4

Using Online Tools

Learning Objectives

By the end of this section of the SPIVA Toolkit, you should be able to:

- Identify at least two publicly available online data sources to effectively design community-based surveys;
- Describe at least four features to consider when choosing to use or purchase an online survey tool; and
- Discuss the purpose of a relational database management system and its use in managing and analyzing data

There is an array of tools available online to assist us with data collection and analysis. A brief synopsis of some useful tools appears in the following section. Each of these tools corresponds to a certain stage in the survey design, data-collection, and data analysis process (Figure 4.1).

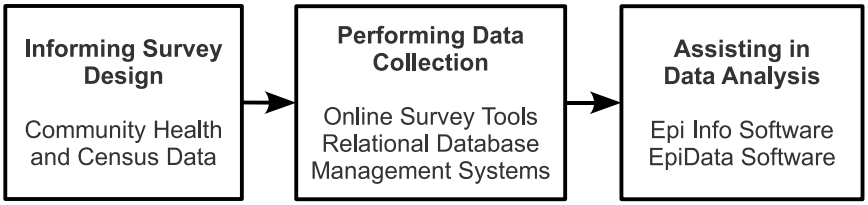

Fig. 4.1. Online tools are available for each survey stage

4.1 Using Community Health and Census Data

Having access to timely, accurate, and properly formatted data sets can be valuable to public health practitioners. The following tools are useful resources that can provide us with the background information necessary for effective survey design:

- Community Health Data Sets
- United States Census Data
- Data.gov

4.1.1 Community Health Data Sets

On June 2, 2010, the Department of Health and Human Services (HHS) launched their Community Health Data initiative, with the goal of releasing more health data in formats that are accessible and easy to use. HHS is seeking to promote the creation of software applications that will make use of the data to produce more effective health outcomes. While this initiative is in its early stages at the time of this writing, it has the potential to become a useful resource to an array of public health practitioners. The HHS Community Health Data website can be found at: http://www.hhs.gov/open/datasets/communityhealthdata.html.

4.1.2 United States Census Data

The U.S. Census Bureau has information from a number of surveys, most notably the decennial census, available through its website, which can be found at: http://factfinder.census.gov/home/saff/main.html.

Information from Census 2000 is searchable down to the city-block level by accessing the "Geography" section of the "Download Center." Data from Census 2010 will be available beginning in April 2011. This tool can help users to access local population data to help inform preparedness and planning activities. It can be found online at: http://factfinder.census.gov/servlet/DCGeoSelectServlet?ds_name=DEC_2000_SF1_U

4.1.3 Data.gov: "Empowering People"

In May 2009, the Federal Chief Information Officer of the United States launched Data.gov "to increase public access to high value, machine readable data sets generated by the Executive Branch of the Federal Government." Data.gov is the central repository for data sets being made available to the public as part of the Open Government Initiative. The website publishes data from a variety of federal agencies and promotes the development of software that takes advantage of these data. The website breaks data into three "catalogs": the "Raw Data Catalog" provides users with access to information in machine-readable files such as .xml, .csv, and .kml formats that can be instantly viewed through different applications; the "Tools Catalog" provides users with a listing of data sets with links to tools

designed to manipulate those data; and the "Geodata Catalog" provides access to data sets related to geospatial and mapping functions. These catalogs, as well as links to applications created by third-party developers, can be found at the following web address: `http://www.data.gov`.

Fig. 4.2. Data.gov: Empowering People—An Official Web Site of the United States Government

4.2 Using Online Survey Tools

Commercial online survey tools such as SurveyMonkey, Zoomerang, and SurveyGizmo provide an easy way to implement electronic surveys. This type of survey may be used at a site with computers available for the public to use. It is also an increasingly convenient avenue by which to reach certain communities in their own homes at relatively low-cost.

There are several important features to consider when using an online survey tool, with cost being among the most important. The three online survey services listed above provide some functionality through their free versions. However, in order to access all of the features and not be limited in survey length or the number of responses,

one would have to purchase their services. The following is a list of other features to consider when choosing an online survey tool.[10]

- Customization of Survey Appearance: The ability to choose different templates and color patterns and incorporate your own fonts, images, and even videos.
- Skip Logic: The mechanism by which survey respondents are given the option to skip certain sections of a survey based on their responses.
- Piping: Piping refers to the ability for survey software to incorporate an answer from one question into questions that follow later in the survey.
- Randomization: Some survey tools will allow for the random ordering of survey questions in order to decrease the bias on the part of respondents due to question order.
- Web site Integration: Certain survey tools will allow surveys and polls to be integrated directly into a user's website.
- Data Reports & Analysis: Most services can produce simple reports summarizing survey responses and will allow for the exportation of data into other programs for further analysis. More advanced (i.e. expensive) services will provide more rigorous tools for analyzing and interpreting data.

Regardless of which software you choose to use, no service can replace the need for properly crafted questions and a well-thought-out survey design. Without the right questions aimed at appropriate respondents, the most expensive and elaborate services will be of very little use.

4.3 Using Relational Database Software

A relational database management system (RDBMS), or simply a "relational database" is a collection of data tables that are related to a unique list (usually individuals). A single table within a relational database resembles a simple spreadsheet, such as those created in Microsoft Excel. What a relational database does that a spreadsheet (or "flat file") cannot, is link data from multiple tables so that users can understand relationships between the varying pieces of information that they have collected. When your data have complex relationships, a flat file may not be sufficient to easily add, delete, and summarize information from a data set. Instead, it is more desirable to use an RDBMS such as Microsoft Access or MySQL.

Creating a relational database to store your data can seem like a complex process, but it will ultimately save time and energy. Most relational database management systems make it easy to export your

data to other software packages designed for statistical analysis, such as STATA, SAS, or R. A well-designed RDBMS is part of the critical infrastructure necessary for efficient collection and analysis of information being gathered to improve public health programs and initiatives, such as vaccination campaigns.

Exhibit 4.1 Guide to Creating a Relational Database Management System Using Microsoft Access

A guide to creating a relational database management system using Microsoft Access is available on the SPIVA Toolkit Web site. The guide can assist users with editing MS Access files, or with creating their own databases from scratch. This guide was developed by Dr. Wayne Enanoria at the UC Berkeley Center for Infectious Diseases and Emergency Readiness.

4.4 Using CDC's Epi Info Software

Epi Info is an MS Windows-specific software package developed by the Centers for Disease Control and Prevention (CDC) and is available free of charge to public health professionals. This software allows users to develop questionnaires and forms, as well as enter and analyze data. It also allows for the production of tables, graphs, and maps. Epi Info is also compatible with Microsoft Access and MySQL relational database software and can serve as a data analysis tool for information housed in such a database.

More information about Epi Info can be found on the CDC's Web site at: http://www.cdc.gov/epiinfo/

4.5 Using EpiData Open Source Software

EpiData is a freely available, open source software program based on the old CDC Epi Info 6 (DOS version). EpiData is used to develop survey questionnaires for printing and rapid field deployment. The computer screen version of the questionnaire is automatically converted into a computer data entry form. Logic checks are easily added to improve accurate data entry. Finally, the accompanying EpiAnalysis package conducts basic analyses on the entered data. The EpiData module will simplify and operationalize the process from survey design, data collection, data entry, and analysis for non-epidemiologists.

More information about EpiData can be found at: http://www.epidata.dk/