Coordinate Reference Systems, an extensive introduction

Daniele Romagnoli,
Ing. Simone Giannecchini
NATO Undersea Research Center
Military Oceanographic Dept.
Viale San Bartolomeo 400
19138 La Spezia (SP), ITALY
+39 0187 527441
giannecchini@saclantc.nato.int

March 19, 2006

Contents

1 Introduction to Coordinate Reference Systems. 2

2 Basic Concepts. 4

2.1 Datum. 4

2.1.1 Geodetic Datum. 4

2.1.2 Vertical Datum. 6

2.1.3 Other Datums. 7

2.2 Coordinate Reference System. 8

2.2.1 Principal sub-types of coordinate reference systems. . . 8

2.2.2 Other sub-types of coordinate reference systems. 9

2.3 Coordinate System. 12

2.3.1 Coordinate System Axis. 13

3 Standardization. 15

4 Appendix A: Terms Definitions 16

5 References. 20

List of Figures

1 Geodetic Terms 6
2 Ellipsoidal Height and Gravity-Related Height 7

1 Introduction to Coordinate Reference Systems.

Geodesy is the applied science that aims at determining the size and shape of the earth. In a more practical and local sense, this may be understood as the determination of the relative positions of points on or near the earth's surface.

Positions on and near the Earth's surface can be described by systems of spatial referencing. These are of two basic types:

1. those using coordinates;
2. those based on geographic identifiers (for example postal addresses).

It is worth to point out that only the first type will be analyzed in this document.

The definition of coordinate can be found in the document "OGC Abstract Specification Topic 2, Spatial referencing by coordinates (Version 3)" [5]. By quoting this definition:

*A **coordinate** is one of n scalar values that define the position of a single point. A coordinate tuple is an ordered list of coordinates. The coordinate tuple is composed of one, two or three spatial coordinates. The coordinates shall be mutually independent and their number shall be equal to the dimension of the coordinate space.*

Coordinates are unambiguous only when the coordinate reference system to which those coordinates are related has been fully defined. A coordinate reference system is a coordinate system that has a reference to the Earth. The concept of a coordinate reference system (CRS) captures the choice of values for the parameters that constitute the degrees of freedom of the coordinate space. The fact that such a choice has to be made, either arbitrarily or by adopting values from survey measurements, leads to the large number of coordinate reference systems in use around the world.

It is a task of specific organizations to provide a set of standards, all those who are involved in the study of these concepts and in the development of the applications related to them, might refer to. The *Open Geospatial Consortium* (**OGC**[1]) and the *Technical Committee 211* (**TC211**[2]) of the International Standards Organization (ISO) accomplish this task.

As far as the topics dealt in this document, we will heavily refer to "OGC Abstract Specification Topic 2, Spatial referencing by coordinates (Version 3)"[5].

2 Basic Concepts.

Some of the following concepts are necessary to understand the basic concepts of Spatial Referencing, and they will be best defined in the specifications reported at the end of this document in the Appendix A. Those specifications are be quoted from the "OGC Abstract Specification Topic 2, Spatial referencing by coordinates (Version 3)"[5].

A spatial reference is the description of positions in the real world. A sequence of numbers (the coordinates) designate the position of each point in n-dimensional space. The set of mathematical rules to specify how coordinates are to be assigned to points is a Coordinate System.

A **Coordinate Reference System** (**CRS**) is a **Coordinate System** which is related to the real world by a **datum**, that is a parameter (or a set of parameters) that define the position of the origin, the scale, and the orientation of a coordinate reference system with respect to.

2.1 Datum.

A datum specifies the relationship of a coordinate system to the earth, thus creating a coordinate reference system. A datum can be used as the basis for one-, two- or three-dimensional systems. In some applications for an Engineering CRS (that we will show better afterwards) the relationship is to a moving platform.

Five subtypes of datum are specified: Geodetic, vertical, engineering, image and temporal. Each of them can be associated only with specific types of CRS.

Presently, we will describe in a slightly detailed manner the Geodetic Datum.

2.1.1 Geodetic Datum.

This kind of Datum is used with 3-dimensional (or horizontal 2-dimensional) CRS. It is needed to describe large portions of the earth's surface as well the whole surface of the earth. The most accurate reference shape approximating the earth is the *geoid*, that is defined as the surface having equal gravity potential which means being approximately at mean sea level.

The Geodetic Datum requires mandatorily the definition of 2 elements: an *ellipsoid* and a *prime meridian*.

1. **Ellipsoid**
   To facilitate easier spatial calculations, the geoid is approximated by the nearest regular body, an oblate ellipsoid, in which the oblateness corresponds to the flattening of the physical earth at the poles due to the earth's rotation. The ellipsoid is a reasonably accurate approximation of the geoid.

   An ellipsoid specification shall not be provided if the datum type is not geodetic but, at the same time, an ellipsoid shall be specified with every geodetic datum, even if the ellipsoid is not used by calculation, due to the fact that the ellipsoid is needed for the determination of the associated geodetic datum. An ellipsoid is defined either by its semi-major axis and inverse flattening (see Appendix A), or by its semi-major axis and semi-minor axis.

2. **Prime Meridian**
   A prime meridian defines the origin from which longitude values are specified. Its definition is mandatory if the datum type is geodetic. Most geodetic datums use Greenwich as their prime meridian. If the prime meridian name is "Greenwich" then the value of Greenwich longitude shall be 0 degrees. As well as the Ellipsoid description, the prime meridian description is not used for any datum type other than geodetic.

The ellipsoid forms the basis of the best-known type of coordinate reference system: the **Geographic CRS**. The position of a point relative to the ellipsoid is expressed by means of geographic coordinates: *geodetic latitude* (φ) and *geodetic longitude* (λ). The height above the ellipsoid (h: *ellipsoidal height*) is an inseparable element of a geographic 3D coordinate tuple.

The **geodetic latitude** it is the angle from the equatorial plane to the perpendicular to the ellipsoid through a given point, northwards treated as positive.

The **geodetic longitude** it is the angle from the prime meridian plane to the meridian plane of a given point, eastward treated as positive.

The **ellipsoidal height** (also known as **geodetic height**) is the distance of the given point, from the ellipsoid measured along the perpendicular from the ellipsoid to this point, positive if upwards or outside of the ellipsoid.

The geodetic height is the sum of the gravity-related height and the geoid height. The first, is defined as the height dependent on the Earth's gravity field, and thus, in most cases, it is the approximations of the distance of the point above the mean sea level. The second, is the height of the geoid above the ellipsoid.

Terms and definitions here discussed can be better understood by looking at Figure 1 and Figure 2.

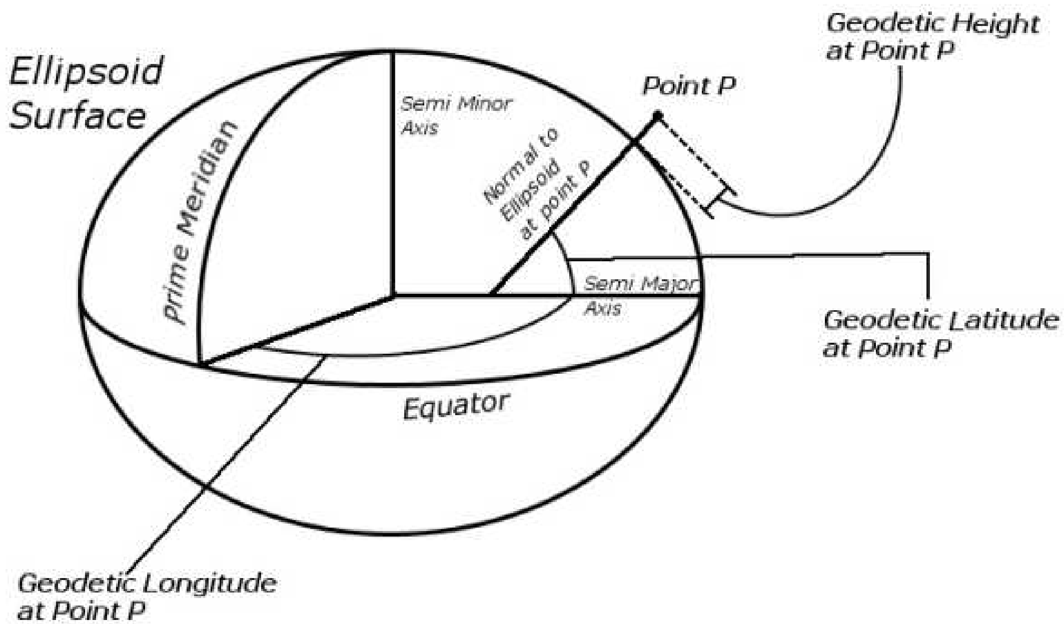

Figure 1: Geodetic Terms

2.1.2 Vertical Datum.

A vertical datum can only be associated with a vertical coordinate reference system. This datum describes the relation of gravity-related heights (or depths) to the Earth.

The most used kind of vertical datum is the Geoidal datum which assume that the zero value of the associated (vertical) coordinate system axis is defined to approximate a constant potential surface (usually the geoid). So, by using the geoid approximation, the vertical datum is related to mean sea level in most cases.

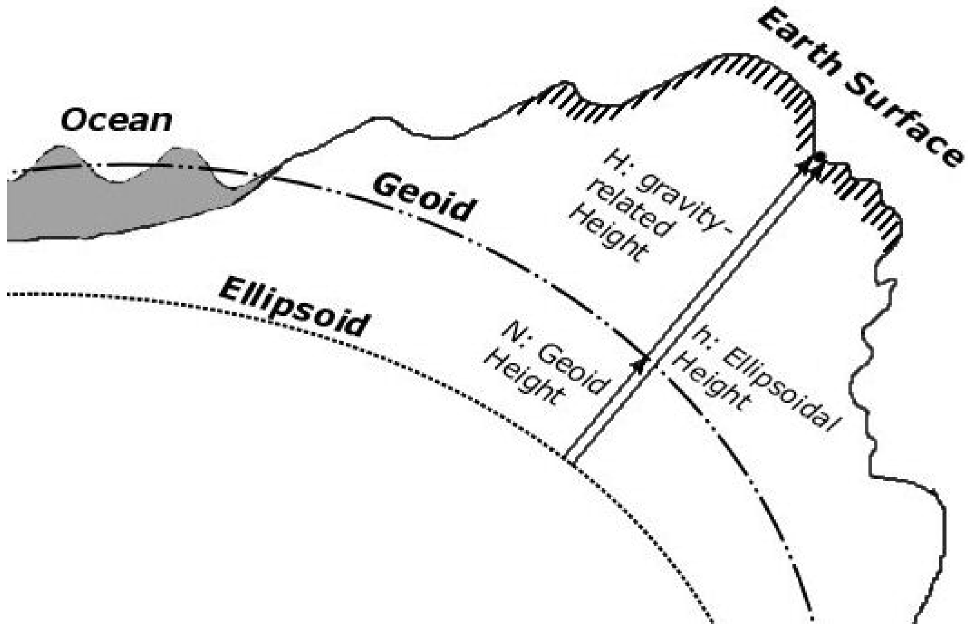

Figure 2: Ellipsoidal Height and Gravity-Related Height

2.1.3 Other Datums.

Image datum and *engineering datum* are both used in a local context only: to describe the origin of an image (by establishing if the image grid lines run through the center of pixels instead of running on the pixels corner) and the origin of an engineering (or local) coordinate reference system. A *temporal datum* is used to define the origin of the time axis in a *temporal coordinate reference system*.

2.2 Coordinate Reference System.

A coordinate reference system consists of one coordinate system that is related to the earth through one datum which specifies the relationship of a coordinate system to the earth. This combination "*coordinate system + datum*" is a coordinate reference system.

Each datum subtype can be associated with only specific types of coordinate systems.

A coordinate reference system shall not change with time, with the exception of engineering coordinate reference systems defined on moving platforms such as cars, ships, aircraft and spacecraft. The date of realisation of the datum shall then be included in its definition.

In several application fields, we need to transpose positions located on a curved surface as that of the ellipsoid, into a planar surface like a 2D map (i.e. a city map). This kind of operation is computed by performing a coordinate conversion called **map projection**. This is a coordinate conversion from an ellipsoidal coordinate system to a plane.

2.2.1 Principal sub-types of coordinate reference systems.

We can distinguish between some sub-types of coordinate reference systems. This division is suitable by different way in which a CRS deal with earth curvature. This has a direct effect on the portion of the earth's surface that can be covered by that type of CRS with an acceptable degree of error. Thus the following principal sub-types of coordinate reference system are distinguished:

Geocentric: Type of coordinate reference system that deals with the earth's curvature by taking the 3D spatial view, which obviates the need to model the earth's curvature. The origin of a geocentric CRS is at the approximate centre of mass of the earth.

Geographic: Type of coordinate reference system based on an ellipsoidal approximation of the geoid. This provides an accurate representation of the geometry of geographic features for a large portion of the earth's surface. Geographic coordinate reference systems can be 2D or 3D. A 2D Geographic CRS is used when positions of features are described on the surface of the reference ellipsoid; a 3D Geographic CRS is used when positions are described on, above or below the reference ellipsoid.

Projected: Type of coordinate reference system that is based on an approximation of the shape of the earth's surface by a plane by way of a map projection. The distortion that is inherent to the approximation is carefully controlled and known.

Engineering: Type of coordinate reference system that is that is used only in a contextually local sense. This sub-type is used to model two broad categories of local coordinate reference systems:

- earth-fixed systems, applied to engineering activities on or near the surface of the earth;
- coordinates on moving platforms such as road vehicles, vessels, aircraft or spacecraft.

Image: An Image CRS is an Engineering CRS applied to images. Image CRSs are treated as a separate sub-type because a separate user community exists for images with its own vocabulary.

The definition of the associated Image Datum contains two data attributes not relevant for other datums and coordinate reference systems.

Vertical: Type of coordinate reference system used for the recording of heights or depths. Vertical CRSs make use of the direction of gravity to define the concept of height or depth, but its relationship with gravity may not be straightforward.

By implication, ellipsoidal heights (h) cannot be captured in a vertical coordinate reference system. Ellipsoidal heights cannot exist independently, but only as inseparable part of a 3D coordinate tuple defined in a geographic 3D coordinate reference system.

Temporal: Used for the recording of time in association with any of the listed spatial coordinate reference systems only.

2.2.2 Other sub-types of coordinate reference systems.

In addition to the principal sub-types, two more sub-types have been defined to permit modelling of certain relationships and constraints that exist between the principal sub-types.

Compound coordinate reference system.

The traditional separation of horizontal and vertical position has resulted in coordinate reference systems that are horizontal (2D) in nature and vertical (1D). It is established practice to combine the horizontal coordinates of a point with a height or depth from a different coordinate reference system.

The coordinate reference system to which these 3D coordinates are referenced combines the separate horizontal and vertical coordinate reference systems of the horizontal and vertical coordinates. Such a coordinate system is called a compound coordinate reference system (Compound CRS). It consists of an ordered sequence of the two or more single coordinate reference systems.

A Compound CRS is thus a coordinate reference system that combines two or more coordinate reference systems, none of which can itself be compound. The Compound CRS contains an ordered set of coordinate reference systems and the tuple order of a compound coordinate set shall follow that order, while the subsets of the tuple, described by each of the composing coordinate reference systems, follow the tuple order valid for their respective coordinate reference systems.

For spatial coordinates, a number of constraints exist for the construction of Compound CRSs. For example, the coordinate reference systems that are combined should not contain any duplicate or redundant axes.

Derived coordinate reference system.

Some coordinate reference systems are defined by applying a coordinate conversion to another coordinate reference system. Such a coordinate reference system is called a Derived CRS and the coordinate reference system it was derived from by applying the conversion is called the Source or Base CRS. A coordinate conversion is an arithmetic operation with zero or more parameters that have defined values. The Source CRS and Derived CRS have the same Datum. The best-known example of a Derived CRS is a Projected CRS, which is always derived from a source Geographic CRS by applying the coordinate conversion known as a map projection.

In principle, all sub-types of coordinate reference system may take on the role of either Source or Derived CRS with the exception of a Geocentric CRS and a Projected CRS. The latter is modelled as an object class under its own name, rather than as a general Derived CRS of type "projected". This has been done to honour common practice, which acknowledges Projected CRSs as one of the best known types of coordinate reference systems.

An example of a Derived CRS: one of which the unit of measure has been modified with respect to an earlier defined Geographic CRS, which then takes the role of Source CRS.

2.3 Coordinate System.

The coordinates of points are recorded in a coordinate system. A coordinate system is the set of coordinate system axes that spans over the coordinate space. This concept implies the set of mathematical rules that determine how coordinates are associated with invariant quantities such as angles and distances.

One coordinate system may be used by multiple coordinate reference systems. A coordinate system is composed of an ordered set of coordinate system axes, the number of axes being equal to the dimension of the space of which it describes the geometry. Its axes can be spatial, temporal, or mixed. The dimension of the coordinate space, the names, the units of measure, the directions and sequence of the axes are all part of the Coordinate System definition. The number of coordinates in a tuple and consequently the number of coordinate axes in a coordinate system shall be equal to the number of coordinate axes in the coordinate system. It is therefore not permitted to supply a coordinate tuple with two heights of different definition in the same coordinate tuple.

On the basis of the geometric properties of the coordinate space spanned and the geometric properties of the axes (which can be straight or curved, perpendicular or not) we can distinguish and select between proper subtypes of coordinate systems. Some of this subtypes of coordinate systems can only be used with specific subtype of coordinate reference systems. For example, a Cartesian coordinate system (explained below) cannot be used in a Geographic Coordinate Reference System.

As follows, we show the most important subtypes of coordinate systems (CS):

Cartesian: 1-, 2-, or 3-dimensional CS. It gives the position of points relative to orthogonal straight axes in the 2- and 3-dimensional cases. In the 1-dimensional case, it contains a single straight axis. In the multi-dimensional case, all axes shall have the same unit of measure.

Ellipsoidal: 2- or 3-dimensional CS in which position is specified by geodetic latitude, geodetic longitude and (in the three-dimensional case) ellipsoidal height, associated with one or more geographic coordinate reference systems.

Spherical: 3-dimensional CS with one distance, measured from the origin, and two angular coordinates.

Polar: 2-dimensional CS in which position is specified by distance from the origin and the angle between the line from origin to point and a reference direction.

Vertical: 1-dimensional CS used to record the heights (or depths) of points dependent on the Earth's gravity field. An exact definition is deliberately not provided as the complexities of the subject fall outside the scope of this specification.

The concept of coordinates may be expanded from a strictly spatial context to include time. Time is then added as another coordinate to the coordinate tuple.

2.3.1 Coordinate System Axis.

A coordinate system is composed of an ordered set of coordinate system axes. Each of its axes is completely characterised by a unique combination of axis name, axis abbreviation, axis direction and axis unit of measure.

The concept of coordinate axis requires some clarification. In the case of a Cartesian Coordinate system describing an Euclidean space, we can define the x-axis as the locus of points with y=z=0, but in the case of a curved space like a surface of an ellipsoid, which geometry is described by an ellipsoidal coordinate system, this definition is not exploitable. In that case, we can establish an analogy to the curvilinear latitude and longitude coordinates: the latitude axis would be the equator of the ellipsoid and the longitude axis would be the prime meridian. And we can say that the "i-th" coordinate axis of a coordinate system is defined as the locus of points for which all coordinates with sequence number not equal to "i", have a constant value locally (whereby $i = 1...n$, and n is the dimension of the coordinate space).

Usage of coordinate system axis names is constrained by geodetic custom in a number of cases, depending mainly on the coordinate reference system type. For example, the names "geodetic latitude" and "geodetic longitude" shall be used to designate the coordinate axis names associated with a geographic coordinate reference system. Conversely, these names shall not be used in any other context, such in a Geocentric CRS using a Cartesian Coordinate System.

So, by doing an example, in a Geographic Coordinate Reference System which use a 3-Dimensional Ellipsoidal Coordinate System, we have coordinate system axis called Geodetic Latitude, Geodetic Longitude and Ellipsoidal height.

3 Standardization.

To implements application or projects using Coordinate Reference System and the topics which is connected to, or to allowing information exchange between this kind of applications, it's highly recommended to follow a set of conventions. In this manner anything will work and everyone could create applications which interact consistently with others application geodesy related.

For example, an application sends the points together with a projected coordinate system code and the application receiving this data it's able to place the points exactly on the earth.

The *European Petroleum Survey Group* (**EPSG**[3]) maintains a freely available database with standard codes for coordinate systems, datums, spheroids, units and such like. The database contains coordinate systems and their components, coordinate transformations and their details as well as a table with units of measure and a list of predefined queries. Additionally, this database contains the parameters for these objects or, if they cannot easily be expressed as values, at least references to where such parameters can be found.

Through the EPSG database, every geographic object (like coordinate system and unit of measure) gets assigned a unique number and, in this manner, different applications can easily communicate and exchange information.

The Specifications are also very flexible. For example, in a Geographic Coordinate System we can see the couple (latitude, longitude) and in another Geographic Coordinate System, the couple (longitude, latitude). A Coordinate System can use degree as unit of measure whereas another Coordinate System can use radians. **GeoTools**[4] provide an OGC OpenGis specifications compliant implementation by supporting every type of this different interpretation. So, for example, GeoTools can handle geographic CS whose axes increase to the west and south and it can handle CS of every possible unit of measure: degree, radians, grad, meter, inch, feet, and all other more uncommon units used in certain countries.

This means that it is always necessary to verify and respect the coordinate system properties, that the specific coordinate system object provides.

4 Appendix A: Terms Definitions

In this section will be quoted only the most important definitions from the TC211-ISO19111 Specifications: OGC Abstract Specification Topic 2, Spatial referencing by coordinates[5].

coordinate
One of a sequence of N numbers designating the position of a point in N-dimensional space. In a coordinate reference system, the coordinate numbers must be qualified by units.

coordinate system
set of (mathematical) rules for specifying how coordinates are to be assigned to points.
NOTE: One coordinate system may be used in many coordinate reference systems.

coordinate reference system
coordinate system which is related to the real world by a datum.
NOTE 1: For geodetic and vertical datums, it will be related to the Earth.
NOTE 2: The geometric properties of a coordinate space determine how distances and angles between points are calculated from the coordinates. For example, in an ellipsoidal (2D) space distances are defined as curves on the surface of the ellipsoid, whereas in a Euclidean plane as used for projected CRS distance is the length of a straight line between two points. The mathematical rules that determine distances and angles are calculated from coordinates and vice versa are comprised in the concept of coordinate system.

datum
parameter or set of parameters that determine the location of the origin, the orientation and the scale of a coordinate reference system.

geodetic datum
datum describing the relationship of a 3D or 2D coordinate system to the Earth.
NOTE: In most cases, the geodetic datum includes an ellipsoid definition.

cartesian coordinate system

coordinate system which gives the position of points relative to N mutually-perpendicular straight axes.
NOTE: In the context of geospatial coordinates the maximum values of N is three.

ellipsoid
surface formed by the rotation of an ellipse about an axis.

semi-major axis
semi-diameter of the longest axis of a reference ellipsoid.
NOTE: This equates to the semi-diameter of the reference ellipsoid measured in its equatorial plane.

semi-minor axis
semi-diameter of the shortest axis of a reference ellipsoid.
NOTE: The shortest axis coincides with the rotation axis of the reference ellipsoid and therefore contains both poles.

flattening (f)
ratio of the difference between the semi-major (a) and semi-minor axis (b) of an ellipsoid to the semi-major axis; f = (a - b)/a
NOTE: Sometimes inverse flattening 1/f = a/(a-b) is given instead; 1/f is also known as reciprocal flattening.

meridian
intersection of an ellipsoid by a plane containing the semi-minor axis of the ellipsoid.

prime meridian, zero meridian
meridian from which the longitudes of other meridians are quantified.

geodetic latitude
angle from the equatorial plane to the perpendicular to the ellipsoid through a given point, northwards treated as positive.

geodetic longitude
angle from the prime meridian plane to the meridian plane of the given point,

eastward treated as positive.

ellipsoidal (geodetic) height
distance of a point from the ellipsoid measured along the perpendicular from the ellipsoid to the point itself, taken as positive upwards or outside the ellipsoid.

coordinate conversion
change of coordinates, based on a one-to-one relationship, from one coordinate reference system to another based on the same datum.

map projection
coordinate conversion from a geodetic coordinate system to a planar surface.

coordinate transformation
computational process of converting a position given in one coordinate reference system into the corresponding position in another coordinate reference system.
NOTE: A coordinate transformation can require and use the parameters of the ellipsoids associated with the source and target coordinate

compound coordinate reference system
coordinate reference system using at least two independent coordinate reference systems describing horizontal positions and/or vertical positions and/or temporal positions or positions.
NOTE: It cannot contain another compound coordinate reference system.

geographic coordinate reference system
coordinate reference system using an ellipsoidal coordinate system and based on an ellipsoid that approximates the shape of the Earth.
NOTE: A geographic coordinate system can be 2D or 3D. In a 3D geographic coordinate system, the third dimension is height above the ellipsoid surface.

geocentric coordinate reference system
3-dimensional coordinate reference system with its origin at the (approximate) centre of the Earth.

projected coordinate reference system
two-dimensional coordinate system resulting from a map projection.
NOTE: Projected coordinate reference system is derived from a 2D geographic coordinate reference system by applying a parameterised coordinate transformation known as a "map projection". Projected coordinate reference system commonly uses a Cartesian coordinate system.

vertical coordinate reference system
one-dimensional coordinate reference system used for gravity-related height or depth measurements.

temporal reference system
reference system against which time is measured.

temporal datum
datum defining the origin of a temporal reference system.

time coordinate system
one-dimensional coordinate system containing a time axis used in temporal reference systems to describe the temporal position of a point in the specified time units from a specified temporal datum.

5 References.

References

[1] $http : //www.opengeospatial.org/$

[2] $http : //www.isotc211.org/$

[3] $http : //www.epsg.org/$

[4] $http : //www.geotools.org/$

[5] $https : //portal.opengeospatial.org/files/?artifact_id = 6716$