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

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

It has frequently been suggested that the most important component to successful GIS operations is data and that the data will usually be the most costly input into individual GIS projects. It should, therefore, be a fundamental aim for all GIS operatives to secure high-quality data because without this the output from project work may lack reliability and decision-makers will not have confidence in using the information produced. This chapter seeks to give guidance towards achieving appropriate data needs. With respect to data for marine fisheries projects, they can be difficult to obtain given the major expenses in gathering data at sea and the sheer volume of data required, i.e. given that marine environments may be very expansive and that they have both horizontal and vertical dimensions. Additionally, the mobility of all aquatic environments imposes greater temporal constraints on the validity of the data collected. Until the late 1980s, virtually all data collected was in hard copy format, but since then rapid technological progress has occurred so that today almost all data gathering is via electronic instruments, meaning that data are usually in a format suitable for direct GIS use. References are made in this chapter to hard copy data collection techniques and formats because there is a legacy of older data that may need to be digitally converted and because there are still circumstances where data collection is more conveniently accomplished using traditional methods.

The use of GIS for any fisheries or aquaculture project will almost certainly be a response to a desire to improve the likely success of these activities. Success for any production activity can be measured in terms of four basic criteria: (i) economic; (ii) social; (iii) physical or biological; and (iv) sustainability. It is important to point out that when the forerunner to this technical paper was published (Meaden and Kapetsky, 1991), the criterion for success was economic (in terms of profit maximization), but it is now recognized that this criterion has entirely changed, such that success is now based on concepts surrounding sustainability, with this being achieved through either an ecosystem approach to fisheries (EAF) or through an ecosystem approach to aquaculture (EAA) (FAO, 2003; Carocci *et al.*, 2009; FAO, 2010a; Aguilar-Manjarrez, Kapetsky and Soto, 2010). This change may have a profound effect on the types of data collected as the data have to mainly support sustainability rather than economic objectives.

Personnel organizing and/or managing GIS projects will need to give detailed considerations as to what data are best required for project success, including factors such as the spatial area for the project and the scale and resolution or detail at which the study is carried out. The technical paper provides advice on all main questions that need to be asked of any data to be collected.

For their success, all production activities are controlled by the availability of "production functions", i.e. those factors or variables that, in combination, influence the success of any production activity. Some production functions will be "one-off" considerations such as the initial land purchased for aquaculture ponds, while others will be akin to operating costs as their consideration may be continual, e.g. access to markets or target fish species to catch. Production functions may be highly variable in an area, such as soil quality, bottom-sediment types or population density, while others may be highly uniform, such as sea

BOX 1
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 storm 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 take account of local biodiversity, waste deposition and benthos issues.
- Distance from other cage farms – Because of disease problems, cages should be 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 with 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.

water temperature or salinity. Production function deficiencies can be increasingly overcome, such as providing electrical energy inputs through the use of mobile generators or the supply of any oxygen deficiency through artificial oxygenation. Each production function will be of variable importance to production success, so it is important that the functions can be ranked or measured in some way. Once this is done, then some measure of weighting can be given to each production function, i.e. as part of the GIS processing procedures. The technical paper describes the main production functions controlling the success of inland aquaculture, marine fisheries and marine cage culture (Box 1). The functions controlling marine fisheries are far more varied than those influencing aquaculture because fisheries are highly variable in terms of the methods used, species targeted, scale of the activity and degree of technology used. It would also be true to say that the thematic range of GIS projects applied to marine fisheries is wider than the range currently applied to aquaculture. Once the production functions have been clearly identified, then it is the spatial data that best describes the disposition of each of these functions that will be needed as source data for any GIS work.

If data are to be of use in a GIS, they must have certain minimum requirements. Because many of these requirements are general to all computing, e.g. data should be in a uniform format, follow classification rules, have an accepted measure of accuracy, then requirements are not discussed here. Other requirements are much more specific to GIS work and they include: (i) having a temporal facet so that the user knows what the data relates to – this has implications for when data might need updating and when comparisons might be made between specified times or time periods; (ii) relating to a thematic area or to some accepted classification system. Data classifications may be very general, e.g. "marine", "fisheries" or "aquaculture", or they may be more specific, e.g. "mature cod", "mature plaice", "mature whiting",, and there can be any subclassification hierarchy within themes; (iii) data having a quantitative facet involving some kind of count or measurement – this will involve considerations of precision and accuracy; and (iv) considerations on the scale and resolution applicable to any GIS project. The scale of the project means considering the area to be covered and therefore where the boundaries are so that data needs can be properly assigned. It is often sensible to make the GIS project area coincide with national or regional boundaries, i.e. simply because data are often collected with respect to named political areas. Resolution implies a consideration of what can be shown on a map relative to the area being mapped, and this will have large implications with respect to the volumes of data that need to be collected.

The most important facet of data that are being collected and used for GIS work is the spatial facet, i.e. exactly where is the data referring to? This locational data may be referring to features that can be drawn on maps as points, lines or polygons, and the data gathered must accurately have a georeference for all objects or events that are of interest to a project. Georeferenced data are typically recorded in terms of latitude and longitude or by using some form of Cartesian coordinate system. Latitudes and longitudes are recorded either in degrees, minutes and seconds or in decimal degrees. While this information is useful, it suffers from the fact that longitudinal lines on maps all converge at the North and South Poles and thus distances between lines of longitude vary with distance from the poles. Cartesian coordinates try to overcome this problem by covering the mapped area with a regular grid of horizontal and vertical parallel lines, with these lines usually being based on the internationally recognized Universal Transverse Mercator (UTM) projection system and its unique alphanumeric referencing system. However, this georeferencing method also suffers because it is impossible to impose a true grid of parallel lines over what is a spherical object (the Earth). For mapping smaller areas, the distortion is very little, but at the world scale the UTM map projection greatly exaggerates areal and horizontal distances that are nearer to the poles. Whatever mapping projection or georeferencing system is used, it is important to use the same system for all mapping within a single project. The spatial data used for mapping can also be in the form of named areas or a descriptive code. These coded data are mainly used when doing GIS work involving polygons, i.e. spatial areas that might be named counties, land-use zones, geology types, for example. The technical paper provides various worked examples that explain georeferencing requirements and methods in more detail.

Before examining the collection of data for GIS, it is important to be familiar with considerations regarding data quality, as this largely determines the reliability of the final output from any GIS. The main factor influencing data quality is the resources of money, time and effort that can be put into data collection. However, data quality is largely scale dependent. Thus, if data were being collected for a small-scale (large area) project, then the use of the same data for a large-scale project would almost certainly be inappropriate because the resolution of the data would be insufficient for accurate mapping at the larger scale. Consideration must also be given to:- (i) the accuracy of data used and to their precision (how precisely has a measurement been recorded); (ii) standardizing the methods used for data collection; (iii) the use of appropriate classification systems and thematic categories; (iv) the timeliness of the data; and (v) possible sources of error in any data collected. In order to try to improve the quality of data, especially for GIS use, various attempts have been made to establish so-called "data standards". These are mainly internationally recognized standards managed by the International Organization for Standardization (ISO), though there are also geographic standards pertaining to individual countries or organizations. Standards set by ISO and others mainly relate to any data products produced, to the format and manner in which data are transferred, to data quality standards, and to metadata standards (the recording of information about any data collected).

The actual data to be collected are conventionally categorized as being in one of two classes: (i) primary data, which are "raw" (unprocessed) data that are directly collected for a specific project; and (ii) secondary data, which are data that have been previously collected, usually for totally unrelated projects. The collection of primary data entails considerations of the availability of time, funding, trained personnel and any equipment needed, as well as whether there is existing secondary data that might be used. It will also be necessary to establish a sampling strategy. This is because it will almost certainly be impossible to collect all data and therefore considerations of what will be a representative sample are important, including how best to define the sampling strategy so as to achieve statistically significant results. At its most basic level, primary data may be collected using a range of methods involving no equipment. These methods include: (i) direct sketch mapping, e.g. fishers might be asked to sketch in on a local map areas where they prefer to fish for particular species; (ii) face-to-face interviewing, which may be useful for obtaining socio-economic information on fisher group activities; (iii) questionnaires, which can usefully be carried out by mail, telephone or face-to-face on any subject; and (iv) filling in pre-printed forms. All data collected in the ways described here will have to be converted into digital formats for storage and use in a GIS.

A vast range of equipment exists that allows for primary data collection, and nowadays nearly all of the data can be directly captured in digital formats that are easily structured for direct GIS use. The equipment can be best described in an approximate hierarchy going from basic or simple to highly complex. At the simplest level, there is a range of electronic "read-out" equipment that comprises small devices, frequently handheld, for measuring a variety of parameters, including water temperature, light intensity, water flow rates, pH levels, or the size and weight of objects. These devices vary in their complexity and utility according to their accuracy, ability to store data, ease of use, range of functions, etc., and their cost usually reflects this diversity. At a slightly more sophisticated level, there are digital cameras. They have the ability to accurately record in a convenient visual format any localized scene or layout of objects of interest, and they are useful in recording temporal rates of change, e.g. the shrinkage rates of glaciers or river water levels. Cameras can also be aircraft mounted to give direct overhead (mapping) imagery.

Data loggers and personal digital assistants (PDAs) – the next level of equipment –have a greater range of functionality. Data loggers typically allow for a wide range of data to be captured and stored on any thematic area of interest. They are generally battery powered, portable and equipped with a microprocessor, plus internal memory for data storage, and they may have various sensors. They may be either handheld or utilized in situ, perhaps in a river to capture water-flow rates, or they may be attached to free floating or tethered buoys in the sea, or installed on board autonomous underwater vehicles (AUVs). They can usually be directly interfaced with computers for the downloading of data. PDAs may best be conceived as handheld computing and communication devices: they have access to the Internet, a small colour screen, cell phone capability and Wi-Fi connectivity. They deploy touch-screen technology and can be loaded with a wide range of software. In many senses, PDAs replicate the functionality of desktop computers, though their small screen size is far from optimum for GIS use. However, technological progress in the field of PDAs is so rapid that it is difficult to conceive of their possibilities within even the next decade.

At a much higher level of sophistication are global positioning systems (GPS) and acoustic sonar devices. GPS are highly sophisticated because they rely on complex satellite systems. However, here, the interest is in their ability to precisely ascertain a wide range of information all of which can be given an exact georeference. Thus, for any location given in latitude/longitude or as UTM coordinates, etc., details can be captured on a thematic object or event, plus any associated attributes such as altimetry, speed or direction of travel, and local time. Having instant access to georeferenced data is a huge advantage when gathering data for GIS purposes. GPS may, additionally, be either handheld or they may be incorporated into other technology such as vehicle navigation, vessel monitoring, cell phones and even digital watches and cameras. Finally, acoustic sonar devices are especially relevant to fisheries GIS because this radar-based technology allows for the real-time capture of a variety of underwater data, including the location of fish shoals and sea-bottom topography and substrates. The technology is usually deployed as attachments to the hull of vessels, but it may be incorporated into AUVs or onto the headrope of large trawlnets. The data received from acoustic sonar use are in the form of echolocations, whose patterns are translated into identifiable features by trained users, and which can form the basis of much underwater mapping.

Data collection methods and data sources for secondary data have both undergone fundamental changes over the past few decades. Changes have involved the types of data sources, the form that data takes, delivery mechanisms for data, and the breadth and volume of data available. The move away from largely fragmented paper-based data sourcing to more centralized digital data sourcing means that almost none of the data sources or the media and formats used 20 years ago are now relevant. It is likely that 95 percent of fisheries or aquaculture-related data are now available in digital format, though there remains important "historical" mapping and tabular data that awaits digitizing when the need occurs. Most useful secondary data are stored in databases on powerful server computers awaiting delivery when required via the Internet, though delivery may also be via CD–ROM or DVDs. There are a vast number of fisheries and aquaculture secondary data sources supplying anything from highly generalized small-scale data to highly specialized large-scale data. The sources are best located via online searching, but the technical paper provides details on some of the main fisheries and aquaculture access portals and on sources of data on selected themes in inland fisheries, i.e. as examples of the types of secondary data available. For GIS-related work in inland fisheries or aquaculture, a major source of secondary data will be from the national mapping or hydrographic agencies in any specific country, whose addresses can be obtained by online searching. A problem to obtaining much secondary data is that the data may be costly, especially if they are in a digital format, and there may also be copyright rules that apply. GIS users seeking mapping may find that some larger private companies may allow free access to mapping, e.g. Google Earth's satellite imagery and the Environmental Systems Research Institute (ESRI) provide a range of ready-to-use, high-quality data for GIS visualization and analysis. A number of cautionary questions that need addressing with respect to obtaining secondary data are, for example: (i) How old is the data?; (ii) Does the data come with usage rules attached?; (iii) What is the scale or resolution of the data?; (iv) Might there be cheaper sources of data required?; and (v) Does the data provide exactly what the project needs? Secondary data for capture fisheries purposes are generally more scarce than the data for aquaculture. This is because most fisheries data have been gathered on very specific parameters, for specific projects covering relatively small spatial areas. Such data may also be hard to locate. This means that GIS projects may concentrate only on spatial areas or themes where it is known that sufficient data exists, or that projects aims are only defined in terms of what data are known to be available.

Suitable data for GIS projects may sometimes be difficult to locate or acquire. A means of coping with this can frequently be found through the use of proxy data; this is data that may not be directly related to the exact data required but can nevertheless be a substitute for the real data. For instance, if water temperature data cannot be easily acquired, then data on air temperatures for an area of interest usually shows a high correlation with water temperatures, and 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. Sometimes two proxy data sets can be combined in order to gain a realistic required map; for example, to obtain a land price or value map, using GIS capabilities it might be sensible to combine a map of population density with a soil quality map. This is based on the fact that population density and soil quality are strong indicators of the demand for, and thus cost of, land. In the marine or freshwater realms, existing aquatic habitats often act as a close proxy of the likely fish species to be found, though in these realms the strength of any relationship would be far more generalized and less certain. Notwithstanding this uncertainty, much recent GIS work has utilized proxy data to establish habitat suitability or essential habitats for aquatic species.