

Video 1: Reverse Engineering Your Topic

Jylisa Doney

University of Idaho, Social Sciences Librarian

jylisadoney@uidaho.edu

Transcript

00:00 Hi everyone,
00:01 welcome to our video series on finding data.
00:03 My name is Jylisa Doney
00:05 and I am the Social Sciences Librarian
00:06 here at the University of Idaho.
00:09 Throughout this video series,
00:10 we're going to focus on strategies
00:12 for specifying your data topics,
00:14 identifying data creators,
00:16 searching within data platforms,
00:18 and evaluating and citing data.
00:21 In Video 1, I'm going to introduce a process
00:24 that can help us ask more precise questions
00:26 about our research topic and data needs.
00:29 When we begin our research, we usually
00:31 have a general idea about our topic,
00:33 such as investigating whether there is a relationship
00:36 between educational attainment and poverty in Idaho.
00:39 A general topic is a great start,
00:41 but we need to get more specific when looking
00:43 for existing data related to this topic.
00:46 One way to do this is to
00:48 reverse engineer our topic and identify
00:50 the most significant aspects of our data need,
00:53 specifically the who, what, when, where, and why.
00:57 Completing this strategy prior to searching for data
01:00 ensures that we have a clear understanding
01:02 of what a data set needs to include
01:04 in order to meet our needs.
01:06 Let's go through each of these parts in more detail.
01:09 The "who" portion of a topic reveals the sample
01:12 or population you are interested in.
01:14 It could be general, such as households,
01:16 or specific, such as college students
01:18 who work part-time.
01:20 Defining the "what" of a topic allows you to determine
01:23 the type of data you are looking for.
01:25 The two main types of data are
01:27 aggregate data and microdata.
01:29 Aggregate data is raw data
01:31 that compiles sets of separate records.
01:34 Aggregate data is often used to protect
01:36 the anonymity of survey respondents.

01:39 Aggregate data often appears in a spreadsheet,
01:41 such as this one, where the separate
01:43 or individual-level responses from all respondents
01:47 are compiled based on the county
01:48 where the respondents live.
01:50 When the raw data appears as separate records,
01:53 it is considered microdata.
01:56 In social science research, microdata can often reflect
01:59 an individual or a household.
02:01 As you can see in this example,
02:03 each row of data is from a single record
02:06 and represents one individual's responses
02:08 to specific survey questions.
02:11 When considering the "who" and "what"
02:14 of your research topic and related data need,
02:16 it is also important to keep your
02:19 unit of observation and unit of analysis in mind.
02:22 In research the unit of observation
02:25 is the level at which the data is collected.
02:28 While the unit of analysis
02:29 is the level at which the data is analyzed.
02:32 The unit of observation and the unit of analysis
02:35 may be the same or different
02:37 depending on the questions you ask
02:38 and how you analyze your results.
02:40 For example, a unit of observation may be the individual,
02:44 but if we combine individual responses
02:47 and analyze our results based on county of residence,
02:50 our unit of analysis would be the county-level.
02:53 As you progress through your project,
02:55 keep these units in mind
02:57 as they can influence the data you select
02:59 and how you choose to analyze your data.
03:02 The "when" question asks us to consider
03:04 the age of the data.
03:06 Depending on your data need,
03:08 you may be interested in historic data,
03:10 data from multiple years to investigate
03:12 changes over time,
03:14 or the most current data.
03:16 The "where" question identifies the location
03:18 of your sample or population.
03:20 It could be a specific city, state,
03:23 country, school, organization, etc.
03:26 In Video 4, we'll talk a bit more about how data sets
03:30 often include more than one location
03:32 and how you can focus in on the specific "where"
03:35 once the data is downloaded.
03:36 Lastly we have the "why" question,
03:39 which underpins your entire research topic or question.
03:42 A few examples of "why" include

03:44 testing a hypothesis,
03:46 practicing statistical analyses or methods,
03:49 and comparing your data to another data set.
03:51 Identifying your "why" is especially important
03:54 because it can impact which search strategy
03:57 you use to find relevant data.
03:59 On the screen right now is an example
04:01 of what a reverse engineered topic can look like.

[Who: People 25 and older

What: Microdata about poverty and educational attainment

When: Current or most recent

Where: Idaho, preferably with the county included

Why: Test a hypothesis that poverty and education attainment are negatively correlated]

04:04 A reverse engineered topic is a good starting point,
04:07 but it is important to be flexible
04:09 when you start searching for data.
04:11 For example, if we were to encounter
04:13 date limitations with this sample topic,
04:16 we would need to be open to either changing our "when"
04:19 and utilizing older data,
04:21 or reworking our topic
04:22 to reflect the data that is actually available.
04:25 Coming up in Video 2, I'm going to share
04:28 three main strategies for finding data
04:31 and highlight a few useful data resources.