that can be placed on a server, or as a file or collection of files that can be accessed via a point of contact such as a specific librarian
❖ The formats in which the data can be used, specifically whether those formats are ones that can be readily used in statistical software and/or GIS applications, or can be readily converted into such formats

There are other important criteria and considerations that the subject librarian should consider:
❖ Is the dataset available for the campus, a school, a department, or an individual faculty member? Answer will dictate terms of license AND funding source (e.g., if the dataset can support multiple researchers across the university, then the Library will be more likely to purchase):
   o Encourage the use of research and grant funds to share costs of datasets of potential use to the entire Emory community, as well as allocated subject funds
   o The option of campus-wide access for a data resource should always be pursued. When data can only be available for usage by specific

individuals or departments, purchases should be funded through the resources of those parties.

❖ Price, and whether it's an outright purchase or if access is time-limited or "leased"
  o If retaining of data or destruction of data is a requirement (note that this is unenforceable by Emory, and can not be supported)
❖ How many researchers are likely to use the data product
❖ Can data that were used for analysis done by a researcher (faculty and/or student) in the research process be publicly shared to comply with journals that have a data sharing policy? Note that the data availability policies of some journals now require authors of submitted manuscripts to provide not just the dataset that was actually analyzed in the manuscript but also the source data that were cleaned and assembled into that analyzed dataset. Journals will make exceptions for data that cannot be shared for terms-of-use reasons, so it will be helpful to have this matter clarified so that the information can be shared with people who wish to use the data.
❖ Can the data be shared with folks at another university also working on the project? Is this a need for the primary users of the data set?
❖ Does the vendor/provider require any user of the data to sign a confidentiality agreement or the like to make use of the data? Are there any other sorts of administrative overhead associated with the resource?

Examples of Data Collection Policies at Other Libraries:

* Brown University: http://library.brown.edu/about/datacenter/ and http://library.brown.edu/about/datacenter/DataCollDevPolicy_08-20-10.docx

* McMaster University:
http://library.mcmaster.ca/maps/Library_Data_Service_Collection_Policy.pdf

* James Madison University:
http://www.lib.jmu.edu/faculty/datasetcdpolicy.aspx

* MIT: http://libraries.mit.edu/guides/subjects/data/access/acquisition.html
and http://libguides.mit.edu/ssds/suggest

* Michigan State University:
http://libguides.lib.msu.edu/dataservicescollectiondevpolicy

* Northwestern University: http://www.library.northwestern.edu/libraries-collections/evanston-campus/government-information/social-science-data-services/policies

* University of British Columbia:
http://data.library.ubc.ca/guide/collections_policy.html

* University of North Carolina at Chapel Hill:
http://library.unc.edu/services/data/purchase/

* Yale University: http://web.library.yale.edu/collection-development/science-and-social-science-data