4 Collecting and Organizing Existing Data

Collecting and organizing data is a key part of a watershed assessment. It makes sense to begin by collecting existing data. If you identify data gaps and if you have the resources, you might decide to collect new data as well. This chapter addresses the issues of how to collect and manage all types of data. Specific approaches for collecting new data in association with a watershed assessment will be discussed in detail in Volume II of the Manual.

Chapter Outline

4.1 Overview and Key Considerations

Data are "known facts or things used as a basis for inference or reckoning." Data pertinent to watershed assessment may include quantitative measurements, qualitative information (e.g., observations of a species' presence), maps, anecdotal information, photographs, and other "facts" relevant to the watershed assessment.

Collecting data can be both time consuming and labor intensive. Therefore, you should focus your efforts on the specific questions you've identified for your watershed. The extent of your data collection will hinge on how complex and detailed your assessment will be. There should be a direct path from the questions you are asking about the watershed to the types and amounts of data you collect within the project timeframe. Gaps in knowledge—questions that cannot be answered within the project timeframe and currently available budget—should be identified to indicate the nature of any remaining uncertainty and to prioritize data needed for future assessment.

4.1.1 Types of Data

Data types can be broadly classified as numeric and spatial. Numeric data can be highly quantitative, semi-quantitative, or qualitative. Examples of quantitative data include most water quality and flow data. Semi-quantitative data might include data that involves scoring habitat characteristics based on standards that might be interpreted differently by different people. For example, characterizing stream bank stability might involve an estimate of the percentage of the bank covered with vegetation. Frequently, this is estimated visually, not by actual measurements. Biological surveys often fall into this category. Qualitative data is descriptive. For example, anecdotal data describes conditions based on observation of a single or a few individuals. The observation that the pools in the stream have filled up with sediment over the past 20 years is one example. This is very useful information, even though it is not quantitative. Spatial data is data that has a geographic reference point. It can be quantitative, semi-quantitative, or descriptive data that is located at a point or area on a map.

As data are collected, they should be organized in a manner that suits the questions being asked and the users' needs. Because watershed assessment usually involves the collection of several different types of data, consider developing file organizational systems for each type of data that conform to a single standard for categories (e.g., wildlife habitat, water quality, land use). One way to keep track of information collected is to make a database of the category types. If you will be collecting data for aquatic and terrestrial systems and of various different types (i.e., from text to digital spatial data), then keeping track of the types of data and the areas they cover will help in both organizing the data and describing how much of the watershed they cover.

4.1.2. Evaluation of data

Once you find a few reports or data sets for your watershed, look them over with a critical eye. When examining past reports, ask yourself if the conclusions make sense and whether the supporting evidence justifies them. Check to see whether there was any sort of external or peer review. It is important that collection of data include a critical analysis of the methods used to obtain the information. Sample design, frequency of sampling, exact type of measurement made, and analytical technique, can all affect the usefulness of data you find and want to use. Not infrequently, streamflow gauges are not serviced regularly and therefore, produced unreliable data on flow rate. Verify their accuracy with knowledgeable people. Compare findings from different reports to see if they corroborate each other. There may be a legitimate reason why two reports may seem to be in conflict, based on different methods, objectives, assumptions, dates, biases, etc.

Consider establishing a record/database of information attributes that describe its usefulness (e.g., date of collection, source, purpose, scale, peer-reviewed, etc.). An example of this is in the North Coast Watershed Assessment Program Mattole watershed assessment report (http://www.ncwatershed.ca.gov, select Mattole watershed and the table is in the Appendix) and an excerpt is shown in Table 4.1.

When citing information for your assessment, check the original document for its methods and conclusions. Avoid contributing to the all-too-common rumor chain of many environmental documents that sequentially misrepresent an original reference by citing a secondary source that misinterpreted the original. Do not ignore or

Table 4.1 An example of archiving data (Downie et al., 2002)

Mattole River Watershed, NCWAP DATA CATALOG

Name	Source	Description	Data Quality	Metadata	Analytical Use in NCWAP
Stream habitat surveys	DFG	Numerous stream habitat surveys of major Mattole River tributaries conducted during 1938 and 1985.	Not digital. Quality unknown	No	Comparison with current in-stream habitat surveys.
Stream fish inventories	DFG	Stream fish inventory data for Mattole river, north fork Mattole river, squaw, mill, Thompson, baker, and bridge creeks.	Not digital. Quality unknown	No	Comparison with current in-stream habitat surveys.
Stream flow data	DFG	Stream flow data for the Mattole river, 1976.	Not digital. Quality unknown	No	Comparison with current in-stream habitat surveys.
Historic photos	Froland	Historic photographs of the Honeydew slide. 1983	Not digital. Quality unknown	No	Historic geological information
Erosion data	MSG	Aerial photo interpretation of erosion in the Mattole river basin	Not digital. Quality unknown	No	Historic geological information
In-stream habitat data	DFG	In-stream habitat data for yew, barnum and dream stream creeks.	Not digital. Quality unknown	No	Comparison with current in-stream habitat conditions
Fisheries data	DWR	Unspecified fisheries data, north coast basins. 1962.	Not digital. Quality unknown	No	Comparison with other fish population data.
Erosion and sedimentation data	MRC	Sediment source and erosion data for the Mattole watershed. 1989.	n/a	n/a	Historic geological information

disregard data just because they don't seem to "fit" or because they run counter to preconceived expectations of what the data should say. During the Data Analysis process, the various data sources can be evaluated for quality and deficient sources can be discarded.

Data obtained over the Internet, a common source of information, has particular limitations. Until the mid-1990s, most agencies and organizations did not put their

Sources for Watershed Assessment Information

Previous watershed assessments

Local Government

County or city
- General plan
- Master environmental assessments
- Specific plans and environmental impact reports for projects
- Local enhancement or restoration plans
- Departmental documents and data (planning, public works, agriculture, environmental health)
- Flood control plans, documents, and data

Water supply, irrigation, and other special districts—reports and data

Local Organizations

Resource Conservation Districts (RCDs) —annual reports and documents
Watershed groups—documents
Conservation groups (e.g., Friends of the XXX Creek/River) —reports

State of California Agencies

Regional Water Quality Control Boards
- Basin plans
- Monitoring studies
- Total maximum daily load (TMDL) studies and determinations

State Water Resources Control Board
- Water quality reports and data
- Water rights reports and data

California Department of Forestry and Fire Protection—timber harvest plans (THPs), maps, data, and documents
State Park Management Plans
California Department of Transportation
- Highway environmental impact reports
- Monitoring data on road runoff

California Department of Fish and Game—documents and data, land management plans
California Geological Survey (Division of Mines & Geology) —documents and data
California Department of Water Resources—documents and data
California Coastal Commission and Coastal Conservancy—documents and data
California State Lands Department—documents and data

data and reports online. Only a small number of older reports have been scanned in, such as for special digital libraries (e.g., Shasta College's WIM project, http://wim.shastacollege.edu/default.aspx; U.C. Berkeley's Digital Library Project http://elib.cs.berkeley.edu/). The University of California's MELVYL online catalog does not go back earlier than 1971 for the documents it lists, although its vast collection of historic documents and maps for California goes back to Spanish land grant days. You should consider seeking out data by visiting and looking through traditional libraries and agency filing cabinets. They frequently contain historical data and information that can be invaluable.

4.1.3 General Sources of Data

One good place to start collecting data is to use an Internet search engine like Google to search for the names of rivers and their tributaries in your watershed, as well as other relevant place names and topics. Your search results may return thousands of links so it is important to selectively narrow your search terms until you find items of interest. You may find that reports you really want are only available in libraries and at local agency offices. The U.S. Geological Survey has an online database of publication abstracts that can be searched by theme or location (http://usgspubs.georef.org/usgsns.htm). Many state agencies and federal agencies post reports online—use the search engines for their sites to search for the name of your watershed or waterway.

4.1.4 Spatial and Temporal Scale of Data

Data collected at scales that you cannot control constrains the use of that data in new analyses. Your task as the assessor is to know the scales at which data were collected and determine the kinds of analyses you can conduct and the decisions you can make based on these data. The scale of data will be associated with any well-developed set of data or obvious from the data distribution (e.g., monthly water quality samples from the same 10 sites in the watershed over 3 years). Be aware that, some of the information that an assessment relies upon as "data" may actually be products of computer models (e.g., for wildlife habitat, fire hazard, and landslide risk), which potentially increases the range of uses of data, but not necessarily the reliability.

Frequently, you will encounter a mix of data scales. Some spatial data will be low resolution and tell you only generally what kinds of plant community, land ownership/use, hydrology, and geology are present. Some temporal data, say for water quality, may be high resolution if they were collected continuously or at frequent regular intervals over a long time. Your knowledge of flow in your watershed will probably be based on very few regularly maintained streamflow gaging stations, due to the gradual loss of this critical data resource in California. On the other hand, water quality monitoring may have taken place in your watershed, with regular collection of periodic and storm event measurements. Most information about non-point source pollution, natural condition, and decisions about combined land and water management comes from the combined measurement of flow and water quality constituents. Sometimes water quality data alone are the basis for management decisions.

In general, when you combine scales, the resolution of your analysis product is the same as the lowest resolution input data.

Provided below are some guidelines for your actions related to watershed assessment based on the more established principles in the scientific disciplines where scale is important.

- *Spatial Scales*

The term spatial scale refers to two characteristics. In everyday usage, the term refers to the extent of the area covered by a dataset, which is useful when talking about the "scale" of a project or analysis. However, more technically, spatial scale refers to the fineness of the resolution of the data. For example, the Natural Resource Conservation Service (NRCS) is gradually mapping soil formations across the country at a resolution of 1:24,000, which translates to a six-acre minimum unit size for a soil area (http://soils.usda.gov/technical/manual/). The spatial scale at which this data should be used is a minimum of six acres. In ecology, this characteristic is referred to as the "grain" of the spatial data. You could use the data at finer scales (e.g., one acre), but uncertainty would increase because the data was designed only to provide a coarse overview of soil types. To use this relatively coarse data to describe soils in smaller units could lead to misrepresentation and errors. Therefore, it is important to consider the grain of the data when you collect spatial data so you can use it appropriately when you get ready to analyze your findings.

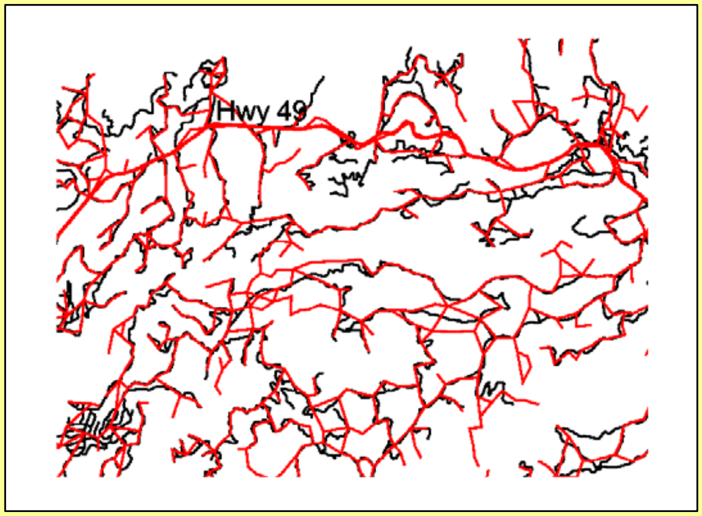

Figure 4.1 Contrasting scales of spatial data

The NRCS data would provide good resolution for a large watershed assessment where the goal was decision-support at the creek sub-watershed, agricultural parcel, or U.S. Forest Service road-segment scale. However, at finer spatial resolution, say for hillslope stabilization, small timber harvest planning area, and protection of rare plant communities in developing areas, you might want a new field survey of soils that has a finer spatial scale.

In some cases, the accuracy of commonly-used spatial data can be a concern. For example, the quality of road data is generally poor in California, in terms of actual road position, road type and condition, and road length. At the scale of a creek watershed, it might be critical to know the points of interactions between the roads and the creek, riparian areas, and slopes.

Figure 4.1 shows the data available from the TIGER database (1:100,000 TIGER data from the Teale Data Center, in red or light color) and those developed by the Tahoe National Forest (TNF) (in black, from a combination of GPS and aerial photos). Within the TNF, road system and watershed analysis can take advantage of higher resolution data, potentially leading to higher resolution decision making. Outside the TNF boundary within the same watersheds, only TIGER data is available, so the data scale coarsens and the analysis suffers resolution, leading to lower resolution decision-making.

If you are fortunate enough to be able to collect new data using a monitoring or assessment program, you can set the scale of data collection to match your analysis and decision needs. Selecting the correct scale for monitoring isn't easy. Some of the material referenced in this manual might be of use, but you should seriously consider consulting with an expert so you can get the most from your investment of time and money.

- ***Temporal Scales***

Most data that you will use will have a time element. It will either be a single point in time for an observation or measurement, multiple points in time (e.g., daily or weekly) over a certain period, or continuously measured. A water quality measurement or grab sample is a snapshot of conditions, which may be accurate for that hour or day, but loses accuracy as you generalize across time (e.g., to a month).

Measurements of aquatic insect communities may be accurate for estimating conditions over the population and individual life spans (months to years), but only in extreme lethal conditions will it tell you much about shorter timeframes (e.g., days). A satellite image that is used to model likely plant communities captures conditions today and will be relatively stable (potentially over years) unless there is an abrupt removal or replacement of the vegetation due to plant community restoration, logging, fire, agriculture, or urban development. The time scale over which new data are collected is another consideration. A one-time grab sample may be all you are able to collect, and it is better than nothing. However, it will provide a limited picture of reality. For example, in some eastern Sierra Nevada streams, episodic events of low pH are associated with snowmelts that are of short duration, but may be lethal to trout. In this case, collecting samples that reflect the extreme conditions, not the average pH, is crucial. Sampling over a short timeframe could totally miss these events. Data should be collected over the period of time in which you would expect to see changes and/or when there is the greatest likelihood of detecting infrequent, but important, events.

The sample list in Table 4.2 is an easy way to keep track of key considerations when evaluating the scale of your data. It will be particularly helpful later in the assessment when you begin data analyses and integration. These kinds of metadata (information about your data) are important when deciding on appropriate uses of the assessment and additional data collection.

4.2 Sources of Numerical Data and Information

The availability of hydrologic and water-quality data on the Internet has thoroughly transformed the hunt for watershed data in recent years. Rather than conducting long searches and even longer copying and transcription sessions in libraries of distant universities and archives of local agencies, an amazing amount of material is now accessible from any computer connected to the Internet. Online resources do not eliminate the need to seek specialized and unpublished information from a variety of sources, but your office computer is now a convenient place to start your search for data.

Table 4.2 Scale reference list (WQ = water quality; LFA = limiting factor analysis, a way of discovering environmental conditions that are limiting survival of a species (e.g., salmon); GIS = geographic information system)

Data type	Use	Spatial scale Spatial units	Temporal scale Time units
Dissolved Oxygen	WQ, LFA	Station	Monthly, hourly, event
Temperature	WQ, LFA	Station	Continuous
Road map	GIS, erosion	1:24,000	5-year update
Employment Sectors	GIS, community characterization	census blocks	10-year update

The U.S. EPA's "Surf Your Watershed" Web site (http://cfpub.epa.gov/surf/state.cfm?statepostal=CA) is a good place to begin. An interactive map will direct to your region. This site provides links to a variety of other Web sites and resources that may be useful in your search for watershed data. The number of links on EPA's site varies tremendously, depending on the watersheds. Also, smaller watersheds tend to be combined together with other ones to form larger units. Data records may be incomplete.

The U.S. Geological Survey (USGS) is a repository for the nation's water data. Your search for USGS data can begin at http://ca.water.usgs.gov or http://waterdata.usgs.gov/ca.nwis/nwis. The USGS Web site contains primarily flow information, although for some sites there is water quality data as well.

4.2.1 Water and Sediment Quality Data

Water quality data include information on both the water column and sediment. Categories of data include: suspended sediment, temperature, dissolved oxygen, nutrients (e.g., phosphates), organic and inorganic chemicals such as pesticides and pharmaceuticals, dissolved organic carbon, pH, etc. Changes in these water quality parameters can occur in response to natural processes, human activities, and the interaction between human activities and natural processes. The data can be collected through grab sampling (i.e., scooping up a volume of water) and continuous sampling/measuring.

The importance of data on the water column is clear—fish and most other aquatic organisms live in and breathe the water. Water quality problems can affect these organisms' respiratory (gills), excretory (kidney), reproductive, nervous, and cardiovascular systems. Often, the effects of poor water quality on the organism(s) of interest are not obvious. If poor conditions persist, however, or if the concentration of the harmful constituents is high enough, mortality will result. Many toxic chemicals found in surface and ground water are carcinogenic, can cause tumors in fish, and/or attack the nervous or endocrine systems.

Sediment contaminants have the potential to affect all aquatic organisms as well. Frequently, invertebrates living in sediment will accumulate contaminants. These animals are subsequently eaten by fish, which concentrate the contaminants, so the fishes' exposure is actually greater than the chemical's presence in the environment. In this way, fish (or things that eat the fish) that normally wouldn't be exposed to a contaminant can be harmed.

Many non-water soluble contaminants are primarily found in sediment. Pesticides, such as DDT (now illegal), persist in sediment for very long periods of time. Sediment contaminants can diffuse into pore-water, the water between streambed particles. Depending on the contaminants' solubility, diffusion occurs at different rates. Compounds found primarily in sediment include PCBs, DDT, dioxins, and many polycyclic aromatic hydrocarbons, byproducts of fossil fuel combustion. Pyrethroid insecticides, such as esfenvalerate, and metals are also primarily found in the sediment. Measurements of these chemicals in the water column do not give a true picture of their presence in the waterway because their highest concentration is in sediment or pore-water, the water found between the streambed material.

Water and sediment quality data will generally be used to evaluate conditions by comparing data from your watershed to standards. These standards, promulgated by the U.S. EPA, USGS, and/or the State Water Resources Control Board, contain benchmark values for acute and chronic exposure to dissolved oxygen, nutrients and other conventional water quality parameters as well as contaminants such as organics, pesticides, and metals. These benchmarks reflect the highest concentration of a contaminant to which an aquatic organism can be exposed without risking adverse effects and will allow you to estimate whether water and sediment conditions pose a risk to the important ecological resources included in the watershed assessment.

4.2.1.1 Sources of Information on Water and Sediment Quality

1. STORET

The primary source for water quality data is the U.S. EPA's STORET system (http://www.epa.gov/storet/dbtop.html). STORET has recently been split into two separate entities: the Legacy Data Center (LDC), a static archive containing historical data collected through 1998, and STORET, the modern system.

STORET contains data collected since January 1, 1999, and older data that have received current levels of documentation. The current STORET program is available to users who wish to organize their water quality data under the STORET protocols on their own computers. Users can upload these local files to the central STORET system and make the data available through the STORET Web site. As of July 2004, the "Modernized STORET Data" site did not contain many California data, in contrast, the "Legacy STORET Data" site did.

Both the LDC and STORET contain a wide range of chemical, biological, and physical data under the broad heading of water quality. A diverse assortment of people, agencies, and organizations collected these data for a myriad of purposes with the common goals of making the data available to the public. Each water quality value is associated with information on where the sample was taken (geographic coordinates, state, county, USGS Hydrologic Unit Code, and a brief site identification), when the sample was collected, the medium sampled (e.g., water, sediment, fish tissue), and the name of the organization that collected the sample. In addition, STORET contains information on why the data were obtained; sampling and analytical techniques used; the laboratory that analyzed the samples; the quality control checks used when sampling, handling the samples, and analyzing the data; and the people responsible for the data. EPA's general description of the STORET database (http://www.epa.gov/storet/descript.html) provides additional details. The Pennsylvania Department of Environmental Protection's Web site offers detailed instructions for navigating through both the LDC and STORET (http://www.dep.state.pa.us/dep/deputate/watermgt/Wqp/WQStandards/STORET-Access.htm).

Depending on your data requirements and computing options, there may be some advantages to obtaining the EPA STORET (both legacy and modern systems) on CD instead of via the EPA website. Two companies publish CDs of the EPA water quality data (http://www.earthinfo.com and http://www.hydrosphere.com).

2. USGS

Water quality data from discrete samples are accessible from the USGS's Web site http://waterdata.usgs.gov/ca/nwis/qw. Under the "Tutorial" button are instructions for accessing water quality data in watersheds. You can get to your watershed of interest quickly if you already know the USGS hydrologic unit code. Otherwise, navigate to your watershed by starting with your county. Depending on what, if any, data are available for sites in your watershed, you can specify a variety of output formats to meet your needs. The USGS has a good summary of its procedures for sample collection and onsite measurements at http://ca.water.usgs.gov/archive/waterdata/text/onsite.html. More detailed information can be found in the USGS reports, Techniques of Water-Resource Investigations, at http://water.usgs.gov/pubs/twri/.

USGS has 35 sites in California equipped with continuous monitoring sensors and telemetry gear that allow reporting of near-real-time measurements and the record of the past 31 days (http://waterdata.usgs.gov/ca/nwis/current/?type=quality). Under the menu for "Available data from site", select "Recent daily" to obtain daily values of the onsite measurements for up to two years. Most of these sites only report water temperature, but some also measure conductance, dissolved oxygen, pH, and turbidity.

3. State Water Resources Control Board (SWRCB) and Regional Boards

Water quality data from the Central Coast Regional Water Quality Control Board's Central Coast Ambient Monitoring Program is available at http://www.ccamp.org. Here you can navigate to a sampling location or water body and then view a summary table of attributes and values. Alternatively, you can first choose "Monitoring Data" for dozens of water quality parameters organized under categories of conventional water quality, freshwater invertebrates, hydrocarbons, metals, and organic chemicals. This path allows you to compare summary values among sampling locations.

The SWRCB is developing the Surface Water Ambient Monitoring Program (SWAMP) (http://www.swrcb.ca.gov/swamp), a Web-based source of water quality data that aims to coordinate and compile data programs of several agencies within California into a common database. SWAMP and a broader Water Information Network are anticipated to be active in the near future.

Water quality standards for each stream in California, based on Regional Board basin plans, can be found on a GIS-based inventory developed by Caltrans' Environmental Program (http://endeavor.des.ucdavis.edu/wqsid/). Queries can be by waterbody name, beneficial uses, regional board, county, or keyword search.

Finally, the Regional Boards perform toxicity tests in accordance with U.S. EPA guidelines on rivers and streams throughout the state. Although not presently published anywhere, this data is public information.

4. Additional Sources of Water Quality Information

- Water Data Library Web site (http://wdl.water.ca.gov/)—California Dept. of Water Resources (DWR). Water quality data, access to hydrologic data (searchable by station and county) collected by the Division of Planning and Local Assistance and other groups within the department, and information on a wide variety of chemical contaminants, such as pesticides and metals, as well as conventional parameters, such as conductivity and hardness.

- The California Digital Conservation Atlas (http://www.legacy.ca.gov) —the Legacy Project at the California Resources Agency. An Internet map-making site with a wide variety of water quality data (under the Environmental Stressors tab) for certain waterbodies in California; data from the Musselwatch Program, Toxic Substances Monitoring Program, and others can be identified on stream and waterbody maps of streams and waterbodies throughout the state; information on contaminants in water and data on toxicity tests, which involve exposing model organisms, such as a water flea or amphipod, to water or sediment samples and measuring mortality or other biological endpoint. These tests do not identify the reason for the toxicity, but they are an excellent way to determine if anything in the water/sediment might pose a risk to aquatic life. You will need to determine the cause of the toxicity independently.

- Surface Water Monitoring Program database (http://www.cdpr.ca.gov/docs/sw/surfdata.html) —The Department of Pesticide Regulations. Data on pesticide concentrations in waterways throughout California, not just agricultural regions.

- Drinking Water Program (http://www.dhs.ca.gov/ps/ddwem/technical/dwp/dwpindex.htm) and Drinking Water Source Assessment and Protection Program (http://www.dhs.ca.gov/ps/ddwem/dwsap/DWSAPindex.htm)—the California Department of Health Services. Information and data on the quality of water used for drinking water supplies. The Web site of the Drinking Water Program also has a fairly thorough directory of other potential sources of data (http://www.dhs.ca.gov/ps/ddwem/dwsap/DWSAP_directory.htm). Contact your county health department and local water-supply utilities for water quality data for local water supplies (surface water and groundwater).

- United States Forest Service (local offices)—water quality studies and monitoring on local water bodies for watersheds that include any national forest lands.

- Data from aquatic bioassessment work (http://www.dfg.ca.gov/cabw/cabwhome.html)—California Department of Fish & Game. As of 2003, a centralized results database did not exist, but SWRCB has recognized the need for such a database (Tetra Tech, 2003).

4.2.1.2 Hydrological and Climate Data

Information and data about your watershed's hydrology are critical components of any watershed analysis, although data have not been collected for many, if not most, watersheds. Streamflow is particularly important for addressing concerns about flooding, aquatic communities, and water quality in your watershed assessment. The availability of hydrologic data largely depends on whether some agency thought the water in your stream had some utility, either locally or for export. If there are major or formerly proposed water engineering projects in or near your watershed, then there is a high likelihood of current or historic stream gauging stations. If your stream is an un-dammed tributary that doesn't supply water to municipalities, hydroelectric facilities, or irrigation districts, the stream is unlikely to have a gauging station. The main exception is if flooding concerns exist. Many small streams and creeks near urban areas contain gauging station for flood control purposes. In general, where there is concern about the waterway, there are more gauges and measurements.

Flow Data

Streamflow data come from three basic types of measurements: 1) continuous records of stage (water level) and discharge at a calibrated cross-section, 2) spot measurements, and 3) crest-stage gauges (where only the highest water level is recorded). From a watershed assessment perspective, the most useful data comes from a long-term continuously recorded stream gauge (#1 above) that has measured daily or monthly streamflow for the past few decades. Gauging stations typically record stream stage at 15-minute intervals. Corresponding discharge is then calculated from a rating curve based on manual measurements of instantaneous discharge across the channel cross-section and stage at time of the measurement. Some stream gauges may not have a recording device, but instead rely on manual observation of a staff gauge (basically a well-anchored ruler from which an observer can read the water level) approximately once a day. Such gauges are common in irrigation ditches and sites with full-time staff. Chapter 3 presents fundamentals of stream gauging. For more detailed information, see Dunne and Leopold 1978:594-598;Hornberger, et al. 1998:100-103) or the USGS Web sites: http://ca.water.usgs.gov/archive/waterdata/text/explain.html and http://water.usgs.gov/pubs/twri/. If the gauging station is in close proximity to a dam or diversion, both the actual flow at the location and a calculated value that includes diverted water or changes in reservoir storage may be reported.

Spot measurement data is generated from occasional measurements with a current meter (device that measures the local speed of the flowing water) or other technique (e.g., Herschey 1985). Unless there is little streamflow variability, such measurements do not reveal much information about the long-term characteristics of streamflow. If such data are all you have, interpret them with caution because a lot can happen in between the measurements. Spot measurements can be valuable longitudinally along a stream when streamflow is measured at several places up and down a channel on the same day. In this case, you can learn how much streamflow changes in the downstream direction as the contributing area increases. These longitudinal studies are also useful for measuring the effects of diversions, irrigation return flow, and subsurface hydrogeology (where the stream is naturally gaining or losing water to the alluvial aquifer).

Maximum water levels (high water marks) from crest-stage gauges generate data useful in flood studies where installation and maintenance of many recording stream gages is not feasible. Crest-stage gauges can be as simple as a vertical pipe containing a ruler and a marker with a few holes in it to allow water to flow in and out. This marker can be something that floats, such as cork particles, and adheres to the ruler at the highest water level or something that dissolves in water and has been applied to the ruler. After an increase in streamflow, a hydrographer (one who measures water) removes the ruler and records how high the water rose in the pipe (and therefore the stream). Estimating the corresponding discharge is difficult and involves considerable uncertainty (often +/-50% or more). Nevertheless, such data are often the only estimates of peak flows.

Although not strictly streamflow, another common type of hydrologic data is estimates of volume in lakes and reservoirs. Similar to stream gauging stations, the lake's or reservoir's water level is observed (either visually on a staff gage or with a water-level sensor and recorder), and the volume is calculated from an equation or table relating water level and volume.

The USGS is primarily responsible for the nation's water data, including streamflow. Navigating USGS streamflow data is easiest when you have the USGS gauge numbers, which have a structure similar to USGS hydrologic unit codes for watersheds. From USGS entry portals http://ca.water.usgs.gov or http://waterdata.usgs.gov/ca.nwis/nwis, you can easily get to a site selection page and enter your county or a pair of latitudes and longitudes to begin your search for stream gages in your watershed. USGS publishes a series of schematic diagrams of the major river basins in California that are very helpful for understanding the geographic arrangement of tributaries, stream gages, and major dams and diversions.

Keeping a copy of the relevant schematic(s) for reference while navigating USGS data archives will aid your search. These schematics can be found within four regional volumes (Vol. 1 – Southern California, Vol. 2 – Central and North Coast; Vol. 3 – Southern Central Valley and Great Basin from Walker to Truckee; Vol. 4 – Northern Central Valley and Northern Great Basin) for California. Volumes for 1999 through 2002 can be downloaded as .pdf files from http://ca.water.usgs.gov/archive/waterdata/. Older volumes can be found at most university libraries and in some offices of public agencies.

Another important piece of information within these volumes is the list of discontinued gauges (online at http://ca.water.usgs.gov/archive/waterdata/99/disc_sw.html). Information about stream gages, or what the USGS calls a "station manuscript" (precise location, drainage area, period of record, summary of extremes, etc.), that were in service between 1994 and 2001 can also be found at http://ca.water.usgs.gov/archive/waterdata/ The search function here works best if you have the gauge number or if the name is unique. Common names like "Clear Creek" will yield thousands of results.

Although the USGS portal (http://ca.water.usgs.gov/archive/waterdata/) provides annual tables (flat files) of daily discharge values for gauges operated between 1996 and 2001, you may wish to obtain more data in a format that can be manipulated on your computer. For access to more thorough data after you know what sites and periods of record are available, go to: http://waterdata.usgs.gov/ca.nwis/sw or http://nwis.waterdata.usgs.gov/ca/nwis/discharge.

Graphical output is also an option. A tutorial on accessing historical streamflow data is available at http://nwis.waterdata.usgs.gov/tutorial/historical_streamflow.html.

In addition to daily streamflow values, USGS also publishes data of the highest annual flows over the period of record for selected sites. Access the peak streamflow database at http://waterdata.usgs.gov/ca/nwis/peak. Again, your progress will be fastest if you already have the station ID number. You can specify the output format and file information. Depending on your data requirements and computing options, there may be some advantages to obtaining USGS streamflow on CD instead of via the USGS Web site. Two companies publish CDs of USGS daily streamflow and peak flow data (http://www.earthinfo.com and http://www.hydrosphere.com).

Other sources of streamflow data include water districts, municipal utility districts, irrigation districts, hydroelectric generating companies, the U.S. Forest Service, and any other local entity that needs to measure streamflow. Data from such sources may not be available online and may require a personal inquiry to the agency or company. Some records are not considered public information and may not be available. In cases where the data are not public records, be prepared to make a good case for your need, demonstrate that release of the data will not be harmful to the supplier, and be prepared to pay for staff time to copy or otherwise prepare the data for you.

Local or regional flood control agencies may also have streamflow and stage data. This data is often available on the city/county websites and serves as a warning system for local residents. Because some stations may not be maintained carefully due to local budget constraints, ask local public utilities department engineers about the data's accuracy.

The California Department of Water Resources' Division of Flood Management maintains the California Data Exchange Center (CDEC) at http://cdec.water.ca.gov, which provides very recent and real-time data in response to flooding (i.e., real-time river stages) and for water project operations. If you need current daily and monthly streamflow and reservoir data, see this site.

Climate Data

Climate data are also available from CDEC at http://cdec.water.ca.gov. Select "Precipitation/Snow" from the menu "CDEC Resource Directory" for a list of precipitation stations where you can select an individual station and obtain the latest data. At the bottom of a page for a particular station, select "Historical Data" to get to the "Bulk Data Selector". Using the three-letter station code from the previous page, specify the data, period of record, and output format. A pair of interactive maps for locating stations with available data can be found at http://cdec.water.ca.gov/cgi-progs/mapper.

The Western Regional Climate Center in Reno, Nevada, is the other major source for precipitation, temperature, and other climate data. Begin your search at http://www.wrcc.dri.edu/CLIMATEDATA.html. Interactive maps with locations of climate stations are available for Northern California (http://www.wrcc.dri.edu/summary/climsmnca.html) and Southern California (http://www.wrcc.dri.edu/summary/climsmsca.html). After selecting a particular site, you will obtain a monthly summary over the period of record and a menu for accessing more detailed information for the site.

4.2.2 Riparian Vegetation and Wetlands Data

The riparian zone is where the aquatic and terrestrial landscapes come together and where species from both environments benefit. As a result, riparian data collection generally reflects both aquatic and terrestrial ecosystems. Description of riparian habitat generally involves interpretation of aerial photography for large watersheds (>100,000 acres) and field surveys for small watersheds (10,000 acres) or parts of large ones.

Commonly collected characteristics of riparian data include:

- Plant species (native and introduced),
- Plant community (e.g., cottonwood riparian, mixed conifer riparian),
- Tree canopy cover (average as a percentage),
- Tree canopy closure over stream (or other index of stream shading)
- Tree size,
- Large woody debris (including potential for more wood to enter the stream channel),
- Stream bank erosion,
- Average width of riparian zone,
- Disturbances (type, extent, and intensity),
- Fragmentation of riparian habitat,
- Water impoundments,
- Stream bank structure and stability

Data about riparian vegetation and other characteristics of riparian areas are likely to be scarce for your watershed. In the past, there has been little demand for systematic surveys of riparian areas (National Research Council 2002). The California Riparian Habitat Conservation Program was created within the Wildlife Conservation Board (WCB) at the Department of Fish and Game in 1991 (http://www.dfg.ca.gov/wcb/california_riparian_habitat_conservation_program.htm.) with the objective of assessing the current amount and status of riparian habitat throughout the state. However, this extensive mapping effort has not yet occurred. This program is partnering with the state-federal-private Riparian Habitat Joint Venture, which is riparian bird habitat protection effort. The Joint Venture also aims to identify riparian areas in the state, but it has not yet compiled a database (http://www.prbo.org/calpif/htmldocs/rhjv). Challenges include the difficulty of mapping land cover statewide with sufficient resolution to characterize the narrow riparian zone and the cost of doing this high resolution work.

Riparian areas have been mapped for parts of the state through several projects conducted by CDF's Fire and Resource Assessment Program (FRAP) (http://frap.cdf.ca.gov). It has developed a California Hardwood Rangeland Riparian Vegetation Database, where riparian is one of the seven database fields mapped in 1990 using satellite imagery.

http://frap.cdf.ca.gov/foofoo2/veg_data/riparian_metadata.html. It also mapped and assessed several north coast watersheds for riparian vegetation conditions through the North Coast Watershed Assessment Program (http://www.ncwatershed.ca.gov). Using existing U.S. Forest Service and CDF vegetation maps, FRAP updated species, canopy cover, and tree size using DFG stream habitat survey data (see above), aerial photograph interpretation, and new field data where needed. FRAP's Forest and Range Assessment 2002 developed a single GIS data layer for vegetation that includes riparian categories by merging multiple sources of data.

The most complete information and data about riparian vegetation are usually found for areas downstream of hydroelectric projects that have been through or are currently in the relicensing process. As part of the relicensing effort's environmental review, the Federal Energy Regulatory Commission (FERC) and U.S. Forest Service (if the project is on national forest land) generally require a thorough assessment of riparian resources that have been affected by the hydroelectric operation to date or that could be impacted if the project continues. The basic information is usually found in an appendix to the Environmental Impact Statement. If this description is not detailed enough, you may be able to obtain raw survey data from the personnel or consultants who worked on the riparian assessment.

Environmental documents for proposed new projects are another source of information, though in very limited areas. Environmental Impact Reports (EIR) for new project require surveys of existing conditions and potential impacts of the project. All EIRs are to be filed in the State Clearinghouse maintained by the state Office of Planning and Research (http://www.opr.ca.gov). These EIRs could be a source of useful information on riparian and other habitat. The local county planning department often keeps an archive of such reports as well. You can also search for documents prepared under the California Environmental Quality Act at http://www.ceqanet.ca.gov.

If your watershed includes some federal land, there may evaluation or monitoring data for selected riparian areas at your local Forest Service or Bureau of Land Management (BLM) office. Most national forests have conducted stream surveys for parts of their forest, both in a systematic process as well as for specific projects. Prior to the mid-1990s, little of this data was collected or archived in a consistent format. In the past few years however, a procedure

The San Joaquin River Riparian Habitat Restoration Program

(http://www.usbr.gov/mp/cvpia/sjr/) serves as a case study for using riparian data sources. The program evaluated historical riparian habitat conditions and changes through the use of soil surveys, historical maps, and historical aerial photographs going back to 1937 (Jones & Stokes 1998a). Geographically corrected maps at a scale of 1:24,000 were compiled for riparian soils, habitat, and land use for several dates during that period of time. Data were entered into a GIS database, which allowed the changes in areas and types to be quantified for those years. A second riparian study analyzed how physical processes, such as flow conditions and sediment regimes, shaped the San Joaquin River and affected the riparian habitat patterns along 150 miles of the river below Friant Dam to the confluence with the Merced River (Jones & Stokes 1998b). Low-altitude aerial reconnaissance flights were made to record current vegetation patterns, as well as geomorphic and hydrologic features. Ground-level surveys recorded riparian vegetation structure, condition, dominant species, and relation of species to channel geometry, in addition to physical channel measurements.

known as Stream Condition Inventory have been used to characterize riparian conditions. Contact hydrologists, botanists, and fisheries biologists in the local district or supervisor's office to check on the availability of stream and riparian survey data for your watershed. The BLM also has a continuing program for assessing riparian conditions on the lands it administers. The BLM approach is known as "proper functioning condition" (PFC), which can be evaluated on the adequacy of vegetation, landform, or large woody debris to serve certain functions (U.S. Department of the Interior BLM 1995). This assessment method tends to be used on public lands and some private rangelands.

There is a chance that some sort of riparian research may have been conducted in your watershed. Many of the riparian studies in California during the 1970s and 1980s were reported on or referred to in a series of conference proceedings (Sands 1977, Warner & Hendrix 1984, Abell 1989). An online bibliographic search may yield leads to other scientific papers, such as the user-friendly riparian research Web site by the University of Washington with over 8,000 citations (http://riparian.cfr.washington.edu/).

Aerial photography is a potential source of raw data about riparian vegetation (primarily vegetative cover and human disturbances), although it requires a lot of effort to interpret the images (e.g., Nelson & Nelson 1984, Grant 1988). Sources of archived aerial photography include offices of the USFS, the BLM, the CDF, county planning departments, the Earth Science and Map Library at U.C. Berkeley, and the Map and Imagery Library at U.C. Santa Barbara. Videotapes of a few California river corridors under study for Wild and Scenic River status filmed from helicopters by the National Park Service are available at the Water Resources Center Archives at U.C. Berkeley. There are also "digital ortho quarter quads" (DOQQs) and "digital raster graphs" (DRGs) available for the state, which are types of satellite or aerial photographs that you can use in GIS. These can be downloaded from the state GIS repository (California Spatial Information Library, http://gis.ca.gov/data.epl) in various forms. This site also has a variety of other statewide data.

The National Wetlands Inventory (NWI) of the U.S. Fish and Wildlife Service provides information about wetlands throughout the United States. About 90% of the wetlands of the continental U.S. have been mapped, with much of the information available in digital form online (http://www.nwi.fws.gov). With the exception of southern desert areas, most of California has been mapped as of November 2003, with digital information available for about half the state. Other potential sources of wetlands information and data include county planning departments, the Natural Resources Conservation Service, and the U.S. Army Corps of Engineers.

4.2.3 Physical, Channel, and Habitat Conditions

Geologic and topographic information are essential components of watershed assessment. California maps are readily available from a variety of agency and private sources—and are always useful in an overview-scale assessment. Data that provide a physical overview of watershed characteristics are available from USGS topographic maps (http://ask.usgs.gov/maps.html) and geologic maps from various agencies. The USGS site also provides numerous thematic maps, such as geology, groundwater resources, seismic, and vegetation maps that may or may not be available for your watershed. The California Geological Survey has specialized geology maps at various scales (http://www.consrv.ca.gov/CGS/information/geologic_mapping/index.htm). Other important geologic maps that are available statewide but are not online include the

California Resources Agency Geologic Map Sheets (scale of 1:250,000).

Historic and current aerial photographs are another valuable source of watershed-scale data that helps characterize a watershed's physical character (as well as vegetation and land use character). Topographic data and photographs are available online from http://terraserver-usa.com/ sponsored by USGS and other groups. County planning departments, museums, Caltrans, NRCS, and a variety of agencies and private firms can sometimes provide historic and current aerial photographs that may be used to evaluate trends and changes in physical watershed characteristics over time.

Collecting data on physical channel and habitat conditions at the scale of individual river reaches or habitat units involves primary reconnaissance and field work, as these features cannot easily be discerned from aerial photographs or maps and very little online or existing information is typically available. Field surveying for these data is very involved and can be comprehensive only for small watersheds. For large watersheds, field work could be used to calibrate a more generalized (e.g., GIS-based) description of classes of sub-watersheds.

Stream channels and associated habitat vary tremendously throughout California as a result of the variability in climate and tectonics that differentiate the landscape, and specific sediment erosion, deposition, and morphologic conditions cannot be generalized without very specific regional and local knowledge. Moreover, stream channel conditions vary longitudinally from the headwaters to the lowlands and laterally from the low flow-high flow channels to the floodplain, and data collection in one area of the watershed system may not represent conditions in another area. For example, measurements of suspended sediment or bed material load at one gauging station will not represent downstream conditions if a dam traps sediment or a tributary is a source of sediment between the two areas. But because of the wide range of natural spatial and temporal differences, extrapolating information from short-term data sets needs qualification. An overview of the USGS sediment collection program is summarized at http://ca.water.usgs.gov/projects02/ca004.html. Geomorphic and sediment-related data can be stored in spreadsheets where it can be easily accessed to report as a table or in a graph.

4.3. Archiving and Managing Numerical Data

Use of a database or spreadsheet is very helpful in managing the data you have collected. Spreadsheets are easy to use, but lack the ability to conduct a focused search or queries that are the hallmark of a database. For example, you may wish to examine particular attributes, locations, sampling dates, or numeric values within or above a certain range. Databases are made to support these types of analyses. A useful introductory reference on environmental data management is Michael (1991). Environmental databases: design, implementation, and maintenance. Lewis Publishers. Boca Raton FL. A more comprehensive reference on this topic is Michener, et al. (1994).

If you are able to take advantage of a reliable online source of data (e.g., USGS daily streamflow records) or a CD archive of data, it is often best not to copy the data on to your computer. A great advantage of these vast databases is their easy and long-term accessibility. In most cases, a link to the data is as good as having it on your hard drive. There are a lot of exceptions to this general recommendation, the main one being that if you want to integrate or overlap data among "disciplines" (e.g., land-use and water quality), then it makes sense to have it all in one place.

Table 4.3 Example of tabular file structure

Location	Year	Day	Discharge (cfs)	Suspended sediment concentration (mg/L)
Clear Creek	2001	1	10	4.0
Clear Creek	2001	2	11	2.2
Clear Creek	2001	3	19	12.1
Clear Creek	2001	4	12	22.5

Existing environmental data will be stored in some sort of file structure or database, with the particular details dependent on the type of data, the agency archiving the data, the needs of data users, and the dynamic nature of the variable being assessed. The simplest data structure is usually a single column of values, such as daily average streamflow for a year at a single gage station. In such a file, there would be 365 records (lines or rows) with a single number in each line or row. Metadata (information about data) would be associated with this file to allow you to interpret the data. In this example, to make sense of the single column of numbers, you must know the location of the stream gauge, the year, the starting date (so you could assign a date to each successive entry), and the units of the values.

A more common file structure is a table with multiple rows and columns (or records and fields in database terminology). Simple data tables are often called "flat files". In the streamflow data from table 4.3, the metadata is organized by columns. The day of the year is unique for each record, but the location, year, and units are common for all records. The metadata still needs to include the start date of the year (calendar year on January 1, water year on October 1, rainfall year on July 1) and the headings for the columns (which are usually not part of the table).

A wide variety of database management systems have been developed, primarily for business needs. Environmental data archiving and retrieval needs can usually be accommodated by these systems, although there may not be an ideal match. Most common database management systems (Microsoft Access, FileMaker Pro, Oracle etc.) can result in relational databases and use Standard Query Language (SQL). The relational structure allows use of many data tables that are linked together by one or more common fields. These linked data tables reduce the need to enter redundant data, such as location, in every record. Questions or queries can be asked of the database through SQL (usually via some simple-to-use interface) without writing a program to perform the search. In addition, the database management software facilitates data entry, updating, deletion, simultaneous processing of commands from multiple users, import and export of large amounts of data, and recovery from system crashes.

4.4 Anecdotal Information

Anecdotal information may provide knowledge about watershed processes and conditions that might otherwise be lacking. Common types of anecdotal data include:

1) The change in the extent of salmon runs in river before and after building of a dam or bridge,
2) Change in turbidity of streams and rivers over time,
3) Changes in land use,
4) Changes in riparian vegetation,
5) Changes in the frequency or size of floods

Anecdotal information has been useful in many assessment efforts. For example, fisheries scientists in the Central Valley have re-constructed the history of salmon runs in the Valley over the past 170 years from archaeological site findings (i.e, fish bones) and written accounts in newspapers, diaries, technical reports, and letters to government agencies (Yoshiyama et al. 1996 & 2001). The South Yuba River Citizens League has provided a more detailed account for the Yuba River basin based on a similar approach (http://www.syrcl.org/majorissues/SalmonReport.htm). In both cases, the results of visual observations by anglers and others were combined with more contemporary fish surveying methods to assess the condition of the salmon populations in the Central Valley watersheds relative to historical distributions.

If you plan to include historical or contemporary anecdotal information, your main tasks will be finding, organizing, and interpreting the data. Public libraries (including holdings at universities and agencies) have extensive archives of newspapers and other written records. People who have lived in the area for a long time and enjoy the rivers or streams are walking resources for certain kinds of information. Neither resource may give you quantitative information. But in the absence of or in combination with other information, it is useful to at least know the presence or absence of watershed processes and the potential range of natural processes (e.g., flooding). This information is best recorded with some associated index or coding for the type of information and where in the watershed it is relevant. The data can be organized using the index or codes, allowing for more efficient retrieval later on. Data interpretation will depend heavily on the source of information, the type of information it is, and the questions you are addressing. You will probably use these data for "yes/no" questions about the watersheds more often than "how much" kinds of questions. The interpretation is probably best done by someone who has familiarity with both the data source and the watershed feature being discussed. It is important for credibility with all stakeholders to qualify your findings and statements when you use of anecdotal data in your analysis, reports, and presentations.

4.5 Types and Sources of Landscape Data

The majority of the area in any California watershed is the terrestrial landscape. Terrestrial landscape data are often collected for entire watersheds. It may also have been collected initially at individual survey sites (e.g., soil or vegetation) and then subsequently generalized to areas. Currently, most contemporary data about a watershed landscape are collected with a geographic reference point. In contrast, historic data may be very valuable, but lack easily usable or identifiable reference points.

Historic maps inform analyses of change and historic condition. Unless these maps have been transferred to an electronic format, they are on paper and may be challenging to use to compare with present day conditions.

Local and regional governments maintain data on land uses. They reflect past or current zoning ordinances and present a reasonable estimate of activities occurring within the watershed. Land use categories include industrial, open space/parks, low density residential, and commercial. Land use maps are usually maintained as GIS files. They can be very helpful in a watershed assessment because they suggest possible practices that could impact the waterways. The State's online GIS center is a good source of land-use maps (http://gis.ca.gov). The Department of Conservation and county agriculture commissioners have maps for rural and agricultural regions of the state.

4.5.1 Data Types

Spatial data comes in a variety of types, including paper maps, digital spatial data ("GIS data"), tables of attributes of areas (such as vegetation types), the results of summarizing or condensing data (e.g., by computer models), and anecdotal data (e.g., un-mapped descriptions of place). Each data type requires different storage and organizational strategies (described in more detail below). Typically, each data type is associated with different levels or complexity of analysis. For example, in a simple and rapid assessment, you might rely on paper maps and anecdotal information, while for a comprehensive assessment involving computer models and multiple questions, you might rely on digital spatial and water quality data.

4.5.1.1 Non-Digital Spatial Data

Historically, spatial data was non-digital. These maps still have useful functions and often can be converted into a digital format. Many local, state, and federal agencies have archives of paper maps. You can collect copies of these maps and make digital versions of them. Alternatively, the information can be summarized and interpreted for later analysis. Paper maps are probably most useful when they show high-resolution information about local occurrences. For most watershed assessments, collecting all paper maps available isn't feasible. You may want to collect maps for select areas where, for example, you are interested in the historical condition or for features where digital information is not available.

Older data can be digitized to create digital maps, which allows historical data to be incorporated into a contemporary database. According to the Information Center for the Environment (ICE, U.C. Davis, http://ice.ucdavis.edu), digitizing costs about $0.50 per acre. Another way to store data from paper maps is to code the data and store that information. Coding refers to attaching a value or code to a particular map feature. For example, if a feature's position has not changed (e.g., a public roadway), but its type, use, or condition has (e.g., paving the public roadway), then the feature can be identified and recorded in a table, and the corresponding coding can be recorded with it. For landscape features that may not have precise position information, you can use nearby positional information (e.g., property name, township name, road number, etc.) to code the data. If data is available on clear acetate sheets, then it is possible to overlay various kinds of data for analysis. As always, you should record the process you use to extract or summarize the information from the paper map source to inform future users of your extracted information.

4.5.1.2 Digital Spatial Data

Digital spatial data are electronic versions of a paper map and shows the relative position of mapped features (e.g., roads, rivers) in a geographical location (e.g., a watershed). These data are often used in watershed mapping for watershed assessments and plans. However, they are useful for much more than just cartography (mapping). They can be used in modeling and understanding the distributions of features across the landscape and how things interact with each other.

4.5.1.3 Non-Spatial Data

Some of your watershed data may have been collected for a project where the exact position of surveying was not critical and where sub-sampling of an area was used. For example, a botanist may have surveyed native and exotic plant communities across a ranch or sub-watershed and provided lists of species present and estimates of percent cover for each of them. This is valuable information for this place, but the data cannot be attributed to particular places except the area being assessed. You can collect data like these and record them by survey area. In cases like this, collecting data and placing it in tables might be the best way to organize it.

There may be other kinds of data that are more suitable to list over a timeframe. For example, if you are interested in flooding occurrence but not in the area that is inundated, then just recording the approximate location of the flooding, the date, and perhaps a ranking for severity may be enough. Certain data will be best collected only on a time "map", for example, number of returning salmon per year, flood frequency, number of landslides per year, number of storm events per month summed over the last four decades, etc.

4.5.1.4 Condensed or Summarized Data

Many agencies and academic institutions have conducted analyses and models of natural and human processes. Condensed or summarized data is new information generated from analysis or modeling. For example, maps of vegetation types, precipitation, fuels moisture, and temperature permit estimates to be made of fire risk, which can result in a new map of fire risk. The values in the fire risk map are derived from the values in the other maps and are a form of data. When recording these data, it is important to record the origin of the values so they are not treated as measurements or the results of surveys, which may have been the case for the base information used in the modeling. Again, describing the data can be as important as the data values themselves. This "metadata" (description of the data) is its own form of information that can inform decision making about data quality, quantity, and sufficiency. The condensed or derived data that results from modeling or analysis may be summarized for specific geographic locations spread throughout the watershed (e.g., erosion risk for all sub-watersheds), or may be summarized only for certain features of the watershed (e.g., stormwater runoff from developed areas).

4.5.2 Sources of Landscape Data

There are a wide variety of sources for data about landscapes. However, these sources are not always easy to find. Large agencies or institutions may hold data, which may be available online. Local agencies or private organizations are other data sources. However, these organizations and agencies may require a more direct approach to explore their databases and retrieve information (i.e., negotiating in person). A list of potential data sources is below, but this list is not exhaustive—there are hundreds of different possible sources of data in California. In all cases, when you access and organize data, maintain a log of where you got the data, make sure you get the metadata, and set up a filing system that is easy to use and intuitive. A list of data source websites is posted at: http://cwam.ucdavis.edu.

4.5.2.1 Federal Sources

Common federal sources of land cover/land use data are land-management agencies, such as the US Forest Service, Bureau of Land Management (BLM), National Park Service (NPS), Bureau Indian Affairs (BIA), and US Fish and Wildlife Service (USFWS), regulatory agencies, such as NOAA-Fisheries, and the Army Corps of Engineers, and research and technical agencies, such as US Geological Survey, Natural Resource Conservation Service (NRCS), and National Oceanographic and Atmospheric Administration (NOAA).

The local offices of these federal agencies are a good place to begin assembling data about the watershed landscape. These data may be in map (spatial) or non-map format. There are usually several natural science professionals (e.g., a soil scientist or a wildlife biologist) who can answer questions about the data and direct you to the appropriate staff for retrieving data. However, in larger offices, there may not be a central directory of all available data, and you may have to dig around a bit before you find everything you want. Some data may be called 'draft' or described as unavailable to the public. In some cases, you may be able to gain access to these data anyway for analysis purposes, especially if the agency has pledged cooperation to your watershed group.

4.5.2.2 State Sources

State land management and regulatory agencies also have a variety of types of data available online or upon request. Examples are the California Department of Forestry and Fire Protection (CDF) Fire and Resource Assessment Program, which has data for vegetation and fire risk, the Department of Conservation (DOC), which has information about landscape conditions and disturbance (e.g., mined areas, agricultural preserve lands), and the Department of Fish and Game (DFG), which has data about wildlife and plant communities. As with the federal agencies, technical staff are familiar with the data the agency has, but you may need to talk with several people to discover everything available. Local offices are generally the best places to begin.

4.5.2.3 University Resources

Academic institutions can have rich mines of information, but it is not always easy to find the right people because they are often spread out among departments. If the university has a directory of faculty research interests on their website, this is a good place to identify experts. If such a directory does not exist, an alternative approach is to go to the webpage for the relevant department website, The departments may be divided among different colleges or schools (e.g., "Agriculture and Environmental Science" at U.C. Davis). Departmental websites usually contain web pages for each faculty member or research program. Reviewing these can help you find people who may have conducted research in your watershed or nearby. Call or email them—they will usually be happy to provide you with information or forward you to someone who can help. If you are lucky, they may have already conducted analyses in your watershed, the products of which may be useful to your assessment. Student theses and dissertations are another source of detailed information about a specific topic and may be found in campus libraries.

4.5.2.4 Local and Regional Agencies

Local agencies are sometimes the richest sources of local data, though not always in the format you want. Only recently has GIS become common among county, municipal, and district agencies. The data are not always free (e.g., if the local agency has bought it from a data provider) and may have less predictable barriers for access than you might encounter with state or federal agencies. However, these data can sometimes be of finer resolution and more current than larger-scale efforts by state or federal agencies. In this case, it is important to check the standards used to generate and update data so that even if these standards don't meet state or federal standards, you know what they are.

4.5.2.5 Private Sources

Private companies, utilities, or nonprofit organizations in your watershed may collect and maintain their own databases about their holdings, the region, the county, or the watershed. These data may be easy to access (e.g., from a nonprofit organization) or nearly impossible (e.g., from a large landowning company), depending on the data owner's level of trust as to how the data will be used and what the owner may lose from certain uses. The watershed assessor should work with the stakeholder watershed group (if one is present) to access certain privately held data that are deemed critical or important to answering questions about watershed processes or disturbance. Data sharing may improve if the data are not associated directly with landowner names. Even if you think getting the data is a long-shot, it can't hurt to ask.

4.6 Geographic Information Systems and Spatial Data

The term "GIS" (geographic information system) gets used a lot in the watershed world. To some it means a single digital map; to others it refers to a series of maps on a computer and includes analysis of spatial data. The spatial data often originates from remote sensing of the earth, from digitization of features from paper maps, or from using global positioning system (GPS) units to geo-reference points or lines on the ground. The history of GIS includes people taking pictures of battlefields from balloons (remote sensing), putting pins in maps for the locations of features (geo-referencing), and the development of rapid automated calculations with computers. Computer-operated GIS was created when these capacities were refined and paper maps could no longer capture processes on earth. If you are responsible for conducting a watershed assessment that involves GIS, you should become familiar with the terms and system descriptions below in order to understand the opportunities and limitations of this approach.

4.6.1 Definitions

GIS One older definition of GIS is "*a system of hardware, software, data, people, organizations, and institutional arrangements for collecting, storing, analyzing, and disseminating information about areas of the earth*". (Dueker & Kjerne 1989, from Lillesand & Kiefer, 1999). Another is "*a GIS combines layers of information about a place to give you a better understanding of that place*" (http://www.GIS.com, a project of Environmental Systems Research Institute). These two definitions capture the uses of the term GIS in this Manual. One important point is that a GIS is not simply a repository of maps or a software package for making and printing maps, which might be better termed "digital cartography." It is an analytical environment, populated by data that you provide, data that are spatially referenced to a place on the earth (ESRI, Inc. 1995). A GIS for your watershed is something you can use to find out things that are otherwise difficult or laborious to do manually (e.g., which roads lie on steep slopes).

Digital map An electronic version of a paper map and shows visually the relative positioning of mapped features (e.g., roads) in a place (e.g., watershed).

Spatial data Data about a space, such as distribution and kinds of vegetation growing in a watershed, that can be displayed in a digital map.

Remote sensing The "science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area, or phenomenon under investigation" (Lillesand & Kiefer, 1999). Satellite imagery is an obvious example of remote sensing information. Remote sensing relies on the detection of electromagnetic radiation (e.g., visible light and radar) that is either originating or reflecting from surfaces (e.g., leaves on trees in the Amazon). Properties and changes in the properties of the electromagnetic radiation can be used as surrogates for measuring properties of the thing under investigation (e.g., plant type).

Digitizing The process of making a paper map usable in an electronic GIS by manually drawing and transferring map features into an electronic form using mapping software. This process attributes real-world coordinates to the points or lines on a map, which are stored as an electronic data table, which, in turn, can then be visualized as a digital map. Digitizing could take place on the computer screen, where a drawing is created on top of a base map of some kind, such as an aerial photograph of a place, or the digitizing could be done by tracing the points and lines on a map using a desktop digitizing tablet.

GPS A global positioning system is a satellite-based system that allows someone holding a GPS receiver unit to determine latitude and longitude of their location on the earth's surface. The unit receives distance information from at least four satellites, which allows it to determine its position on the globe. Positional data from the unit allows the user to "GPS" the location of features of interest on the ground (e.g., landslide) for later entry into a GIS.

Geo-reference Information referenced to the earth, or "geo-referenced" by relating points, such as the corners of a map, to a coordinate system for the earth's surface, such as latitude and longitude. Points are thus associated with "coordinate pairs" or a value for each of the north-south and east-west axes (e.g., longitude and latitude). Once information about a place is geo-referenced, it can be overlaid by other maps, enabling analysis to be done that relates to a specific location or area.

4.6.2. Using Spatial Data

Non-digital spatial data

Many maps available to an assessment team will not be in digital form. Some of the most useful maps of historical activities or conditions may only be in paper form. These can be valuable resources even if they are never digitized and used in a GIS. Useful maps will relate to common location schemes, such as the quadrangle, township, range, and section maps of the USGS, or latitude and longitude coordinates. In order to analyze data contained on these maps, you may need to transfer some of it to digital form (e.g., a table of logging activity by sub-watershed), or to acetate overlays, a cheap and non-digital way to create a GIS. What such a system loses in precision (lost during copying) and accuracy (actual representation of ground features) may be made up for in speed and cost of acquisition. However, just as with the most expensive GIS, the quality of the analysis will be determined by the quality of the data and process of assessment. It is important to record actions and results so that others may easily tell how something was done and why a certain result was obtained.

Digital spatial data

Once information is entered into a computer database with spatial descriptors, it becomes digital spatial data and can be used in a computer GIS. This provides many opportunities both to simplify your assessment and to make it more complicated. Analyzing sets of data in various combinations from visual overlays to calculated or modeled relationships becomes somewhat easier. Because different types of information about a place are being used in such a way that the location is the commonality, you may discover relationships and qualities about a place that otherwise weren't obvious. For example, if you had a database of street addresses for parcels within your watershed and a map of the streams, and you wanted to find all of the riparian owners, you could use your GIS to combine these two sets of digital data that have a common spatial referencing scheme (e.g., latitude and longitude) and find the landowner addresses within a certain distance from a stream.

Development of new spatial data should be done in accordance with Federal Geographic Data Committee (FGDC; http://www.fgdc.gov) guidelines. This committee is responsible for coming up with common standards for data development, storage, and description. If the guidelines are not used, then data sharing becomes limited, and the utility of the data declines.

Types of digital data

There are two main types of digital information: 1) "spatial," which tell you where something is, and 2) "descriptive," which tell you about the something. There are also two primary types of data used in

GIS: Figure 4.2) raster data, which refers to information that is distributed into "cells" arranged in a grid pattern across the earth's surface, and 2) vector data, which is information occurring in a series of coordinate points termed "point" (single coordinate pair), "line" (two coordinate pairs), and "polygon" (one coordinate pair per angle/corner) features.

Each data type allows for different types of data distribution and calculations. It is possible to convert between raster and vector forms, though some information may be reduced in value or resolution depending on the scales of the original data type and new data type.

4.6.3 Data Scale and Resolution

Not all spatial data have been collected at the same spatial resolution. In this case, "spatial resolution" refers to the ability of a sensor (e.g., a camera) to separate detail on the ground (Lillesand & Kiefer, 1999). If data for a map were collected by digitizing features from aerial or satellite photographs, then the spatial resolution of the recording device is important, since this determines the limits of the data's use. The resolution of particular films and digital recording devices, the height of the camera, and variables, such as atmospheric conditions, determine the actual resolution of the resulting photographs and derived maps. In addition, remotely collected data may be aggregated into blocks of multiple pixels of a similar land-cover type or that are dominated by one land-cover type. This process decreases the data's resolution and affects the data's subsequent use.
Scale refers to the relationship between distance on a map or photograph compared to the actual distance on the ground. An example of scale is "1:24,000", which means that one inch on a map is equivalent to 24,000 inches, or 2,000 feet, on the ground. If map data are derived from photographs, then the resolution of the photographs will determine the map resolution or scale. In general, the smaller the scale ratio, the higher the resolution, assuming similarities in the resolution of the recording devices. A map derived from a 1:5,000 aerial photograph may have higher resolution than one derived from a 1:500,000 satellite photograph. However, if the resolution of the satellite camera is 100 times that of the aerial camera, then the resulting data will have similar spatial resolution. In other words, the analyst can discriminate between objects just as well in the satellite photograph as in the aerial photograph.

As you collect data into a database and GIS, it is likely that they will vary in their scales and resolutions. You may have high-

The Federal Geographic Data Committee

"The FGDC develops geospatial data standards for the National Spatial Data Infrastructure only when there are no externally developed standards appropriate for Federal use. FGDC standards are developed in consultation and cooperation with State, local, and tribal governments, the private sector and academic community, and, to the extent feasible, the international community. FGDC standards are intended to be national in scope and go beyond individual agencies and the Federal government enterprise. Federal agencies are required to use FGDC standards.

State and local agencies are not required to use FGDC standards, but are encouraged to do so to promote data sharing between different levels of government."

The standards are online at: http://www.fgdc.gov/publications/documents/standards/endorsed.html

(FGDC Web site, 2003)

resolution aerial photographs (1:2,000) of your county and low-resolution maps of soil series or geological formations (1:250,000) that you want to combine in analyses (e.g., percent of housing development on erosive areas). If data sets have been collected at different scales, use caution in any analyses involving two or more with different scales. In an extreme case, 1:500,000 scale data should not be referenced to a 1:24,000 data set because this represents an artificial increase in accuracy. However, it is possible to surrender data accuracy and use high-resolution data (e.g., 1:24,000) at a reduced scale of 1:500,000 if that is the only way to perform an analysis.

4.6.4 Metadata

Metadata refers to the information describing the data. Various information about scale, how the data were created, how they are stored, purpose of the data, source of the data, originating agency/organization/individual, and updates should be included in the metadata. The FGDC has developed standards for documenting spatial data. Most state and federal agencies follow these or similar standards and formats when describing data. It is critical to standardize these metadata so that the data can be used beyond a single person and that limitations and opportunities for uses are understood. An example of this would be the Calwater 2.2 system.

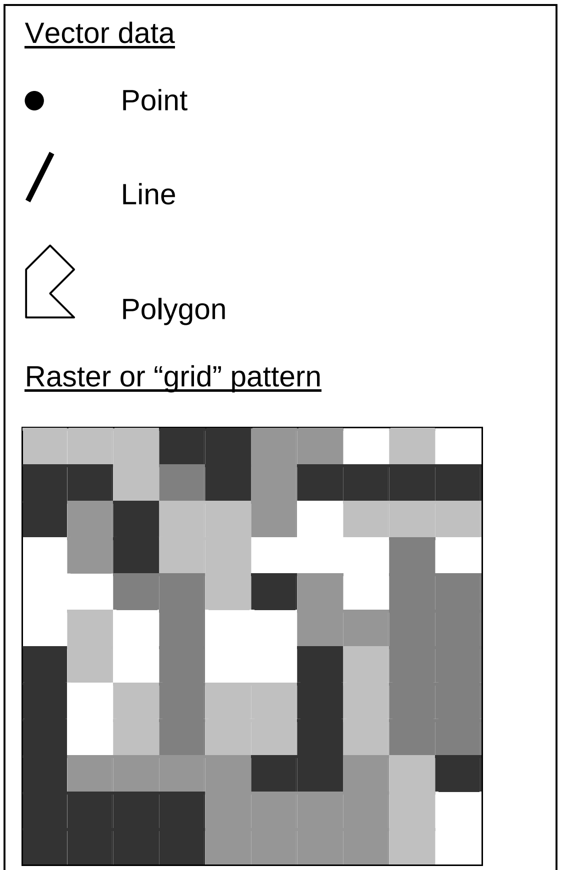

Figure 4.2 Illustration of the difference between vector and raster data

4.6.5 Developing your Watershed GIS

When developing a new GIS, a good starting point is to ask yourself two questions: "What questions do I intend to answer with this GIS?" and "How much do I have to spend?" Answers to the first question will tell you the scope of your GIS project and help inform the second question. Costs for a GIS can vary widely and this is where there may be the least amount of information for the watershed group to make fiscal decisions. For example, you may decide you want to collect digital spatial data and do simple analyses (e.g., where roads cross streams) with an emphasis on visual presentation of map information. In that case, your cheapest route is to use free GIS software on a donated computer, taking advantage of spatial data online and printing on a color inkjet printer. Being able to do this requires a basic education in GIS, which you can get cheaply or for free online. At the other end of the spectrum, you may want to spend $10,000 to $100,000 hiring staff or a consultant to do all of this for you on a purchased computer, using licensed software, and presenting your maps in large printed format and online using a map server. A likely outcome of hiring a consultant is that a GIS professional will do a good job more quickly than someone local having to learn GIS. At the same time, if

Questions for Your Watershed GIS

1) Location information (e.g., zip code, watershed, county)

2) Location of important features you plan to characterize (e.g., road-less areas, fragile soils, roads crossing streams)

3) Trends in a place over time

4) Spatial patterns and distributions (e.g., fish populations and water storage)

5) What would happen if you changed things on the landscape (e.g., added a road or other disturbance)

(ESRI, 1997)

GIS is likely to be part of your planning, monitoring, and management work for several years, it may make more sense to train a volunteer or staff person to carry out the GIS in order to increase local capacity.

4.6.6 Data Organization

Most people who have used GIS for a while are familiar with the problems of storing, sorting, and updating the files associated with a GIS. It is possible to accumulate thousands of files and take up gigabytes of disk space. The key to managing this is organization of your files in a structure. Once you have created a logical architecture for your GIS, organizing it should be relatively straightforward.

Filing systems

Just as with the filing system in your local library, the best GIS filing system is one that has an intuitively obvious structure and can be understood by users. There is no one perfect way to create such a system, but there are sets of rules that can be used to guide the construction.

Develop categories that match your expectations for the types of data you will be collecting, analyzing, and storing. The FGDC (http://fgdclearhs.er.usgs.gov/) uses the following categories: administrative and political boundaries; agriculture and farming, atmospheric and climatic data; base and scanned maps and charts; biologic and ecologic information; business and economic information; cadastral and legal land descriptions; earth surface characteristics and land cover; elevation and derived products; environmental monitoring and modeling; facilities, buildings and structures; geodetic networks and control points; geologic and geophysical information; human health and disease; imagery and aerial photographs; inland water resources and characteristics; ocean

Free GIS Software

"a complete index of Open Source / Free GIS-related software projects"
http://www.opensourcegis.org

"GRASS GIS (Geographic Resources Analysis Support System) is an open source, Free Software Geographical Information System"
http://grass.itc.it

"MapServer is an OpenSource development environment for building spatially enabled Internet applications."
http://mapserver.gis.umn.edu/index.html

"open-source GIS-based applications you can download written for a variety of platforms and in various languages"
http://gislounge.com/ll/opensource.shtml

"IDRISI, developed by Clark Labs, is an innovative and functional geographic modeling technology that enables and supports environmental decision making for the real world"
http://www.clarklabs.org/

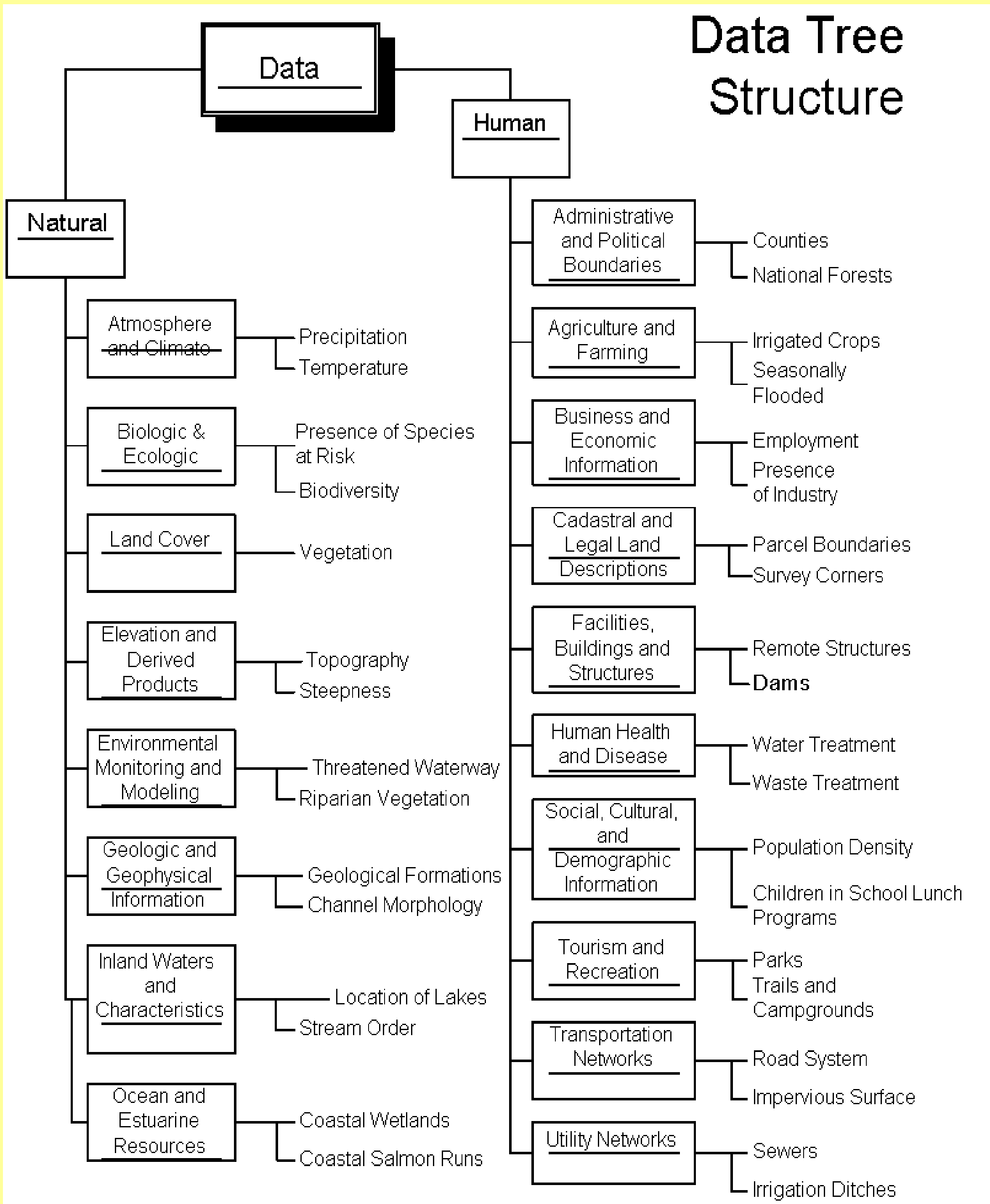

Figure 4.3 FGDC data structure

and estuarine resources and characteristics; society, cultural, and demographic information; tourism and recreation; transportation networks and models; and utility distribution networks (Figure 4.3).

If you don't have limitations on electronic storage space, there are several ways to deal with using and modifying digital maps after their initial acquisition or creation. One is to maintain the original versions of maps in separate backup directories so that you can always go back to the original. Later, or “final”, versions of maps that are the most recent can be retained in a unique directory. This will make updating maps easier as modified versions can replace the most recent version. Any modification should of course be accompanied by an update of the metadata describing the changes and giving the date.

Develop a structure rationale that fits people’s expectations for the system. The usual structure is a hierarchical, or tree structure. The diagram below shows an example of this type of structure, using FGDC categories and samples of data types you might have.

In conclusion, the nature of your watershed assessment should guide your collection and organization of the various kinds of data. The questions you ask in your assessment determine the kinds of data you need (e.g., water quality, mapped sources of pollution, projected human activities). The level of complexity and nature of the analyses determine the data types (e.g., digital vs. paper maps), which, in turn, determine your strategy for data organization.

4.7 References

Abell, D.L. (ed.) 1989. Proceedings of the California riparian systems conference: Protection, management, and restoration for the 1990s. General Technical Report PSW-110. Berkeley: U.S. Forest Service, Pacific Southwest Forest and Range Experiment Station.

Downie, Scott T., C.W. Davenport, E. Dudik, F. Yee, and J. Clements (2002) Mattole River Watershed Assessment Report. North Coast Watershed Assessment Program, p. 441 plus Appendices. California Resources Agency, and California Environmental Protection Agency, Sacramento, California.

Dueker, K.J. and D. Kjerne 1989. Multipurpose Cadastre Terms and Definitions. American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping. Falls Church, VA.

Dunne, T. and L.B. Leopold. 1978. Water in environmental planning. San Francisco: W.H. Freeman and Company.

ESRI, Inc. 1997. Understanding GIS – The ARC/INFO Method. 4th ed. Environmental Systems Research Institute Inc. Redlands, CA.

Grant, G. 1988. The RAPID technique: a new method for evaluating downstream effects of forest practices on riparian zones. General Technical Report PNW-220. Corvallis, OR: U.S. Forest Service, Pacific Northwest Forest and Range Experiment Station

Herschey, R.W. 1985. Streamflow measurement. London: Elsevier.

Hornberger, G.M., J.P. Raffensperger, P.L. Wiberg, and K.N. Eshleman. 1998. Elements of physical hydrology. Baltimore: The Johns Hopkins University Press

Jones & Stokes Associates, Inc. 1998a. Historical riparian habitat conditions of the San Joaquin River. San Joaquin River Riparian Habitat Restoration Program. Sacramento.

Jones & Stokes Associates, Inc. 1998b. Analysis of physical processes and riparian habitat potential of the San Joaquin River. San Joaquin River Riparian Habitat Restoration Program. Sacramento.

Lillesand, T.M. and R.W. Kiefer 1999. Remote Sensing and Image Interpretation. 4th ed., John Wiley & Sons, Inc., NY. Pp. 724.

Michael, G. 1991. Environmental data bases: design, implementation, and maintenance. Lewis Publishers. Boca Raton FL. 98 p.

Michener, W.K., J.W. Brunt, and S. Stafford (eds.). 1994. Environmental information management and analysis: Ecosystem to global scales. New York: Taylor and Francis.

National Research Council. 2002. Riparian areas: functions and strategies for management. National Academy Press. Wash. DC. 428 p. [//www.nap.edu]

Nelson, C.W. and J.R. Nelson. 1984. The Central Valley riparian mapping project. In California riparian systems: Ecology, conservation, and productive management, edited by R.E. Warner and K.M. Hendrix, 307-13. Berkeley: University of California Press.

Sands, A. (ed.) 1977. Riparian forests in California: Their ecology and management. Publication 15. Davis: University of California, Institute of Ecology.

Tetra Tech. 2003. The status and future of biological assessment for California streams,final report. Sacramento: California State Water Resources Control Board, Division of Water Quality.

U.S. Department of the Interior Bureau of Land Management (BLM). 1995. Riparian area management: process for assessing proper functioning condition. Tech. Ref. 1737-9, Denver CO. 60 p.

University of Washington. 2003. Stream riparian bibliography. Center for Water and Watershed Studies, with USFS Stream System Technology Center, http://riparian.cfr.washington.edu/

Warner, R.E. and K.M. Hendrix. 1984. California riparian systems: Ecology, conservation, and productive management. Berkeley: University of California Press

Yoshiyama, R., E. Gerstung, F. Fisher, and P.B. Moyle. 2001. "Historical and present distribution of chinook salmon in the Central Valley." CDFG Fish Bulletin 179.

Yoshiyama, R.E. and P.B. Moyle. 1996 in Sierra Nevada Ecosystem Project: Final Report to Congress, Vol. III. University of CA, Davis, Centers for Water and Wildland Resources