A Flowchart for the Typical Process of Content Analysis Research

from *The Content Analysis Guidebook*, by Kimberly A. Neuendorf
1) **Theory and rationale**  What content will be examined, and why? Are there certain theories or perspectives that indicate that this particular message content is important to study? (e.g., Studies on violent television have shown that children may be affected; hence, we analyze the amount and type of aggression shown on TV.) Library work is needed here. Will you be using an integrative model, linking content analysis with other data to show relationships with source or receiver characteristics? Do you have research questions? Hypotheses?

2) **Conceptualization decisions.** (Remember, you are the boss! There is no one right way). What variables will be used in the study, and how do you define them conceptually? You may want to screen some examples of the content you’re going to analyze, in order to make sure you’ve covered everything you want.

3) **Operationalization measures**  Your measures should match your conceptualizations (this is called internal validity). What unit of data collection will you use? You may have more than one unit (e.g., a by-utterance coding scheme and a by-speaker coding scheme). Are the variables measured well (i.e., at a high level of measurement, with categories that are exhaustive and mutually exclusive)? An “a priori” coding scheme describing all measures must be created. Both face validity and content validity may also be assessed at this point.

   Human Coding

   ![Diagram](http://academic.csuohio.edu/kneuendorf/content/resources/flowc.htm)

   Computer Coding

4a) **Coding schemes**  You need to create the following materials:

   1. **Codebook** (with all variable measures fully explained)

   2. **Coding form**

4b) **Coding schemes**  With computer text content analysis, you still need a codebook of sorts—a full explanation of your dictionaries and method of applying them. You may use standard dictionaries (e.g., those in Hart’s program Diction) or originally created dictionaries. When creating original dictionaries, be sure to first generate a frequencies list from your text sample, and examine for keywords and phrases.
5) **Sampling**  Is a census of the content possible? (If yes, go to #6). How will you randomly sample a subset of the content? This could be by time period, by issue, by page, by channel, etc.

6) **Training and initial reliability.** During a training session in which coders work together, find out whether they can agree on the coding of variables. Then, in an independent coding test, note the reliability on each variable. At each stage, revise the codebook/coding form as needed.

7a) **Coding.** Use at least two coders, in order to establish intercoder reliability. Coding should be done independently, with at least 10% overlap for the reliability test.

    Human Coding

7b) **Coding.** Apply dictionaries to the sample text to generate per-unit (e.g., per-news story) frequencies for each dictionary. Do some spot checking for validation.

    Computer Coding

8) **Final reliability.** Calculate a reliability figure (percent agreement, Scott’s pi, Spearman’s rho, or Pearson’s r, for example) for each variable.

9) **Tabulation and reporting.** See various examples of C.A. results to see the ways in which results can be reported. Figures and statistics may be reported one variable at a time (univariate), or variables may be cross-tabulated in different ways (bivariate and multivariate techniques). Over-time trends are also a common reporting method. In the long run, relationships between C.A. variables and other measures may establish criterion and construct validity.