Document Analysis:

Using the Written Record

Political scientists have three main methods of collecting the data they need to test hypotheses: interviewing, document analysis, and observation. Of these, interviewing and document analysis are the most frequently used. In Chapter 8 we discussed observation techniques; here we describe how empirical observations can be made using the **written record,** which is composed of documents, reports, statistics, manuscripts, and other written, oral, or visual materials.

Political scientists turn to the written record when the political phenomena that interest them cannot be measured through personal interviews, with questionnaires, or by direct observation. For example, interviewing and observation are of limited utility to researchers interested in large-scale collective behavior (such as civil unrest and the budget allocations of national governments), or in phenomena that are distant in time (Supreme Court decisions during the Civil War) or space (defense spending by different countries).

The political phenomena that have been observed through written records are many and varied—for example, judicial decisions concerning the free exercise of religion, voter turnout rates in gubernatorial elections, the change over time in Soviet military expenditures, and the incidence of political corruption in the People's Republic of China.[1] Of the examples of political science research described in Chapter 1 and referred to throughout this book, Steven C. Poe and C. Neal Tate's investigation of governments' violation of human rights, Jeff Yates and Andrew Whitford's investigation into Supreme Court justices' decisions in cases involving presidential powers, Jeffrey A. Segal and Albert D. Cover's investigation of the ideology of Supreme Court justices, and Kim Fridkin Kahn and Patrick J. Kenney's study of national elections all depended on written records for the measurement of important political concepts.[2] Not all portions of the written record are equally useful to political scientists. Hence we discuss the major components of the written record of interest to political scientists and how researchers use those components to measure significant political phenomena.

Generally speaking, use of the written record raises fewer ethical issues than either observation or interviewing. Research involving the collection or study of existing data, documents, or records often does not pose risks to individuals, because the unit of analysis for the data is not the individual. Also, issues of risk are not likely to arise where records are for individuals, as long as individuals cannot be identified directly or though identifiers linked to them, or where the records are publicly available, as in the case of the papers of public figures such as presidents and members of Congress. However, allowing researchers access to their private papers may pose some risk to private individuals. Thus access to private papers may be subject to conditions designed to protect the individuals involved.

Types of Written Records

Some written records are ongoing and cover an extensive period of time; others are more episodic. Some are produced by public organizations at taxpayers' expense; others are produced by business concerns or by private citizens. Some are carefully preserved and indexed; other records are written and forgotten. In this section we discuss two types of written records: the episodic record and the running record.

The Episodic Record

Records that are not part of an ongoing, systematic record-keeping program but are produced and preserved in a more casual, personal, and accidental manner are called **episodic records.** Good examples are personal diaries, memoirs, manuscripts, correspondence, and autobiographies; biographical sketches and other biographical materials; the temporary records of organizations; and media of temporary existence, such as brochures, posters, and pamphlets. The episodic record is of particular importance to political historians, since much of their subject matter can be studied only through these data.

The papers and memoirs of past presidents and members of Congress could also be classified as part of the episodic record, even though considerable resources and organizational effort are invested in their preservation, insofar as the content and methods of organization of these documents vary and the papers are not available all in the same location.

To use written records, researchers must first gain access to the materials and then code and analyze them. Gaining access to the episodic record is sometimes particularly difficult.[3] Locating suitable materials can easily be the most time-consuming aspect of the whole data collection exercise.

Researchers generally use episodic records to illustrate phenomena rather than as a basis for the generation of a large sample and numerical measures

for statistical analysis. Consequently, quotations and other excerpts from research materials are often used as evidence for a thesis or hypothesis. Over the years, social scientists have conducted some exceptionally interesting and imaginative studies of political phenomena based on the episodic record. We describe three particular studies that used the episodic record to illuminate an important political phenomenon.

DEVIANCE IN THE MASSACHUSETTS BAY COLONY. In the 1960s the sociologist Kai T. Erikson studied deviance in the Puritans' Massachusetts Bay Colony during the seventeenth century.[4] He was interested in the process by which communities decide what constitutes deviant behavior. In particular, he wished to test the idea that communities alter their definitions of deviance over time and use deviant behavior to reaffirm and establish the boundaries of acceptable behavior. Contrary to the conventional view that deviant behavior is uniformly harmful, Erikson believed that the identification of and reaction to deviant behavior serve a useful social purpose for a community.

Obviously no one is still alive who could be interviewed about the Puritan form of justice in the colony. Consequently Erikson had to search existing historical documents for evidence relating to his thesis. He found two main collections germane to his inquiry: *The Records of the Governor and Company of the Massachusetts Bay in New England* and *The Records and Files of the Quarterly Courts of Essex County, Massachusetts, 1636–1682.*[5] With these documents Erikson was able to weave together a fascinating tale of crime and punishment, Puritan style, during the mid-1600s.

Erikson's primary concern was with the identification of acts judged deviant in the Massachusetts Bay Colony. From the records of the Essex County courts, he was able to collect information on all 1,954 convictions reached between 1651 and 1680. These data allowed Erikson to investigate the frequency of criminal behavior and to calculate a crude crime rate for the Bay Colony during this period.

Erikson's analysis of the historical records was not altogether straightforward. For example, he discovered that the Puritans were extremely casual about how they spelled people's names. One man named Francis Usselton made many appearances before the Essex County Court, and his name was spelled at least fourteen different ways in the court's records. This did not present insurmountable difficulties in his case because his name was so distinctive. However, Erikson had a more difficult time deciding whether Edwin and Edward Batter were the same man and whether "the George Hampton who stole a chicken in 1649" was the same man as "the George Hampden who was found drunk in 1651."[6]

A second problem with the Puritans' record keeping was that they often passed the same name from generation to generation. Hence it was some-

times unclear whether two crimes twenty years apart were committed by the same person or by a father and a son. Between 1656 and 1681, for example, John Brown was convicted of seven offenses. However, since John Brown's father and grandfather were also named John Brown, it was unclear who committed which crimes.

Despite these difficulties, Erikson's research is a testimonial to the ability of historical records to address important contemporary issues. Without the foresight of those who preserved and printed these records, an important aspect of life in Puritan New England would have been measurably more difficult to piece together.

ECONOMICS AND THE U.S. CONSTITUTION. In 1913 the historian Charles Beard published a book about the U.S. Constitution in which he made imaginative use of the episodic record.[7] Beard's thesis was that economic interests prompted the movement to frame the Constitution. He reasoned that if he could show that the framers and pro-Constitution groups were familiar with the economic benefits that would ensue upon ratification of the Constitution, then he would be able to argue that economic considerations were central to the Constitution debate. If, in addition, he could show that the framers themselves benefited economically from the system of government established by the Constitution, the case would be that much stronger. This thesis, which has stimulated a good deal of controversy, was tested by Beard with a variety of data from the episodic record.

The first body of evidence presented by Beard measured the property holdings of those present at the 1787 Constitutional Convention. These measures, which Beard admits are distressingly incomplete, are derived largely from six different types of sources: biographical materials, such as James Herring's multivolume *National Portrait Gallery* and the *National Encyclopedia of Biography;* census materials, in particular the 1790 census of heads of families, which showed the number of slaves owned by some of the framers; U.S. Treasury records, including ledger books containing lists of securities; records of individual state loan offices; records concerning the histories of certain businesses, such as the *History of the Bank of North America* and the *History of the Insurance Company of North America*; and collections of personal papers stored in the Library of Congress.

From these written records Beard was able to discover the occupations, land holdings, number of slaves, securities, and mercantile interests of the framers. This allowed him to establish a plausible case that the framers were not economically disinterested when they met in Philadelphia to "revise" the Articles of Confederation.

Beard coupled his inventory of the framers' personal wealth with a second body of evidence concerning their political views. His objective was to

demonstrate that the framers realized and discussed the economic implications of the Constitution and the new system of government. By using the existing minutes of the debate at the convention, the personal correspondence and writings of some of the framers, and the *Federalist Papers*, by James Madison, Alexander Hamilton, and John Jay—which were written to persuade people to vote for the Constitution—Beard was able to demonstrate that the framers were concerned about, and cognizant of, the economic implications of the Constitution they wrote.

A third body of evidence allowed Beard to analyze the distribution of the vote for and against the Constitution. Where the data permitted, Beard measured the geographical distribution of the popular vote in favor of ratification and compared this with information about the economic interests of different geographical areas in each of the states. He also attempted to measure the personal wealth of those present at the state ratification conventions and then related those measures to the vote on the Constitution. These data were gleaned from the financial records of the individual states, the U.S. Treasury Department, and historical accounts of the ratification process in the states.

Through this painstaking and time-consuming reading of the historical record, Beard constructed a persuasive (although not necessarily proven) case for his conclusion that "the movement for the Constitution of the United States was originated and carried through principally by four groups of personal interests which had been adversely affected under the Articles of Confederation: money, public securities, manufactures, and trade and shipping."[8]

PRESIDENTIAL PERSONALITY. A third example of the use of the episodic record may be found in James David Barber's *The Presidential Character.* Because of the importance of the presidency in the American political system and the extent to which that institution is shaped by its sole occupant, Barber was interested in understanding the personalities of the individuals who had occupied the office during the twentieth century. Although he undoubtedly would have preferred to observe directly the behavior of the fourteen presidents who held office between 1908 and 1984 (when he conducted his study), he was forced instead to rely on the available written materials about them.

For Barber, discerning a president's personality means understanding his style, world view, and character. Style is "the President's habitual way of performing his three political roles: rhetoric, personal relations, and homework." A president's world view is measured by his "primary, politically relevant beliefs, particularly his conceptions of social causality, human nature, and the central moral conflicts of the time." And character "is the way the President orients himself toward life." Barber believes that a president's style, character, and world view "fit together in a dynamic package understandable in psychological terms" and that this personality "is an important shaper of his

Presidential behavior on nontrivial matters." But how is one to measure the style, character, and world view of presidents who are dead or who will not permit a political psychologist access to their thoughts and deeds? This is an especially troublesome question when one believes, as Barber does, that "the best way to predict a President's character, world view, and style is to see how they were put together in the first place . . . in his early life, culminating in his first independent political success." [9]

Barber's solution to this problem was to use available materials on the twentieth-century presidents he studied, including biographies, memoirs, diaries, speeches, and, for Richard Nixon, tape recordings of presidential conversations. Barber did not use all the available biographical materials. For example, he "steered clear of obvious puff jobs put out in campaigns and of the quickie exposés composed to destroy reputations." [10] He quoted frequently from the biographical materials as he built his case that a particular president was one of four basic personality types. Had these materials been unavailable or of questionable accuracy (a possibility that Barber glosses over in a single paragraph), measuring presidential personalities would have been a good deal more difficult, if not impossible.

Barber's analysis of the presidential personality was exclusively qualitative; the book contained not one table or graph. He used the biographical material to categorize each president as one of four personality types and to show that the presidents with similar personalities exhibited similar behavioral patterns when in office. In brief, Barber used two dimensions—activity-passivity (how much energy does the man invest in his presidency?) and positive-negative affect (how does he feel about what he does?)—to define the four types of presidential personality. (See Table 9-1.)

Barber's research is a provocative and imaginative example of the use of the episodic record—in this case, biographical material—as evidence for a series of generalizations about presidential personality. Although Barber did not empirically test his hypotheses in the ways that we have been discussing in this book, he did accumulate a body of evidence in support of his assertions and presented his evidence in such a way that the reader can evaluate how persuasive it is. [11]

The Running Record

Unlike the episodic record, the **running record** is more likely to be produced by organizations than by private citizens; it is carefully stored and easily accessed; and it is available for long periods of time. The portion of the running record that is concerned with political phenomena is extensive and growing. The data collection and reporting efforts of the U.S. government alone are impressive, and if you add to that the written records collected and preserved by state and local governments, interest groups, publishing

TABLE 9-1
Presidential Personality Types

Positive-Negative Affect	Activity-Passivity	
	Active	Passive
Positive	Franklin D. Roosevelt Harry S. Truman John F. Kennedy Gerald Ford Jimmy Carter	William Howard Taft Warren Harding Ronald Reagan
Negative	Woodrow Wilson Herbert Hoover Lyndon Johnson Richard Nixon	Calvin Coolidge Dwight Eisenhower

Source: Based on data from James David Barber, *The Presidential Character,* 3d ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1985). Courtesy of James David Barber, James B. Duke Professor of Science, Emeritus, Duke University, Durham, N.C.

houses, research institutes, and commercial concerns, the quantity of politically relevant written records increases quickly. Reports of the U.S. government, for example, now cover everything from electoral votes to electrical rates, taxes to taxi cabs, and, in summary form, fill one thousand pages in the *Statistical Abstract of the United States,* published annually by the U.S. Bureau of the Census. What makes the running record especially attractive as a resource is that many data sets are now housed online. The *Statistical Abstract,* for example, can be found at www.census.gov/compendia/statab.

There are far too many sources of the running record to list them here, but a quick look at some popular sources will help you understand the array of sources that are available. If you are interested in elections and campaigns you can visit the *Federal Election Commission* at www.fec.gov and find financial records filed by candidates, interest groups, and political parties, or you can visit privately operated Web sites, like www.opensecrets.org, that offer processed reports in an easy to read and use format. Or you might visit the Web sites administered by the secretaries of state to find state-level election returns or summaries of election law changes over time, or the *America Votes* series to find election results for national and some state and local elections. Alternatively, if you are interested in the lawmaking process, Congress makes the text and legislative histories of bills, committee reports, hearings, congressional votes, and the *Congressional Record* available in print or online at www.thomas.gov with a useful search engine to find needed documents. Or you can search for similar material through nongovernmental sources like

the Inter-university Consortium for Political and Social Research archive or in print in the *CQ Almanac* or in CQ Press's *Politics in America*. Finally, you can find a wealth of information related to foreign affairs in *World Resources*, published by the World Resources Institute in collaboration with the United Nations Environment Programme and the United Nations Development Programme, or in the Central Intelligence Agency's *World Fact Book* at https://www.cia.gov/library/publications/the-world-factbook/index.html. As you can imagine, the references listed here represent only a small fraction of the available records. Each reference has its own advantages and disadvantages, and you should take care to understand exactly what is and what is not included in each reference before using it.

THE POLICY AGENDAS PROJECT. In this section we provide a detailed example of how you can use sources of the running record in your own research projects by focusing on one such reference, the Policy Agendas Project. Available at www.policyagendas.org and maintained by Bryan Jones and John Wilkerson at the University of Washington and Frank Baumgartner at Pennsylvania State University, the Policy Agendas Project Web site offers many different sources of data linked together by public policy topics. The project seeks to provide users with an easy one-source way to track long-term policy changes at the national level of government across many different arenas. At the heart of this project is a comprehensive list of public policy topics (Table 9-2) and subtopics that includes a numeric code for virtually every public policy issue on the national agenda. As demonstrated in Figure 9-1, each topic is divided into dozens of subtopics to better organize the broad policy areas. Finally, each topic and subtopic is assigned a unique identification number that is used in each of the data sets available on the Web site.

The Web site is updated continually with new and more recent data. As of this writing, it included ten distinct data sets with links to several others. The currently available data sets are as follows: Budget, Congressional Quarterly Almanac, Congressional Hearings, Executive Orders, New York Times Index, Gallup's Most Important Problem, Public Laws, State of the Union Speeches, Supreme Court Cases, and Congressional Roll Call Votes. A partner site, *The Congressional Bills Project* (available at www.congressionalbills.org) is maintained by E. Scott Adler and John Wilkerson at the University of Washington and includes data on every congressional bill from 1947 to 2000. This Web

TABLE 9-2
Policy Agendas Project Policy Topics
Agriculture
Banking & Commerce
Civil Rights/Liberties
Defense
Education
Energy
Environment
Foreign Trade
Government Operations
Health
Housing & Community Development
International Affairs & Aid
Labor, Employment, & Immigration
Law, Crime, & Family
Macroeconomics
Public Lands
Science & Technology
Social Welfare
Transportation

273

FIGURE 9-1

Sample Codes for the Policy Agendas Project

4. Agriculture

400: General (includes combinations of multiple subtopics)

Examples: DOA, USDA and FDA appropriations, general farm bills, farm legislation issues, economic conditions in agriculture, impact of budget reductions on agriculture, importance of agriculture to the U.S. economy, national farmland protection policies, agriculture and rural development appropriations, family farmers, state of American agriculture, farm program administration, long range agricultural policies, amend the Agriculture and Food Act, National Agricultural Bargaining Board.

401: Agricultural Trade

Examples: FDA inspection of imports, agriculture export promotion efforts, agricultural trade promotion programs, tobacco import trends, agricultural export credit guarantee programs, impact of imported meats on domestic industries, country of origin produce labeling, USDA agricultural export initiatives, value added agricultural products in U.S. trade, establish coffee export quotas, effects of Mexican produce importation, international wheat agreements, livestock and poultry exports, amend Agricultural Trade Development and Assistance Act of 1954, reemphasize trade development, promote foreign trade in grapes and plums, prohibit unfair trade practices affecting producers of agricultural products, extend Agricultural Trade Development, enact the Agriculture Trade Act of 1978, establish agricultural aid and trade missions to assist foreign countries to participate in US agricultural aid and trade programs, Food, Agriculture, Conservation and Trade Act Amendments.

See also: 1800 general foreign trade; 1502 agricultural commodities trading.

402: Government Subsidies to Farmers and Ranchers, Agricultural Disaster Insurance

Examples: agricultural price support programs, USDA crop loss assistance, farm credit system financial viability, federal agriculture credit programs, agricultural disaster relief programs, subsidies for dairy producers, farm loan and credit issues, reforming federal crop insurance programs, credit assistance for family operated farms, federal milk supply and pricing policies, renegotiation of farm debts, USDA direct subsidy payments to producers, establishing farm program payment yields, peanut programs, wheat programs, evaluation of the supply and demand for various agricultural commodities, beef prices, cotton acreage allotments, shortages of agricultural storage facilities, agricultural subterminal storage facilities, financial problems of farm banks, Agricultural Adjustment Act, farm vehicle issues, Wool Act, Sugar Act, feed grain programs, cropland adjustment programs.

See also: 1404 farm real estate financing.

403: Food Inspection and Safety (including seafood)

Examples: FDA monitoring of animal drug residues, consumer seafood safety, budget requests for food safety programs, food labeling requirements, grain inspection services, regulation of health and nutrition claims in food advertising and labeling, sanitary requirements for food transportation, regulation of pesticide residues on fruit, food irradiation control act, regulation of artificial food coloring, federal control over the contamination of food supplies, meat grading standards, meat processing and handling requirements, improvement of railroad food storage facilities, shortage of grain storage facilities, food packaging standards, food buyer protection, regulation of food additives, federal seed act, definition and standards of dry milk solids.

See also: 401 inspection of food imports.

404: Agricultural Marketing, Research, and Promotion

Examples: soybean promotion, research, and consumer information act, USDA commodity promotion programs, cotton research and promotion, wheat marketing problems, livestock marketing, new peanut marketing system, establishing a national commission on food marketing, fruit and vegetable marketing, industrial uses for agricultural products, meat promotion program, national turkey marketing act, federal marketing quotas for wheat.

405: Animal and Crop Disease and Pest Control

Examples: USDA regulation of plant and animal mailing to prevent the spread of diseases, control of animal and plant pests, pork industry swine disease eradication program, virus protection for sheep, grasshopper and cricket control programs on farmland, USDA response to the outbreak of citrus disease in Florida, eradication of livestock diseases, brucellosis outbreak in cattle, USDA integrated pest management program, toxic contamination of livestock, fire ant eradication program, proposed citrus blackfly quarantine, predator control problems, biological controls for insects and diseases on agricultural crops, eradication of farm animal foot and mouth diseases.

See also: 704 for pollution effects of pesticides; 403 for pesticide residues on foods.

498: Agricultural Research and Development

Examples: condition of federally funded agricultural research facilities, USDA nutrition research activities, USDA agricultural research programs, regulation of research in agricultural biotechnology programs, organic farming research, potential uses of genetic engineering in agriculture, agricultural research services, research on aquaculture.

499: Other

Examples: methodologies used in a nationwide food consumption survey, agricultural weather information services, federal agricultural census, designate a national grain board, home gardening, redefinition of the term "farm", farm cooperative issues.

site uses the same policy codes used on the Policy Agendas Project Web site for easy combination of data between the two sites.

Each of these data sets is a useful source of data in its own right, but the policy codes linking these data make this Web site especially important. In the paragraphs that follow we briefly discuss the contents of the data sets. Each data set is available in Microsoft Excel or text format along with a codebook. You can open the files using Excel or a program that will be able to read the text format data and transform them into a spreadsheet or table. The codebooks are important because they explain what the data include, how the data are formatted, and any special instructions that must be followed in using the data. If you wish to use the public policy codes, you will also need the public policy codebook, which provides the identification numbers for each policy topic and subtopic.

The Budget data set contains data on the budget of the U.S. government from financial years 1947 through 2006. These data allow you to track government spending across time and to compare spending in different policy areas.

The Congressional Quarterly Almanac data set includes data on each article appearing in the main chapters of the *CQ Almanac* from 1948 through 2003. These data include information about the policy topics, congressional committees and members, bills, and laws mentioned in each article. This data set is linked through identification numbers with the public laws data set maintained by the Policy Agendas Project. Since the *CQ Almanac* follows congressional action quite closely, the data set provides an accessible and easy way to track congressional action in different policy areas.

The Congressional Hearings data set includes data on congressional hearings from 1946 to 2005. The data set includes information about the committee or subcommittee holding the hearing and the policy topics discussed. These data will be useful to students who are, for example, interested in determining how active congressional committees are in responding to policy failures with hearings to determine what went wrong and how they might fix problems.

The Executive Orders data set includes information about every executive order issued between 1945 and 2003. This file includes data on the policy content of executive orders as well as characteristics such as the party of the president and whether the order came during a period of unified or divided government.

The New York Times Index data set includes a systematic random sample of *New York Times* articles from 1946 to 2003 coded for policy content. The data can be used to measure media attention to a policy issue over time or at specific points in time or to determine which issues were most important in a given time period.

The Gallup's Most Important Problem data set includes aggregated responses to the Gallup opinion poll's "most important problem" question from 1947 to 2004. The data allow students to measure the policy issue that was most important to the public during a specific time frame and include a relative ranking of each issue mentioned by respondents. Used in combination with other data available through the Policy Agendas Project, these data might be useful in determining how Congress or the president react to public opinion on the most important issues on the national agenda.

The Public Laws data set contains information about the laws passed between 1948 and 2004. Each law is coded for policy topic and includes an identification variable that links the law to the Congressional Quarterly Almanac data set. The Web site includes a subset of this data set, called the Most Important Law data set, which identifies the 576 most important laws passed between 1948 and 2004, based on the number of lines CQ dedicated to coverage of the laws. These data offer the opportunity to link policy demands with governmental outputs based on policy topic.

The State of the Union Speeches data set contains information on each State of the Union address delivered from 1946 to 2005. The speeches are coded on the statement level for policy content and other variables. These data offer a way to determine the issues the president views as important, based on the amount of attention given to the issues within a State of the Union speech.

The Supreme Court data set includes data on every Supreme Court case from 1953 until 1998. The data set includes measures of important characteristics such as the decision to accept or reject a case and a summary of the decision if the case was accepted for review. Used in combination with the Public Laws, Executive Orders, or Congressional Hearing data sets, these data could be used to explore inter-branch relations on public policy issues.

The Congressional Roll Call Votes data set codes every congressional roll-call vote from 1946 to 2000. Combining these data with the Gallup's Most Important Question data set would allow students to estimate the effect of public opinion on congressional voting.

As indicated earlier, combining these data sets will allow you to answer a variety of research questions about the development and passage of public policy. The Policy Agendas Project Web site offers a single source for many kinds of policy-related data, and the unique policy identification numbers allow for a seamless combination of the data.

The Running Record and Episodic Record Compared

There are three primary advantages to using the running record rather than the episodic record. The first is cost, in both time and money. Since the costs of collecting, tabulating, storing, and reporting the data in the running record are generally borne by the record keepers themselves, political scientists are

usually able to use these data inexpensively. Researchers can often use the data stored in the running record by photocopying a few pages of a reference book, purchasing a government report or data file, or downloading data into a spreadsheet. In fact, the continued expansion of the data collection and record-keeping activities of the national government has been a financial boon to social scientists of all types.

A second, related advantage is the accessibility of the running record. Instead of searching packing crates, deteriorated ledgers, and musty storerooms, as users of the episodic record often do, users of the running record more often handle reference books, government publications, and computer printouts. Many political science research projects have been completed with only the data stored in the reference books and government documents of a decent research library.

A third advantage of the running record is that by definition it covers a more extensive period of time than does the episodic record. This permits the type of longitudinal analysis and before-and-after research designs discussed in Chapter 5. Although the episodic record helps explain the origins of and reasons for a particular event, episode, or period, the running record allows the measurement of political phenomena over time.

The running record presents problems, however. One is that a researcher is at the mercy of the data collection practices and procedures of the record-keeping organizations themselves. Researchers are rarely in a position to influence record-keeping practices; they must rely instead on what organizations such as the U.S. Census Bureau, Federal Election Commission, and the Policy Agendas Project decide to do. A trade-off often exists between ease of access and researcher influence over the measurements that are made. Some organizations—some state and local governments, for example—do not maintain records as consistently as researchers may like. One colleague found tracing the fate of proposed constitutional amendments to the Delaware State Constitution to be a difficult task. Delaware is the only state in which voters do not ratify constitutional amendments. Instead, the state legislature must pass an amendment in two consecutive legislative sessions in between which a legislative election has occurred. Thus constitutional amendments are treated like bills and tracking them depends on the archival practices of the state legislature. Even when clear records are kept, such as election returns for mayoral contests, researchers may face a substantial task in collecting the data from individual cities, because the returns from only the largest cities are reported in various statistical compilations.

Another related disadvantage of the running record is that some organizations are not willing to share their raw data with researchers. The processed data that they do release may reflect calculations, categorizations, and aggregations that are inaccurate or uninformative. Access to public information is

not *always* easy. More problems may be encountered when trying to obtain public information that shares some of the characteristics of the episodic record, for example, such as information on the effect of specific public programs and agency activities. Emily Van Dunk, a senior researcher at the Public Policy Forum, a nonpartisan, nonprofit research organization that conducts research on issues of importance to Wisconsin residents, notes that obtaining data from state and local government agencies can be difficult at times and offers tips for researchers.[12]

Finally, it is sometimes difficult for researchers to find out exactly what some organizations' record-keeping practices are. Unless the organization publishes a description of its procedures, a researcher may not know what decisions have guided the record-keeping process. This can be a special problem when these practices change, altering in an unknown way the measurements reported.

Although the running record has its disadvantages, political scientists often must rely on it if they wish to do any empirical research on a particular topic. To illustrate some of the problems with using written records, we conclude this section with a description of PollingReport.com, one of many Web sites dedicated to providing users with national- and state-level public opinion data.

Presidential Job Approval

PollingReport.com is a popular source of public opinion polling data, as evidenced by Time.com's inclusion of the site on its list of the 50 best Web sites in 2007.[13] PollingReport.com provides national poll results, free of charge, from well-known polling organizations such as Gallup, Pew, and Quinnipiac and news organizations such as CNN, CBS, and the *Los Angeles Times*. The Web site also offers state-level poll results to paid subscribers. In this section we focus on the data available for free.

PollingReport.com organizes its poll results into categories for Elections, the State of the Union, National Security, In the News, and Issues. As shown in Figure 9-2, each of these categories offers a number of subtopics of interest. The State of the Union category, for example, includes subtopics covering each branch of the federal government, President Bush, Congress, and the Supreme Court, as well as the "Direction of the country," "National priorities," and "Consumer confidence." Likewise, the Issues category is divided into several subtopics. A great deal of useful public opinion data may be found among these many subtopics.

Let's assume that you are interested in studying public support for the president. The obvious place to start would be by finding data on President Bush's job approval ratings. By clicking on the President Bush subtopic under State of the Union, you will find several different kinds of presidential

FIGURE 9-2

Categories of Poll Results on PollingReport.com

PollingReport.com ™

An independent, nonpartisan resource on trends in American public opinion.

Subscription Information

One of TIME.com's 50 Best Websites
● PollingReport.com ●

ADVERTISE HERE

XML RSS feed

SITE DIRECTORY
POLITICS & POLICY
BUSINESS/ECONOMY
AMERICAN SCENE
SEARCH PollingReport
TABLE OF CONTENTS
INFO ON SUBSCRIBING
ARTICLE ARCHIVES
RANDOM SAMPLES
JOB BANK 9/11

PollingReport.com is updated when new polls are released. Click a subject area from the site directory above, use the CONTENTS page or site SEARCH engine, or access featured topics via the links on the right.
Data are from national surveys of the American public. State polls and articles by leading pollsters are available to subscribers. INFO ON SUBSCRIBING

Editor
@PollingReport.com

DIRECT RELIEF INT'L

SUBSCRIBER PAGES

[Subscriber log-in] 9/12

Elections
2008: President 9/12
2008: Congress
State of the Union
President Bush <job trend>
Congress <job trend>
Supreme Court
Direction of the country
National priorities
Consumer confidence 9/12
National Security
Iran
Iraq 9/11
Terrorism 9/11
In the News
General David Petraeus
Senator Craig
Oprah/Obama
Candidate traits
Grading the schools
School safety
Union influence
Products from China
After Katrina
Sports scandals
Favorite sport
Hillary Clinton
John Edwards
Rudy Giuliani
John McCain
Barack Obama
Mitt Romney
Fred Thompson
Airline passengers
Sam Brownback
Al Gore
Mike Huckabee
Karl Rove
Schools and race
Bridges
Vice President Cheney
Rating the parties
Ds vs. Rs on the issues
Surveillance cameras
Speaker Pelosi
Majority Leader Reid
Immigration reform
Issues
Abortion 9/11
Energy
Environment
Evolution
Gun laws
Health policy
Immigration
Minimum wage
Same-sex marriage
Stem cell research
Taxes
More issues . . .

Details on accessing STATE polls

CNN/OPINION RESEARCH POLL
Presidential Preference

Giuliani 28%
Thompson 27%
McCain 15%
Romney 11%
Huckabee 5%
Other 6%
Unsure 8%

Asked of Republicans/leaners

Clinton 46%
Obama 23%
Edwards 16%
Richardson 5%
Kucinich 3%
Other 3%
Unsure 3%

Asked of Democrats/leaners

9/7-9/07   pollingreport.com

CBS NEWS/NEW YORK TIMES POLL
"In general, do you think the United States is adequately prepared to deal with another terrorist attack, or not?"

■ Prepared ■ Not prepared

64%
29%
49%
44%
39%
56%

3/03   3/06   9/07

9/4-9/07   pollingreport.com

More graphics:
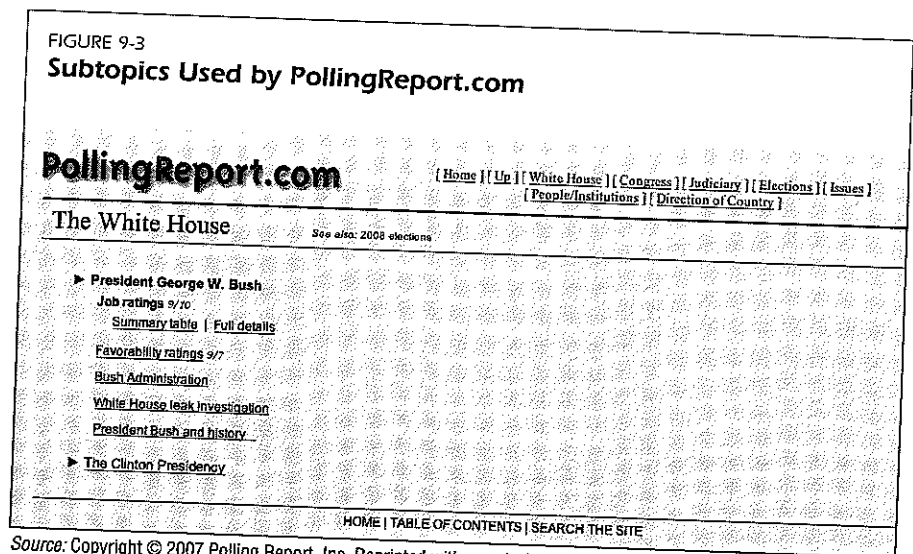• National Barometer • Poll Gallery

Random Samples 9/12
Polls, truth sometimes at odds • Clinton widens lead, Republicans split • Most see Iraq war as failure • Giuliani tumbles, GOP race tightens • Survey of Iraq finds worsening attitudes • Americans feel military best at ending the war • Wisdom of crowds • Brain study finds political divide • Majority wants troops out of Iraq within a year • Religion in campaign '08 • MORE ...

Copyright © 2007 POLLING REPORT, INC.

Source: Reprinted with permission of Polling Report, Inc.

support data (Figure 9-3). PollingReport.com provides data on "President Bush's Job ratings," "Favorability ratings," the "Bush Administration," the "White House leak investigation," "President Bush and history," as well as similar data for the Clinton presidency. Since you are interested in President Bush, not his administration, the "Job ratings" and "Favorability ratings" might be useful. We will explore President Bush's job ratings poll results in

FIGURE 9-3

Subtopics Used by PollingReport.com

PollingReport.com   [Home] [Up] [White House] [Congress] [Judiciary] [Elections] [Issues]
[People/Institutions] [Direction of Country]

The White House   *See also: 2008 elections*

▶ **President George W. Bush**
    Job ratings 9/10
        Summary table | Full details
    Favorability ratings 9/7
    Bush Administration
    White House leak investigation
    President Bush and history
▶ The Clinton Presidency

HOME | TABLE OF CONTENTS | SEARCH THE SITE

Source: Copyright © 2007 Polling Report, Inc. Reprinted with permission.

this example. You can click on "Summary table" to find poll results from various polling organizations that answer a question similar to Gallup's job approval rating question: "Do you approve or disapprove of the way George W. Bush is handling his job as president?" (Figure 9-4).[14] As you can see, the results span a two-year period from September 2005 to September 2007 on an irregular basis as poll data were available. Clicking on "Complete trend" at the top of the page takes you to a page with a more complete set of poll results from seven select organizations from the beginning of the Bush administration in 2001 until the present.

PollingReport.com has many advantages for students using the written record. First, and perhaps most important, PollingReport.com offers free, high-quality data with an easy-to-use Web site. The results found at Polling Report.com are from the same professional polling organizations that news organizations around the country rely on. Second, students have access to multiple surveys administered during different time periods using very similar question wording. But as valuable as PollingReport.com is, it shares some disadvantages with other examples of the running record. Perhaps most glaring is the lack of consistency and regularity in the poll results provided on the Web page. This is not an indictment of PollingReport.com but a reality of the kind of information the Web site provides. PollingReport.com can report only the data made available by other polling organizations. Although those other organizations provide a great deal of data, sometimes a large number of surveys are administered at the same time whereas no data are available for other time periods. And, although a great many organizations are listed on the Web site, it does not include all polling organizations.

FIGURE 9-4
Survey Data from PollingReport.com

PollingReport.com   [Home] [Up] [Bush: Job Ratings] [Bush: Favorability] [Bush Administration]
[Leak Probe] [Clinton] [Presidents & History]
▼Advertise on PollingReport.com▼

PollingReport.com

PRESIDENT BUSH – Overall Job Rating in recent national polls

See also: Complete trend

	Survey Dates	Approve %	Disapprove %	Unsure %	Approve minus Disapprove
USA Today/Gallup	8/7-8/07	33	62	5	-29
CBS/New York Times	9/4-8/07	30	64	6	-34
ABC/Washington Post	9/4-7/07	33	64	3	-31
FOX/Opinion Dynamics RV	8/21-22/07	33	56	11	-23
Pew	8/1-18/07	31	59	10	-28
Gallup	8/13-16/07	32	63	5	-31
Quinnipiac RV	8/7-13/07	29	64	7	-35
CBS	8/8-12/07	29	65	6	-36
AP-Ipsos	8/6-8/07	35	62	•	-27
CNN/Opinion Research Corp.	8/6-8/07	36	61	3	-25
USA Today/Gallup	8/3-5/07	34	62	4	-28
Newsweek	8/1-2/07	29	63	8	-34
NBC/Wall Street Journal	7/27-30/07	31	63	6	-32
Pew	7/25-29/07	28	61	10	-32
CBS/New York Times	7/20-22/07	30	62	8	-32
Diageo/Hotline RV	7/19-22/07	33	63	4	-30
ABC/Washington Post	7/18-21/07	33	65	2	-32
FOX/Opinion Dynamics RV	7/17-18/07	32	61	7	-29
CBS/New York Times	7/9-17/07	29	64	7	-35
Gallup	7/12-15/07	31	63	6	-32
Newsweek	7/11-12/07	29	64	7	-35
AP-Ipsos	7/9-11/07	33	65	•	-32
USA Today/Gallup	7/6-8/07	29	66	5	-37
Newsweek	7/2-3/07	26	65	9	-39
CBS	6/26-28/07	27	65	8	-38
FOX/Opinion Dynamics RV	6/26-27/07	31	60	9	-29
CNN/Opinion Research Corp.	6/22-24/07	32	56	3	-34
Newsweek	6/18-19/07	26	65	9	-39
Gallup	6/11-14/07	32	65	3	-33
NBC/Wall Street Journal	6/8-11/07	29	66	5	-37
Quinnipiac RV	6/5-11/07	28	65	7	-37
L.A. Times/Bloomberg	6/7-10/07	34	62	4	-28
FOX/Opinion Dynamics RV	6/5-6/07	34	57	9	-23
AP-Ipsos	6/4-6/07	32	66	•	-34
USA Today/Gallup	6/1-3/07	32	62	6	-30
Pew	5/30 - 6/3/07	29	61	10	-32
ABC/Washington Post	5/29 - 6/1/07	35	62	3	-27
CBS/New York Times	5/18-23/07	30	63	7	-33
Diageo/Hotline RV	5/16-20/07	32	64	4	-32
FOX/Opinion Dynamics RV	5/15-16/07	34	56	10	-22
Gallup	5/10-13/07	33	62	5	-29
AP-Ipsos	5/7-9/07	35	61	•	-26
CNN/Opinion Research Corp.	5/4-6/07	38	61	1	-23
USA Today/Gallup	5/4-6/07	34	63	3	-29
Newsweek	5/2-3/07	28	64	8	-36
WNBC/Marist RV	4/26 - 5/1/07	33	61	6	-28
Quinnipiac RV	4/25 - 5/1/07	35	60	6	-25
Diageo/Hotline RV	4/26-30/07	35	62	3	-27
NPR LV	4/26-29/07	37	59	5	-22
CBS/New York Times	4/20-24/07	32	61	7	-29

Second, even though the president's job approval rating question is one of the most frequently asked questions in national political surveys, a single organization is not likely to have a survey in the field every week of the year, so there will be gaps in the results. Relying on multiple surveys from different organizations can introduce bias to the conclusions because the results of each survey are influenced by question wording—and all organizations use slightly different wording to ask the same question about presidential job approval. Third, although PollingReport.com provides the results from the most basic measure of presidential approval, representing the whole population, the results for sometimes more interesting questions, like presidential support divided into subgroups by party identification, race, or gender, are not consistently available on the Web site.

Finally, PollingReport.com provides poll results from the Bush and Clinton administrations, but results from previous administrations are not available. If you wish to compare President Bush's approval ratings with those of previous presidents, you will have to search for those results elsewhere. These are only some of the potential problems that might be encountered with PollingReport.com or other examples of the running record. Problems like these are generally not going to prevent you from using such sources, but they can be a nuisance depending on the purpose of your research project.

Content Analysis

Sometimes researchers extract excerpts, quotations, or examples from the written record to support an observation or relationship. Those who rely on the episodic record, such as Charles Beard and James David Barber, often use the written record in this way. Alternatively, a researcher might measure the number of times the president references the economy in his State of the Union address. This use of the written record is an example of **content analysis.** We can think of document analysis, much like other forms of analysis, as taking both a qualitative and a quantitative form. In the earlier examples, we can label Beard and Barber's use of the written record as qualitative because they are using the words of others, or interpreting written documents without the use of a numeric coding scheme, to provide evidence for their arguments. Alternatively, we can label the use of State of the Union addresses as quantitative because quantitative content analysis enables us to "take a verbal, nonquantitative document and transform it into quantitative data." A researcher "first constructs a set of mutually exclusive and exhaustive categories that can be used to analyze documents, and then records the frequency with which each of these categories is observed in the documents studied."[15] This is exactly what Segal and Cover had to do with newspaper editorials to produce a quantitative measure of the political ideologies of Supreme Court justices.[16] In this section we focus primarily on quantitative content analysis for use in statistical analyses, but remember that a qualitative approach to documents can be just as useful, if not more so, depending on the purpose of a research project.

Content Analysis Procedures

The first step in content analysis is deciding what sample of materials to include in the analysis. If a researcher is interested in the political values of candidates for public office, a sample of political party platforms and campaign speeches might be suitable. If the level of sexism in a society is of interest, then a sample of television entertainment programs and films might be drawn. Or if a researcher is interested in what liberals are thinking about, liberal

opinion magazines might be sampled. Actually, two tasks are involved at this stage: selecting materials germane to the researcher's subject (in other words, choosing the appropriate sampling frame) and sampling the actual material to be analyzed from that sampling frame. Once the appropriate sampling frame has been selected, then any of the possible types of samples described in Chapter 10—random, systematic, stratified, cluster, and nonprobability—could be used.

The second task in any content analysis is to define the categories of content that are going to be measured. A study of the prevalence of crime in the news, for example, might measure the amount of news content that either deals with crime or does not. Content that deals with crime might be further subdivided into the kinds of crime covered. A study of news coverage of a presidential campaign might measure whether news content concerning a particular presidential candidate is favorable, neutral, or unfavorable. Or a study might measure the personality traits of various prime-time television characters—such as strength, warmth, integrity, humility, and wisdom—and the sex, age, race, and occupation of those characters. This process is in many respects the most important part of any content analysis because the researcher must measure the content in such a way that it relates to the research topic, and he or she must define this content so that the measures of it are both valid and reliable.

The third task is to choose the recording unit. For example, from a given document, news source, or other material, the researcher may want to code (1) each word, (2) each character or actor, (3) each sentence, (4) each paragraph, or (5) each item in its entirety. To measure concern with crime in the daily newspaper, the recording unit might be the article. To measure the favorableness of news coverage of presidential campaigns in news weeklies, the recording unit might be the paragraph. And to measure the amount of attention focused on different government institutions on television network news, the recording unit might be the story.

In choosing the recording unit, the researcher usually considers the correspondence between the unit and the content categories (stories may be more appropriate than words in determining whether crime is a topic of concern, whereas individual words or sentences rather than larger units may be more appropriate for measuring the traits of political candidates). Generally, if the recording unit is too small, each case will be unlikely to possess any of the content categories. If the recording unit is too large, however, it will be difficult to measure the single category of a content variable that it possesses (in other words, the case will possess multiple values of a given content variable). The selection of the appropriate recording unit is often a matter of trial and error, adjustment, and compromise in the pursuit of measures that capture the content of the material being coded.

Finally, a researcher has to devise a system of enumeration for the content being coded. The presence or absence of a given content category can be measured or the "frequency with which the category appears," or the "amount of space allotted to the category," or the "strength or intensity with which the category is represented."[17] For example, suppose we were coding the presence of Hispanics in televised entertainment programming, with the program as the recording unit. For each program we could count (1) whether there was at least one Hispanic present, (2) how many Hispanics there were, (3) how much time Hispanics were on the screen, and (4) how favorable or how important the portrayal of Hispanics was for the overall story.

The validity of a content analysis can usually be enhanced with a precise explanation of the procedures followed and content categories used. Usually the best way to demonstrate the reliability of content analysis measures is to show intercoder reliability. **Intercoder reliability** simply means that two or more analysts, using the same procedures and definitions, agree on the content categories applied to the material analyzed. The more the agreement, the stronger the researcher's confidence can be that the meaning of the content is not heavily dependent on the particular person doing the analysis. If different coders disagree frequently, then the content categories have not been defined with enough clarity and precision.

The following example of a content analysis may be helpful. Suppose you were interested in studying coverage of the 2004 presidential campaign by *Time* and *Newsweek* (the sampling frame). You could decide to analyze every article about the campaign from September 1 to election day, taking the article as the recording unit. The content categories could be (1) the subject of the article (that is, the "who"); (2) the topic of the article (that is, the "what"); and (3) the tone of the article (was it unfavorable or favorable?). To encode the content you could devise a coding sheet like the one presented in Table 9-3. It shows the content variables, the categories for each variable, the recording unit, and the system of enumeration. This is the type of sheet that would be used to quantify the data.

Among the drawbacks of content analysis are the time involved and the need to avoid mistakes when analyzing a large collection of written records. Suppose, for example, you wanted to see how the use of political symbols had changed over the last fifty years. You might take a sample of presidential addresses or campaign speeches and simply count the number of occurrences of certain phrases, such as "it is not the role [job, responsibility, etc.] of government to. . . ." You might then calculate and plot the proportion of such ideas over time. Doing so, though, requires that you—or, preferably, a coder or coders—read the material, look for phrases that meet the selection criteria, and make tallies. This is a time-consuming process; if a coder becomes fatigued, he or she might overlook instances that should be recorded or count phrases twice or make other mistakes.

TABLE 9-3
Coding Sheet for Hypothetical Content Analysis of Presidential Campaign Coverage

Magazine   1. *Time* _____   2. *Newsweek* _____

Date _____

Page no. _____

No. of paragraphs _____

No. of paragraphs devoted   Democrat   Republican

to each candidate: 1. Pres. _____   3. Pres. _____

                2. V.P. _____   4. V.P. _____

Primary focus
of article:

     1. Candidate prospects _____   3. Policy issues _____

     2. Campaign events _____   4. Personalities _____

Overall tone
of article

Republican Ticket:	Negative	1	2	3	4	5	6	7	Positive
Democratic Ticket	Negative	1	2	3	4	5	6	7	Positive

Today many social scientists are conducting content analyses with the help of computer programs. This software can read and store text and search for various patterns of words or even look for ideas implied by the text. Many of the programs have become quite sophisticated. Besides doing the actual content analysis, they write reports and calculate summary statistics. Because there are now so many of these programs, we cannot explain their operation here. But a good source for further information is *Text Analysis Resources,* compiled by Harald Klein and available at www.textanalysis.info.

Although political scientists have used content analysis relatively sparingly, it is a useful technique in some areas of inquiry. Content analysis may be used to analyze the content of a large number of lengthy, semi-structured interviews after they are transcribed.

News Coverage of Presidential Campaigns

A frequent subject of content analysis is press coverage of election campaigns. Given the importance of how candidates are presented and how the electoral process is treated in the news, political scientists have been interested for some time in accurately and systematically describing and explaining campaign news coverage. Most of these studies have investigated whether candidates receive favorable or unfavorable coverage; whether news coverage relays useful information to the American electorate; whether the press accurately presents the complex and lengthy presidential nomination process; and whether journalists are, in general, objective, accurate, fair, and informative.