3. Spatial data for fisheries and aquaculture: characteristics, quality and data sources

G.J. Meaden (FAO consultant, Canterbury, United Kingdom)

3.1 INTRODUCTION

This chapter first looks at the types of data needed for fisheries or aquaculture GIS work; it then describes the main characteristics of geographic data and how they are recorded in map form. An examination is also made of issues surrounding data quality and data standards, and the chapter concludes with the important subject of data sources and data collection. Aspects relating to the preparation of data and mapping for GIS use are examined in Chapter 5. For any GIS work, the quality of the data that can be acquired (captured) will have a major bearing on the success of the project, and the main aims of this chapter are to make certain that the GIS user fully understands data requirements in order that success can be achieved in terms of having the appropriate inputs for any GIS project.

Here, data are defined as "purposeful observations that have been recorded and stored", and data differ from information in that the latter represents data (or other material) that have usually been summarized, organized and processed.[65] As mentioned in Chapter 1, data now represent the largest cost element in a GIS project, where the often repeated figure of 80 percent of project costs is given (Clark, 1993; Bernhardsen, 2002). Data for use in fisheries-related work (especially marine fisheries) are frequently more expensive than data used for terrestrial projects because there may be major expenses in gathering data at sea (e.g. vessel purchase or hire, crew costs) and additional costs related to the 3D or 4D environments (vertical and horizontal data needs). There is also the sheer size of marine areas over which data may need to be gathered. Additionally, marine or river environments are mobile in terms of both the water and the objects in them, and this may create a need for more frequent data collection.

The format in which data has been traditionally obtained has undergone a rapid transition during the last two decades. Until the end of the 1980s, virtually all data were procured in a hard copy format, usually paper. This consisted of data represented by maps, tabular and graphical data, and numerous data obtained from a range of documents and reports. For much of the world, the move to digital information systems has entirely reversed this situation, so nowadays it is rare to obtain data that are not in a digital format. However, in this technical paper hard copy is discussed because some data are still collected and stored in this format in many developing countries and thus there is a legacy of existing data that may need to be converted to a digital format.

3.2 EVALUATING DATA NEEDS FOR FISHERIES OR FOR AQUACULTURE GIS PROJECTS

In Section 1.2 some background was provided on the use of GIS and it was made clear that GIS is a tool for allowing the various spatial aspects of a project to be combined for a variety of analytical purposes. The GIS project itself might be ameans of ensuring success for a proposed future activity, or it might be used to improve

[65] Readers who require knowledge on marine or fisheries information and information systems (rather than data) should peruse any Internet search engine under "fishery" or "marine" information systems.

the management of existing capture fisheries or aquaculture in a particular area, or fishery, or with a particular species. Essentially, GIS is being used as a tool to aid the planning, managing or monitoring of fisheries or aquaculture in order to bring about improvement and success. It is important to keep in mind that "success" for any production activity ideally needs to be measured in terms of four criteria:

- Economic. The activity might aim to minimize costs or to maximize profits.[66]
- Social. The activity should aim to best foster the social ideals of those participants in the production activity. This is best thought of in terms of successful social well-being and cohesion, including health, education and job satisfaction.
- Physical/biological. The activity needs to consider production in terms of successfully meeting physical and biological input requirements to the production process.
- Sustainability. The production activity needs to function successfully in combination with other activities for the indefinite future.

It is apparent that there is likely to be a clash in perceptions of "success" between these four criteria. Thus, profit maximization as a goal for success is quite likely to be incompatible with the goal of sustainability, at least in the short term. It is of interest to note that in Meaden and Kapetsky (1991), a forerunner of this present publication, it was noted "we must take profit maximization as our production rationale since failure to aim for this objective will mean that there is no point in seeking suitable locations for aquaculture or inland fisheries, i.e. with enough capital it is possible, through the manipulation of production functions, to produce almost anything anywhere." (p. 7). However, in 1991 sustainability was not a concept underlying much production activity, but since then there has been a flowering of sustainability ideals in many facets of productive life. With the demise of fisheries, as outlined in Chapter 1, it is now recognized that the only way in which the activity can have an assured future is to place sustainability as the core rationale (or modus operandi) for all fisheries-related activities. Profit maximization as the rationale for fishery production is, therefore, being replaced by concepts of sustainability, and it is hoped that sustainability will be achieved through the ecosystem approach to fisheries (EAF) (FAO Fisheries Department, 2003) or the ecosystem approach to aquaculture (EAA) (Soto, Aguilar-Manjarrez and Hishamunda, 2008; FAO, 2010b).

In terms of data needed for any GIS-based project on capture fisheries or aquaculture, what is required are data on how to ensure that any production activity thrives bearing in mind four criteria for success (economic, social, physical/biological and sustainability), or, in other words, bearing in mind the EAF or EAA approaches. In practical terms, for any capture fishery or aquaculture GIS project, it is likely that the people identifying and organizing the project will need to convene meetings with relevant "participants" to identify a number of project parameters.[67] For instance, initially, careful attention should be given to the extent and boundaries of the geographic (spatial) area[68] and the scale (resolution or detail) at which the study is carried out (scale and resolution are examined in detail in Section 3.3.5). Box 3.1 gives some recent suggestions on spatial area and scale for potential GIS applications in aquaculture.[69] It will also be essential to decide on project partners and to decide on work partitioning. It will then be necessary to identify what the spatially based production functions are which in combination can best lead to success. Here, "production functions" are defined as those factors or variables that, in various combinations, influence the success

[66] Whether cost minimization or profit maximization is aimed for may depend upon the nature of the political economy of the country, society or the individuals who are organizing the production activity.

[67] "Participants" here might include local or national experts in the field of concern or perhaps a focus group of local fishers or aquaculturists.

[68] In EAA or EAF terminology, the first phase of the ecosystem approach is to "define ecosystem boundaries".

[69] It is also important to note that the data needs of any project might be dependent on the issue being addressed, e.g. assessment of aquaculture potential, zoning and site selection.

BOX 3.1

Scales and levels of analysis for potential GIS applications to aquaculture

When evaluating data needs for fisheries and aquaculture, an important first step is to define the ecosystem boundaries. In this regard, important considerations are (i) the scales and (ii) the levels of analysis.

Scales

Experts at the FAO workshop on "Building an ecosystem approach to aquaculture (EAA): initial steps for guidelines" (Soto, Aguilar-Manjarrez and Hishamunda, 2008) identified four scales and/or levels of EAA application: the farm; the waterbody and its watershed; the aquaculture zone or region; and the global, market-trade level.

Kapetsky, Aguilar-Manjarrez and Soto (2010) demonstrated that EAA scales are easily accommodated by GIS, remote sensing and mapping as applied to aquaculture because GIS is capable of being applied at any scale. Practically, many spatial applications in aquaculture primarily deal with a natural or an artificial waterbody in its entirety or in part. Otherwise, the geographic extent of many applications is often defined by some extranational, national or subnational level of administration, or national level of administration, or subnational clusters of administrations (see the following "scale definitions"). The scale chosen will dictate data needs and will have a direct influence on the selection of an appropriate methodology and on the level of effort and costs for data purchasing and processing.

Scale definitions

Scale	Scale description
Local	Generally, a natural or artificial ecosystem or a third-level administrative area
State or province	The second level of administration below national
Region within a country	Generally, an area occupying an appreciable part of a country and/or including more than one state or province
National	An application covering the entire country
Region among countries	Covering two or more countries
Continental	Covering all of the countries of a continent
Global	Including all countries with aquaculture

Levels of analysis

Here, the consideration is with differing degrees of assessments that dictate a number of characteristics of data needs. Common levels of analysis from an aquaculture viewpoint are: assessments of the potential for aquaculture; local zoning of aquaculture; siting of aquaculture; and monitoring of aquaculture operations. Assessments of the potential deal with data at relatively low resolution using a few key data sets to generate indicative results, while siting and monitoring require more detailed data at higher resolutions and are aimed at producing more accurate results.

An important point is that there are no hard and fast rules on the numbers and kinds of data at any scale or level of analysis; one uses as much as one can to achieve the desired objective.

Source: Kapetsky, Aguilar-Manjarrez and Soto (2010).

of the production activity.[70] Section 1.4 gave clues to the likely production functions that controlled aquaculture success in the Democratic Socialist Republic of Sri Lanka. These will obviously be production functions that relate to the specific functioning of the fishery or aquaculture activity in a particular area and at a particular scale – no two areas are likely to have the same set of production controls. Some of the functions may be "one-off" considerations that only need apply when, for instance, a site is

[70] It is also important to note that the data needs of any project might be dependent on the issue being addressed, e.g. assessment of aquaculture potential, zoning and site selection.

being acquired for aquaculture facilities. This would be synonymous with fixed costs such as land purchase. Other functions will be akin to operating costs as their consideration may be continual, e.g. access to markets. Production functions may be highly variable within an area or region (e.g. soil suitability, bottom sediment types or population density), while others will be much more uniform in their distribution (e.g. water temperature, salinity and some climatic factors). Increasingly, functions are able to be manipulated in the sense that deficiencies can be overcome. For instance, if there are no electricity supplies nearby, then a mobile generator may be employed, or the deficiencies in oxygen supplies to water can be redressed through artificial oxygenation. However, overcoming function deficiencies can add significantly to costs.

The various production functions will each have their own degree of importance and, if only fairly crudely, it is essential that these can be ranked or measured for specific locations[71]. Practising fishers or aquaculturists are often the best people to do this. Table 3.1 shows the production functions that pertain to cage culture in the Mediterranean Sea, the Kingdom of Spain (Forget, Stuart and Platt, 2009), and it can be seen that both objective and relative suitability are indicated. The importance of measuring the relative importance of production functions relates to the fact that an "importance score" (or "weighting") can be incorporated into GIS-based scoring systems that might be utilized in the search for optimum production locations. Box 3.2 outlines the main production functions that apply to most inland aquaculture. As well as the functions listed, there may be others that apply to individual locations and functions that are less important.

TABLE 3.1
Production functions for fish cage site selection and environmental monitoring in the Mediterranean Sea, Spain

Production function	Good	Medium	Bad
Coastal exposition	Partial	Sheltered	Non-sheltered
Wave height (m)	1 to 3	<1	>3
Water depth (m)	>30	15–30	<15
Water current speed (cm/s)	>15	5–15	<5
Pollution level	Low	Medium	High
Max. temperature (°C)	22–24	24–27	>27
Min. temperature (°C)	12	10	<8
Salinity (average)(L')	25–27	15–25	<15
Salinity fluctuations (L')	<5	5–10	>10
Dissolved oxygen (%)	100	70–100	<70
Turbidity/suspended solids	Low	Moderate	High
Sediment type	Sand/gravel	Mixture	Mud
Water classification	Oligotrophic	Mesotrophic	Eutrophic
Fouling	Low	Moderate	High
Predators	No	Few	Abundant

Source: IOCCG (2009).

[71] It should be mentioned that it is often impossible to obtain data on all production functions, but data on at least the main functions should be included in the GIS project work.

BOX 3.2
Main production functions applying to inland aquaculture

- **Land availability** – Usually refers to so-called "greenfield" sites that might be available for development.
- **Topography** – Flat sites are far cheaper to develop, though they can be liable to flooding.
- **Water temperature** – Most species are only tolerant to fairly narrow ranges of temperature.
- **Water quality** – Salmonids are far less tolerant of poor-quality water than, for instance, cyprinids.
- **Water quantity** – Most culture systems require a turnover of water to ensure oxygen availability and to dilute certain chemical parameters.
- **Water access rights** – Access rights to, and availability of, sufficient water is becoming an increasingly important spatial consideration. Aquaculture may be competing with other water users.
- **Soil chemistry and structure** – Many species cannot, for instance, tolerate high or low pH values, and porous soils may be unsuitable for pond construction and for water retention.
- **Adjacent land uses** – Certain neighbouring land uses may be adverse for aquaculture facilities.
- **Proximity to supporting infrastructure** – It is advantageous to be near to essential suppliers or sources of advice, e.g. feed suppliers and veterinarians.
- **Access to roads** – Most inputs and outputs must be brought in or out by road transport.
- **Climatic disturbance** – Care and consideration must be taken in areas subject to tropical cyclones, hurricanes or other severe weather factors.
- **Population density** – People living nearby can provide a workforce and market outlets, but they also provide a source of disturbance or poaching.
- **Distance from other fish farms** – This can be positive (for sharing resources), but negative from a disease-spreading perspective.
- **Access to electricity** – Electrical supplies may be vital to certain production activities.
- **Environmental constraints** – Increasingly, environmental constraints to aquaculture developments are being applied, e.g. mangrove clearance for shrimp ponds has mostly been prohibited.
- **Access to main markets** – It will be important to establish the main markets for output initially, though these might vary over time. They could be wholesale or customer direct.
- **Availability of fertilizer or agricultural by-products** – These are often used to obtain maximum productivity in freshwater ponds.

Source: Adapted and updated from Meaden and Kapetsky (1991).

The spatially variable production functions that influence fisheries are far more varied than those that exert a control on aquaculture, and they vary enormously from fishery to fishery according to, for example, whether the activity takes place in a developed or less-developed country; the species being targeted; and the size of the fishing vessels being used. An additional factor adding to a complex situation is that, whereas most GIS used for aquaculture projects are for optimizing site selection or for monitoring the aquaculture environment (see Table 8.1), in capture fisheries the uses to which GIS are put are far more varied (see Table 10.1). There is, therefore, a far wider variety of production functions about which data might need to be collected. Box 3.3 gives some pointers to the range of spatially variable production functions that may influence the relative success of marine fisheries. For inland fisheries, the production function list would be similar except that some of the water quality factors would vary, the functions would generally be much more crucial to fishery success, and there are likely to be seasonal variations, e.g. in water quantity. The functions shown in Box 3.3 have been ordered to show a progression from natural physical functions, through biological functions, and on to economic and then social functions.

BOX 3.3

Some important spatially variable production functions influencing marine fisheries

Only a limited range of the more important functions can be illustrated here. The functions controlling fisheries on larger inland lakes will be similar to those influencing marine fisheries.

- Bottom sediments – The distribution of many demersal and benthic species is defined by bottom sediment types. This will affect the types of fishing gear used.
- Bathymetry – Different species are physiologically adapted to live at different water depths.
- Salinity – Different species are also adapted to different salinity levels, though some living in tidal zones can tolerate wide variations.
- Chlorophyll – The abundance of these algae can be a good indicator of water productivity.
- Bed shear stress – This indicates the current speeds on the seabed and it strongly influences marine benthic assemblages.
- Water temperature – Again, physiological adaptations are made to water temperature. Seasonal variations can encourage migrations.
- Thermal fronts – These occur where bodies of relatively warm water meet colder water. High productivity along these fronts encourages feeding by pelagic species.
- Species distributions – These are highly variable both spatially and temporally, and will affect the distribution of fishing effort.
- Nursery or spawning grounds – Once identified, it is often essential to give some protection to these areas.
- Marine vegetation – Kelp forests, seagrass beds, coastal mangroves, etc., may all offer unique and important habitats.
- Migration routes – Species at higher trophic levels often make major migrations in response to water temperatures, spawning and/or food needs.
- Predators – Some areas have significant numbers of bird, fish or mammal natural predators, e.g. seals and marine sea birds.
- Distance from port – Fishers will try to minimize fuel costs by fishing as close as possible to home ports. Some ports act as major markets.
- Fishing systems – These may vary greatly at macro- and microscales.
- Fishing effort – The amount of fishing taking place in a given area over a given time period. It may be measured in fishing days at sea, or engine capacity, etc.
- Catch distribution – This is typically measured by species per unit area for a temporal period.
- Fish values – Prices for different catch species may vary significantly and this can affect the fish species targeted.
- Prevailing regulations – Most marine fisheries operate within defined areas that have prevailing regulations that may significantly affect fishery activities.
- Conservation areas – Various by-laws will operate in these areas that will usually restrict fishing effort and catches.

Box 3.4 examines production functions in marine cage culture. As this is a rather specialized activity, it is restricted to a fairly narrow range of potential sites, although in total a considerable marine area has the potential for cages to be sited, especially if consideration is given to aquaculture using submerged sea cages. The functions controlling successful production are much more restricted for this activity. It should be remembered that production functions will also control a range of other aquatic culturing activities, most of which are carried out on a very small-scale in highly selected locations, for example, mussel culture, seaweed culture, and culturing of sponges, sea cucumbers, crabs and lobsters.

BOX 3.4
The main spatially variable production functions influencing marine cage culture

Here the aim is to consider off-the-coast or offshore mariculture.

- Distance from shore (ports) – This is important in respect to frequent observation, feeding, stocking and harvesting activities.
- Water depth – Sea cages must be tethered and deep water presents a challenge.
- Water temperature – Species will have developed preferred temperatures and temperature tolerance ranges.
- Availability of shelter – Cages in open waters are vulnerable to storrn conditions that can cause cages to break free, or break up with subsequent stock losses.
- Distance from competing water activities – It is essential to avoid siting cages in busy sea areas, or areas liable to pollution.
- Water quality – Many near-coastal sites may suffer from various forms or sources of pollution, e.g. oil leakages, sewage outfalls, or sources of disease. Dissolved oxygen levels are also important.
- Turbidity and suspended solids – Some species have preferences for clearer waters.
- Interactions of farm sites with immediate environments – Cage sites need to aware of local biodiversity, waste deposition and benthos issues.
- Distance from other cage farms – Because of disease problems, cages are preferably located in relatively isolated and well-dispersed locations.
- Prevailing wave heights – Where long fetches prevail (usually around open oceans), prevailing waves may be too high for conventional cages, though completely submerged cages may be possible.
- Availability of inputs – The location of marine cages should be chosen in respect to important inputs such as extension services, veterinarians and feed suppliers.
- Predators – In some areas, predation from cages is a problem.
- Visual impacts – Cage locations should not be visually intrusive.

This section has discussed in some detail the spatially variable production functions determining the success of fisheries or aquaculture as productive activities, and it is clear that it will be the spatial data that can best describe the disposition of these functions that will be needed as source data for any GIS work. It will be the job of the project personnel (e.g. IT staff, aquaculturists, fishery experts) to decide on what production functions are essential to achieving success for the fishery or aquaculture planning, monitoring or management project. This is an extremely important task in GIS project planning, as the nature of the data to be gathered will depend on it.

3.3 SOME CHARACTERISTICS OF DATA FOR GIS USE

For GIS purposes, all data must minimally have five facets: (i) be temporal (indicate a date); (ii) be thematic (relate to a theme or an attribute); (iii) have a quantity (a number of data objects or events); (iv) be spatial (have a georeference or location); and (v) it will be important that scale and resolution are relevant to the tasks being undertaken. There are other facets relating to data captured for GIS purposes, but these will not be essential requirements with respect to data collection.[72] There is also the important GIS consideration of how data are mapped, but this will be looked at in Chapters 5 and 7. Each of the main facets of data can now be briefly examined.

[72] Examples of other data facets include accuracy, temporary or permanent, data classes – some of these facets are examined later.

3.3.1 The temporal facet

The first facet, temporal, needs little explanation. Knowing when data were collected is both important and useful. Data gradually age and knowing their age is highly relevant in both an absolute and a sequential or relative sense. Because different natural or human processes operate at hugely varying temporal rates, the user may wish to allocate differing time spans over which different data sets retain their usefulness, though the user may constantly face a dilemma over the costs versus benefits of when data sets should be updated. The great majority of data can be used for time series analyses, which themselves point to rates of change or temporal trends within or between thematic areas.

3.3.2 The thematic facet

The second facet of data, thematic, refers to what it is that the data are concerned with. This may be in the form of a thematic area, subject, object or a process and/or event. Themes may be perceived in a wide, general sense (e.g. "marine", "fisheries" or "aquaculture"), or they may be much more specific (e.g. "freshwater fish of the Federal Republic of Nigeria"). There can also be a hierarchy of data within a theme, for example:

- marine fauna
- fish
- demersal fish
- plaice
- female plaice
- juvenile female plaice

Any hierarchy may be important when it comes to setting up a database of associated objects (see Section 5.4). As well as being hierarchical, themes can be simply subdivided into a range of different classes (or classifications). Data can be collected on areas, themes, objects or process and/or events at the same time. For instance, in the sphere of fisheries (theme), data may be collected on the number of salmon (object) in British Columbia, Canada (area) passing a counting station (event), or the number of times a fisher (object) casts a net (event or process) per day. For all thematic data, the main subclassifications (or categories) are referred to in GIS as attributes. So the salmon could be juvenile or mature; large, medium or small; coho, pink, chinook, chum, or sockeye, etc.; and the net being cast could have various dimensions, various mesh sizes, be made of various materials, etc., or it could be undergoing various activities, such as mending, storing, repairing and changing. Although it might be useful with respect to sourcing data, it is beyond the remit of this technical paper to explore all the possible themes that data being used for fisheries or aquaculture work might take. It is suggested that the use of an Internet search engine will be the most convenient means of locating whether data on specific thematic areas exist or not.

3.3.3 The quantitative facet

For the third facet of data, all thematic data (about objects or events) can be quantitative as well as descriptive. This simply means that they can be counted or measured. It is essential to know that counting or measuring can be carried out at various levels, as explained in Box 3.5. It is also important to mention here that quantitative factors relative to data should include a consideration of precision or accuracy – these are discussed in Section 3.4.

BOX 3.5
Levels of measurement that can be assigned as an attribute of collected data

On most occasions, data collection will involve counting or measuring (i.e. quantitative data will be recorded), and these numerical data become an attribute of the objects or events being sampled or surveyed. There are basically four "levels of measurement", as follows:

- **Nominal** – This really only signifies the name given to an object or to a class of objects. It can, however, include numbers, e.g. a fishing vessel registration number. Arithmetic operations cannot be applied to these numbers.
- **Ordinal** – This measurement characterizes objects or events by their rank or order in a sequence. For instance, the suitability of soils for the construction of aquaculture ponds in any area could be ranked from best to worst. Although arithmetic operations cannot be performed on ordinal data, it is possible to establish the median and/or the rank.
- **Interval** – This measurement derives a scale along which readings relative to the object can be placed. Data on water temperatures are measured on an interval scale. But note that there may be no absolute zero on this level of measurement (though there may be an arbitrary zero), and that, for instance, a temperature of 20 °C is not twice as hot as 10 °C. A wider range of arithmetic operations can be applied to interval data.
- **Ratio** – This measure is where an absolute zero is possible, such as the number of fish in a particular aquatic area. It is called "ratio" because it is possible to say that 50 fish is exactly twice as many as 25 fish. Ratio data are the most useful for statistical work.

Source: Adapted from DeMers (2009b).

3.3.4. The spatial facet

The fourth facet of data, i.e. spatial, is a record of where an object is located on the planet or where an event took place, and from a GIS perspective this is arguably the most important facet of data.[73] The spatial facet of data refers to a location that may be specific, for example, a point, or it may be more widespread, for example, an area such as a county. Points are frequently located in terms of a georeference,[74] and there are various methods of georeferencing. For spatially locating exact points on a map, it is common either to record the latitude and longitude for that point or to use a Cartesian coordinate system whereby a grid reference is assigned that is based on an artificial placement of parallel "x" and "y" grid lines across a flat map. Figure 3.1 shows the familiar earth coordinate system provided by lines of latitude and longitude. Lines of latitudes are measured as being from 0 degrees at the equator to 90°N or –90°S at the North and South Poles, respectively. Lines of longitude are measured from the 0° line (meridian), which passes through Greenwich, London (the Prime Meridian), and are recorded in degrees from 0 to 180°E or from 0 to –180°W. For greater precision, both latitudes and longitudes can be divided into degrees, minutes and seconds. Thus, a georeference using latitude and longitude might state that the Statue of Liberty near New York is located at 40° 68' 92"N and 74° 04' 45"W. Computers usually store latitude and longitude data in a decimal format.[75]

---

[73] All facets of data are important, but it is the spatial facet that distinguishes data as being suitable to be employed in a GIS. The other four facets of data mean that the data could be employed for other purposes but not by themselves for GIS analyses.

[74] A georeference is usually an alphanumeric means of giving a location on a map.

[75] For details on converting degrees, minutes and seconds to decimal degrees, see de Graaf *et al.* (2003) (available at www.fao.org/DOCREP/006/Y4816E/y4816e0e.htm#bm14). For conversion tools, see GPS Visualizer "calculator" (www.gpsvisualizer.com/calculators).

FIGURE 3.1
Illustration of lines of latitude and longitude

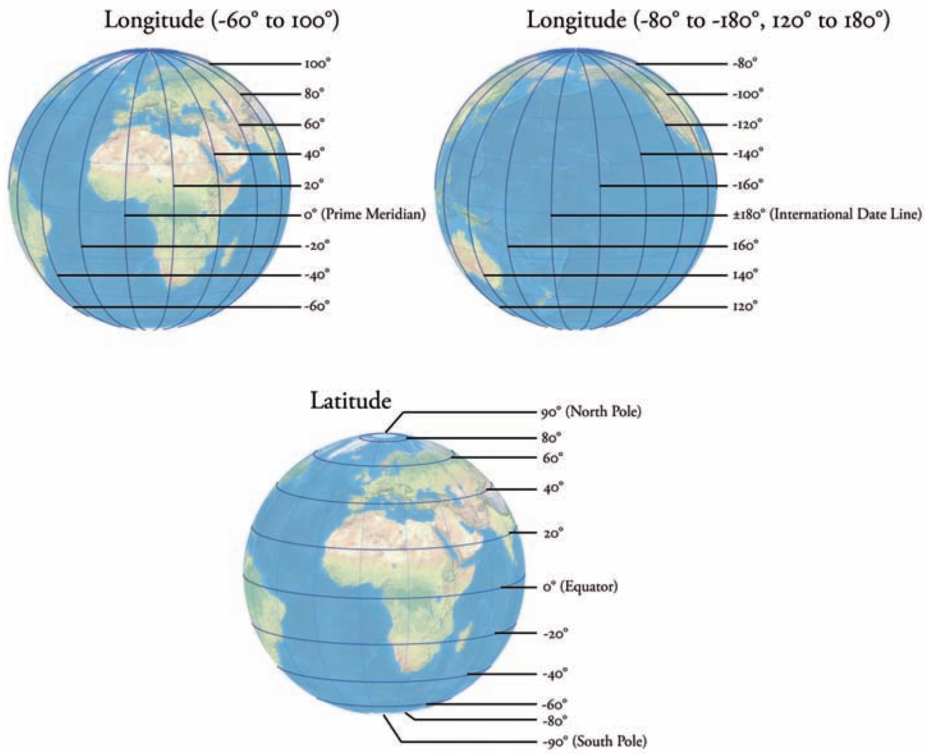


Figure 3.2 shows an example of the use of latitude and longitude designations on a map illustrating a newly designated marine protected area in eastern Canada (Canada Gazette, 2005). Here, the map border uses latitude and longitude designations in degrees, minutes and seconds (to the nearest one minute), though in the legend individual points on the map are shown as latitude north and longitude west to the nearest degree, minute and second.

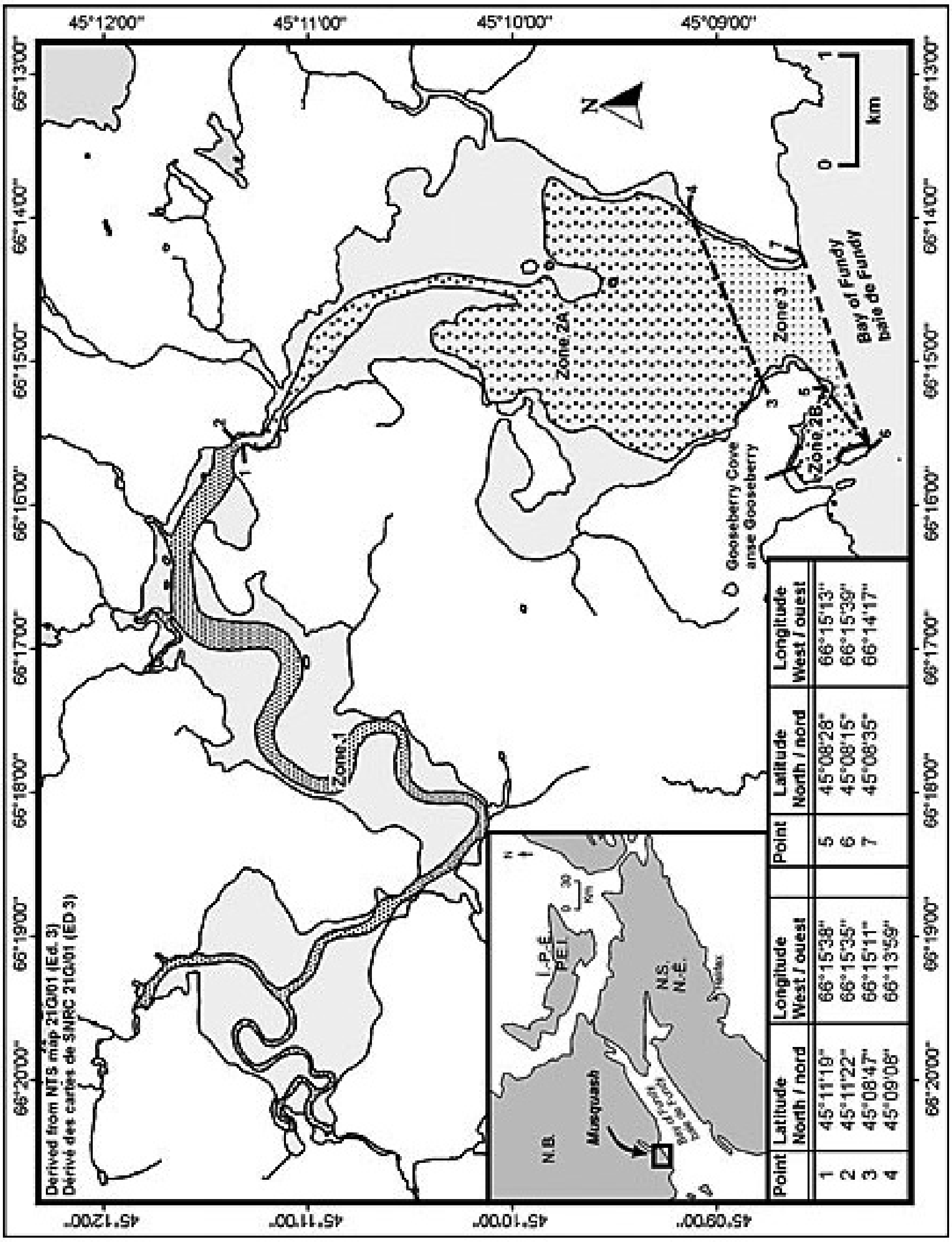
Although exact latitude and longitude georeferences are readily obtained, e.g. from GPS, and they are useful as points from which to measure exact distances or directions across a curved surface, they are not always used in GIS work. This is because many technologies and mediums for working with geographic data are inherently flat, e.g. paper, square grids, and because the grids (graticule) produced by latitudes and longitudes are non-standard in size. Because people are used to viewing maps, plans, atlases, etc., displayed as flat surfaces, it is necessary to transcribe the ellipsoid shape[76] of the earth onto a flat surface. This requires the use of a so-called "map projection", which is an attempt to convert a spherical shape into a flat surface. There are many different map projections,[77] but none of them can give a true representation of the actual surface of the earth. Different projections have advantages or disadvantages in terms of how accurately they portray mapped shapes, areas, directions or distances, and Table 3.2 gives the best projection for each. For a single GIS project, it is important that all the data sets used conform to the same projection, but most GIS software have the ability to accept a wide range of projections and projections can be readily changed when required. Detailed information on map projections can be obtained from Robinson *et al.* (1995) or Harvey (2008).

[76] An ellipsoid resembles a flattened sphere. The earth is slightly flattened, such that the diameter across the equatorial axis is 23 km more than the diameter across the polar axis. The ellipsoid itself is not smooth having various bumps and hollows that depart from the smoothed average surface by up to 60 m, and on top of this irregular ellipsoid mountains or valleys may be imposed.

[77] For a range of examples, see Robinson *et al.* (1995).

FIGURE 3.2

The Musquash marine protected area in New Brunswick, Canada



Source: Department of Justice, Canada (2012).

TABLE 3.2
Suitable map projections for highlighting different map attributes

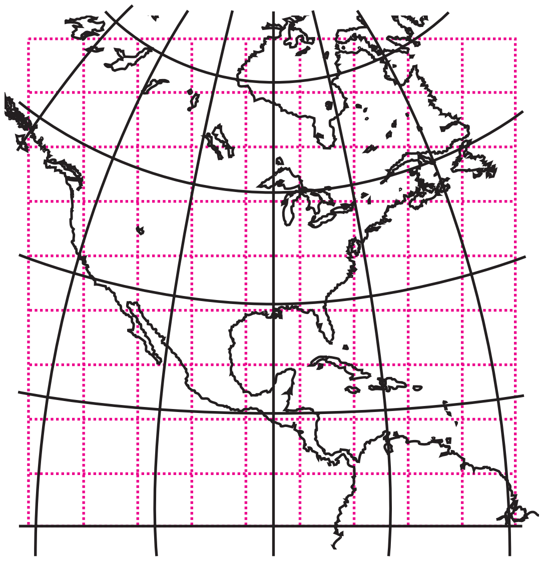
Attribute preserved	Projection	Applications	Examples
Area	Equal area	Many thematic maps use an equal area projection. Maps of the United States of America commonly use this projection.	- Albers Equal Area Conic
Shape	Conformal	Useful for navigational charts and weather maps. Shape is preserved for small areas, but the shape of a large area such as a continent will be significantly distorted.	- Lambert Conformal Conic - Mercator
Distance	Equidistant	No projection can preserve distances from all points to all other points. Instead, distance can be held true from one point (or a few points) to all other points or along all meridians or parallels. If using a map to find features that are within a certain distance of other features, an equidistant map projection should be used.	- Equidistant Conic
Direction	Azimuthal	Projections preserve direction from one point to all other points. This quality can be combined with equal area, conformal and equidistant projections.	- Lambert Equal Area Azimuthal - Azimuthal Equidistant

Source: de Graaf *et al.* (2003).

Once the projected map has been produced, it is still not in a convenient form for georeferencing. This is because the projected latitudes and longitude lines may not be evenly spaced, the lines may curve and the lines may not be parallel. To circumvent these problems, plane Cartesian coordinates are used. This means that a regular square grid is superimposed on the projected map such that the "y" axis is pointing north, or strictly speaking, to the top of the map (Figure 3.3). The most commonly used map projection upon which Cartesian coordinates are used is the Universal Transverse Mercator (UTM). This projection, at least for smaller areas, keeps distances, directions, shapes and areas reasonably accurate. The UTM system has a unique alphanumeric referencing system that allows any location in the world between latitudes 84°N and 80°S to be allocated to a numbered zone to allow for more detailed grid-referencing (Figure 3.4 shows the UTM alphanumeric referencing on the x and y axes[78]). The regular plane coordinate square grid can now be used as the basis for giving georeferences (commonly called grid references) based on linear measurements along the "x" and "y" axes. If any single project is using a variety of mapping sources, it is important that all map layers are utilizing a common georeferencing system.

[78] The yellow arrow in Figure 3.4 identifies the UTM grid cell discussed in Box 3.6.

FIGURE 3.3
A plane coordinate grid (dotted lines) superimposed onto a projected map



Source: U.S. Geological Survey (2012).

When collecting most geographic data, it is essential that the user knows the basics of reading grid references, even though this task is now usually done by the use of GPS. Box 3.6 explains how a grid reference is derived for the red star located in a small section of a United States of America topographic map.

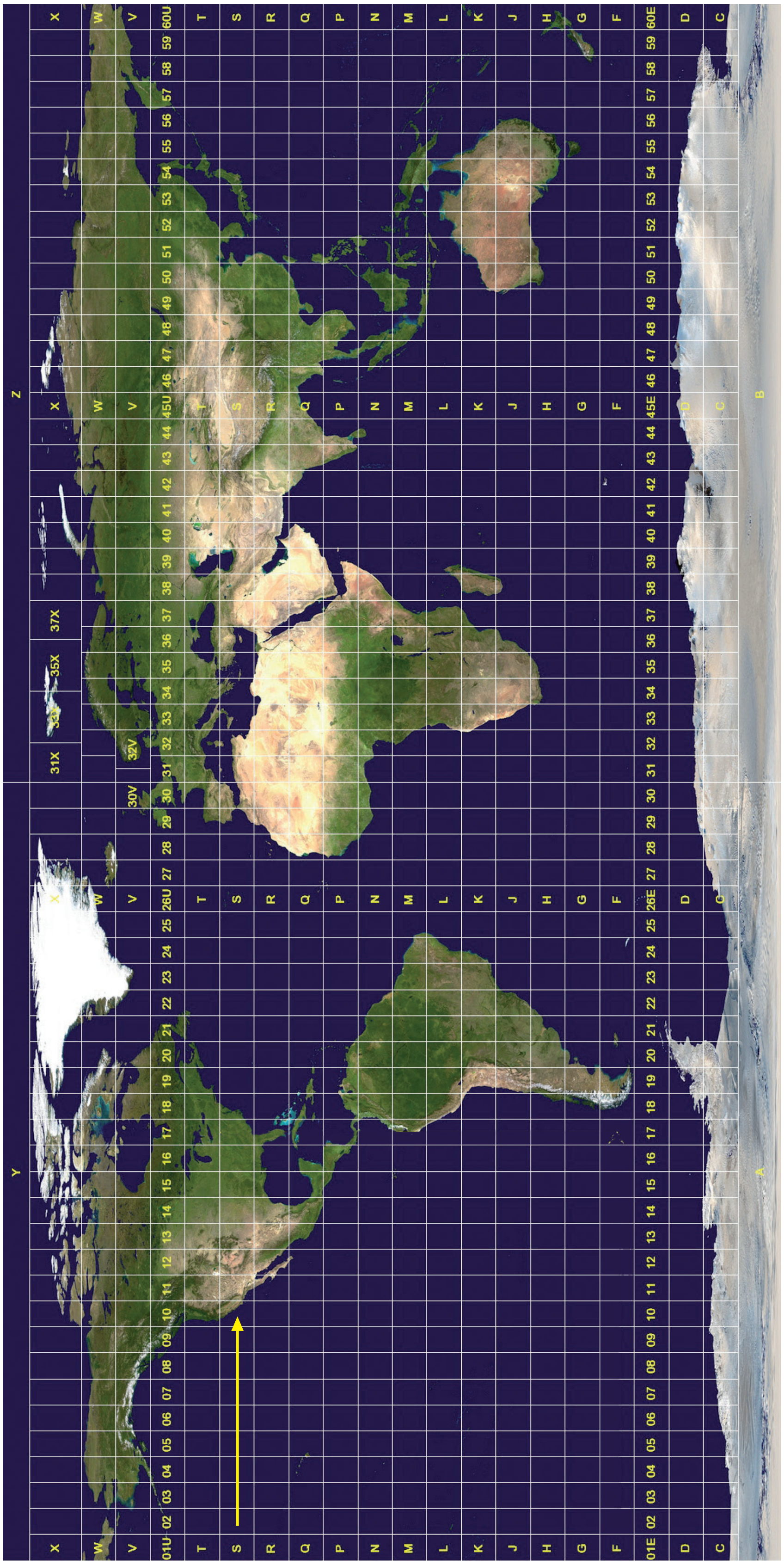
As well as recording the spatial facet of data through the use of grid reference coordinates, a named area or descriptive code can be used. These codes would normally be allocated to polygons, i.e. spatially extensive or non-point locations such as a county, post code area or census enumeration district. The GIS will have the ability to, for instance, draw a map of county boundaries and then match any data collected for each county to the correct county area. Just as thematic facets of data are stored in a hierarchical referencing system, so too are the descriptive codes given to spatially extensive polygons. For further details on the principles of mapping and georeferencing, see Butler *et al.* 1987; Robinson *et al.* (1995); Van Sickle (2004); Harvey (2008), Petersen (2009) or Crampton (2010), or the following Web sites:

- Harvard University, Graduate School of Design, Boston, United States of America: (www.gsd.harvard.edu/gis/manual/data_basics/index.htm);
- Free Geography Tools: (http://freegeographytools.com/2009/google-earth-coordinate-system-grids)
- University of Colorado, Boulder, Colorado, United States of America: (www.colorado.edu/geography/gcraft/notes/coordsys/coordsys_f.html)

Readers who do not feel confident about handling, finding, interpreting and using geographic information should refer to Johnson (2003).

FIGURE 3.4

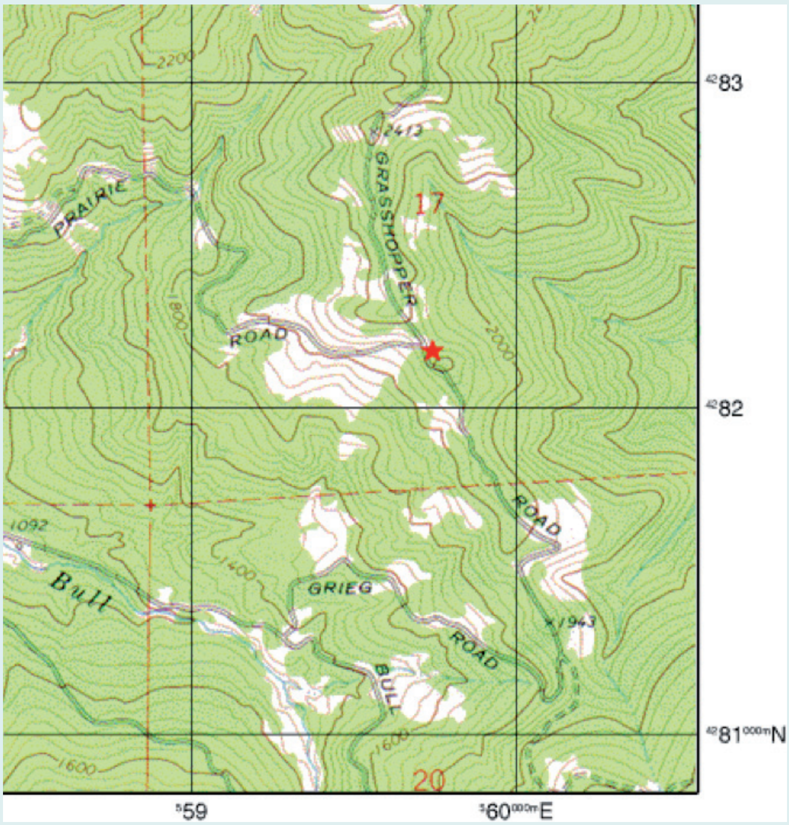
The basis of the Universal Transverse Mercator grid zone designations



Source: Rambler-Info (2012).

BOX 3.6

How to read a grid reference from a topographic map



Standing at the road junction marked with a red star on the section of a United States of America topographic map shown here, a very precise grid reference using Cartesian coordinates based on a Universal Transverse Mercator (UTM) projection would give a location of:

10 S 0559741 (Easting) 4282182 (Northing)

The 10 S represents the UTM zone number (see Figure 3.4) showing zone 10 to pass through the northwest of the United States of America – this is where horizontal zone S is located; this is a unique global zone number.

The number 0559741 shows that the location is 559 741 metres east of the Cartesian based origin line for zone 10 S. The 559 in the metre measurement is read along the bottom of the map, and the 741 in the measurement is an estimation of the number of metres between the 559 and 560 Easting lines on the topographic map, which themselves are 1 000 metres apart.

The number 4282182 represents a measurement of the north/south position within the 10 S zone in metres from the Cartesian coordinate origin line. The 4282 can be read off from the vertical (northing) axis, and the 182 in the measurement is the estimated number of metres between the 4282 and the 4283 northing lines.

So the full UTM coordinate 10 S 0559741 4282182 gives the grid reference for a location on the earth to the nearest one square metre. Some countries use variations on the above coding method to give their full UTM grid reference.

Source: U.S. Geological Survey (2010).

3.3.5 Scale and resolution facets

Although data themselves might not have properties of scale or resolution, these are essential features to consider when the data need to be mapped or used for any type of GIS purpose. First, when data are mapped, the map itself will have to be reproduced at some scale. This means that there must be a proportional relationship between the size of the mapped area compared with the real area being mapped (or the ratio between distance on the map and distance on the ground). Figure 3.5 gives an illustration showing maps from North America at three very different scales. There is no “correct” scale to use and the scale that is chosen will be related to the size of the area that a GIS project is concentrating upon. A scale may be recorded as a statement such as “one centimetre equals one kilometre”, or in the form of a linear bar that is marked off in mapped distances, or as a representative fraction such as 1:50 000 – meaning that one distance unit on the map is equal to 50 000 units in the real world.

FIGURE 3.5
Portions of North America at three different scales

More detail			Less detail
Larger scale	←	→	Smaller scale
1:24 000	1:100 000	1:500 000	1:1 000 000

Resolution refers to the degree of detail that can be shown at any specific scale. For instance, in the first map in Figure 3.5, the city of Chicago and its surrounding towns can be generally shown, but this amount of detail could not be shown in the last map because the resolution is inappropriate. If data are mapped using pixels (rasters – see Section 5.8.2), then the real world length of the side of one pixel represents the resolution of the map. For example, a spatial resolution of 30 m means that one pixel represents an area 30 m by 30 m on the ground. If the pixel is rectangular, it will be represented by a length and width dimension (e.g. 50 m × 80 m).[79] Resolution is a very important consideration when using data in GIS. For instance, if measurements were to be taken to compile a surface water temperature map of the Pacific Ocean, a very large number of sampling points would be needed. If data were only obtained from 100 points, then this resolution would be entirely inappropriate, i.e. it would be possible to construct a map but this map would be very unreliable and it would have very low statistical significance. Therefore, when collecting data for GIS use, thought must be given as to how many data samples will be necessary for the production of reliable mapped information.

3.4 DATA QUALITY AND DATA STANDARDS

Before examining the collection of data for GIS, it is imperative that those who are engaged in data collection are familiar with considerations regarding data quality. GIS output will only be acceptable, and thus useful and successful, if high-quality data are input to the system. Many users of GIS output, and of mapping products more generally, readily accept that these products are always of a high quality. Perhaps this is because many maps are sourced from reputable companies or government agencies, most of

[79] For details on scale and resolution, see remote sensing resources at: (http://biodiversityinformatics.amnh.org/index.php?section_id=31&content_id=119).

which do indeed produce qualitative products, but it is also the case that mapping output should be treated with a degree of caution. This section exemplifies various aspects concerning the basic quality of the raw data used to compile mapping output and, hence, hints at the wide range of circumstances that may cause mapping deficiencies.

What are the main components contributing to data being of high, or at least acceptable, quality? Box 3.7 explains the main qualitative components in brief; further information is available in most GIS textbooks. The overriding factors influencing data quality are the resources of money, time and effort that can be put into data collection or gathering, and the amount spent can significantly affect the final quality of GIS output achieved. It is important to mention that data quality is scale-dependent. For instance, any map made at a world scale will be significantly generalized compared with a town map. It would be impossible to zoom into the dot representing a city at the world scale to obtain better quality information or data on that city. Accuracy is mentioned several times in Box 3.7, and it is important to state that there is an important distinction between accuracy and precision. Accuracy expresses how close a measurement is to the real measurement being quantified; greater care taken should result in greater accuracy, though it is very difficult to say how accurate any measurement is. Precision relates to how precise one needs to be in recording the measurement, e.g. to the nearest millimetre, centimetre or perhaps metre. It is also important that data are collected in a consistent manner. Thus, all data collectors should use the same methods, the same degree of precision and the same means of classification, including object descriptions and class boundaries. A further consideration about data quality is that the GIS user should be aware that there are a large number of possible sources of error in geographic data. Because in some ways these are almost infinite, it is not possible or relevant to list them here. However, most GIS texts devote sections to this (e.g. Burrough and McDonnell, 1998; Lo and Yeung, 2002; Bernhardsen, 2002; Heywood, Cornelius and Carver, 2006).

BOX 3.7

Some important factors to consider regarding data quality

The most important factors or measures contributing to data quality are:

- Positional accuracy – This refers to the locational accuracy to which data are placed on a map. Map scale can affect accuracy, e.g. the width of line shown on a small-scale map may of necessity be shown as being a real world width of several hundred metres. Additionally, in crowded areas of a map, it may be necessary to show objects in approximate positions only.
- Attribute accuracy – This refers to the description given to mapped objects. While it may seem obvious what most things are, difficulties may arise, e.g. in defining the edge between one marine ecosystem and another, and frequently two individuals would categorize objects differently. These problems are multiplied when objects need to be classified into numeric classes. This is a particular problem in remote sensing image analyses.
- Data completeness – Have all the required data been collected in a consistent manner across an entire study area? Note that completeness might not require all data if a sampling strategy is employed that only requires a representative sample of data.
- Data timeliness – Timeliness refers to how up-to-date data may be. Some data may only change very slowly, e.g. bathymetry, while other data may be constantly changing, and for these data timeliness may be almost impossible to comply with. This is a particular problem in fisheries where many objects are in perpetual motion.
- Data lineage – Refers to a knowledge of how the data was collected, e.g. the instrument types used, or how the data may have been transformed or treated in some way.
- Data accessibility – This describes where data can be found and obtained. Data may be more or less accessible according to price, copyright, format, etc., and inaccessible data can result in a need to use perhaps lower-quality substitute data.

Given that qualitative data are of great importance to the success of any GIS project, then one means of ensuring quality and consistency is through the establishment and maintenance of certain data standards.[80] Bernhardsen (2002) notes that standards will provide a definition of data structures, data content and rules that will:

- increase mutual understanding of the geographic data among users;
- eliminate the technical problems of exchanging geographic data between different GISs;
- increase integration and combination of geographical data and related information.

The components of geographic data standards include:

- standard data products – whereby many data sets will conform to standard georeferencing systems, projections, legend categorization, syntax, symbolism, etc.;
- data transfer standards – also known as data exchange standards, ensure that data is readily exchanged among different users;
- data quality standards – see Box 3.7;
- metadata standards – see Section 5.6.

International standards are managed by the International Organization for Standardization[81] (ISO) through a series of technical committees (TCs). The ISO/TC 211 coordinates all matters regarding the development and setting of standards on geographic information and geomatics. Many individual countries or regions have their own geographic standards organizations, such as the European Umbrella Organisation for Geographic Information in the European Union[82] or the Federal Geographic Data Committee in the United States of America.[83] A general overview of standards is given on www.opengis.org, and www.isotc211.org provides further details on international geographic standards.

The FAO GeoNetwork open source (www.fao.org/geonetwork/srv/en/main.home) allows spatial data to be easily shared among different FAO units, other United Nations agencies, non-governmental organizations and other institutions. The FAO GeoNetwork has been developed to connect spatial information communities and their data using a modern architecture, which is at the same time powerful and low cost and is based on the principles of free and open source software (FOSS) (see Section 2.3.3) and international and open standards for services and protocols (i.e. from ISO/TC 211 and the Open Geospatial Consortium).

3.5 DATA COLLECTION AND DATA SOURCES

It has been emphasized that at the same time as data needs are established, for each project the geographic boundary must also be agreed. It is often wise to confine projects to national or state boundaries because many data sources will be restricted – obtaining equivalent data for a neighbouring state may be problematic. When assessing data needs, it must also be remembered that many projects, especially those concerned with fisheries or aquaculture, will require data that are collected and registered in four dimensions. A consideration of the scale or resolution for the project also needs to be made. It might be argued that these do not matter, i.e. because the GIS program will contain a zoom facility that allows the user to work at almost any scale. However, if the data have only been collected using relatively few data sampling points, then it will be increasingly inaccurate to zoom into the mapped information. Finally, it may occasionally be necessary to modify the project aims in order to suit data availability - this may theoretically be a poor basis on which to proceed, but it could be the only practical solution for acquiring certain desired information.

---

[80] GIS “standards” apply not only to data but to other aspects of computing such as applications areas, technology and professional practice. This allows for interoperability and integration among disparate components of the whole IT system. The term “standards” also has two connotations, i.e. that of “quality” and that of “uniformity”.

[81] See International Organization for Standardization (www.iso.org).

[82] See European Umbrella Organisation for Geographic Information (www.eurogi.org).

[83] See Federal Geographic Data Committee (www.fgdc.gov).

Given this paramount importance of data to GIS project success, it is important to ask a range of fundamental questions on data – Box 3.8 provides such a list (as obtained from www.gis.com/content/selecting-right-data).[84] Having determined data needs and other essential facts relating to data requirements, the data themselves must be obtained. Basically there are two broad options for collecting data: (i) collecting one's own data; or (ii) securing data that have previously been collected. These types of data are known respectively as primary data and secondary data.

BOX 3.8
Some questions that should be asked of any data to be collected

- What are the data needed for?
- Where the data might come from and is the data source reliable?
- What are the specific geographic features required?
- What attributes of those features are required?
- What scale and resolution are needed for the data?
- What is the geographic extent of the area of interest?
- What is the level of geography to be examined within the area of interest?
- How current must the data be?
- What type of computing environment will be used?
- What GIS software will be used?
- How many concurrent users will be accessing the data, at how many locations?
- When are the data required?
- Will periodic data updates be required and, if so, how frequently?
- Is the plan to start small, and then expand?
- How much might the data cost?

Source: ESRI (2012a).

3.5.1 Primary data collection

Primary data collection will depend on a number of limiting factors such as:

- the time available;
- the capital outlay planned;
- the skills and number of personnel available;
- the availability and usefulness of any existing data;
- the size of the area being studied;
- the equipment available;
- the purpose for which the data are required.

Any primary data collected could be in various formats, e.g. photographic (video or still), written textual, recorded counts or measurements, labelled pictorial, or digitally captured, but, for GIS purposes, all data collected will eventually need to be transformed into digital formats (discussed in Chapter 5). Additionally, almost all data collected will be the result of a sample survey. Thus, it is rarely possible to collect the total data on any aspect of a project, so some kind of sampling strategy (considering mainly sampling frequency and data resolution) will need to be devised, such that data are collected from a representative range of all possible sites within an area under study. As data sampling strategies is a wide topic, it is not covered here, but there are many texts and Web sites covering this theme (e.g. see Gunderson, 1993; National Research Council, 1998; Stamatopoulos, 2002; Gregoire and Valentine, 2004; Sabatella and Franquesa, 2004). The collection of primary data on fisheries and aquaculture can be conveniently examined under the headings of:

[84] This Web site can provide further detail on these questions.

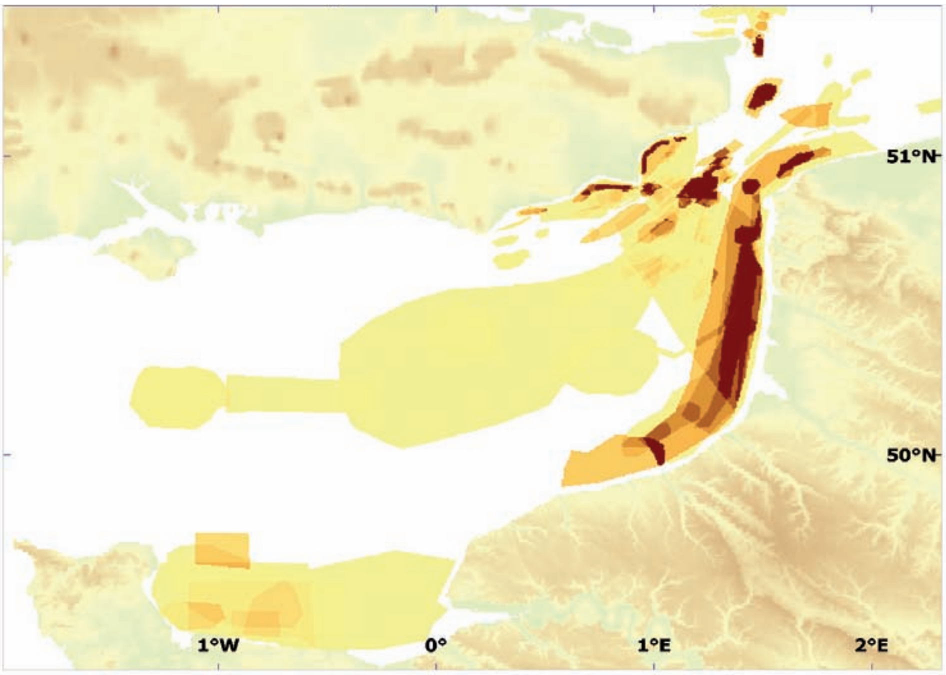
- manual methods using no equipment;
- methods using a variety of equipment.

Manual methods of primary data collection using no equipment

Manual methods of data collection might be conceived as being rather basic but, under many conditions and circumstances, these will be the most practical ways of obtaining the required data. Each of these methods entails varying considerations and complexities – space will preclude a detailed examination of these.

- **Direct sketch mapping**. This will be useful to establish both specific point-based distributions of any variable or more general areas in which an activity takes place. Sketch mapping may also be necessary in instances where available base maps may need updating, and in instances where initially plans are being sketched out, e.g. for the siting of aquaculture facilities. Figure 3.6 shows areas in the eastern English Channel (between the United Kingdom of Great Britain and Northern Ireland and the French Republic) where surveyed local fishers perceive that they mainly fish. Each fisher was given a base map on which to sketch in perceived main fishing areas (perhaps for a particular species), and the results were aggregated to produce this map. Areas of darkest brown show most fishing effort (see www.ifremer.fr/charm, for further details).

FIGURE 3.6
Aggregated fishers' perceptions of their fishing locations for the main commercial species in the eastern English Channel



Source: IFREMER (2011).

- **Interviewing.** Large amounts of data are gathered by face-to-face interviews or by groups of people – often called "focus groups". Data gathered can be both qualitative and quantitative. Interviewing is frequently the only way of obtaining a range of socio-economic information about fisher groups and their activities, and in many areas fishers will have a more detailed knowledge of local conditions than will fishery experts. A potential problem with interviewing is that it may be difficult to ascertain the degree of objectivity prevailing in respect to the data obtained.

- **Questionnaires.** Probably the most common form of primary survey, this usually relies on pre-printed forms. Questioning may be on a face-to-face basis or through the mail or via e-mail or telephone surveys. Questionnaires are very useful for collecting data that may be open-ended or of a preferential response type; aggregations of answers of this latter type can usually allow for statistical analyses. Before carrying out a large-scale questionnaire survey, a small pilot-scale survey should be performed in order to gauge the validity of each question.
- **Form filling.** Many government departments, for instance, have pre-printed forms that are utilized to gather factual data, but any organization can develop forms for any use. Forms are frequently structured so as to allow data to be readily converted into a digital format. Figure 3.7 illustrates typical forms used for recording fishing vessel trips and fishing catch details.

FIGURE 3.7
A basic fishing-vessel logsheet (Form A1) recording details of vessel activity. This may be accompanied by many fishing catch logsheets (Form A2)

FORM A1. TRIP LOGSHEET (LOGBOOK)

Logsheet Serial No:		Vessel ID.:		IRCS:		Trip No:		Date:	
License No:		Vessel name		Logsheets attached		From:		To:	

	PORT	DATE	TIME
Departure			
Arrival			

LANDING PLACE	BUYER	DATE	LANDED QUANTITY (KG)

Captain's signature: ______________________ Date: __/__/__

FORM A2. LOGSHEET (Many A2 forms for each A1 form)

Logsheet Serial No:		Vessel ID.:		IRCS:		Trip No:		Date:	

FILL IN A NEW PAGE FOR EACH DAY, CHART AREA, GEAR OR MESH SIZE FISHED

Chart area:		Gear type:		Quantity gear:		Gear size:	
Mesh size:		No. of hauls/sets:		Average tow/soak time:			

STATISTICAL AREA	SPECIES	LANDINGS (kg)	DISCARDS (kg)

Captain's signature: ______________________ Date: __/__/__

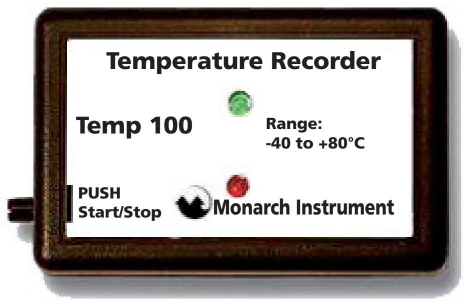
Source: Sparre (2000).

Primary data collection methods using a variety of equipment

A vast range of equipment has been developed for almost every type of data gathering. Over the past few decades this equipment has gradually changed from being mechanically or analogue[85] based to being digitally based, and today very few data are gathered by non-digital means. Fisheries-related examples of data still collected using the more basic, non-digital equipment include the use of various nets to either catch fish samples, or various plankton nets, and the use of simple length measuring devices. In non-fishery fields, instruments such as rain gauges, barometers, soil sieves and clinometers are still in frequent use. Most of this equipment is relatively easy to use, but it is likely that it will be slowly phased out.

In digital data collection, the range of relatively simple instruments is extremely wide and here only a basic range is discussed.[86] The full range can easily be ascertained via the use of Internet search facilities. Here, our discussion is ordered in an approximate hierarchy moving from the more basic towards very complex equipment and systems. The use of an even more complex system – remote sensing – is discussed in Chapter 6.

FIGURE 3.8
Simple electronic device for collecting temperature data



Source: Monarch Instrument (2008).

- **Electronic "read-out" equipment.** This category covers basic pieces of mainly portable equipment that allows the user to read or record measurements on a variety of parameters, e.g. temperature, water-flow rate, pH, weight, humidity and light strength. An Internet search will quickly reveal a more complete range of this equipment. The most basic devices simply give instant readings that could be recorded manually. With increasing sophistication, these devices will record sequences of reading that are typically aligned to time or place. Figure 3.8 illustrates a simple device, weighing only 28 grams, that records temperatures between –40 °C and +80 °C. It can store more than 30 000 readings taken at intervals from 30 minutes to 2 days and to a resolution of 0.1 °C, and these data can be downloaded to a computer for further analysis. There is a close association between functionality and cost and thus care needs to be taken in deciding the quality of the data required.
- **Digital cameras.** These cameras can be usefully employed to capture data on an almost infinite variety of subjects. For instance, McKnight (2009) recently took over 1 500 photographic records showing Pacific oyster densities of this invasive species at sampling points on structures built along the northeast Kent coast in the United Kingdom of Great Britain and Northern Ireland. This baseline information will allow for differential rates of population change to be calculated. The price of cameras has dramatically reduced over the past decade, and their memory storage has exponentially increased. Many cameras have the added capacity of video recording. Airborne digital photography has become a source of huge inputs of data for GIS purposes, and digital underwater photography is capturing increasing data holdings relating to underwater topography, rugosity,

[85] Analogue means electrical output information that is represented by, for example, a voltage reading or a pointer rather than through use of a digital display.
[86] Note that some of this digital data collection equipment was more suitably discussed in Section 2.2.1.

sedimentary structures and vegetation distributions. Airborne digital photographs can be readily incorporated into most raster-based GIS (see Chapters 5 and 6) as an aid to updating mapped data, or as the basis for additional analyses.

- **Data loggers and personal digital assistants (PDAs).** This heading includes a wide range of usually handheld digital devices for communications, basic computing and data capture. Data loggers are handheld, or in situ, devices developed for capturing a range of data on specific themes. They are generally battery powered, portable, and equipped with a microprocessor, internal memory for data storage and sensors. Some data loggers interface with personal computers and utilize software to activate the data logger and view and analyse the collected data, while others have a local interface device (keypad, LCD) and can be used as a stand-alone device. Data loggers are frequently developed and programmed for specific purposes. For instance, nearly all household electricity meter readings are made via a person calling at the house who punches in the current meter readings using a pre-programmed form. Figure 3.9 shows that data loggers may be specially waterproofed (and ruggedized) for use in more hostile natural environments. For all environments (e.g. marine, freshwater and brackish water) for fisheries and aquaculture purposes, data loggers may be deployed to capture data on water quality parameters, on water flow regimes, on fish movements, etc., and instrument deployment may be in the form of free floating or tethered buoys, on instruments attached to vessels, or on a variety of programmed autonomous underwater vehicles, etc.

  PDAs commonly have colour screens and audio capabilities, enabling them to be used as mobile phones (smartphones), Web browsers, portable media players or computers. Many PDAs can access the Internet, intranets or extranets via Wi-Fi or wireless wide area networks (WWANs). Most PDAs employ touch-screen technology and can be synchronized to a PC allowing for two-way information or data flows. They may be loaded with their own proprietary software, or they may be synchronized with a range of external programs.[87] Figure 3.10 shows a PDA that combines communications functionality with data collection and mobile computing functions. Most PDA-type devices today are pushing the limits on storage capacity, although only a few have so far ventured into the micro-hard drive territory, which greatly expands the storage capacity, making PDAs

FIGURE 3.9
A typical waterproofed and ruggedized handheld data logger

Source: MobileDataforce Australia (2011).

[87] An important program that can be loaded to mobile computers is ArcPad. This is mobile GIS software, produced by ESRI, that has been purposefully designed so that field data can be collected in a variety of ways; most include inserting a zoomable backdrop map so that data can be recorded at required georeferenced points as an "overlay" layer to the map. ArcPad also has an integrated GPS, range finder and digital camera (see www.esri.com/software/arcgis/arcpad/index.html).

potentially every bit as powerful and flexible as a desktop PC. There is also a strong demand to increase the integration of functionality so that any one piece of equipment will perform a wide range of tasks. From the data collection perspective, this might be successful, but where there is a need to display and visualize mapped data then a penalty is adopted because of the micro nature of much of the visual display capabilities. However, some PDAs, e.g. Blackberrys, have managed to marginally increase display size. Finally, it is important to mention that developments in this area of data capture equipment are occurring at a very rapid rate, so any PDA equipment purchased may quickly become obsolete.

FIGURE 3.10
An integrated computing and communications-based personal digital assistant

Source: PC Cubed Ltd. (2012).

- **Global positioning systems (GPS).** Although GPS are complex satellite-based systems, for the purposes of this technical paper it is not necessary to be concerned with exactly how the system works – the concern is simply with the capability of the user segment of the system. Those wishing to find out more details on the total system should refer to Steede-Terry (2000), Kennedy (2005), Taylor and Blewitt (2006) or El-Rabbany (2006). Figure 3.11 shows a typical handheld, battery-operated GPS receiving device. This receiver is used for capturing location information or a georeference at the point where each reading is taken. Along with georeferences, which may be captured as a latitude/longitude set of coordinates or as a UTM-based set of Cartesian coordinates relating to the area and/or region in which the readings are based, GPS receivers can capture and store data on a wide range of variables including route lines, altimetry, travelling speeds and direction, distances and time. This functionality makes GPS a very useful data collection device for any geographic purpose. GPS functionality has now been integrated to a wide range of separate devices or equipment, such as watches, digital cameras, mobile telephones, vehicle navigation systems and in vessel monitoring systems (VMS). Data from VMS are being extensively used to monitor the trajectories of fishing vessels, and it is relatively simple to estimate from the recorded vessel locations and speed of travel exactly where fishing activities have taken place. Handheld GPS may record location accuracy to within 5 m of a precise location, and this is adequate for most fishery and aquaculture purposes. More expensive "differential" GPS equipment can be accurate to approximately 5 cm.[88]

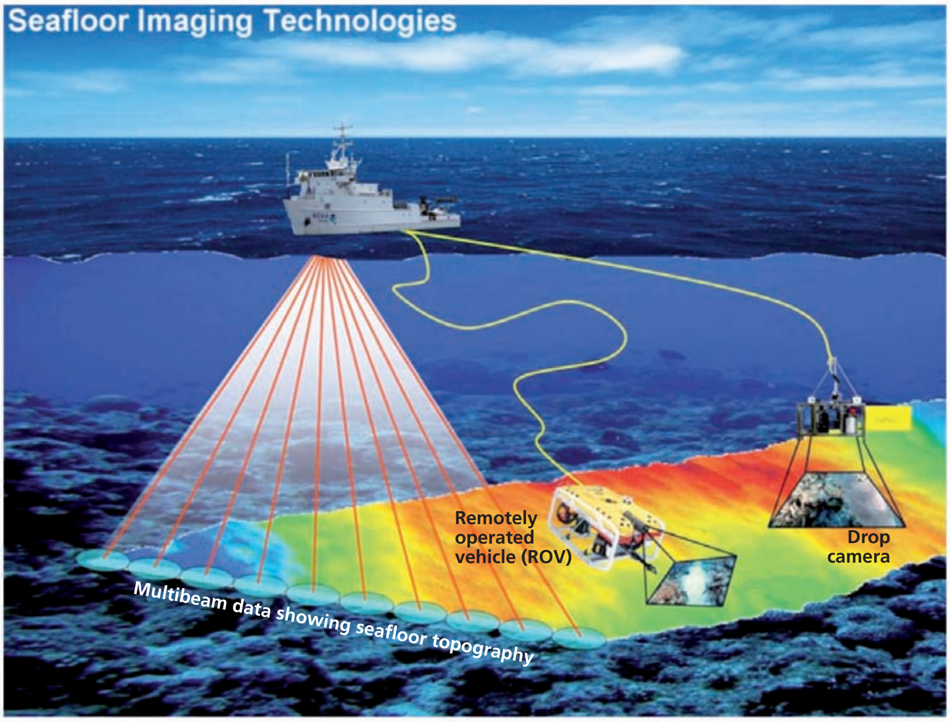
FIGURE 3.11
A handheld GPS receiver

Source: Garmin Ltd. (2012).

[88] Differential GPS makes use of ground stations having exact known fixed locations to enhance the accuracy of the GPS positions as suggested by any space-based satellites that are fixing the location.

- **Acoustic sonar and other underwater devices.** There is a wide range of acoustic sonar and other devices that capture data relating mainly to depth and to the bottom characteristics of a waterbody, e.g. seafloor substrates or rugosity, or relating to fish locations. There are different types of acoustic sonar that vary according to the trajectories of their scanning functions. The systems work on the principle of having a transducer underwater, which may be mounted on the hull of a vessel, or which may be towed attached to a "towfish", or which may be integral to an autonomous underwater vehicle, or which may be attached to the headrope of a large trawl fishing net. The transducer sends out pulses that, when they encounter objects such as the seafloor or fish, bounce off the object and are echoed back to the transducer. Knowing the time taken for the echoes to be received indicates the distance of objects from the known location of the transducer. Different echo patterns are received on a display unit and this gives an indication as to what the object is. Figure 3.12 illustrates the operation of multibeam acoustic scanning. Here, the remotely operated vehicle (ROV) and the drop camera are being used for ground truthing purposes,[89] but they may be used independently to gather seafloor data.

FIGURE 3.12
Seafloor imaging using multibeam acoustic sonar, a drop camera and a remotely operated vehicle



Source: NOAA National Centers for Coastal Ocean Science (2004)

Sound waves travel differently through fish than through water because a fish's air-filled swim bladder has a different density than seawater. This density difference allows the detection of schools of fish by using reflected sound. While it is difficult to identify individual fish species, it is relative easy to deduce the species of larger shoals that may be present and to estimate the biomass of a shoal. Acoustic technology is especially well

[89] Ground truthing is the process of verifying the captured imagery digital data with what is actually recorded at the site. It enables the captured data to be calibrated so as to aid the interpretation.

suited for underwater applications because sound travels farther and faster underwater than in air and because the use of multifrequencies and omnidirectional sonars has greatly improved the ability to understand the complexities surrounding fish stock behaviour and distributions. Today, larger commercial fishing vessels and fisheries research vessels rely almost completely on acoustic sonar and sounders to detect fish. While acoustic sonar has been mainly used in the sea or large inland waterbodies, Hateley and Gregory (2005) report on the monitoring of fish quantities in freshwater river environments through the use of multibeam sonar techniques. Useful information on acoustic sonar and similar data collection techniques can be found in Simmons and MacLennan (2005) and Foote (2009).

3.5.2 Secondary data acquisition and sources

With respect to both data collection methods and data sources, the acquisition of secondary data has undergone rapid changes over the last few decades. These changes have involved the types of data sources, the form that the data takes, the delivery mechanisms for data and the breadth and volumes of data available. Twenty-one years ago, Meaden and Kapetsky (1991) reported 21 different sources of secondary data, including encyclopaedias, directories, textbooks, reports, maps and atlases. They further noted that the production format of secondary data included written accounts, diagrams, figures, tables, graphs, maps and aerial photographs. Today, very few of their data sources and formats would be considered as applicable. The only sources that still apply are maps, some aerial photography, digitally encoded material and computerized databases, and, in fact, these headings more or less encapsulate all of today's secondary data sources. There has been a move from largely fragmented paper-based data sourcing to more centralized digital data sourcing. Probably more than 95 percent of today's fisheries-related secondary data already exist in digital format, with the main exceptions being some "historical" tabular data held in various paper-based archives and a range of cartographic source materials that have yet to be digitized. The proportion of available digital mapping varies significantly from country to country. A major problem that has been largely overcome in the past 20 years is that of the digital format of any data set obtained. Thus, through standardization of data formats and through the ability to change formats, it is more likely that any digital data obtained will be usable by any GIS software that may be deployed.

Today, secondary data are mostly stored in databases that themselves are held at numerous portals[90] on powerful computer servers awaiting delivery to "clients" via the Internet. Some digital data would still be delivered via CD–ROMs or DVDs, but with the rapid rise of broadband Internet services these delivery mediums would be of decreasing importance. What are the main sources of secondary data? There are a number of major data sources that might be used as initial starting points in any search for marine and fisheries data. These sites include those that supply very general geographic data that could be utilized for cartographic or socio-economic needs, those sites that specialize on broad themes such as the environment or oceanography, and those that concentrate on fisheries (marine or inland) or aquaculture related subjects. Box 3.9 lists some useful starting points. This section does not attempt to list sources of more basic and general data, such as climatic data, soils data, protected areas, administrative regions and biodiversity regions, even though some of these data could well be very useful. Good sources for information on this data are presented in Box 3.10. As well as these major general data repositories, there are a large number of more specific data (and information) sources beginning to accumulate. Some of the means of accessing these are presented in Box 3.10,[91] and Table 3.3 provides an example of the more specific databases that are available for inland fisheries in North America.

[90] Portals are usually access points to the Internet allowing connections to specialized information and data.
[91] The sources listed here represent major FAO compilations.

BOX 3.9
FAO portals to major marine, fisheries and aquaculture data providers

Aquaculture
Kapetsky, Aguilar-Manjarrez and Soto (2010) –
Available at: www.fao.org/docrep/012/i1359e/i1359e00.htm

- Chapter 3: Spatially defined global ecosystems, their issues and their relevance to the ecosystem approach to aquaculture
- Chapter 4: Spatial data to support the ecosystem approach to aquaculture

Kapetsky and Aguilar-Manjarrez (2007) –
Available at: www.fao.org/docrep/009/a0906e/a0906e00.HTM

- Table 4.3: Freely downloadable spatial data and their application to assess marine aquaculture potential: cultured organisms (CO), offshore culture facilities (OF) and transport and maintenance trips from shore facilities to offshore culture facilities (TM)
- Chapter 5: Data availability

Marine and fisheries
Carocci *et al*. (2009) –
Available at: www.fao.org/docrep/012/i1213e/i1213e00.htm

- Annex: Major fisheries and marine data providers on the Internet

Although the material included in general data repositories may cover marine, fisheries and aquaculture data sources, for GIS work directed specifically to inland fisheries or aquaculture it is likely that access will be needed to terrestrial topographic mapping at various scales. Most countries have government mapping agencies that are responsible for map distribution, and the location of these agencies are readily discovered via Web-based searching. Most mapping agencies apply modest charges for paper-based products and much higher charges for the same information in digital formats. However, the latter products will not require digital conversion, a process that can be expensive. On the other hand, paper mapping products can be selectively digitized (see Section 5.2.1) ensuring that only required digital information is obtained. Mapping agencies in some countries provide for interactive map deliveries of cartographic data covering selected areas at selected scales, though this data may only be obtained under certain licensing and copyright conditions.[92] As well as national mapping agencies, many countries have "hydrographic" agencies that can supply charts and other marine information, e.g. bottom sediment, ocean currents, bathymetry, and locations of shipwrecks and other obstructions. As well as government mapping agencies, there are a large number of private companies and other organizations that provide maps or mapped data, with each usually concentrating on specific themes or geographic areas. One such private source of a large range of data is shown in Box 3.10.

[92] See http://nationalmap.govsed or www.ordnancesurvey.co.uk/oswebsite, which provide, respectively, mapping agency data for the United States of America and the United Kingdom of Great Britain and Northern Ireland. Many countries now provide similar mapping services.

TABLE 3.3
Some sources of GIS databases for inland fisheries in North America

Characteristic	Database	Source and Internet address**	Scale	Comments
Geology (bedrock and quaternary)	Geological surveys	Association of American State Geologists (www.stategeologists.org); Geological Survey of Canada www.gsc.nrcan.gc.ca) Global geology (www.portal.onegeology.org)	1:250 000	Small-scale allows for landscape-level analyses
Soils	Soil surveys	USDA, Natural Resource Conservation Service (www.nrcs.usda.gov); Global soil regions (www.soils.usda.gov/use/worldsoils/mapindex/order.html); Agriculture and Agri-Food Canada, National Soils Database (sis.agr.gc.ca/cansis/nsdb/intro.html)	1:20 000 –1:250 000	Large-scale data are useful to characterize the watershed or miscellaneous conditions that influence water quality
Elevation and watershed boundaries	Digital elevation model (DEM); Digital terrain elevation data (DTEM)	USGS, Earth Resources Observation and Science Center (eros.usgs.gov/#/Guides/dem); Natural Resources Canada, Canadian Digital Elevation Data (www.geod.nrcan.gc.ca); NASA, ASTER Global Digital Elevation Model (asterweb.jpl.nasa.gov/gdem.asp)	1:20 000 –1:250 000	Most DEMs currently are small-scale
Hydrography (lakes, reservoirs, streams, rivers)	Digital line graph (DLG); TIGER/line files	USGS, Earth Resources Observation and Science Center (eros.usgs.gov/#/Guides/dlg); U.S. Census Bureau, TIGER (www.census.gov/geo/www/tiger/index.html); Natural Resources Canada, Geometrics Canada; (geonames.nrcan.gc.ca) Global Hydrographic Data GGHYDRO (www.geodiscover.cgdi.ca)	1:24 000 – 1:2 000 000	Provides a base map for planning, mapping and analysis
Dam locations	National Inventory of Dams (NID)	USACE (http://nid.usace.army.mil); Global Reservoir and Dam (GRanD) Database (http://atlas.gwsp.org/)	Not available	A georeferenced database of dam locations in the United States of America and the world
Transportation (roads and railways)	Digital line graph (DLG); TIGER/line files	USGS; U.S. Census Bureau; Natural Resources Canada, the Atlas of Canada (atlas.nrcan.gc.ca/site/index.html)	1:24 000 – 1:2 000 000	Intersections with streams identify access and impact sites; compute road density in watersheds
Land cover/land use	Land use and land cover (LULC); National Wetlands Inventory (NWI); digital raster graphics (DRG); digital orthophoto quads (DOQ)	USGS, Seamless Data Warehouse (seamless.usgs.gov/); USGS, Land Cover Institute (http://landcover.usgs.gov/landcoverdata.php); USFWS, National Wetlands Inventory (www.fws.gov/wetlands/); Natural Resources Canada, the Atlas of Canada (atlas.nrcan.gc.ca/site/index.html)	1:100 000 –1:250 000; 1:24 000	Useful for watershed and riparian analyses; DRGs are digitized topographic maps; DOQs are rectified aerial images matched to DRGs
Place names	Geographic names information system (GNIS)	USGS (nhd.usgs.gov/gnis.html); Natural Resources of Canada, Geomatics Canada (geonames.nrcan.gc.ca); National Geospatial Intelligence Agency, GEOnet Names Server (GNS) (http://earth-info.nga.mil/gns/html/)	1:24 000	Names of physical and cultural places, features and areas

***Acronyms for data sources:* National Aeronautics and Space Administration (NASA); United States Department of Agriculture (USDA); United States Geological Survey (USGS); Department of Defense (DOD); United States Army Corps of Engineers (USACE); United States Fish and Wildlife Service (USFWS).
Source: Modified and updated from Fisher and Rahel (2004b).

BOX 3.10

Online and mobile solutions to GIS mapping from the Environmental Systems Research Institute

The Environmental Systems Research Institute (ESRI) has a number of new and evolving solutions for online and mobile GIS, many of them free that can be used to complement and extend GIS use. These new tools are:

ArcGIS Online provides a common platform to find, share and organize geographic content and to build GIS applications. A wide range of worldwide mapping is available for access and download, and these will be updated twice annually. Online map layers that can be used in your documents include multiresolution imagery, street maps, topographic maps, relief maps and demographic maps. A range of "tasks" is also provided that allows for tasks such as routing, address finding and place locating.

ArcGIS.com is simply the Web interface for ArcGIS. Online, users can access maps, applications and tools published by ESRI and other GIS users, and share their own content with a broad community of users.

Portal for ArcGIS provides the same collaboration and sharing tools as ArcGIS Online, but in the user's own secure environment. It becomes the central repository for authoritative content that users inside an organization can access to quickly create maps and applications using templates and Web mapping APIs*, form groups to collaborate on projects or common activities, share maps and applications with private groups or the entire organization, and embed maps and applications in customized Web pages or blogs.

ArcGIS applications for smartphones allow users to navigate maps, collect and report data, and perform GIS analysis. It is a part of the ArcGIS system and is a new means to discovering content by browsing map galleries from ArcGIS Online or leveraging existing enterprise GIS services. It also displays maps and captured information, develops a custom application or brands own applications specific to business needs, and extends the users GIS to a wider audience.

*Applications Programme Interface.

Source: ESRI (2011a).

Data users will need to check: (i) whether data are free or if they incur a cost; (ii) what the spatio-temporal resolution is; and (iii) what format the data are available in. Data are commonly delivered from any of the Web-based, CD–ROM or DVD data sources in spreadsheet formats that are easy to incorporate into GIS. If the data format is not compatible with the chosen GIS, then conversion to another format may be necessary. The availability of fisheries data tends to be less frequent than marine data mainly because fisheries data are often collected for small-scale projects, where only short time periods are covered, or by a diversity of smaller and perhaps more isolated institutions. Fisheries databases are typically difficult to find on search engines. Valavanis (2002) provides a helpful and extensive commentary on how marine and fisheries data may be integrated and used within a GIS environment.

In order to circumvent any data assembling problems, many groups that are working on fisheries-related GIS projects decide that collection of their own data is beneficial, or they decide that they will only work on GIS projects where they know that good quality appropriate data can readily be obtained. For users, or potential users, of GIS in developing countries, it will be important to check the cost and availability of data before embarking on any GIS work or, indeed, before implementing a GIS per se. Unless the means can be obtained for acquiring appropriate data, then it may not be wise to proceed. However, GIS users could also be encouraged to experiment with data that are readily available, and then to try to assess the viability of any GIS-based analytical output produced against some measure of expectation.

3.5.3 Proxy data

One final means of coping with a deficiency of data needs to be mentioned. This refers to the use of so-called "proxy" data. Proxy data are data that may not be directly related to the exact data required but can nevertheless be a substitute for the real data. For instance, there is plenty of evidence that there is a close relationship between air and water temperatures, certainly in rivers[93] and lakes (Smith and Lavis, 1975; Balarin, 1987), so if average air temperature data can be obtained for a given aquatic site, then these can usually be a good substitute for water temperatures. There are also many relationships between plant and soil types, with all plants showing a preference for more or less acidity, or soils having various structures, e.g. free draining or water retaining. A vegetation map therefore may give clues as to the types of soil prevailing in an area. Much of the GIS-related work on site suitability for aquaculture has been based on these relationships. An example of this work, done by Aguilar-Manjarrez and Nath (1998), showed how soil suitability for the siting of fish ponds in Africa could be inferred from using proxy data obtained from a 1995 FAO-UNESCO Soil Map of the World (see: www.fao.org/docrep/w8522e/W8522E04.htm#P951_45085).[94] Whether to use proxy data or not has to be based on a decision involving considerations of the time taken and costs involved in collecting the real data versus the likely accuracy of any intended proxy data. There are also considerations of scale and of the areal extent of the proxy data. Sometimes it might be better to use two or more proxy data sources, e.g. the cost of land might be a function of both population density and of the quality of the soils. In this case, the latter two distributions would need to be combined in an agreed way. Table 3.4 provides some examples of possible uses and sources of "proxy" data.

TABLE 3.4
Examples of possible uses and sources of "proxy" data for aquaculture site selection

Production function requiring data	Possible proxy data source
Soil quality	Distribution maps of particular plant species
Water quality	Distribution maps of various aquatic fauna or flora
Water temperature	Maps or tables of air temperatures
Catering outlets	Listings of hotels or restaurants
Underground water sources	Hydrogeological or geology maps
River water quantity	Annual rainfall maps
Land costs	Population density map
Wholesale market outlets	Distribution of large towns or cities
Fertilizer inputs	Maps of livestock or poultry farms
Availability of capital inputs	Distribution of large towns and cities

Source: Meaden and Kapetsky (1991).

Most of the proxy data shown in Table 3.4 apply to terrestrial mapping of production functions pertaining to aquaculture production and locations. However, it is also possible to use proxy data for developing maps showing aspects relating to capture fisheries, although generally these maps are likely to be less reliable than proxy terrestrial-based maps. This is simply because in the fisheries environment both the

---

[93] This relationship may not be the case where the main body of water in rivers is coming from underground (spring) sources.

[94] For details on the main properties of soils for freshwater fish culture, see FAO (1995a) and Meaden and Kapetsky (1991).

environment itself and the fauna in it may be constantly moving both at the micro- and macroscales. Nevertheless, most marine and inland waters can be subdivided into numerous biogeographic realms (at a large-scale) or aquatic habitats (at a small-scale), and both of these are classified on the basis that they are aquatic areas exhibiting collections of integrated species living under specific environmental conditions. So, if the details pertaining to the trophic web of a particular marine biogeographic realm or aquatic habitat are known, it is likely that broad assumptions can be made as to the range of species found there and the likely parameters of local waters (depth, salinity, temperature, bottom sediments, etc.). Much of the recent work done on establishing habitat suitability or essential habitats for aquatic species has relied on these relationships (MacCall, 1990; Rubec *et al.*, 1998; Brown *et al.*, 2000; Leathwick *et al.*, 2006; Fisher *et al.*, 2011; Lauria *et al.*, 2011). An example of this is the work done by Le Pape *et al.* (2003), who showed the direct relationship between the growth of juvenile sole in the Bay of Biscay, in the northeast Atlantic Ocean, and the quality of coastal and estuarine nursery habitats. So, a map of juvenile sole abundance could prove a good proxy for a map of sole nursery habitats. There are, of course, an almost infinite number of similar relationships within the aquatic spheres.