Chapter 10: Glossary

10.1 Glossary

An comprehensive guide to the terms used in GIS is provided by:

McDonnell, R. and Kemp, K., 1995, *International GIS dictionary*. Cambridge: GeoInformation International.

Accuracy: The difference between a set of representative values and the actual values. The accuracy of a *point* location would be the difference between the point's coordinates in the GIS and the coordinates accepted as existing in the real world. Section 3.4.

Animated GIF: A *GIF* is a bitmap file format often used on the World Wide Web. An animated GIF is a series of individual GIF frames joined together to create an *animation*. It is perhaps the easiest way to create and view simple animations. Section 6.7.

Animation: A collection of static images joined together and shown consecutively so that they appear to move. Section 6.7 (and also section 1.3).

Arc: See *line*.

ArcGIS: A commonly used GIS package from Environmental Systems Research Institute (ESRI). It updated and merged two of ESRI's earlier products: *ArcInfo* and *ArcView*.

ArcInfo: Was the market leading GIS software package when GIS computing was workstation-based. Is now available for NT but has in some ways been superseded by desktop solutions such as *ArcGIS*, *ArcView*, and *MapInfo*. *ArcGIS*, *ArcInfo* and *ArcView* are produced by Environmental Systems Research Institute (ESRI).

ArcView: A commonly used desktop GIS software package produced by Environmental Systems Research Institute (ESRI). Its sister product *ArcInfo* provides more functionality but is harder to use.

Area cartogram: These are *choropleth maps* that have been distorted so that the size of the *polygons* is not proportional to the polygon's area, but is instead proportional to another of the polygon's variables such as its total population. Section 6.6.

Areal interpolation: The process by which data from one set of source *polygons* are re-districted onto a set of overlapping but non-hierarchical target polygons. Section 4.7 (and also section 7.4).

Areas: See *polygons*.

Attribute data: Data that relate to a specific, precisely defined *location*. The data are often statistical but may be text, images or multimedia. These are linked in the GIS to *spatial data* that define the location. Section 2.2 (and also sections 1.2 and 2.1).

Attribute querying: A *query* that extracts features from a *layer* based on the value of its attribute data: for example, 'select polygons with an unemployment rate greater than 15%' would be an attribute query. Section 4.2.

AVI: A video file format that can be used to publish *animations*. Section 6.7.

Backcloth: A *raster* image, usually a map or aerial photograph, used to provide contextual information behind other *layers* of data. Section 3.3.

Blunder: The introduction of *error* by mistakes. Section 3.5.

Buffering: A buffer is a polygon that encloses the area within a set distance of a spatial features. *Points*, *lines*, and *polygons* can all have buffers placed around them. For example, if a user is interested in all areas within 1km of a church, a buffer would be placed around all the points representing churches. This would create a new layer consisting of polygons representing those areas within 1km of a church. Section 4.4.

Capture: See *data capture*.

Cartogram: See *area cartogram*.

Centroid: A point at the geometric centre of a *polygon*. This can be used to represent a *polygon* as a *point*. Section 2.3.

Choropleth maps: Maps of quantitative data that show patterns by using different colours or different shading for polygons classed in some way. For example, a map of polygon-based unemployment rates (expressed as percentages) might sub-divide rates into 0-5, 5-10, 10-15 and 15-20 and shade the polygons accordingly. Section 6.2.

Coordinate pair: An x and y coordinate used to represent a location in two-dimensional space, for example (6.523,4.910). Section 1.2.

Correlation: A form of statistical modelling that attempts to summarise how one dataset will vary in response to another. A correlation coefficient of +1.0 means that where there are high values in one set there will be high values in the other, while a correlation coefficient of –1.0 means that where there are high values in one set there will be low values in the other. A correlation coefficient of 0.0 means that there is no discernible relationship between the two sets. This is a form of *global analysis* as it only provides a single summary statistic for the entire study area. Section 7.2.

Coverage: See *layer*.

Dangling node: A *node* that should join with another node to join two or more *lines* together, but which does not join. This will result in holes in *topology*. Section 3.3.

Data capture: The process by which data are taken from the real-world (*primary source*), or from a secondary source such as a paper map, and entered into *GIS* software. From primary data this is usually through the use of *Global Positioning Systems* or *remote sensing*. For secondary data it is usually through *digitising* or *scanning*. Chapter 3 (and also section 1.4).

Database Management Systems: Software systems specifically designed to store *attribute data*. Section 1.2.

Date-stamping approach: A way of handling time in GIS where time is treated as an attribute. Each feature has date stamps attached that define the times that it was in existence. Section 5.4.

DBMS: See *Database Management Systems*.

DEM: See *Digital Terrain Model*.

DGPS: See *Differential GPS.*

Diachronic analysis: A form of analysis drawn from systems theory in which change over time is examined by comparing a large number of states, none of which are assumed to be in equilibrium. Section 5.2.

Differential GPS: A way of collecting *Global Positioning Systems* data with increased accuracy. It involves using a fixed base station at a known position to help find the location of a roving receiver. Section 3.8.

Digital Elevation Model: See *Digital Terrain Model*.

Digital Terrain Model: A data model that attempts to provide a three-dimensional representation of a continuous surface. Often used to represent relief. Section 2.5.

Digitiser: In GIS terminology this refers specifically to a *digitising table* or *digitising tablet*. Section 3.3.

Digitising: In GIS this has a more precise meaning than in other disciplines. It usually refers to extracting coordinates from secondary sources such as maps to create vector data. Section 3.3.

Digitising table: A flat table with a fine mesh of wires under the surface used to allow accurate *digitising* of paper maps through the use of a *puck*. Section 3.3.

Digitising tablet: Similar to a *digitising table* only smaller. Section 3.3.

Dissolve: An operation in which adjacent *polygons* are merged if a selected feature of their *attribute data* are the same. An example might be merging polygons representing fields to create a new *layer* containing crop type. Section 4.4.

Drape: Involves laying features over a *digital terrain model* to provide information on features that lie on the terrain. The terrain model provides the shape of the terrain. Draped features may then include a *satellite image* of the terrain to show land-use, and *vector data* to show features such as roads. Section 2.5.

DTM: See *Digital Terrain Model*.

Dublin Core: A *metadata* standard designed to allow digital datasets to be found and evaluated. Section 9.1.

Ecological fallacy: The mistake of assuming that where relationships are found among aggregate data, these relationships will also be found among individuals or households. Section 7.2.

Edge-matching: See *rubber-sheeting*.

Error: In the context of *GIS* this means the difference between the real world and its digital representation. Section 3.5.

Error propagation: As layers of data are integrated through *overlays* the *error* present on the output *layer* will become the cumulative total of the *error* present on all the input layers. Section 4.6 (and also section 7.2).

Exploratory analysis: Statistical or visualisation techniques that attempt to produce a good summary of the data or the patterns with them. Section 7.2.

Fly-through: Often used to view *digital terrain models*. In a fly-through a user is given the functionality to allow him or her to move through the terrain in what appears to be three dimensions, thus giving the illusion of flying. It is an effective way of exploring a virtual landscape from different directions. Section 6.7 (and also section 1.3).

Gazetteer: Often used to standardise place names or to locate place names within a hierarchy. These are often stored in a *Relational Database Management System*. Section 2.2.

GDA: See *Geographical Data Analysis.*

Geary's coefficient: A statistical technique that measures the degree of *spatial autocorrelation* present in the data. It is a form of *global analysis.* Section 7.3.

Geary's G_i: This is a *local analysis* form of *Geary's coefficient* that produces a measure of *spatial autocorrelation* for each *location* in the dataset. Section 7.3.

Geographical Data Analysis (GDA): A way of analysing data that explicitly incorporates information about *location* as well about attribute. This term may be used almost interchangeably with *spatial analysis.* Chapter 7.

Geographical Information Science: Methods of exploring and analysing spatially referenced data that take account of the benefits and limitations of such data. Section 1.2 (and also Chapter 7).

Geographical Information System: A computer system that combines *database management system* functionality with information about *location*. In this way it is able to capture, manage, integrate, manipulate, analyse and display data that are spatially referenced to the earth's surface. Chapter 1.

Geographically weighted regression (GWR): A form of *regression* modelling that explicitly incorporates the role of location. This is a form of *local analysis*. Section 7.3.

Geo-referencing: The process of proving a *layer* of data with a real-world coordinate system such as the British National Grid or *latitude* and *longitude*. Section 3.4.

GIF: A commonly used bitmap file format often used on the World Wide Web.

GIS: See *Geographical Information System.*

GIS data: Data stored in a GIS are represented in two ways: *attribute data* says what the feature is, and *spatial data* says where it is using *points*, *lines*, *polygons*, or *pixels*. Section 1.2.

GISc: *Geographical Information Science.*

Global analysis: Forms of statistical analysis that provide an average measure of a relationship or relationships across the study area. Traditional *correlation* and *regression* techniques do this. From a GIS perspective they are flawed in that they do not allow for any geographical variations in the pattern so *local analysis* techniques are seen as more relevant. Section 7.2.

Global Positioning Systems (GPS): A system based on satellites that allows a user with a receiver to determine precise coordinates for their location on the earth's surface. These are a *primary source* of *spatial data*. Section 3.8.

GPS: See *Global Positioning Systems*.

Graphic primitive: The basic representations of spatial features used in GIS. These are usually *points*, *lines*, *polygons* or *pixels*. Sections 1.4, 2.3 and 2.4.

GWR: See *Geographically weighted regression*.

Head-up digitising: The process by which *vector data* are extracted from *raster* scans using a cursor on-screen. Section 3.3.

Idrisi: A raster based GIS software package produced by Clark Labs, Clark University

Interpolation: A method of reallocating *attribute data* from one spatial representation to another. A simple example is to reallocate data from sample *points* to *polygons* using *Thiessen polygons*. *Kriging* is a more complex example that allocates data from sample points to a *surface*. Section 4.4.

Isolines: A line joining points of equal value. The most common example is the contour line on a map. Isobars showing lines of equal pressure on weather maps are another example. Section 6.2.

Java: A computer programming language often used to create Internet applications. Section 6.7.

JPEG: A commonly used bitmap file format often used on the World Wide Web.

Key: In the context of *Relational Database Management Systems* this refers to a common field that can be used to join two or more tables. Section 2.2.

Key dates approach: A way of handling time in a GIS where the situation at different times is represented by different *layers*. Section 5.4.

Kriging: A form of statistical modelling that *interpolates* data from a known set of sample *points* to a continuous *surface*. Section 7.3.

Latitude: The angle of a location on the earth's surface from the equator expressed in degrees north or south. The Arctic Circle, for example, is at approximately latitude 66° North. Section 3.4.

Layer: The GIS data model represents the world by sub-dividing features on the earth's surface according to a specific theme. Each theme is then *geo-referenced*. Examples of layers for a study area might include: roads, railways, urban areas, coal mines, etc. A layer usually consists of both *spatial* and *attribute* data. Section 2.6 (and also section 2.1).

Line: A spatial feature that is given a precise location that can be described by a series of *coordinate pairs*. In theory a line has length but no width. Sections 1.2 and 2.3.

Local analysis: Forms of statistical analysis that allow relationships to vary across a study area by providing summary statistics for many locations. The results are usually best presented in map form. Examples of this type of technique include *Geary's Gi*

and *Geographically Weighted Regression*. The opposite approach is *global analysis* where only a single summary statistic is provided for the average relationship across the study area. Section 7.2.

Location: The position of a feature on the earth's surface. In GIS this is usually explicitly defined in terms of precise coordinates. Chapters 1 and 2.

Location-allocation models: Models that attempt to find the optimum location for a feature based on information about other features. An example might be to find the best location for an industrial plant based on information about the transport network and the locations of raw materials and markets. Section 7.3.

Longitude: The angle of a location on the earth's surface usually expressed in degrees east or west of the Greenwich Meridian. New York, for example, is at approximately 74° West. Section 3.4.

Map algebra: A form of *overlay* used with *raster data*. In it the values for *pixels* on the output layer is calculated by performing a mathematical operation on the pixels from the input layers. The calculation may be arithmetic (addition, subtraction, multiplication, etc.) or Boolean (and, or, not, etc.). Section 4.6.

MapInfo: A commonly used desktop GIS software package produced by the MapInfo Corporation.

MAUP: See *Modifiable Areal Unit Problem*.

Metadata: Data that describe a dataset to allow others to find and evaluate it. Section 3.10.

Modifiable Areal Unit Problem (MAUP): Where data are published using totals for arbitrary areas such as administrative units, the patterns that they show may be simply the effect of the administrative units rather than genuine patterns among the underlying population. Section 7.2.

Moran's coefficient: A form of statistical modelling that measures the degree of *spatial autocorrelation* present in the data. Section 7.3.

MPEG: A video file format that can be used to publish *animations*. Section 6.7.

Network: A *topological* GIS data structure that uses a series of lines to describe, for example a transport or river network. Section 2.3 (and also section 7.3).

Network analysis: Usually used to analyse flows along a *network*. An example is finding the shortest path between two locations on a road network perhaps taking into account the differing speeds and fuel costs of different types of roads. Section 7.3.

Node: The start or end point of a *line* segment. As such a node is often the point at which lines intersect. Section 2.3.

Non-spatial data: See *attribute data*.

Object-orientated approach: A way of modelling the world that allocates entities to hierarchical classes. Section 2.1.

Overlay: A formal geometric intersection between two or more *layers* of data. A layer produced by an overlay will contain the merged *spatial data* and *attribute data* from both of the input layers. Section 4.6.

Pixels: The small units that sub-divide space to make up a *raster surface*. They are usually small grid squares. Sections 1.2 and 2.4.

Points: Spatial features that are given a precise *location* that can be described by a single *coordinate pair*. In theory a point has neither length nor width. Sections 1.2 and 2.3.

Polygons: Spatial features that are *areas* or *zones* enclosed by precisely defined boundaries. The boundaries of a polygon are formed from one or more *lines*. Sections 1.2 and 2.3.

Polyline: A term for a *line* used by some GIS packages.

Precision: The number of decimal places to which a value is given. This usually far exceeds its *accuracy*. For example, a GIS might give the coordinate of a *point's* location for building to ten decimal places providing a value that is precise to fractions of a centimetre. In reality this value may only be accurate to the nearest ten meters. Section 3.5.

Primary source: In GIS terms this usually means a digital data source that is derived directly from the real world such as through *Global Positioning Systems* or *remote sensing*. Section 3.8.

Projection system: A method by which features on a curved earth are translated to be represented on a flat map sheet. This involves converting from *longitude* and *latitude* to x and y coordinates. Section 3.4.

Proximity measure: Usually an n by n matrix that gives a measure of the influence each location i has on each other location j. This is often expressed as a weighting W_{ij}. Section 7.3.

Puck: A hand held device used with a *digitising table* or *digitising tablet*. It is used to point to an exact location in order to capture its coordinates. Section 3.3.

Quadrat analysis: Analysis where the study area is sub-divided into regular grid squares and the number of occurrences of a phenomenon in each square is counted. The resulting pattern can then be mapped. Quadrat analysis is not a particularly satisfactory technique as the results are too reliant on the size and position of the grid squares. Better techniques such as kernel estimations are described in the literature.

Quadtree: A way of encoding *raster data* that attempts to reduce storage requirements by avoiding sub-dividing homogeneous areas rather than storing values for every *pixel*. Section 2.4.

Quality: In the context of *GIS* data, quality usually refers to how fit the data are for a particular purpose. Section 3.5.

Querying: The process by which data are retrieved from a database in order to gain information from it. Section 4.2 (and also sections 2.2 and 2.3).

Raster data model: A way of representing the earth's surface by sub-dividing it into small *pixels*, usually square cells. Each pixel has values attached to it providing *attribute data* about the pixel. Section 2.4 (and also section 2.1).

Raster-to-vector conversion: The process by which *vector* features (*points*, *lines* and *polygons*) are automatically extracted from *raster data*. This usually requires a large amount of user input and is often error prone. Section 3.7.

RDBMS: See *Relational Database Management Systems*.

Reference points: A small number of points used to *geo-reference* a *layer*. Often the four corners of the layer are used. Once the layer has been *digitised* we know the coordinates of the reference points in inches from the bottom left-hand corner of the *digitising table* or *digitising tablet*. We also know their *locations* in real-world units from the map. This allows us to convert the entire layer's coordinates from digitiser inches to real-world coordinates. Section 3.4.

Regression: A form of statistical modelling that attempts to evaluate the relationship between one variable (termed the dependent variable) and one or more other variables (termed the independent variables). It is a form of *global analysis* as it only produces a single equation for the relationship thus not allowing any variation across the study area. *Geographically Weighted Regression* is a *local analysis* form of regression. Section 7.2.

Relational Database Management Systems: Software systems that store data in such a way that tables can be joined together by linking on a common item of data, termed a *key*. Section 2.2.

Relational join: The way by which two or more tables from a *Relational Database Management System* can be joined together based on one or more common items or *keys*. Section 2.2.

Remote sensing: The process by which *satellite images* are created by scanning the earth's surface using sensors on satellites. Section 3.8 (and also section 2.4).

RMS Error: See *Root Mean Square Error*.

Root Mean Square Error (RMS): A measure of the average *error* across a map. It is used in *digitising* to give an approximate measure of the difference between the real-world coordinates and the registration *points* on the digital *layer*. Section 3.5 (and also section 9.4).

Rubber-sheeting: The process by which a *layer* is distorted to allow it to be seamlessly joined to an adjacent layer. Often this has to be done when layers created from adjacent map sheets are joined together. It is a process that inevitably introduces some *error*. Section 4.3.

Run-length encoding: A way of encoding *raster data* that reduces storage requirements by creating linear groups of identical *pixels* rather than storing the values of each pixel individually. Section 2.4.

Satellite images: *Raster* models of the earth's surface produced from sensors on satellites. Section 3.8 (and also section 2.4).

Scanning: The process by which *raster* data is *captured* from paper maps. Section 3.2.

Segments: See *lines*.

Sliver polygons: Small *polygons* formed as a result of overlaying two or more *layers* of *vector data*. These are formed due to small differences in the way that identical *lines* have been *digitised*. Section 4.6.

Space: In a GIS context this means position on the earth's surface. Its meaning is very similar to *location*. Chapters 1 and 2.

Space-time composite: A way of handling time in GIS that preserves *topology* by sub-dividing space into a small set of areas that can then be re-aggregated into the arrangement that existed at different dates. Section 5.4.

Spans: A *raster* based GIS software package produced by PCI-Geomatics

Spatial analysis: A way of analysing data that explicitly incorporates information about *location* as well about *attribute*. This term may be used almost interchangeably with *geographical data analysis.* Chapter 7.

Spatial autocorrelation: The degree to which a set of features tend to be clustered together (positive spatial autocorrelation) or be evenly dispersed (negative spatial autocorrelation) over the earth's surface. This is often measured using either *Geary's coefficient* or *Moran's coefficient*. When data are spatially autocorrelated the assumption that they are independently random is invalid, so many statistical techniques are invalidated. Section 7.3.

Spatial data: Data that define a *location*. These are in the form of *graphic primitives* that are usually either *points*, *lines*, *polygons* or *pixels*. Sections 1.2, 2.3 and 2.4.

Spatial querying: A *query* that extracts features from a layer based on their location; for example, clicking on a *point* and listing its *attribute data* is a spatial query. Section 4.2.

SQL: See *Structured Query Language*.

Structured Query Language (SQL): A language used by many *Relational Database Management Systems* to manipulate their data. Section 2.2.

Surfaces: A surface is a way of modelling *space* that attempts to treat it as continuous rather than sub-dividing it into discrete features such as *polygons*. Surfaces are usually modelled either as *raster data* or *digital terrain models*. Sections 2.1, 2.4 and 2.5.

Synchronic analysis: A form of analysis drawn from systems theory in which change over time is examined by comparing the situation at two points in time when the system is assumed to be in equilibrium. Section 5.2.

Temporal data: Data that explicitly refer to time. Chapter 5.

Tessellation: A sub-division of *space* into discrete elements. *Raster surfaces* sub-divide space into regular tessellations such as *pixels*. *Polygons* are examples of irregular tessellations. Section 2.1.

Theme: See *layer*.

Thiessen polygons: A method of allocating *space* to the nearest *point*. The input *layer* will contain a set of points. The output layer, containing the Thiessen polygons, will contain *polygons* whose boundaries are *lines* of equal distance between two points. Section 4.4.

Tic points: See *reference points*.

TIN: See *Triangular Irregular Network.*

Topology: The description of how spatial features are connected to each other. Section 2.3 (see also section 5.3 for the problems of creating topology that connects features through time as well as space).

Travelling Salesman Problem: A form of *network analysis* that attempts to find the shortest or cheapest route between a number of locations on a *network.* Section 7.3.

Triangular Irregular Network: A data structure that produces a continuous *surface* from *point* data. Often used to create a *digital terrain model*. Section 2.5.

Uncertainty: A measure of the amount of doubt or distrust with which the data should be used. Section 3.5.

Vector data model: Divides space into discrete features, usually *points*, *lines* or *polygons*. Section 2.3.

Vector-to-raster conversion: The process by which *vector data* are converted to *rasters*. This is usually automated. Section 3.7.

Voronoi diagrams: See *Thiessen polygons*.

Web-based mapping: Maps created for use on the Internet so they often have some interactive functionality. Web-based mapping is not well developed with *vector* file formats. Section 6.7.

Whole-map analysis: See *global analysis*.

Zones: See *polygons.*