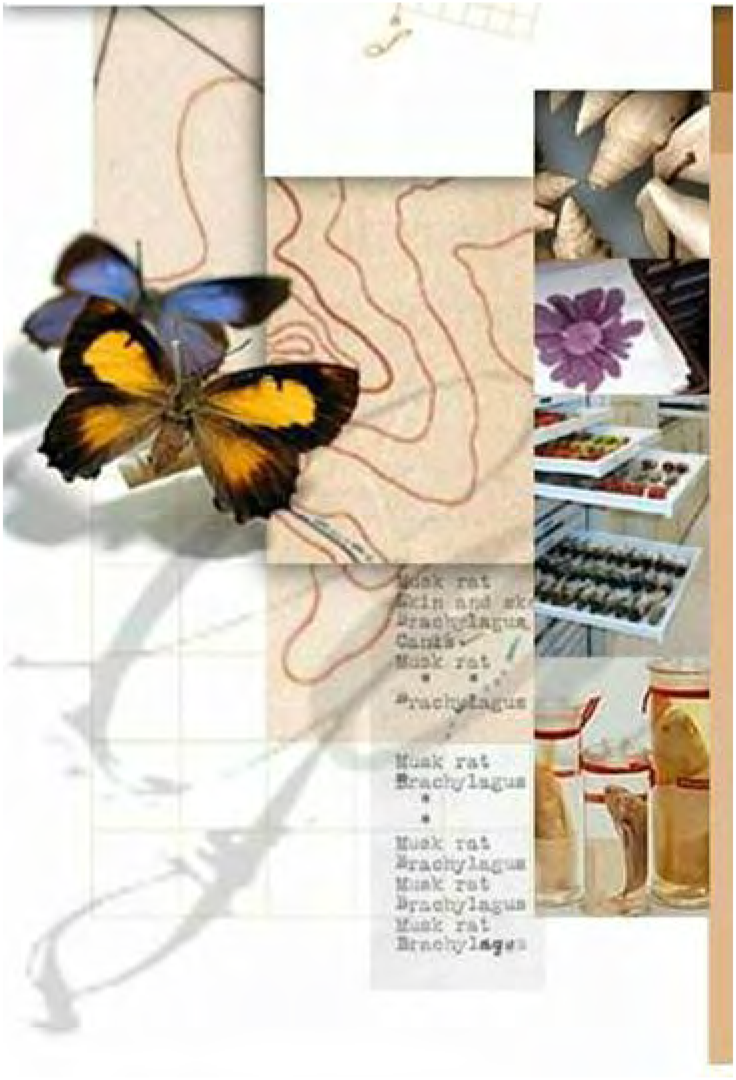


BioGeomancer

Guide to Best Practices for Georeferencing

Guide to

Best Practices

for

Georeferencing

BioGeomancer Consortium

August 2006

Published by:
Global Biodiversity Information Facility, Copenhagen
http://www.gbif.org

The information in this book represents the professional opinion of the authors, and does not necessarily represent the views of the publisher or of the Regents of the University of California. While the authors and the publisher have attempted to make this book as accurate and as thorough as possible, the information contained herein is provided on an "As Is" basis, and without any warranties with respect to its accuracy or completeness. The authors, the publisher and the Regents of the University of California shall have no liability to any person or entity for any loss or damage caused by using the information provided in this book.

Guide to Best Practices for Georeferencing
Includes Index
ISBN: 87-92020-00-3

Recommended Citation: Chapman, A.D. and J. Wieczorek (eds). 2006. *Guide to Best Practices for Georeferencing*. Copenhagen: Global Biodiversity Information Facility.

Edited by: Arthur D. Chapman and John Wieczorek

Contributors: J.Wieczorek, R.Guralnick, A.Chapman, C.Frazier, N.Rios, R.Beaman, Q.Guo.

Contents

Glossary

Accuracy — a measure of how well data represent true values.

Cadastre — a register that defines boundaries of public and/or private land.

Cadastral map — a map showing ***cadastre*** (*q.v.)* boundaries

Coordinates — a sequence of numbers designating the position of a point in n-dimensional space [ISO 19111]. Examples of two-dimensional coordinate systems are Latitude/Longitude and Universal Transverse Mercator (UTM).

Coordinate reference system — a reference system that relates a sequence of numbers or ***coordinates*** (*q.v.*) to the real world via a ***datum*** (*q.v.*).

Coordinate system — a system used to denote direct or relative positions by ***coordinates*** (*q.v.*).

Data Quality — described 'fitness for use' (Juran 1964, 1994, Chrisman 1991, Chapman 2005a) of data. As a collector, you may have an intended use for the data you collect but data have the potential to be used in unforeseen ways; therefore, the value of your data is directly related to the fitness of those data for a variety of uses. As data become more accessible, many more uses become apparent (Chapman 2005c).

Datum — a parameter or set of parameters that serve as a reference or basis for the calculation of other parameters [ISO 19111]. A datum defines the position of the origin, the scale, and the orientation of the axes of a coordinate system. A datum may be a geodetic datum, a vertical datum or an engineering datum. In this document, the term ***datum*** generally refers to a ***geodetic datum*** (*q.v.*).

Decimal degrees — degrees expressed as a single real number (e.g., −22.343456) rather than as a composite of degrees, minutes, seconds, and direction (e.g., 7º 54' 18.32" E). Note that minus (−) signs are used to indicate southern and western hemispheres.

Decimal latitude — the latitude coordinate (in decimal degrees) at the center of a circle encompassing the whole of a specific locality. Convention holds that decimal latitudes north of the equator are positive numbers less than or equal to 90, while those south are negative numbers greater or equal to −90.
<u>Example</u>: −42.5100 degrees (which is roughly the same as 42º 30' 36" S).

Decimal longitude — the longitude coordinate (in decimal degrees) at the center of a circle encompassing the whole of a specific locality. Decimal longitudes east of the Greenwich Meridian are considered positive and less than or equal to 180, while western longitudes are negative and greater than or equal to −180.
<u>Example</u>: −122.4900 degrees (which is roughly the same as 122º 29' 24" W).

Digital Elevation Model (DEM) — a digital representation of the elevation of locations on the land surface of the earth, usually represented in the form of a rectangular grid.

Easting and Northing — within a coordinate reference system (e.g., as provided by a GPS or a map grid reference system), *Easting*s are the vertical grid lines running from top to bottom (North to South) which divide a map from East to West and *Northings* are the horizontal lines running from left to right (East to West) dividing the map from North to South. The squares formed by intersecting eastings and northings are called grid squares. On 1:100,000 scale maps each square represents an area of 100 hectares or one kilometer square.

Elevation — the elevation of a geographic location is its height above mean sea level or some other fixed reference point (*cf.* ***vertical datum***). Elevation may be a negative number in those parts of the earth where the land surface is below mean sea level. Elevation may be recorded on maps in the form of contour lines linking points of uniform elevation, or as spot heights at ***trig points*** (*q.v.*) – usually at the summits of mountains, and rarely at low points. Elevation is used when referring to points on the earth, whereas altitude is used for points above the surface of the earth, such as the altitude of an aircraft, and depth for positions below the surface (of a lake, sea, etc.).

Extent — the geographic range, magnitude, or distance which a location may actually represent. With a town, the extent is the polygon that encompasses the area inside the town's boundaries. In this document, we usually refer to the linear extent – the distance from the geographic center of the location to the furthest point in the representation of the location.

False Precision — occurs when data are recorded with a greater number of decimal places than implied by the original data. This often occurs following transformations from one unit or coordinate system to another, for example from feet to meters, or from degrees, minutes, and seconds to decimal degrees. In general, precision cannot be conserved across metric transformations; however, in practice it is often recorded as such. For example, a record of 10º 20' stored in a database in decimal degrees is ~10.3º. When exported from some databases, however, it will result in a value of 10.3333333333 with a precision of 10 decimal places rather than 1, leading to a metric uncertainty of around 0.02 mm instead of the real uncertainty of ~15 km. This is not a true precision as it relates to the original data, but a false precision as reported from the database.

Feature — a natural or anthropogenic object or observation that can be represented spatially. The term "feature" may refer to categories of objects or ***feature types*** (e.g., mountains, roads, or cities) or to specific ***feature instances*** (e.g., Mount Everest, Interstate 25, or San Fransisco), which are also sometimes referred to as "named places."

Feature Name — a proper name applied to a ***feature*** (*q.v.*); the name of a named place.

Footprint — a spatial representation of a ***feature*** (*q.v.*) as an area. The extent and shape of a footprint may comprise the actual boundaries of a feature, the uncertainty around a point representation of a feature, or some combination of an estimate of the boundaries of a feature and the uncertainty associated with those boundaries.

Gazetteer — a geographic dictionary or index of ***feature names*** (*q.v.*)., usually also including an indication of position on the earth's surface using one of several ***geographic coordinate systems*** (*q.v.*), but most generally ***latitude*** (*q.v.*) and ***longitude*** (*q.v.*).

Geocode — the process of determining the coordinates for a street address. It is also sometimes used as a synonym for **georeferencing** (*q.v.*).

Geodetic datum — a model of the earth used for geodetic calculations. A geodetic datum describes the size, shape, origin, and orientation of a coordinate system for mapping the surface of the earth (NAD27, SAD69, WGS84, etc.). In this document, we use the term to refer to the ***horizontal datum*** (*q.v.*) and not the ***vertical datum*** (*q.v.*).

Geodetic datums are often recorded on maps and in gazetteers, and can be specifically set for most GPS devices so the waypoints match the chosen datum. Use "not recorded" when the datum is not known.

Geographic coordinate system — the net or graticule of lines of latitude (parallels) numbered 0° to 90° north and south of the equator, and lines of longitude (meridians) numbered 0° to180° east and west of the international zero meridian of Greenwich, used to define locations on the Earth's surface (disregarding elevation) with the aid of angular measure (degrees, minutes and seconds of arc)[1].

---

[1] Glossary of Terminology. <http://www.ngi.be/NL/glossary/glossang-inf.htm>

This is the traditional global coordinate system based on latitude and longitude.

Geographic center — the geographic center of a shape is the mean of the extremes of latitude and longitude of that shape. If the result is not within the shape itself, choose instead the point in the shape nearest to the calculated geographic center.

Georeference — to translate a locality description into a mappable representation of a ***feature*** (*q.v.*) (verb); or the product of such a translation (noun).

GPS (Global Positioning System) **—** a satellite-based navigation system that provides 24 hour three-dimensional position, velocity and time information to suitably equipped users (i.e., users with a GPS receiver) anywhere on or near the surface of the Earth. See discussions on accuracy elsewhere in this document.

Heading — the direction from a starting location, given in the form of points of the compass such as E, NW, or N15ºW, etc. Usually used in conjunction with ***offset*** (*q.v.*) to give a distance and direction from a named place. See discussion on true and magnetic north in the **Recording Headings** section of this document.

Horizontal datum — that portion of a ***datum*** (*q.v.*) which refers to the horizontal positions of mapped features with respect to parallels and meridians or northing and easting grid lines on a map as opposed to the ***vertical datum*** (*q.v.*).

Latitude — describes the angular distance that a location is north or south of the equator, measured along a line of ***longitude*** (*q.v.*).

Locality — a) the position of a feature in space; b) The verbal representation of this position (i.e., the ***locality description***).

Location — a position on the earth's surface or in geographic space definable by ***coordinates*** (*q.v.*) or some other geographic referencing system, such as a street address, offset, etc.

Longitude — describes the angular distance east or west of a prime meridian (*q.v.*) on the earth's surface along a line of ***latitude*** (*q.v.*).

Map projection — a method of representing the earth's three-dimensional surface as a flat two-dimensional surface. This normally involves a mathematical model (of which there are many) that transforms the locations of features on the earth's surface to locations on a two-dimensional surface. Such representations distort one or more parameters of the earth's surface such as distance, area, shape, or direction.

Maximum uncertainty estimate — the numerical value for the upper limit of the distance from the coordinates of a locality to the outer extremity of the area (often a circle) within which the whole of the described locality must lie.

Maximum uncertainty units — the units of length in which the maximum uncertainty estimate is recorded (e.g., mi, km, nm, m, ft). The maximum uncertainty distance should be recorded using the same units as the distance measurements in the locality description.

Meridian — the intersection in one hemisphere of the earth's surface with a plane passing through the poles, usually corresponding to a line of ***longitude*** (*q.v.*).

Named place — used to refer not only to traditional ***features*** (*q.v.*), but also to places that may not have proper names, such as road junctions, stream confluences, highway mile pegs, and cells in grid systems (e.g., townships).

Northing — See ***Easting and Northing***.

Offset — a displacement from a reference point, named place, or other feature. Used here as the distance from a named place using the location of the named place as the starting point. Usually used in conjunction with ***heading*** (*q.v.*) to give a distance and direction from a named place.

Precision — with measurements and values, it describes the finest unit of measurement used to express that value (e.g., if a record is reported to the nearest minute, the precision is 1/3600th of a degree; if a decimal degree is reported to two decimal places, the precision is 0.01 of a degree). It is important to always calculate the precision from the original data and units of measurement. See also ***false precision*** (*q.v.*).

Prime meridian — a meridian from which longitude east and west is reckoned, the most recent standard for which passes through Greenwich, England.

Spatial fit — a measure of how well the geometric representation matches the original spatial representation. See discussion elsewhere in this document.

Trig point — a surveyed reference point, often on high points of elevation (mountain tops, etc.) and usually marked by a small pyramidal structure or a pillar. The exact location is determined by survey triangulation and hence the name trigonometrical point or triangulation point.

Uncertainty —a "measure of the incompleteness of one's knowledge or information about an unknown quantity whose true value could be established if a perfect measuring device were available" (Cullen & Frey 1999). Uncertainty is a property of the observer's understanding of the data. Throughout this document we use ***Maximum uncertainty estimate*** (*q.v.*) as the way of recording and documenting uncertainty.

UTM (Universal Transverse Mercator) **—** a standardized coordinate system based on a metric rectangular grid system and a division of the earth into sixty 6-degree longitudinal zones. Zones are numbered consecutively with Zone 1 between 180 and 174 degrees west longitude. UTM only covers from 84º N to 80º S. When citing UTM coordinates, it is essential that the UTM Zone also be recorded.

Vertical datum — that portion of a ***datum*** (*q.v.*) that refers to the vertical position of mapped features with respect to a base measurement point (such as mean sea level at a location) and from which all elevations are determined (e.g., AHD – The Australian Height Datum; NAVD88 – North American Vertical Datum). See comments on accuracy under the section on GPS accuracy in this document.

WGS84 (World Geodetic System 1984) — a ***coordinate reference system*** (*q.v.*) in common use globally to fit the shape of the entire Earth as accurately as possible using a single ellipsoid. Other ellipsoids (*datums*) are commonly used locally to provide a better fit to the Earth in a local region.

Introduction

One of the outputs from the BioGeomancer project is a document on best practice for georeferencing biological species (specimen and observational) data. Several projects (MaNIS, MapSteDI, INRAM, GEOLocate, NatureServe, CRIA, ERIN, CONABIO, etc.) have previously developed guidelines and tools for georeferencing, and these provide a good starting point for such a document.

The document provides guidelines to the world's best practice for georeferencing such data, but it is important that organisations and institutions then produce their own internal document that incorporates the practices outlined in this document into their own working environment.

The document presents examples of how to georeference a range of different location types, and provides information and examples on how to determine the extent and maximum uncertainty distance for locations based on the information provided.

1. Definition

> *"The term best practice generally refers to the best possible way of doing something; it is commonly used in the fields of business management, software engineering, and medicine, and increasingly in government. […] The [qualified] term, 'best current practice', often represents the meaning in a more accurate way, showing the possibility for future developments of 'better practice'."* (Wikipedia: Best Practice[2]).

2. Principles of Best Practice

- **Accuracy** – a measure of how well the data represent true values. It is good practice to quote a percentage area or an uncertainty in meters, or to draw an uncertainty polygon.

  > With georeferencing – this is currently mostly an uncertainty radius, however uncertainty polygons are beginning to be used in some circumstances. Uncertainty probability surfaces are also under consideration.

- **Effectiveness** – the likelihood that a work program achieves its desired objectives.

  > With georeferencing – this is the percentage of records for which the latitude and longitude can be accurately identified through use of BioGeomancer or in some other way.

- **Efficiency** – the ratio of output to input.

  > With georeferencing – this is the amount of effort that is needed to produce an acceptable output. It also refers to the amount of input data the user has to obtain to produce an acceptable result (e.g., gazetteers, collectors' itineraries, etc.).

- **Reliability** – related to accuracy, and refers to the consistency with which results are produced.

  > With georeferencing – it refers to the repeatability with which a georeference can be produced by the user for the same locality.

- **Accessibilty** – how accessible are the results to the users, public, etc.

  > With georeferencing – this is the ease with which users, other institutions, etc., can access the information for a particular locality that has already been georeferenced.

---

[2] Wikipedia: Best Practice <http://en.wikipedia.org/wiki/Best_practice>

- **Transparency** – an annunciation of the procedures for collection, analysis, reporting and update.

> With georeferencing – this refers to the quality of the metadata and methodology by which a georeference was obtained for a particular locality.

- **Timeliness** – relates to the frequency of data collection, its reporting and updates.

> With georeferencing – it largely refers to how often gazetteers are updated, or when the records are georeferenced and made available to others.

- **Relevance** – the data collected should meet the needs of the user – i.e., should fulfill the principle of "fitness for use".

> With georeferencing – it refers to the format of the output (i.e., does it include good metadata on the above topics).

In addition, an effective best practices document should:

- Align the vision, mission, and strategic plans in an institution to policies and procedures and gain the support of sponsors and/or top management.
- Use a standard method of writing (writing format) to produce professional policies and procedures within the institution.
- Satisfy industry standards.
- Satisfy the scrutiny of management and external/internal auditors.

This list is by no means exhaustive, but does cover most of the elements in identifying best practice.

Background

A number of projects have been working for many years on the development of guidelines and tools for improving the georeferencing of primary biodiversity data. This document largely draws on those initiatives and attempts to bring the results of all this previous work into one comprehensive best practices document. Without this background work, such a document would not be possible. For link locations see under 'Key Documents and Links' at the end of this Chapter.

BioGeoMancer Classic

The original BioGeoMancer Classic was developed by Reed Beaman, now at Yale University. This tool provides a georeferencing service for collectors, curators and users of natural history specimens. BioGeoMancer Classic can parse English language place name descriptions and provide a set of latitude/longitude coordinates associated with that description. It provides offset calculations for when a collection is georeferenced a given distance and cardinal direction from the nearest named place. For more details on how it works – see "What it does …[3]".

MaNIS

With support from the National Science Foundation, seventeen North American institutions and their collaborators developed the **Ma**mmal **N**etworked **I**nformation **S**ystem. The original objectives of MaNIS were to 1) facilitate open access to combined specimen data from a web browser, 2) enhance the value of specimen collections, 3) conserve curatorial resources, and 4) use a design paradigm that could be easily adopted by other disciplines with similar needs.

The MaNIS network has developed a number of tools and guidelines for assisting the georeferencing of collections in the MaNIS network. These documents and tools have been heavily drawn upon in this document.

MapSTeDI

The Mountains and Plains Spatio-Temporal Database Informatics Initiative (MaPSTeDI) was a collaborative effort between the University of Colorado Museum, Denver Museum of Nature and Science, and Denver Botanic Gardens to convert their separate collections into one distributed biodiversity database and research toolkit for the southern and central Rockies and adjacent plains. Unlike MaNIS or other projects, which have strong taxonomic focus and a distributed database federation outcome, MaPSTeDI had a regional focus and a distributed GIS mapping system outcome. Like other projects listed here, georeferencing was the essential first step in MaPSTeDI, providing the data that will be eventually analyzed spatially and temporally on the MaPSTeDI online GIS. The MaPSTeDI project also developed detailed guidelines and tools such as the MaPSTeDI Georeferencing Protocols and Guide to Georeferencing, and these have been heavily relied upon in this document.

INRAM

The Institute of Resource Analysis and Management (INRAM) sought to increase the value of New Mexico museum specimen data by supporting an effort to georeference New Mexico specimen localities. Data that are georeferenced haphazardly are of little use to science, so the first goal of the INRAM Georeferencing Team was to develop a detailed, comprehensive protocol describing how best to determine the coordinates and uncertainty estimate to apply to

---

[3] BioGeoMancer Classic – What it does … <http://130.132.27.130/yu/bgm-docs/what-it-does.html>

a given locality. The INRAM team started by evaluating the protocol used by the **M**ammal **N**etworked **I**nformation **S**ystem (MaNIS) in which the Museum of Southwestern Biology (MSB) mammal division was participating, and determined that there were many ways it could be improved. In particular, INRAM created a more detailed list of locality types with a specific rule set for each as to how to determine coordinates and uncertainty. INRAM also sought to maximize the efficiency and accuracy of the georeferencing process. With help from the New Mexico Natural Heritage Program and the Museum of Southwestern Biology, INRAM developed a combined GIS and database system that made implementing the protocol much easier for the students doing the work. Together, the INRAM protocol and georeferencing software system allowed a semi-automated georeferencing process which provided accurate, rapid data capture and which left a detailed record of the methods and assumptions used to georeference each specimen.

GEOLocate

In March of 1995, Dr. Henry L. Bart received funding from the U.S. National Science Foundation to computerize and georeference the Tulane University Museum of Natural History Fish Collection. Georeferencing was accomplished by manually plotting each locality description on hardcopy USGS topographic maps and using a digitizing tablet to register the maps and determine coordinates. Where possible, hand-plotted, hardcopy maps were compared to electronic versions of the same maps (USGS digital line graphs), allowing the technician to use a mouse to electronically capture the coordinates. Using this method, 15,000 locality descriptions for nearly 7 million specimens were georeferenced by one technician over a period of 18 months.

In Febuary of 2002, Dr. Bart and Nelson Rios received funding from the the U.S. National Science foundation to develop a software package to facilitate georeferencing of natural history collections data, using the Tulane Fish Collection as a testbed. The result was GEOLocate, a tool for comprehensive automated georeferencing of North American locality descriptions. Ongoing development involves expanding coverage to the entire world, multi-lingual support, user-defined pattern recognition, and collaborative georeferencing. GEOLocate is also being developed as a webservice for integration into the current development of BioGeomancer.

ERIN

The Environmental Resources Information Network (ERIN) was established in the Australian Department of the Environment in 1989 and began funding the databasing and georeferencing of Australia's museum and herbarium collections. It established methods for assisting georeferencing, including the linking of records to Digital Elevation Models to determine elevation, and sophisticated methods for data checking and validation by searching for outliers in environmental space using niche modeling techniques. These have recently been upgraded in conjunction with the Centro de Referência em Informação Ambiental (CRIA) and Robert Hijmans, the author of the DIVA-GIS software.

Key Documents and Links

- Best Practices Guidelines for GPS Survey (NLWRA, Australia)
  http://www.nlwra.gov.au/toolkit/10/10-2.html
- BioGeoMancer Classic
  http://classic.biogeomancer.org
- Centro de Referência em Informação Ambiental (CRIA)
  http://www.cria.org.br
- DIVA-GIS
  http://www.diva-gis.org

- Environmental Resources Information Network (ERIN)
  http://www.deh.gov.au/erin/index.html
- Examples of Good and Bad Localities
  http://mvz.berkeley.edu/Locality_Field_Recording_examples.html
- GEOLocate – University of Tulane
  http://www.museum.tulane.edu/geolocate/
- Institute of Resource Analysis and Management (INRAM)
  http://biodiversity.inram.org/
- INRAM Protocol for Georeferencing Biological Museum Specimen Records
  http://www.inram.org/modules/UpDownload/store_folder/Documents/INRAM_Biodiversity_Georeferencing_Project/Georeferencing_Guidelines_INRAM-V1.3_2004-03-01.pdf
- Mammal Networked Information System (MaNIS)
  http://manisnet.org/
- MaNIS Documents
  http://manisnet.org/Documents.html
- MaNIS/HerpNet/ORNIS Georereferencing Guidelines
  http://manisnet.org/manis GeorefGuide.html
- Manual de Procedimientos para Georeferenciar, CONABIO, 2004. An internal georeferencing manual produced by the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), Mexico.
- The Mountains and Plains Spatio-Temporal Database Informatics Initiative - MaPSTeDI
  http://mapstedi.colorado.edu/index.html
- MaPSTeDI Georeferencing Protocols
  http://mapstedi.colorado.edu/georeferencing-protocols.html
- MaPSTeDI Guide to Georeferencing
  http://mapstedi.colorado.edu/georeferencing-howto.html
- Museum of Vertebrate Zoology Informatics (MVZ) – University of California, Berkeley
  http://mvz.berkeley.edu/Informatics.html
- MVZ Guide for Recording Localities in the Field:
  http://mvz.berkeley.edu/Locality_Field_Recording_Notebooks.html
- Reasons Why it is Important to Take Good Locality Data (MVZ)
  http://mvz.berkeley.edu/Locality_Field_Recording_important.html
- OGC Recommendations Document Pointer
  http://www.opengeospatial.org/specs/?page=recommendation

Collecting and Recording Data in the Field[4]

Collecting data in the field sets the stage for good georeferencing procedures. Many new techniques now exist that can lead to quite accurately georeferenced locations; however it is important that the locations be recorded correctly in order to reduce the likelihood of error. We recommend that all new collecting events use a GPS for recording coordinates wherever possible, and that the GPS be set to a relevant datum (see below).

1. *The Importance of Good Locality Data Recording*

Good locality descriptions lead to more accurate georeferences with smaller uncertainty values and provide users with much more accurate and high quality data. When recording data in the field, whether from a map or when using a GPS, it is important to record locality information as well as the georeferences, so that later validation can take place if necessary.

One purpose behind a specific locality description is to allow the validation of coordinates, in which errors are otherwise difficult to detect. The extent to which validation can occur depends on how well the locality description and its spatial counterpart describe the same place. The highest quality locality description is one with as few sources of uncertainty as possible. By describing a place in terms of a distance along a path, or by two orthogonal distances from a place, one removes uncertainty due to imprecise headings. Choosing a reference point with small extent reduces the uncertainty due to the size of the reference point, and by choosing a nearby reference point, one reduces the potential for error in measuring the offset distances.

To make it easy to validate a locality, use reference points that are easy to find on maps or in gazetteers. At all costs, avoid using vague terms such as “near” and “center of” or providing only an offset without a distance such as “West of Albuquerque”.

In any locality that contains a named place that can be confused with another named place of a different type, specify the feature type in parentheses following the feature name.

Examples:

Locality example using distance and heading along a path:
E shore of Bolinas Lagoon, 3.1 mi NW via Hwy. 1 from intersection of Hwy. 1 and Calle del Arroyo in Stinson Beach (town), Marin Co., Calif.

Locality example using two cardinal offset distances from a reference point:
ice field below Cerro El Plomo, 0.5 km S and 0.2 km W of summit, Region Metropolitana, Chile.

2. *Recording Localities*

Provide a descriptive locality, even if you have geographic coordinates. The locality should be as specific, succinct, unambiguous, complete, and as accurate as possible, leaving no room for uncertainty in interpretation.

Localities used as reference points should be stable – i.e., places (towns, trig points, etc.) that will remain for a long time after the collection events. Do NOT use temporary locations or waypoints as the key reference location. You may have made an accurate GPS recording for the temporary location and then referenced future collections from that point (e.g., 200 m SE of the Land Rover), and that may make perfect sense for that series of collections. It is

---

[4] See also Museum of Vertebrate Zoology, Berkeley, California (2006) *MVZ Guide for Recording Localities in Field Notes* <http://mvz.berkeley.edu/Locality_Field_Recording_Notebooks.html>

meaningless, however, when those collections are later broken up and placed in a museum under a taxonomic arrangement, and no longer have a link to where the ‘Landrover’ was.

If recording locations along a path (road, river, etc.) it is important to also record whether the distances were measured along the path (‘by road’) or as a direct line from the origin (‘by air’).

> **Hint:** The most specific localities are those described by a) a distance and heading along a path from a nearby and well-defined intersection, or b) two cardinal offset distances from a single persistent nearby feature of small extent.

3. Recording Coordinates

Coordinates are a convenient way to define a locality that is not only more specific than is otherwise possible with a description, but that is also readily usable in GIS applications. Always include as many decimals of precision as given by the coordinate source. A measurement in decimal degrees given to five decimal places is more precise than a measurement in degrees minutes seconds to the nearest second, and more precise than a measurement in degrees decimal minutes given to three decimal places (see Table 4). Some new GPS receivers now provide for recording data in decimal seconds and this (to two decimal places) provides a precision comparable to that of decimal degrees.

Whenever practical, provide the coordinates of the location where collecting actually occurred (see ***Extent***, below). If reading coordinates from a map, use the same coordinate system as the map. The datum is an essential part of a coordinate description; it provides the frame of reference. When using both maps and GPS in the field, set the GPS datum to be the same as the map datum so that your GPS coordinates will match those on the map. Be sure to record the datum used.

Specific projects may require particular coordinate systems, but we find geographic coordinates in decimal degrees to be the most convenient system for georeferencing. Since this format relies on just two attributes, one for latitude and the other for longitude, it provides a succinct coordinate description with global applicability that is readily transformed to other coordinate systems as well as from one datum to another. By keeping the number of recorded attributes to a minimum, the chances for transcription errors are minimized (Wieczorek *et al.* 2004).

> **Hint:** Decimal degrees are preferred when reading coordinates from a GPS, however see Note under *Using a GPS*, below.
>
> **Hint:** If using UTM coordinates, always record the UTM Zone.

4. Using a GPS

GPS (Global Positioning System) technology uses triangulation to determine the location of a position on the earth’s surface. The distance calculated is the range between the GPS receiver and the GPS Satellites (Van Sickle 1996). As the GPS satellites are at known locations in space, the position on earth can be calculated. A minimum of four GPS satellites is required to determine the location of a position on the earth’s surface (McElroy *et al.* 1998, Van Sickle 1996). This is not generally a limitation today, as one can often receive seven or more satellites in most locations on earth, however, historically the number of satellites receivable was not always sufficient. Prior to May 2000, most GPS units used by civilians were subject to “Selective Availability”. The removal of this signal degradation technique has greatly improved the accuracy that can generally be expected from GPS receivers (NOAA 2002).

To obtain the best possible accuracy, the GPS receiver must be located in an area that is free from overhead obstructions and reflective surfaces and have a good field of view to the horizon (for example, they do not work very well under a heavy forest canopy). The GPS receiver must be able to record signals from at least four GPS satellites in a suitable geometric arrangement. The best arrangement is to have “*one satellite directly overhead and the other three equally* *spaced around the horizon*" (McElroy *et al.* 1998). The GPS receiver must also be set to an appropriate datum for the area, and the datum used recorded (Chapman *et al.* 2005a).

GPS accuracy: Most GPS devices are able to report a theoretical horizontal accuracy based on local conditions at the time of reading. For highly specific localities, it may be possible for the potential error in the GPS reading to be on the same order of magnitude as the extent of the locality. In these cases, the GPS accuracy can make a non-trivial contribution to the overall uncertainty in the position given by the coordinates.

Prior to the removal of Selective Availability, the accuracy of *Hand-held GPS* receivers as used by most biologists and observers in the field was around 100 meters or worse (McElroy *et al.* 1998, Van Sickle, 1996, Leick 1995). Since then, however, the accuracy of GPS receivers has improved and today, most manufacturers of hand-held GPS units promise errors of less than 10 meters in open areas when using four or more satellites. The accuracy can be improved by averaging the results of multiple observations at a single location (McElroy *et al.* 1998), and some modern GPS receivers that include averaging algorithms can bring the accuracy down to around five meters or maybe even better. NOAA (2001) suggests that GPSs without differential (see below) may be as accurate as 10-15 meters, depending on the receiver being used, satellite configuration and atmospheric conditions, but that this is at the better end of the scale.

The use of *Differential GPS* (DGPS) can improve the accuracy considerably. DGPS uses referencing to a GPS Base Station (usually a survey control point) at a known location to calibrate the receiving GPS. This works through the Base Station and hand-held GPS referencing the satellites' positions at the same time and thus reduces error due to atmospheric conditions. In this way, the hand-held GPS applies the appropriate corrections to the determined position. Depending on the quality of the receivers used, one can expect an accuracy of between 1 and 5 meters. This accuracy decreases as the distance of the receiver from the Base Station increases. Again, averaging can further improve on these values (McElroy *et al.* 1998). For example, the U.S. Coast Guard's DGPS has a stated horizontal accuracy of ± 10 meters (95%). In other words, 95 percent of the time a position determined using DGPS will be within 10 meters of its true position on the earth. Under certain conditions, mariners may observe better than 10-meter accuracy (NOAA 2001).

The Wide Area Augmentation System (WAAS) is a GPS-based navigation and landing system developed for precision guidance of aircraft (Federal Aviation Administration 2004). WAAS uses ground-based antennae with precisely known locations to provide greater positional accuracy for GPSs. Similar technologies such as Local Area Augmentation System (LAAS) are also being developed to provide even finer precision.

Even greater accuracies can be achieved using either *Real-time Differential GPS* (McElroy *et al.* 1998) or *Static GPS* (McElroy *et al.* 1998, Van Sickle 1996). *Static GPS* uses high precision instruments and specialist techniques and is generally employed only by surveyors. Surveys conducted in Australia using these techniques reported accuracies in the centimeter range. These techniques are unlikely to be extensively used with biological record collection due to the cost and general lack of requirement for such precision.

Note! Set your GPS to report locations in decimal degrees rather than make a conversion from another coordinate system as it is usually more precise, better and easier to store, and saves later transformations which may introduce error.

Note2! An alternative where reference to maps is important, and where the GPS receiver allows it, is to set the recorder to report in degrees, minutes, and decimal seconds.

5. Recording Datum

Except under special circumstances (the poles, for example), coordinates without a datum do not uniquely specify a location. Confusion about the datum can result in positional errors of hundreds of meters.

When using a GPS, it is important to set and record the Datum being used. See discussion below under *Calculating Uncertainties.*

> **Note!** If you are not basing your locality description on a map, set your GPS to report coordinates using the WGS84 datum. Record that fact in all your documentation.

6. Recording Elevation

Supplement the locality description with elevation information if this can easily be obtained. It is preferable to use a barometric altimeter if available. Alternatively, obtain the elevation from a Digital Elevation Model (usually done retrospectively in the laboratory), or by using the contours and spot height information from a suitable scale map of the area. Record the method used in Remarks.

> **Note!** "Elevation markings can narrow down the area in which you place a point. More often than not, however, they seem to create inconsistency. While elevation should not be ignored, it is important to realize that elevation was often measured inaccurately and/or imprecisely, especially early in the 20th century. One of the best uses of elevation in a locality description is to pinpoint a location along a road or river in a topographically complex area, especially when the rest of the locality description is vague." (MaPSTeDI 2004)

Under normal conditions, GPS devices are much less accurate for recording elevation than horizontal distances, and they do not report the altitudinal accuracy. It is important to note that the height displayed by a GPS receiver is actually the height in relation to an ellipsoid as a model of the Earth's surface, and not a height based on mean sea level, or to a standard height datum such as the Australian Height Datum. In Australia, for example, the difference between altitudes reported from a GPS receiver and mean sea level can vary from –35 to +80 meters and tends to vary in an unpredictable manner (Chapman *et al.* 2005, McElroy *et al.* 1998, Van Sickle 1996).

If elevation is a defining part of the locality description, be sure to use a reliable source for this measurement (barometric altimeter, trustworthy map, or Digital Elevation Model at suitable scale), and specify the source under references. It is not recommended that elevation be determined using a GPS.

> **Hint:** A barometric altimeter, when properly calibrated, is much more reliable than a GPS for obtaining accurate elevations. It is not recommended that elevation be determined using a GPS. See remarks above under ***Using a GPS*** about the error inherant in using a GPS to determine elevations.

7. Recording Headings

It is important when using a compass to record headings, that adjustments be made to record True North and not Magnetic North. The differences between True North and Magnetic North vary in different parts of the world, and in some places can vary greatly across a very small distance. The differences also change over time. For example, in an area about 250 km NW of Minneapolis in the United States, the anomolous declination changes from 16.6° E to 12.0° W across a distance of just 6 km (Goulet 2001).

The National Geophysical Data Center (NGDC) in the USA has an on-line calculator[5] that can calculate the anomolous or magnetic declination for any place on earth and at any point in time. If you need to make adjustments, we suggest that you use this calculator to determine the declination for the area in question. Otherwise determine your heading using a reliable map.

8. Recording Extent

The extent is a measure of the size of the area within which collecting events or observations occurred for a given locality. Assuming the locality is recorded as a coordinate, the extent is the distance from that point to the furthest point where collecting or observations occurred in that locality. Extent has not traditionally been recorded with collecting activities, but can be important where activities have taken place over a small range, along a transect, or over an area (for example it is common to record bird observations over a 2 ha area).

Collecting events or observations often take place in an area described collectively by a single locality (e.g., within 1 km of the place described in the recorded locality). Without a measure of the potential deviation from the point provided, a user of the data usually has no way of knowing how specific the locality actually is. The extent is a simple way to alert the user that, for example, all of the specimens collected or observations made at the stated coordinates were actually within an area of up to 0.5 miles from that point. It can be quite helpful at times to include in your field notes a large-scale map of the local vicinity for each locality, marking the area in which the collecting and observations occurred.

> **Hint:** A 1 km linear trap line for which the coordinates refer to the center has an extent of 0.5 km. A 2ha area where the coordinates are given at the center of a circle has an extent of ~80 m.

9. Recording Year of Collection

The year a collection was made can often affect the georeferencing of a location. Towns, roads, counties, and even countries can change names and boundaries over time. Rivers and coastlines can change position, billabongs and ox-bow lakes can come and go, localities (such as towns) can change size and shape, and areas of once pristine environment may become farmland or urban areas. Dated maps may no longer represent the current situation. The date is an important characteristic of the collection and must be taken into account when determining a georeference.

> **Example:** "Collecting localities along the Alaska Highway are frequently given in terms of milepost markers; however, the Alaska Highway is approximately 40 km shorter than it was in 1942 and road improvements continue to re-route and shorten it every year. Accurate location of a milepost, therefore, would require cross-referencing to the collecting date. To further complicate matters, Alaska uses historical mileposts (calibrated to 1942 distance), the Yukon uses historical mileposts converted to kilometers, and British Columbia uses actual mileage (expressed in kilometers)".
>
> (From Wheeler *et al.* 2001).

10. Documentation

Record the sources of all measurements. Minimally, include map name and scale, GPS model, the datum, the source for elevation data, the UTM Zone if using UTM coordinates, and the extent of the location or collecting event.

---

[5] National Geophysical Data Center. 2004. Estimated Value of Magnetic Declination .

Using a GPS. For the best accuracy of a location determined by GPS it is important to document:

- The coordinates obtained from the GPS
- The datum
- The accuracy reported by the GPS
- Make of GPS receiver used

Note! Most GPS devices do not record accuracy with the waypoint data, but provide it in the interface showing current satellite conditions.

Note!: The accuracy reported by most GPS recorders is only a relative accuracy for the instrument on which it is read and not real accuracy. For many GPS recorders, the accuracy reported is almost always smaller than warranted.

Example:

Locality: "Modoc National Wildlife Refuge, 2.8 mi S and 1.2 mi E junction of Hwy. 299 and Hwy. 395 in Alturas, Modoc Co., Calif."

Lat/Long/Datum: 41.45063, −120.50763 (WGS84)
Elevation: 1330 ft
GPS Accuracy: 24 ft
Extent: 150 ft
References: Garmin Etrex Summit GPS for coordinates and accuracy, barometric altimeter for elevation.

(From MVZ Guide for Recording Localities in Field Notes)

11. Recording Data for Small Labels

An issue that often arises with insect collections is the problem of recording locality information on small labels. This should not be as big a problem as previously because new technologies allow for linking information on the label to a database (through bar codes, etc.) with the recording of basic information on the label. See Wheeler *et al.* (2001) on guidelines for preparing labels for terrestrial arthropods, but bear in mind the principles laid out in this document when preparing data for insect labels, especially the recording of datums, etc., which are not covered in that document.

12. New Technologies

A number of new technologies are beginning to make data recording in the field a lot easier. For example, a number of companies have recently released Personal Digital Assistants (PDAs) with built-in GPS receivers that can, depending on the type, record to a relatively high degree of accuracy. While these are excellent for recoding locality information in the field for later transfer to the database and for the preparation of labels, many do not include an exterior aerial for receipt of the satellite data and this is likely to reduce the accuracy of the recorded information. The lack of an exterior aerial makes the need for clear line of site for the satellites more important.

The use of Globally Unique Identifiers (GUIDs) for uniquely identifying individual objects and other classes of data (such as collections and observations) are under discussion. We recommend that these be followed once a stable system is implemented. Further information can be found on the TDWG[6] and GBIF[7] websites.

---

[6] http://www.tdwg.org/TDWG_GUID.htm
[7] http://www.gbif.org

Beginning the Georeferencing Process

1. Introduction

A number of issues must be addressed before one begins to georeference. It may appear to be a daunting task at the beginning, however there are many ways the process can be simplifed and made more practical.

Managers and curators are sure to ask many of the following questions and more:

- How hard is this going to be?
- How long is it going to take?
- What proportion of my collection is already digitized?
- What is the current condition of the collection?
- What are the advantages and disadvantages of georeferencing the collection?
- How will the georeferenced data be used and by whom?
- What kind of expertise am I going to need?
- What supervision will be needed and who will do it?
- To what extent will I have to, or want to change my data model?
- How much is it going to cost and what resources are available for georeferencing?
- What tools exist to help me?
- Can I trust what comes out of these tools?
- How many data entry staff will I need?
- What training will I need to give my data entry staff?
- How much of the established best practices do I really need to follow?

This document will not answer all these questions, as many are institution specific, however, it should provide the answer to some, and provide the means of determining the others.

The first issue that will need to be addressed is the database management system:

- Will my current database cope or do I need to have it modified?
- How will I need to modify my user interface to make it easier for data entry operators to georeference?
- What is the most efficient way to go about data entry, including the georeferencing?

This document does not cover methods of general data entry. There are many ways that this may be conducted. These include direct entry from the label with the specimen or ledger brought to the computer; use of PDA's where the computer is brought to the specimen; the use of scanning or photographic (still or video) equipment to capture the label information so that the data entry operator can enter the information from a screen; or use of handwriting and OCR tools to capture the data, etc. Some of these methods are only just becoming practical, but you should make an active decision on the method that best suits your institution.

The next section will help you decide if your database will need modifying or not, and to what extent. It is often tempting to just include fields for the georeferenced coordinates and ignore any additional fields; however, you (or those who follow after you) are sure to regret taking such an option further down the line. The associated information on methods used to determine the georeference, and on the extent and uncertainty associated with the georeference, are very important pieces of information for the end user. Additionally, these are very important pieces of information for managing and improving the quality of your information.

Good examples of production systems that are well documented are the Mountains and Plains Spatio-Temporal Database Informatics Initiative (MaPSTeDI) program and the Mammal Networked Information System (MaNIS). It is worth looking at the processes these projects go through for georeferencing data.

2. The Resources Needed

Each institution will have needs for different resources in order to georeference their collections. The basics, however, include:

- A database and database software (we do not recommend the use of spreadsheets)
- Topographic maps (electronic, paper or both)
- Access to a good gazetteer – (many are available free via the Internet, either for downloading, or via on-line searching)
- Preferably internet access (as there are many resources on the Internet that will help in georeferencing and locating places)
- Suitable computer hardware

Further information on some of these requirements can be found on the MaPSTeDI site under “What you Need”.

3. Fields to Include in your Database

One of the key aspects to efficient georeferencing is setting up a database correctly.

Some georeferencing projects (e.g., MaPSTeDI) use a separate working database for data entry operators so that the main data are not modified and day-to-day use of the database is not hindered. The data from the working database can be checked for quality, and then uploaded to the main database from time to time. Such a way of operating is institution dependant, and may be worth considering.

a. Determine what fields you need[8]

This step seems self-explanatory but it is surprising how often a database is created and finalized before it is determined exactly what the database is supposed to hold. The supervisors for the georeferencing process should be consulted before the database is created to ensure the required georeferencing fields are included in the data model from the outset. Be sure not to lump together dissimilar data into one field. Always atomize the data into separate fields where possible. For instance, if you are collecting latitude and longitude, your database should at least have a separate field for each. Finally, it is also appropriate to use this discussion to decide which fields the data entry operators should see when they are georeferencing. Fields such as date of collection, collector, specimen ID, and taxonomy are very helpful for georeferencing operators to see along with the more obvious locality data.

> **Note!** When you are atomizing data on entry, always include a field or fields that record verbatim the original data so that atomization and other transformations can later be revealed and checked.

b. Locality fields

What are the fields you need in your database to best store georeferencing information? This can perhaps best be divided into two parts, the first are those fields associated with the locality description. Many institutions are currently breaking down locality descriptions into their component parts, i.e., location name, distance and direction, etc., and include this information in separate fields in their databases. With the development of the BioGeomancer toolkit, however, and its automated parsing of natural language locality descriptions, this is now becoming redundant and unnecessary (see further discussion, below). If this break-up of locality information is done, it is important not to replace the free-text locality field (the data as

---

[8] Modified from the MaPSTeDI Guidelines <http://mapstedi.colorado.edu/GuideToGeoreferencing/Georeferencing1-3_SettingUpYourDatabase.html> .

written on the label or in the field notebook), but to add additional fields, as the written format of the description is often important, and this original information should never be over-written or deleted.

Other fields that may be important and useful to aid in georeferencing are:

- date last modified
- township/section/range/Local Government Area/county/state/country
- elevation
- date of collection
- remarks.

A reference worth checking before developing your own data base system is the *Herbarium Information Standards and Protocols for Interchange of Data* (Conn 1996, 2000), which although set up for herbaria, is applicable to most natural history collection data.

c. Georeferencing fields

The second set of fields are those fields actually associated with the georeference, and the georeferencing process. It is recommended, for best practice in georeferencing, that the following fields[9] be added to your database as a minimum. These are in additional to other fields your database may already have, such as Latitude_Degrees, Latitude_Minutes, Latitude_Seconds, etc. Some databases include a user interface to the database that allows data to be entered as degrees, minutes, second, but then translates it to decimal degrees on entry into the database. If this is the case, then both sets of georeferences should be stored, with the decimal degrees used for data exchange. See also the Geospatial Element Definitions Extension to Darwin Core (TDWG 2005).

Field	Comments
Decimal Latitude	See ***Glossary*** for definition. Positive numbers are north of the equator and are less than or equal to 90, while negative values are South of the Equator and are greater or equal to −90. Example: −42.5100 degrees (which is roughly the same as 42º 30' 36" S).
Decimal Longitude	See ***Glossary*** for definition. Positive values are East of the Greenwich Meridian and are less than or equal to 180, negative values are West of the Greenwich Meridian and greater than or equal to −180. Example: -122.4900 degrees (which is roughly the same as 122º 29' 24" W).
Geodetic Datum	The geometric description of a geodetic surface model (e.g., NAD27, NAD83, WGS84). Datums are often recorded on maps and in gazetteers, and can be specifically set for most GPS devices so the waypoints match the chosen datum. Use "not recorded" when the datum is not known. [See separate discussion on datums in this document].
Maximum Uncertainty Estimate	The upper limit of the distance from the given latitude and longitude describing a circle within which the whole of the described locality must lie.
Maximum Uncertainty Unit	The unit of length in which the maximum uncertainty is recorded (e.g., mi, km, m, and ft). Express maximum uncertainty distance in the same units as the distance measurements in the locality description.
Verbatim Coordinates	The original (verbatim) coordinates of the raw data before any transformations were carried out.

---

[9] From the Museum of Vertebrate Zoology Georeferencing Guidelines <http://manisnet.org/GeorefGuide.html>

Verbatim Coordinate System	The coordinate system in which the raw data were recorded. If data are being entered into the database in Decimal Degrees, for example, the geographic coordinates of the map or gazetteer used should be entered (e.g., decimal degrees, degrees-minutes-seconds, degrees-decimal minutes, UTM coordinates).
Georeference Verification Status	A categorical description of the extent to which the georeference and uncertainty have been verified to represent the location and uncertainty for where the specimen or observation was collected. This element should be vocabulary-controlled. Examples: ‘requires verification’, ‘verified by collector’, ‘verified by curator’, ‘not verified’, etc.
Georeference Validation	Shows what validation procedures have been conducted on the georeferences – for example various outlier detection procedures, revisits to the location, etc. Relates to Verification Status.
Georeference Protocol	A reference to the method(s) used for determining the coordinates and uncertainty estimates (e.g., “MaNIS Georeferencing Calculator”).
Georeference Sources	The reference source (e.g., the specific map, gazetteer, or software) used to determine the coordinates and uncertainties. Such information should provide enough detail so that anyone can locate the actual reference used (e.g., name, edition or version, year). Map scales should be recorded in the reference as well (e.g., USGS Gosford Quad map 1:24000, 1973).
Spatial Fit	A measure of how well the geometric representation matches the original spatial representation and is reported as the ratio of the area of the presented geometry to the area of the original spatial representation. A value of 1 is an exact match or 100% overlap. This is a new concept for use with biodiversity data, but one that we are recommending here. [See section on Spatial Fit later in this document].
Georeference Determined By	The person or organization making the coordinate and uncertainty determination.
Georeference Determined Date	The date on which the determination was made.
Georeference Remarks	Comments on methods and assumptions used in determining coordinates or uncertainties when those methods or assumptions differ from, or expand upon, the methods referenced in the Georeference Protocol field.

d. Ecological data

The georeferencing portion of an ecological data collection should be treated in a similar way to specimen and observation data. Often ecological data are recorded using a grid, or transect, etc., and may have a starting locality and an ending locality as well as start time and end time. Sometimes the center of the transect is used as the locality, and half of the length of the transect used for the extent. The uncertainty is then calculated as for other data. If the data are recorded in a grid, then the locality is recorded as the center of the grid, and the extent from that position to the furthest extremity (i.e., the corner) of the grid. These data should be in addition to the recorded locality data, especially where many different fields are used to record the original data. See comments in Appendix.

e. Applying constraints

One of the key ways of making sure that data are as clean and accurate as possible is to assure that data cannot be put in the wrong field and that only data of a particular type can be put into each field. This is done by applying constraints on the data fields – for example, only allowing values between +90 and −90 in the decimal_latitude field. Many of the errors found when checking databases are needless errors – errors that should not be allowed to occur if the database had been set up correctly in the first instance.

With ecological or survey data etc., one could set boundary limits between the starting locality and ending locality. For example, if your methodology always uses 1 km or shorter transects, then the database could include a boundary limit that flagged whenever an attempt was made to place these two points more than 1 km apart.

4. User Interfaces

Good user-friendly interfaces are essential to make georeferencing efficient and fast, and to cut down on operator errors. The layout should be friendly, easy to use, and easy on the eyes. Where possible (and the software allows it) a number of different views of the data should be presented. These views can place emphasis on different aspects of the data and help the data entry operator's efficiency by allowing different ways of entering the data and by presenting a changing view for the operator, thus cutting down on boredom.

In the same way, macros and scripts can help with automated and semi-automated procedures, reducing the need for tedious (and time-consuming) repetition. For example, if data are being entered from a number of collections by one collector, taken at the same time from the same location, the information that is repeated from record to record should be able to be entered using just one or two key strokes.

5. Using Standards and Guidelines

Standards, standard methodologies, and guidelines can help lead to consistency throughout the database and cut down considerably on errors. A set of standards and guidelines should be established at the start of the process and before any georeferencing begins. They should remain flexible enough to cater for new data and changes in processes over time. Standards and guidelines in the following areas can improve the quality of the data and the efficiency of data entry. It is hoped that this document will provide guidelines for many of these. They include:

- Units of measure. Use a single unit of measure in interpreted fields. For example, do not allow a mixture of feet and meters in elevation and depth fields. Irrespective of this, the original units and measurements should be retained in a verbatim field.
- Methods and formats for determining and recording uncertainty and extent.
- Degree of accuracy in determining points where known. (For much legacy data, this will not be determinable).
- Fields that must be filled in (i.e. required fields).
- Format for recording coordinates (i.e., for lat/long, degrees/minutes/seconds, degrees/decimal minutes, or decimal degrees).
- Original source(s) of place names.
- Dealing with typos and other errors in the existing database.
- Number of decimal places to keep in decimal numbers.
- How to deal with “null” values as opposed to zero values (some databases have problems with this).
- How to deal with mandatory fields that cannot be filled in immediately (for example, because a reference has to be found). There may be need for something that can be put in the field that can allow the database to be filed and closed, but that flags that the information is still required.
- What data validation is to be carried out before a record can be considered complete?

Determining these standards and documenting them can help you to maintain them as well as assist you in training and data quality recording. They should form part of the institution's own georeferencing best practice manuals.

6. Choosing a Methodology

Institutions and many experienced georeferencers develop their own preferences for the order in which they georeference. This may be determined by the nature of the data, the way specimens are stored or documented or on the general preference of the operator.

The MaPSTeDI project makes the following recommendations. Note that these will not suit every institution, but may provide a guide:

Georeferencing Procedures

Step 1 - Locate and plot the locality point
The actions involved in this step are described in Finding Coordinates.

Step 2 - Assign a confidence value to the locality
The actions involved in this step are described in Assigning Confidence Values.

Step 3 - Record the georeferenced locality data
This is an important but often under-appreciated step. Most of the mistakes in georeferenced data come from incorrectly recorded data. It is important that all required database fields be filled in as completely as possible in the correct format. The database administrator should place constraints upon some fields to force correct format.

Step 4 - Document the georeferencing rationale for each record
This step is critical because it documents the decision making process for each georeferenced record. For problem records, as well as confusing or detailed records, this information is very important to permit quality checking personnel and museum database users to understand the rationale behind the locality point and confidence value selection. This information also serves as a daily log which permits georeferencing personnel to communicate ideas and report problems. This documentation should be databased with the georeferenced data. If databasing this information is not possible due to database software limitations, it should be kept in electronic documents.

Step 5 - Mark record for further review, if necessary
If the locality cannot be found or is confusing, it should be marked for review by quality checking personnel. This can occur in the database itself or however it is most convenient, but the georeferencer should attempt to complete the record if possible to expedite the quality checking process. The georeferencer should also collect as much relevant locality data as possible to aid the quality checker.

From MaPSTeDI (2004).

a. Sorting records for batch georeferencing

Another set of questions revolves around whether you are best georeferencing each record as you enter the data into the database or if it is better to georeference in a batch after the information on the label has been entered. There are arguments for each method, and again the circumstances of your institution should dictate the best method for you. If your data are stored taxonomically and not geographically (as is the case in the majority of instances) it is often best to georeference in a batch mode by sorting the locality data electronically, and in this way you can deal with many records on one map sheet or area at a time and not be jumping back and forth between map sheets. In other cases, there may be less wear and tear on collections, you may wish to database collections as they are received and before distributing duplicates, or sending on loan, or there may be other good practical reasons to georeference as you go. One advantage of georeferencing as you go is that you may be able to do all the collections of one collector at a time, and virtually follow his/her path, thus reducing errors from not knowing which of several localities may be correct.

Often there is value in georeferencing in batch (tools such as BioGeomancer, work better this way) or in collaboration (MaNIS and MaPSTeDI found that collaborative georeferencing resulted in great efficiency gains), but then reviewing the records using collector and date, or looking at the records taxonomically to check for outliers, and other such data quality flags, afterwards. It usually boils down to what is the best method for your institution, but first, you should consider each of the alternatives before deciding which to use.

The data, once entered into the database, may be sorted using the locality field itself, or some other field such as region, state, nearest named place, etc. You may be able to sort the data into:

- map squares (C-squares[10] often used for marine data, map sheets, UTM zones, etc.)
- geographic regions (country, state, local government area, etc.)
- named place (town, river)
- collector, collector number, and date of collection.

Note! Major efficiency gains can usually be made by georeferencing in batch mode. Consider also, georeferencing collaboratively with other researchers or institutions with similar goals and complementary resources.

b. Using previously georeferenced records

It may be possible to use a look-up system that searches the database for similar localities that may have already been georeferenced. For example, if you have a record with the locality "10 km NW of Campinas", you can search the database for all records with locality "Campinas" and see if any records that mean the same thing as "10 km NW of Campinas" have been georeferenced previously.

An extension of this method could use the benefits of a distributed data system such as the Global Biodiversity Information Facility (GBIF) Portal. A search could be conducted to see if the locality had already been georeferenced by another institution. At present, we quite often find that duplicates of the one collection have been given different georeferences by different institutions. The problem is knowing which of the several georeferences may be the correct one, and one needs to put a lot of faith in another institution's georeferencing methodologies and accuracy determination. This gives strength to the arguments for good documentation with georeferencing, collaboration, and the recording of maximum uncertainty.

Care! This method can add error, if a mistake was made the first time, it will be perpetuated through all later instances.

c. Using BioGeomancer

The BioGeomancer Consortium has developed an online workbench, web services, and desktop applications that will provide georeferencing for collectors, curators and users of natural history specimens, including software tools to allow natural language processing of archival data records that were collected in many different formats and languages. The BioGeomancer Workbench will be launched in September 2006 and is founded on the pioneering efforts of four existing applications, BioGeoMancer Classic, GEOLocate, DIVA-GIS, and the MaNIS Georeferencing Calculator, as well as a number of innovations such as machine learning, spatial data editing, data validation and outlier detection.

BioGeomancer allows the submission of locality descriptions, either singly or in batch mode, and reports back the georeference, along with information on uncertainty. It also passes the data (and other data submitted by the user) through a number of validation tests to check for possible errors in already georeferenced data and to provide further information where several options exist from the locality information.

---

[10] C-Squares <http://www.marine.csiro.au/csquares/about-csquares.htm>

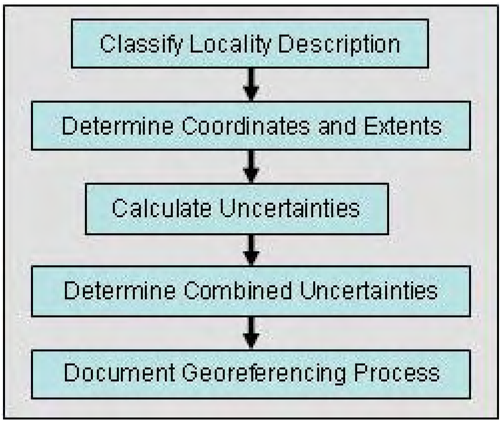
7. Data Entry Operators

The choice and training of data entry operators can make a big difference to the final quality of the georeferenced data. As mentioned earlier, the provision of good guidelines and standards can help in the training process and allow for data entry operators to reinforce their training over time. One of the greatest sources of georeferencing error is the data entry process. It is important that this process be made user-friendly, and be set up so that many errors cannot occur (e.g., through the use of pick lists, field constraints, etc.).

Georeferencing Legacy Data

By far the most difficult issue in georeferencing primary species occurrence data is the massive amount of legacy data held in the world's museums, herbaria, universities, etc. Most modern collectors are now using GPSs or large scale maps to locate their collection events, and thus most of the new data entering institutions already include georeferences. Most museums beginning to database their collections, however, are faced with the massive task of georeferencing the huge backlog of data in their collections, much of it with very little or vague location information. This document aims to assist these institutions with georeferencing their legacy data.

Wieczorek *et al.* (2004) identified five key steps to georeferencing. These have been modified slightly here to include:



Note! These steps should be considered in conjunction with the Appendix to this document.

Refer to the original document for a detailed explanation. We have extracted key points and elaborated on those below.

1. Classifying the Locality Description

Locality descriptions of primary species occurrence data encompass a wide range of content in a vast array of formats, but mostly are cited as a free text description. There are a limited number of categories that locality descriptions can be placed into for georeferencing purposes. The locality type determines the best method of calculating coordinates and uncertainties (see Appendix).

A locality description can contain multiple clauses and can match more than one category. If any one of the parts falls into one of the four categories, '*dubious*' '*cannot be located*', '*demonstrably inaccurate*', or '*captive or cultivated*' (see Appendix), then the locality should not be georeferenced. Instead, an annotation should be made to the locality record giving the reason why it is not being georeferenced.

If the locality description does not fall into one of those four categories, the most specific part of the locality description should be used for georeferencing. For example, a locality written as

'bridge over the St. Croix River, 4 km N of Somerset'

should be georeferenced based on the ***bridge*** rather than on ***Somerset*** as the named place with an offset at a heading. The locality should be annotated to reflect that the bridge was the locality that was georeferenced. If the more specific part of the locality cannot be unambiguously identified, then the less specific part of the locality should be georeferenced and annotated accordingly.

2. Finding the Latitude and Longitude

As discussed elsewhere in this document, geographic coordinates can be expressed in a number of different coordinate systems (decimal degrees, degrees minutes seconds, degrees decimal minutes, UTM, etc.). Conversions can be made readily between coordinate systems, but decimal degrees provide the most convenient coordinates to use for georeferencing for no more profound a reason than a locality can be described with only two attributes - decimal latitude and decimal longitude (Wieczorek 2001). Decimal Degrees are also the coordinate system used in most Geographic Information Systems (GIS).

The first step in determining the coordinates for a locality description is to identify the most specific named place within the description. Coordinates may be retrieved from gazetteers, geographic name databases, maps, or from other locality descriptions that have coordinates. We use the term '***feature***' to refer to not only traditional features, but also to places that may not have proper names, such as road junctions, stream confluences, highway mile pegs, and cells in grid systems (e.g., townships). The source and precision of the coordinates should be recorded so that the validity of the georeferenced locality can be checked. The original coordinate system and the geodetic datum should also be recorded. This information helps to determine sources and degree of maximum uncertainty, especially with respect to the original coordinate precision.

3. Using Offsets

An offset is a displacement from a reference point, named place, or other feature, and is generally accompanied by a direction (or heading). Some locality descriptions give a method for determining the offset ('by road', 'by river', 'by air', 'up the valley', etc.). In such cases, follow the path designated in the description using a map with the largest available scale to find the coordinates of the offset from the named place. It is sometimes possible to infer the offset path from additional supporting evidence in the locality description. For example, the locality

'58 km NW of Haines Junction, Kluane Lake'

suggests a measurement by road since the final coordinates by that path are nearer to the lake than going 58 km NW in a straight line. At other times, you may have to consult detailed supplementary sources, such as field notes, collectors' itineraries, diaries, or sequential collections made on the same day, to determine this information.

4. Finding the Extent

Every named place occupies a finite space, or 'extent'. The extent is usually measured as the distance from the geographic center of the shape that defines the feature, to the furthest extremity of that shape.

If the locality described is an irregular shape (e.g., a winding road or river), there are two ways of calculating the coordinates and determining the extent. The first is to measure along the vector (line) and determine the mid point as the location of the 'named place'. This is not always easy, so the second method is to determine the geographic center (i.e., the midpoint of the extremes of latitude and longitude) of the named place. This method describes a point where the uncertainty due to the extent of the named place is minimized. The extent is then determined as the distance from the determined position to the furthest point at the extremes of the vector. If the geographic center of the shape is used and it does not lie within the locality described (e.g., the geographic center of a segment of a river does not actually lie on the river), then the point nearest the geographic center that lies within the shape is the preferred reference for the named place and represents the point from which the extent should be calculated.

Many localities are based on named places that have changed in size over time; current maps might not reflect the extents of those places when specimens were collected. If possible, extents should be determined using maps contemporary with the events. In most cases, the current extent of a named place will be greater than its historical extent.

5. Calculating Uncertainties

Calculating uncertainties in georeferenced data provides a key provision in determining the data's fitness for use and thus their quality. There are many methods of determining maximum uncertainty; however most of these are complicated, difficult to simply record in most current natural history databases, and are often more sophisticated than necessary for the level of data being used. Over time, it is likely that the recording of uncertainty will be by way of geographic polygons; however, at this stage we recommend the use of a simple point-radius method (see Wieczorek *et al.* 2004) to record the error. The point-radius method is designed to not underestimate the true error. The introduction of polygons will allow, for example, clipping a circle where it overlaps the ocean for terrestrial data, and thereby provide a much more accurate representation of the locality.

Whenever subjectivity is involved, it is preferable to overestimate the maximum error or uncertainty. The following six sources of uncertainty are the most common encountered and these are elaborated below and in the Appendix:

- the extent of the locality
- unknown datum
- imprecision in distance measurements
- imprecision in direction measurements
- imprecision in coordinate measurements
- map scale.

a. Calculating uncertainties due to an unknown datum

Seldom do natural history collections have geographic coordinates recorded together with geodetic datum information. Even with modern collections using a GPS to record coordinates, the geodetic datum is typically ignored. A missing datum reference, however, introduces ambiguity, which varies geographically and adds greatly to the error inherent in the georeferencing.

It is important to record the datum used for the coordinate source (GPS, map sheet, gazetteer) if it is known, or to record the fact that it is not known.

Differences between datums may cause an error in true location from a few centimeters to around 1000 meters (US Navy *n. dat.*), or even, in some extreme instances, up to 3.552 km (Wieczorek *et al.* 2004). Some known average and/or maximum differences between datums are cited in Table 1. Note that the difference between datums is not a linear relationship and they do not always vary in the same direction. For example, the difference between NAD27 and WGS84 in the conterminous USA varies between 0 and 104 m (Wieczorek *et al.* 2004).

Datum from	Region or Location	Datum to	Difference
AGD66	Australia	AGD84	Max ± 0-5 m
AGD66/84	Australia	GDA94	Max ± 200 m
AGD66/84	Australia	WGS84	Max ± 200 m
GDA94	Australia	WGS84	Max ± <1 m
NAD 1983	North America	WGS84	Max ± <1 m
NAD27	North America	WGS84	Max ± 200 m
NAD 27	Contiguous USA	WGS84	Max ± 105 m
NAD 27	Aleutian Islands, Alaska	WGS84	Max ± 235 m
NAD 27	Hawaii	WGS 84	~ 500 m
TOKYO	Japan	WGS84	Max ± 750 m
ED-50	Europe	WGS84	Max ± 175 m
ARC-50	Africa	WGS84	Max ± 265 m
INDIAN 1975	Bangkok, Thailand	WGS84	~ 405 m
INDIAN 1956	Delhi, India	WGS84	~ 135 m
INDIAN 1956	Mumbai, India	WGS84	~ 120 m
HONG KONG 1973	Hong Kong	WGS84	~ 320 m
LUZON	Manila, The Philippines	WGS84	~ 225 m
TOKYO-KOREA	Seoul, South Korea	WGS84	~380 m
KERTAU 1948	Singapore	WGS84	~190 m

Table 1: Shows the maximum differences over total range, or approximate differences at a location for a number of common datums. Data derived from US Navy (*n. dat.*), Srivastava and Ramalingam (2006) and Wieczorek *et al.* (2004). All except the very small values have been rounded to the nearest 5 m.

b. Calculating uncertainty from distance

Precision can be difficult to gauge from a locality description as it is seldom, if ever, explicitly recorded. Further, a database record may not reflect, or may reflect incorrectly, the precision inherent in the original measurements, especially if the locality description in the database has undergone normalization, reformatting, or secondary interpretation of the original description.

There are a number of ways of calculating uncertainty from distances. In this document, we have taken a conservative approach. The form in which a distance is written can often give an indication of the precision and hence the uncertainty. One method is to use half of the precision (for example, 10.5 mi N of Bakersfield could reasonably be expected to mean 10½ mi and thus be between 10.25 and 10.75 mi N, or 10.5 ±0.25 mi N of Bakersfield). The uncertainty in the measurement is thus 0.25 mi.

A second method, and that recommended here, is one proposed by Wieczorek *et al.* (2004) and assumes that many records have undergone a certain amount of interpreptation or transformation when being entered into the database, and thus a record of 10¼ mi may be entered into the database as 10.25 mi. The precision implied in the value 10.25 is thus a false precision (see glossary) and should not be assumed to be between 10.24 and 10.26. The method of Wieczorek *et al.* (2004) bases the estimate of uncertainty on the fractional part of the distance – i.e. calculated by dividing 1 by the fractional denominator. Thus:

- for 9 km, the fraction is 1/1 and thus the uncertainty estimate is 1 km;
- for 9.5 km, the fraction is ½ and the uncertainty estimate 0.5km;
- for 9.25 km, the fraction is ¼ and the uncertainty estimate 0.25 km;
- for 9.6 km, the fraction is 1/10 and the uncertainty estimate 0.1 km.

For distance measurements which are positive integer powers of 10, the uncertainty estimate is based on 0.5 times ten to that power (see Table 2).

A third method, suggested by Frazier *et al.* (2004), is for distances that are given as multiples of 10, or fractions of 100 such as 25 and 75. This method recommends using 15% of the distance as the uncertainty. Thus, for 10 km, the uncertainty would be 1.5 km; and for 75 km it would be 11.25 km. This gives a smaller uncertainty than recommended by Wieczorek *et al.* for distances between 10 and 30 km, and a greater value for distances between 40 and 90 km (Table 2).

Example	**Uncertainty** (Wieczorek *et al.* 2004)	**Uncertainty** (Frazier *et al.* 2004)
10.6 km N of Bakersfield	0.1 km	
10.5 mi N of Bakersfield	0.5 mi	
10 km N of Bakersfield	5 km	1.5 km
30 km N or Bakersfield	5 km	4.5 km
140 mi N of Bakersfield	5 mi	21 mi
200 mi N of Bakersfield	50 mi	30 mi
2000 m N of Bakersfield	500 m	300 m

Table 2. Calculating uncertainy using the precision in a distance recording.

Precision can also be masked or lost when measurements are converted, such as from feet to meters, or from miles to kilometers.

Care! Be careful that the value you are using for precision when calculating the uncertainty is a true precision and not a false precision. For example, converting a collector's recording of 16 miles (with a precision of 1 mile) to 25.6 km (with a precision of 0.1 km) leads to a level of precision that is more than 10 times as precise as the original.

Note! Further details of calculations used to determine uncertainties from distance precision can be found in Wieczorek (2001) and Wieczorek *et al.* (2004)

c. Calculating uncertainties from extents of localities

The extents of named places are an important source of uncertainty. Points of reference for named places may change over time – post offices and courthouses are relocated, towns change in size, the courses of rivers change, etc. Moreover, there is no guarantee that the collector paid attention to any particular convention when reporting a locality as an offset from a named place. For example,

'4 km E of Bariloche'

may have been measured from the post office, the civic plaza, or from the bus station on the eastern edge of town, or anywhere else in Bariloche. When calculating an offset, we generally have no way of knowing where the collector started to measure the distance.

We recommend uncertainty be determined by measuring the distance from the point marked by the coordinates to the point in the named place furthest from those coordinates. The magnitude of the uncertainty will be smallest if the coordinates mark the geographic center of the named place and the maximum uncertainty is then the distance from that point to the furthest point in the locality. In most cases, the current extent of a named place will be greater than its historical extent and the uncertainty may be somewhat overestimated if current maps are used. When documenting the georeferencing process, it is recommended that the named place, its extent, and the source of the information all be recorded.

d. Calculating uncertainty from direction

The calculation of uncertainty from the precision in which a direction is recorded depends on distance from the reference point. The uncertainty will increase as one moves further from the source. For simple calculations of precision due to direction – see Table 3.

> **Note!** The uncertainty due to directional imprecision increases with distance, so it can only be calculated from the combination of distance and direction (see below).

Precision	Interpretation	Example	Directional Uncertainty
N	Between NW and NE	10.6 km N of Bakersfield	45°
NE	Between NNE and ENE	10.5 mi NE of Bakersfield	22.5°
NNE	Between N of NNE and E of NNE	10 km NNE of Bakersfield	11.25°

Table 3. Calculating uncertainy using the precision of the recorded direction (derived from Wieczorek *et al.* 2004).

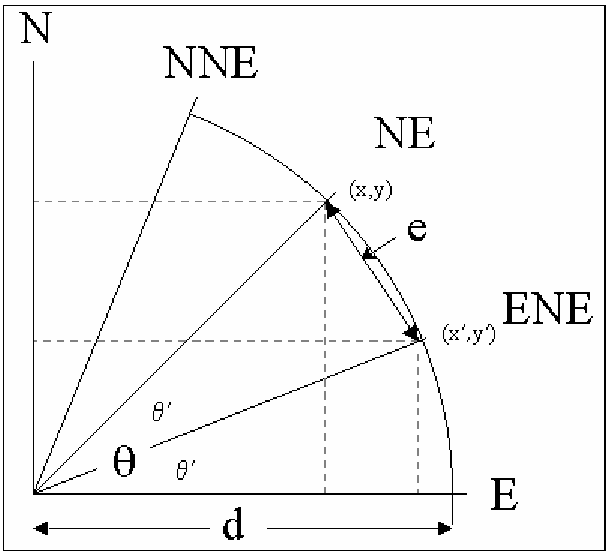


Fig. 1. A simple diagram showing directional precision where $x = d\cos(\theta)$, $y = d\sin(\theta)$, $x' = d\cos(\theta')$, and $y' = d\sin(\theta')$. From Wieczorek *et al.* (2004).

Using the example

'10 km NE of Bakersfield'

if we ignore distance imprecision, the uncertainty due to the direction imprecision (Figure 1) is encompassed by an arc centered 10 km (**d**) from the center of Bakersfield (at **x,y**) at a heading of 45 degrees (**θ**), extending 22.5 degrees in either direction from that point. At this scale the distance (**e**) from the center of the arc to the furthest extent of the arc (at **x',y'**) at a heading of 22.5 degrees (**θ'**) from the center of Bakersfield can be approximated by the Pythagorean Theorem.

e=sqrt($(x'-x)^2 + (y'-y)^2$) the uncertainty in the above example is 3.90 km

This shows just one simple example. For details and formulae for calculating more complicated uncertainties, see Wieczorek (2001) and Wieczorek *et al.* 2004. Because of the complicated nature of these calculations, it is often best to use the MaNIS Georeferencing Calculator - see discussion below.

e. Calculating uncertainty from coordinate precision

Geographic coordinates should always be recorded using as many digits as possible; the precision of the coordinates should be captured separately from the coordinates themselves, preferably as a distance, which conserves its meaning regardless of location and coordinate transformations. Recording coordinates with insufficient precision can result in unnecessary uncertainties. The magnitude of the uncertainty is a function of not only the precision with which the data are recorded, but also of the datum and the coordinates themselves. This is a direct result of the fact that a degree does not correspond to the same distance everywhere on the surface of the earth.

Table 4 shows examples of the contributions to uncertainty for different levels of precision in original coordinates using the WGS84 reference ellipsoid. Calculations are based on the same degree of imprecision in both coordinates and are given for several different latitudes. Approximate calculations can be made based on this table, however, more accurate calculations can be obtained using the MaNIS Georeferencing Calculator (see further discussion, below).

From Table 4, it can be seen that an observation recorded in degrees, minutes, and seconds (DMS) has a minimum uncertainty of between 32 and 44 meters.

Precision	0 degrees Latitude	30 degrees Latitude	60 degrees Latitude	85 degrees Latitude
1.0 degree	156,904 m	146,962 m	124,605 m	112,109 m
0.1 degree	15,691 m	14,697 m	12,461 m	11,211 m
0.01 degree	1,570 m	1,470 m	1,246 m	1,121 m
0.001 degree	157 m	147 m	125 m	112 m
0.0001 degree	16 m	15 m	13 m	12 m
0.00001 degree	2 m	2 m	2 m	2 m
1.0 minute	2,615 m	2,450 m	2,077 m	1,869 m
0.1 minute	262 m	245 m	208 m	187 m
0.01 minute	27 m	25 m	21 m	19 m
0.001 minute	3 m	3 m	3 m	2 m
1.0 second	44 m	41 m	35 m	32 m
0.1 second	5 m	5 m	4 m	4 m
0.01 second	1 m	1 m	1 m	1 m

Table 4. Table showing metric uncertainty due to precision of coordinates based on the WGS84 datum at varying latitudes. Uncertainty values have been round up in all cases. From Wieczorek (2001).

Care! ***False precision*** can arise when transformations from degrees minutes seconds to decimal degrees are stored in a database (see ***Glossary*** for expanded discussion).

Never use precision in a database as a surrogate for the coordinate uncertainty; instead, record the uncertainty explicitly, preferably as a distance.

Note! Details of calculations used to determine uncertainties in coordinate precisions can be found in Wieczorek (2001) and Wieczorek *et al.* (2004).

Example:

Lat: 10.27º **Long:** −123.6º **Datum:** WGS84

In this example, the lat/long precision is 0.01 degrees. Thus, latitude error = 1.1061 km, longitude error = 1.0955 km, and the uncertainty resulting from the combination of the two is 1.5568 km.

f. Calculating uncertainty by reading off a map

One of the most common methods of finding coordinates for a location is to estimate the location from a paper map. Using paper maps can be problematic and subject to varying degrees of inaccuracy. Unfortunately, the accuracy of many maps, particularly old ones, is undocumented. Accuracy standards generally explain the physical error tolerance on a printed map, so that the net uncertainty is dependent on the map scale. Map reading requires a certain level of skill in order to determine coordinates accurately, and different types of maps require different skills. Challenges arise due to the coordinate system of the map (latitude and longitude, UTM, etc.), the scale of the paper map, the line widths used to draw the features on the maps, the frequency of grid lines, etc.

The accuracy of a map depends on the accuracy of the original data used to compile the map, how accurately these source data have been transferred onto the map, and the resolution at which the map is printed or displayed. For example, USGS maps of 1:24,000 and 1:100,000 are different products. The accuracy is explicitly dependent on scale but is due to the different methods of preparation. When using a map, the user must take into account the limitations encounted by the map maker such as accuity of vision, lithographic processes, plotting methodologies, and symbolization of features (e.g., line widths) (NOAA 2001).

With paper topographic maps, drawing constraints may restrict the accuracy with which lines are placed on the map. A 0.5 mm wide line depicting a road on a 1:250,000 map represents 125 meters on the ground. To depict a railway running beside the road, a separation of 1-2 mm (250-500 meters) is needed, and then the line for the railway (another 0.5 mm or 125 meters) makes a total of 500-750 m as a minimum representation. If one uses such features to determine an occurrence locality, for example, then minimum uncertainty would be in the order of 1 km. If thicker lines were used, then appropriate adjustments would need to be made (Chapman *et al.* 2005).

Note! A digital map is never more accurate than the original from which it was derived, nor is it more accurate when you zoom in on it. The accuracy is strictly a function of the scale and digitizing errors of the original map.

Table 5 shows the inherent accuracy of a number of maps at different scales. The table gives uncertainties for a line 0.5 mm wide at a number of different map scales. A value of 1 mm of error can be used on maps for which the standards are not published. This corresponds to about three times the detectable graphical error and should serve well as an uncertainty estimate for most maps.

Scale of Map	Map Horizontal Uncertainty (Geosciences Australia[11])	Map Horizontal Uncertainty (USGS[12])	NIMA Product	NIMA Product Accuracy (US Navy)[13]
1:1000	0.5 m	2.8 ft		
1:10,000	5 m	28 ft		
1:25,000	12.5 m	70 ft	City Graphic	>50 m
1:50,000	25 m	139 ft	Topo	50 m
1:75,000			Nautical	75 m
1:100,000	50 m	278 ft		
1:250,000	160-300 m	695 ft	JOG	250 m
1:500,000			TPC	1,000 m
1:1 million	500 m	2,777 ft	ONC	2,000 m

Table 5. Horizontal uncertainty and accuracy associated with a 0.5 mm line on maps of different scales.

The table uses data from several sources. The TOPO250K Map series is the finest resolution mapping that covers the whole of the Australian continent. It is based on 1:250,000 topographic data, for which Geoscience Australia (2003) defines the accuracy as "*not more than 10% of well-defined points being in error by more than 160 meters; and in the worst case, a well defined point is out of position by 300 meters*". The USGS Map Horizontal Uncertainty is calculated from US Bureau of Budget (1947) which states that "*for maps on publication scales larger than 1:20,000, not more than 10 percent of the points tested shall be in error by more than 1/30 inch, measured on the publication scale; for maps on publication scales of 1:20,000 or smaller, 1/50 inch*." These values need to be taken into account when determining the uncertainty of your georeference. The third set of values was obtained from the US Navy with reference to various NIMA[14] (US National Image and Mapping Agency) products.

If you are using phenomena that do not have distinct boundaries in nature to determine a locality (such as soils, vegetation, geology, timberlines, etc.) then err vastly on the side of conservatism when determining an uncertainty value as such boundaries are seldom accurate, often determined at a scale of 1:1 million or worse and would have a minimum uncertainty of between 1 and 5 km. Also be aware that coastlines vary greatly at different scales (see Chapman *et al.* 2005) and rivers are often straightened on smaller scale maps, and can thus include uncertainties far greater than are generally recorded on maps whose accuracies are determined from "well-defined" points such as buildings, road intersections, etc. In addition, coastlines and river paths can change greatly over time (Bannerman 1999) and thus the date of the map needs to be taken into account when determining uncertainty.

For elevation where contours are drawn on a map, the vertical uncertainty is usually described as being half of the contour interval.

---

[11] Based on 0.5mm of accuracy per unit of scale, except for the 1:250,000 map series where the figure supplied with the data has been used.

[12] Derived from United States National Map Accuracy Standards (US Bureau of Budget 1947) http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/nmas/NMAS647.PDF

[13] Navigator of the Navy https://www.navigator.navy.mil/navigator/accuracy_0009.ppt

[14] US National Image and Mapping Agency (NIMA) http://erg.usgs.gov/nimamaps/

Care! Care must be used when using a digital map that records the scale in the form of text (1:100,000, etc.) rather than by using a scale bar, as the resolution of the computer screen, and the level of zooming will change the apparent scale of the map being viewed. (It does not change the scale at which the map was prepared). This also applies to maps printed from a digital map. When preparing digital maps, always include scale as a scale bar and do not just record scale in textual form (e.g., 1:20,000).

g. Calculating combined uncertainties

When combining uncertainties from different sources, it is not as simple as taking the average or adding them together. Uncertainties inherent in the location of the named place, in its extent, in the direction of the offset, and the distance of the offset, are just four sources that need to be combined to get an overall uncertainty. A detailed discussion of the calculations involved can be found in Wieczorek (2001) and Wieczorek *et al.* (2004), and for a practical way of calculating uncertainties in locations, we recommend use of the MaNIS Georeferencing Calculator. In the Appendix to this document, we provide a number of examples.

h. Using the MaNIS Georeferencing Calculator

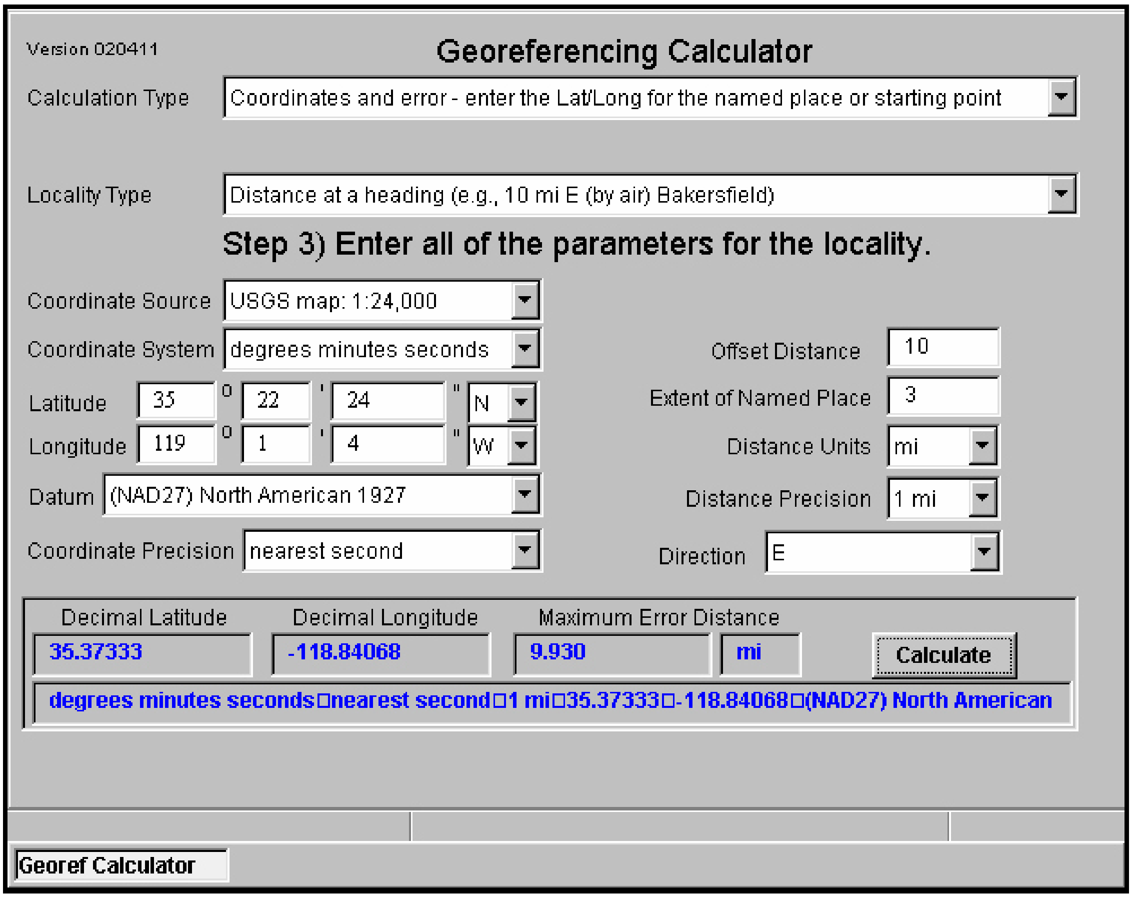
The MaNIS Georeferencing Calculator[15] (Figure 2), is a java applet created as a tool to aid in the georeferencing of descriptive localities such as those found in museum-based natural history collections. It was specifically designed for the Mammal Networked Information System (MaNIS) Project and has been adopted as well by both HerpNet, ORNIS, and other collaborative database initiatives.

The application makes calculations using the methods described in the Georeferencing Guidelines (Wieczorek 2001). We recommend its use generally by all natural history institutions to calculate uncertainty in location data without the need for a detailed understanding of the complicated underlying algorithms. The more institutions that use this one method, the more consistent will be the quality of data across and between institutions, making it easier for users to evaluate the quality of the data. We recommend reading both Wieczorek (2001) and the MaNIS Georeferencing Calculator Manual (Wieczorek 2002) for an understanding of the calculations involved and an understanding of how the calculator works.

The algorithms developed for the Georeferencing Calculator have also been incorporated in the the uncertainty calculations used in the BioGeomancer georeferencing tools. This too will serve to standardize the determination of this important attribute of data quality documentation.

---

[15] MaNIS Georeferencing Calculator <http://www.manisnet.org/gc.html>



Version 020411

Georeferencing Calculator

Calculation Type: Coordinates and error - enter the Lat/Long for the named place or starting point

Locality Type: Distance at a heading (e.g., 10 mi E (by air) Bakersfield)

Step 3) Enter all of the parameters for the locality.

Coordinate Source: USGS map: 1:24,000

Coordinate System: degrees minutes seconds

Latitude: 35 ° 22 ' 24 " N

Longitude: 119 ° 1 ' 4 " W

Datum: (NAD27) North American 1927

Coordinate Precision: nearest second

Offset Distance: 10

Extent of Named Place: 3

Distance Units: mi

Distance Precision: 1 mi

Direction: E

Decimal Latitude	Decimal Longitude	Maximum Error Distance	
35.37333	-118.84068	9.930	mi

degrees minutes seconds□nearest second□1 mi□35.37333□-118.84068□(NAD27) North American

Fig. 2. A snap shot of the MaNIS Georeferencing Calculator showing maximum uncertainty calculation for the locality: '*10 mi E (by air) Bakersfield*'. From Wieczorek (2002).

6. Determining Spatial Fit

Spatial fit is a new georeferencing concept designed to allow for a measure of how well a given geometric representation matches the original spatial representation. This is useful when spatial transformations change the way a locality is represented, either to mask its detail, or to match an agreed upon schema for data sharing (such as fitting locations to a grid cell).

A spatial fit with a value of 1 is an exact match or 100% overlap. If the geometry given does not completely encompass the original spatial representation, then the spatial fit is zero (i.e., some of the original is outside the transformed version, which we interpret as not being a fit). If the transformed shape does completely encompass the original spatial representation, then the value of the spatial fit is the ratio of the area of the transformed geometry to the area of the original spatial representation. Special case: If the original spatial representation is a point and the geometry presented in not a point, then the spatial fit is undefined. The range of values of spatial fit is 0, 1, greater than 1, or undefined.

An example of the applicability of the spatial fit is where a point representing a terrestrial collection lies close to the coast, and the calculated uncertainty radius encompasses some marine area. In this case the Spatial Fit would be greater than 1 as it represents an area greater than the real uncertainty.

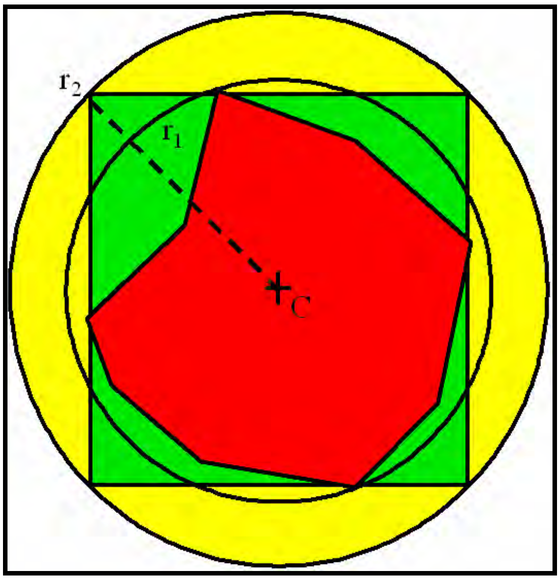


Fig. 3. A diagram illustrating the spatial fit of a number of locations that can be described by a polygon, a grid, or a point.

Figure 3 illustrates a few examples of the definition of spatial fit and these are elaborated below:

1) Suppose the original spatial representation of a locality was given by the **red polygon** with area **A**.

The spatial fit of the **yellow circle** would be	**$(PI*r_2^2)/A$**
The spatial fit of the **green bounding box** would be	**$(2*r_2^2)/A$**
The spatial fit of the **black circle** (r_1) would be	**$(PI*r_1^2)/A$**
The spatial fit of the **red polygon** would be	**1**
The spatial fit of the **point C** would be	**0**

2) Suppose the original spatial representation of a locality was given as the **green bounding box** with area **$2*r_2^2$.**

The spatial fit of the **yellow circle** would be	**$(PI*r_2^2)/(2*r_2^2)$**
The spatial fit of the **green bounding box** would be	**1**
The spatial fit of the **black circle** (r_1) would be	**0**
The spatial fit of the **red polygon** would be	**0**
The spatial fit of the **point C** would be	**0**

3) Suppose the original spatial representation of a locality was given as the **black circle** with area **$PI*r_1^2$.**

The spatial fit of the **yellow circle** would be	**r_2^2/r_1^2**
The spatial fit of the **green bounding box** would be	**0**
The spatial fit of the **black circle** (r_1) would be	**1**
The spatial fit of the **red polygon** would be	**0**
The spatial fit of the **point C** would be	**0**

4) Suppose the original spatial representation of a locality was given as the **point C**.

The spatial fit of the **yellow circle** would be	**Undefined**
The spatial fit of the **green bounding box** would be	**Undefined**
The spatial fit of the **black circle** (r_1) would be	**Undefined**
The spatial fit of the **red polygon** would be	**Undefined**
The spatial fit of the **point C** would be	**1**

Maintaining Data Quality

Data that have been incorporated into the database and georeferenced need to be maintained and checked for quality. The quality checking process involves a number of steps, including receiving feedback from users, providing feedback to collectors, and running various validation tests. For more information on ***data quality*** and what it means for primary species collection data see Chapman (2005b). Two major principles associated with data quality and data cleaning are:

- Error prevention is preferrable to error correction.
- The earlier in the information chain that you can detect an error, the cheaper it will be to correct it.

1. Feedback to Collectors

Maintaining the quality of the data may require giving feedback to others. For example, if you find that a particular collector is not recording his collection information correctly (e.g., not recording the datum with the georeference information), then you need to provide feed back to him so that future records have a lower level of error and thus a higher quality. See the earlier chapter on ***Collecting and Recording Data in the Field***. Key issues that may require feedback to collectors include:

- Making sure the datum is recorded with all GPS readings
- Encouraging consistent use of a standard coordinate system (e.g., encourage collectors to use decimal degrees wherever possible)
- Recording localities in a consistent and clear manner
    - Using nearest named place and offsets
    - Recording ‘by road’ or ‘by air’
- Using a barometric altimeter for recording elevation.

2. Accepting Feedback from Users

Feedback from users can be one of the most valuable resources for maintaining the quality of one’s collections. For this to work, however, the institution needs to set up a good feedback mechanism. There needs to be a process whereby all feedback related to quality are checked and the results documented (see Chapman 2005a, b). Feedback may be be from other institutions holding duplicates of some of your collections, from users who are carrying out analyses on large amounts of data and find records that are either wrongly georeferenced, or wrongly identified, or from users who are carrying out data quality checking on related records. All feedback is important, and should not be ignored. Checks carried out should also always be documented so that the same ‘error’ is not checked over and over again.

3. Data Checking and Cleaning

An important but often overlooked aspect to any georeferencing project is the checking of the georeferenced data that goes into the database. This aspect is often ignored because of lack of funds or personnel. However, because the point of any georeferencing project is to produce geographic coordinates linking a specimen to a place on a map or environmental data, it is important that the coordinates chosen are truly the best ones for the location. Not only does it improve the quality of data, but it also identifies trends and habits in georeferencing that may need to be corrected. Often a graduate assistant, intern, or someone with more experience will do most of the quality checking.

a. Data entry

One of the major sources of error in georeferencing is at the stage of data entry. Errors can be reduced by the establishment of good data entry procedures – use of pick lists, field constraints, etc., to reduce the possibility of error. However, once these are in place and working, then regular checks need to be carried out on the data entry operators and on the process of data entry. Quality checking can take several forms, but we recommend that it use the following two taken from the MapSTeDI Georeferencing Guidelines.

The first is to check the accuracy of the georeferencing. This process involves checking a certain number of each georeferencer's records. Based on various trials, it is recommended that the first 200 records that a new georeferencer completes be checked for accuracy. Not only is this initial checking beneficial to the accuracy of the data, but also it is essential to allow the georeferencer to improve and learn from making mistakes. If significant problems still exist after the initial 200 records, an additional batch of 100 records should be checked. After the quality checker, usually a highly experienced georeferencer, is satisfied with the new georeferencer's abilities, the quality checking is reduced to 10 randomly selected records out of every 100 completed. If more than two records are found to be incorrect within that 10, an additional 20 records should be checked. The quality checker may ask the georeferencer to redo the entire 100 if enough problems exist. After a period of few mistakes, the checking is reduced to five records for every 100 or at the quality checker's discretion.

To summarize:

- Initial 200 records should be checked. If problems remain, check groups of 100 until satisfied with georeferencer's abilities.
- Regular checks of 10 randomly selected records for every 100
- If more than 2 incorrect records, quality checker should check 20 more records and can ask georeferencer to redo entire 100.
- After awhile, the regular checks can be reduced to 5 records for every 100.

The second purpose of quality checking is to allow georeferencers to refer difficult or confusing records to the quality checker for help or advice. The quality checker will then resolve these ‘problem records’ as well as possible. Checking problem records can be like detective work. Historical records often have locality descriptions with named places that do not appear on modern maps or gazetteers. To find these localities, it is often necessary to consult several different sources of information. These sources include, but are not limited to catalog books, field notes, other records with similar localities, other collections, scientific and other publications, websites, online databases, specialty gazetteers, and historical maps. Bits of information from several places can often be used to establish the correct coordinates for a historical locality. In addition, some problem records do not make sense because of contradictions or missing or garbled information (see locality type categories in Appendix). These problem records may be the result of mistakes in data entry made in either the paper catalog or the database. It may also be necessary to consult the curatorial staff or even the original collector.

b. Data validation

Data validation (checking for errors) can be a time-consuming process, however, it is one of the most important processes you can carry out with your data. It is not practical to check every record individually, so the use of batch processing techniques and outlier detection procedures, etc. are essential. Fortunately, a number of these have been developed and are available in software products or on-line. Most of these are elaborated in the document *Principles and Methods of Data Cleaning. Primary Species and Species Occurrence Data* published by GBIF (Chapman 2005b) and the information therein will not be repeated here. We recommend that you download and use that document as an adjunct to this one.

There are many methods of checking for errors in georeferenced data. These can involve

- using external databases (collector's itineries, gazetteers, etc.),
- checking against other fields in your own database (making sure the georeference falls within the correct state, country, region, etc.),
- using a GIS to look for records that fall outside polygon boundaries such as bioregions, local government areas,
- using statistical methods such as box plots, reverse jackknifing, cumulative frequency curves, and cluster analysis to identify outliers in latitude or longitude,
- using modelling software in conjunction with statistical analysis to identify outliers in environmental (e.g., climate) space.

Some of these techniques will shortly be available on-line through the GBIF Portal, and the BioGeomancer website, and yet others are available through the stand-alone GIS software DIVA-GIS (Hijmans *et al.* 2005).

c. Making corrections

When making corrections to your database, we strongly recommend that you always add and never replace or delete. For this to happen you will usually require additional fields in the database. For example, you may have 'original' or 'verbatim' georeference fields in addition to the main georeference fields. Additionally, the database may require a number of 'Remarks' fields. Fields that can be valuable are those that describe validation checking that has been carried out – even (and often especially) if that checking has led to confirmation of the georeference. These fields may include information on what checks were carried out, by whom, when and with what results.

d. Truth in labelling

'Truth in Labelling' is an important consideration with respect to documenting data quality. This is especially so where data are being made available to a wider audience, for example, through the GBIF data nodes. We recommend that documentation of the data and their quality be up-front and honest. Error is an inescapable character of any dataset, and it should be recognized as a fundamental attribute of those data. All databases have errors, and it is in no-one's interest to hide those errors. On the contrary, revealing data actually exposes them to editing, validation and correction through user feedback, while hiding information almost guarantees that it remain dirty and of little long-term value.

4. Reponsibilities of the Manager

It is important that the manager maintain good sets of documentation (guidelines, best practice documents, etc.), ensure that there are good feedback mechanisms in place, and ensure that data quality procedures are maintained, are up-to-date, and are being implemented. For further responsibilities, we refer you to the document Principles of Data Quality (Chapman 2005a) which should be read as an adjunct to this document.

5. Responsibilities of the Supervisor

The georeferencing supervisor has the principle responsibility for maintaining the quality of the data on a day-to-day basis. Perhaps their key responsibility is to supervise the data-entry procedures (see *Data Entry*, above), and the data validation, checking and cleaning processes. This role is the key role in any georeferencing process, along with that of the data entry operators. It is important that the duties and responsibilities be documented in the institution's best practice manuals and guidelines.

6. Training

Training is a major responsibility of any institution beginning or conducting the georeferencing of their collections. Good training can reduce the level of error, reduce costs and improve data quality. A Georeferencing and Data Cleaning Training Kit is being planned, and hopefully will be developed over the next couple of years. This will aid institutions in training their data entry operators and supervisors in all aspects of the georeferncing and data quality control processes.

7. Performance Criteria

The development of performance criteria is a good way of ensuring a high level of performance, accuracy and quality in the database. Performance criteria can relate to an individual (data entry operator, supervisor, etc.) or to the process as a whole. It can relate to the number of records entered each week, but we would recommend that it relate more to the quality of entry. Where possible, performance criteria should be finite and numeric so that performance against the criteria can be documented. Some examples may include

- 90% of records will undergo validation checking within 6 months of entry,
- any suspect records identified during validation procedures will be checked and corrected within 30 working days,
- feedback from users on errors will be checked and the user notified of the results within two weeks,
- all documentation of validation checks will be completed and up-to-date.

8. Index of Spatial Uncertainty

An Index of Spatial Uncertainty may be developed and documented for the dataset as a whole to allow for overall reporting of the quality of the dataset. This index would supplement a similar index of other data in the database, such as an index of Taxonomic Uncertainty and would generally be for internal use. Currently, no such universal index exists for primary species occurrence data, but institutions may consider developing their own and testing its usefulness. Such indexes should, wherever possible, be generated automatically and produced as part of a data request from the database and packaged with the metadata as part of the request. Such an index could form the basis for helping users determine the quality of the database for their particular use. The authors of this document would be interested in any feedback from institutions that develop such an index. The index should form an integral part of the metadata for the collection and may include for the georeferencing part of the database:

1. Completeness Index

- percentage of records with georeference fields that have values
- percentage of records with extent fields that have values
- percentage of records with uncertainty fields that have values
- percentage of records with coordinate-precision fields with a value
- percentage of records with datum fields that have a known datum value

2. Uncertainty Index

- average and standard deviation of ‘uncertainty’ value for those records that have a value
- percentage of records with a maximum uncertainty value in each class
    a. <100 m
    b. 100-1,000 m
    c. 1,000-2,000 m
    d. 2,000-5,000 m
    e. 5,000-10,000 m
    f. >10,000 m
    g. not determined

3. Currency Index

- time since last data entry
- time since last validation check

4. Validation Index

- percentage of records that have undergone validation test x
- percentage of records that have undergone validation test y, etc.
- percentage of records identified as suspect using validation tests
- percentage of suspect records found to be actual errors

9. Documentation

Documentation is one of the key aspects of any georeferencing process. Documentation involves everything from record-level documentation such as

- how the georeference was determined,
- what method was used to determine the extent and error,
- what modifications were made (for example, if an operator edits a point on the screen and moves it from point ‘a’ to point ‘b’ it is best practice to document "why" the point was moved and not just record that location was moved from point ‘a’ to point ‘b’ by the operator),
- any validation checks that were carried out, by whom and when,
- flags that may indicate uncertainy, etc.

through to the metadata related to the collection as a whole which may include:

- the overall level of data quality,
- the general checks carried out on the whole data set,
- the units of measurement and other standards adopted,
- the guidelines followed,
- the Index of Uncertainty (see earlier discussion, this chapter).

A second set of documentation relates to

- the institution’s ‘Best Practice’ document which we recommend should be derived from this document and tailored to the specific needs of the institution,
- training manuals,
- standard database documentation,
- guidelines and standards.

We recommend that documentation be made an integral part of any georeferencing process.

Acknowledgements

Many people have contributed ideas to this document, either directly or indirectly through discussions at meetings, publications, or in email correspondence. We would particularly like to thank Larry Speers and Hannu Saarenmaa from the Global Biodiversity Information Facility (GBIF), members and participants in meetings of the Taxonomic Databases Working Group (TDWG) and the Environmental Resources Information Network, Canberra, Australia. Individuals who have supplied suggestions and comments, including on earlier drafts, include Rebecca Shapley, Lynn Kutner (NatureServe), Lee Belbin (TDWG) and Simon Bennett (ERIN).

We would particularly like to thank the members of the BioGeomancer Consortium enumerated below, who, in the course of the project, have delved deep into some of the pressing issues of georeferencing and the use of georeferenced data. We would also like to acknowledge the Gordon and Betty Moore Foundation, whose sponsorship of the BioGeomancer Project led directly to this document as one of the stated outputs.

- Australian Biodiversity Information Services: Arthur Chapman.
- Australian Museum: Paul Flemons, Ajay Ranipeta.
- Centro de Referência em Informação Ambiental: Renato de Giovanni, Alexandre Marino, Sidnei de Souza.
- Ecosystem Associates: Philip Dibner.
- International Rice Research Institute: Robert Hijmans.
- Tulane University: Hank Bart, Nelson Rios, Andy Tseng.
- University of California, Berkeley: John Deck, Stephen Long, Michelle Koo, Craig Moritz, Carol Spencer, Aaron Steele, John Wieczorek, Kristina Yamamoto.
- University of California, Merced: Qinghua Guo, Yu Liu.
- University of Colorado, Boulder: Robert Guralnick, Greg Hill, Dave Neufeld.
- University of Illinois Urbana-Champaign: Eugene Chung, P. Bryan Heidorn, Wengsheng Wu, Hong Zhang.
- University of New Mexico: Chris Frazier, Tomas Giermakowski, Teri Neville.
- Yale University: Reed Beaman, Ashley Green, Youjun Guo.

BioGeomancer logo by Ruby Blessing.

References

ADL (Alexandria Digital Library). 2001. *Alexandria Digital Library Gazetteer Server* <http://www.alexandria.ucsb.edu/gazetteer/> [Accessed 27 Jan. 2006].

Bannerman, B.S. 1999. *Positional Accuracy, Error and Uncertainty in Spatial Information*. Geoinnovations, Howard Springs, NT, Australia. <http://www.geoinnovations.com.au/posacc/patoc.htm> [Accessed 21 Jan. 2006].

Chapman, A.D. 2005a. *Principles of Data Quality*. Report for the Global Biodiversity Information Facility 2005. 61pp. Copenhagen: GBIF. <http://circa.gbif.net/Public/irc/gbif/pr/library?l=/webfiles/digit_documents/dataquality_pdf/_EN_1.0_&a=d> [Accessed 10 Jan. 2006].

Chapman, A.D. 2005b. *Principles and Methods of Data Cleaning*. Report for the Global Biodiversity Information Facility 2005. 75pp. Copenhagen: GBIF. <http://circa.gbif.net/Public/irc/gbif/pr/library?l=/webfiles/digit_documents/principlesmethods/_EN_1.0_&a=d> [Accessed 10 Jan. 2006].

Chapman, A.D. 2005c. *Uses of Primary Species-Occurrence Data*. Report for the Global Biodiversity Information Facility 2005. 111pp. Copenhagen: GBIF. <http://circa.gbif.net/Public/irc/gbif/pr/library?l=/webfiles/digit_documents/uses_primary_data/_EN_1.0_&a=d> [Accessed 10 Jan. 2006].

Chapman, A.D., Muñoz, M.E. de S. and Koch, I. 2005. Environmental Information: Placing Biodiversity Phenomena in an Ecological and Environmental Context, *Biodiversity Informatics* **2:** 24-41. <http://jbi.nhm.ku.edu/viewarticle.php?id=9&layout=abstract> [Accessed 21 Jan 2006].

Chrisman, N.R. 1991. The Error Component in Spatial Data. pp. 165-174 **in**: Maguire D.J., Goodchild M.F. and Rhind D.W. (eds) *Geographical Information Systems* Vol. 1, Principals: Longman Scientific and Technical.

Conn, B.J. (ed.) 1996. *HISPID3. Herbarium Information Standards and Protocols for Interchange of Data.* Version 3 (Draft 1.4). Sydney: Royal Botanic Gardens. http://www.bgbm.org/TDWG/acc/hispid30draft.doc [Accessed 27 Jan. 2006]

Conn, B.J. (ed.) 2000. *HISPID4. Herbarium Information Standards and Protocols for Interchange of Data.* Version 4 – Internet only version. Sydney: Royal Botanic Gardens. <http://plantnet.rbgsyd.nsw.gov.au/Hispid4/> [Accessed 27 Jan. 2006].

Cullen, A.C. & Frey, H.C. 1999. *Probabilistic Techniques in Exposure Assessment. A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. New York: Plenum Press. 335 pages.

DSDM. 2005. *MoSCoW Rules.* Dynamic Systems Development Method Consortium. <http://na.dsdm.org/en/about/moscow.asp> [Accessed 10 Jan. 2006].

Frazier, C., Neville, T., Giermakowski, T. and Racz, G. 2004. *The INRAM Protocol for Georeferencing Biological Museum Specimen Records*. Version 1.3. <http://www.inram.org/modules/UpDownload/store_folder/Documents/INRAM_Biodiversity_Georeferencing_Project/Georeferencing_Guidelines_INRAM-V1.3_2004-03-01.pdf> [Accessed 12 Mar. 2006].

Geoscience Australia. 2003. *Geodata TOPO-250K* (Series 1) Topographic Data [online]. Geoscience Australia, Canberra. <http://www.ga.gov.au/nmd/products/digidat/250k.htm> [Accessed 21 Jan. 2006].

Geoscience Australia. 2005. *NATMAP Raster Premium. 1:250 000 scale topographic maps of Australia.* On DVD. Canberra: Geoscience Australia.

Goulet, C.M. 2001. *Magnetic Declinations. Frequently Asked Questions.* Version 4.4. <http://www.geocities.com/magnetic_declination/> [Accessed 7 Jun 2006].

Gustafson, D.L. and Wefald, M. 2003. *TRS-data*. <http://www.esg.montana.edu/gl/trs-data> [Accessed 31 Jan. 2006].

Hall, B. 1994. *A Review Of The Environmental Resource Mapping System and A Proof That It Is Impossible To Write A General Algorithm For Analysing Interactions Between Organisms Distributed At Locations Described By A Locationally Linked Database and Physical Properties Recorded Within The Database.* Australian Digital Thesis Program: University of Western Sydney. <http://library.uws.edu.au/adt-NUWS/public/adt-NUWS20040506.151314/index.html> [Accessed 28 Jan. 2006].

Hijmans, R.J., Guarino, L., Bussink, C., Mathur, P., Cruz, M., Barrentes, I. and Rojas, E. 2005. *DIVA-GIS Version 5.2 A geographic information system for the analysis of biodiversity data*. <http://www.DIVA-GIS.org> [Accessed 27 Jan. 2006].

Juran, J.M. 1964. *Managerial Breakthrough*. New York: McGraw-Hill. 396 pp.

Juran, J.M. 1994. *Managerial Breakthrough*. Rev. edn. New York: McGraw-Hill. 451 pp.

McElroy, S., Robins, I., Jones, G. and Kinlyside, D. 1998. *Exploring GPS, A GPS Users Guide*: The Global Positioning System Consortium.

Morton, A. 2006. *UTM Grid Zones of the World*. Digital Mapping Software (DMAP). <http://www.dmap.co.uk/utmworld.htm> [Accessed 31 Jan. 2006].

Museum of Vertebrate Zoology. 2006. *MVZ Guide for Recording Localities in Field Notes.* <http://mvz.berkeley.edu/Locality_Field_Recording_Notebooks.html> [Accessed 24 Jan. 2006].

National Geophysical Data Center (NGDC). 2004. *Estimated Value of Declination*. <http://www.ngdc.noaa.gov/seg/geomag/jsp/struts/calcDeclination>. [Accessed 7 Jun. 2006].

National Mapping Council of Australia. 1975. *Standards of Map Accuracy*. First Edition February 1953. Second Edition December 1975.

NIMA, 2000. *Department of Defense World Geodetic System 1984. Its Definition and relationships with local geodetic systems.* TR8350.2, Third Edition, 3 Jan 2000.

NOAA (2001). *Nautical Charts and DGPS – A warning from NOAA.* http://www.boatwashington.org/nautical_charts_and_dgps.htm [Accessed 21 Jan. 2006].

Srivastava, B.K. and Ramalingam, K. 2006. Error Estimates for WGS-84 and Everest (India-1956) Transformation. *GIS Development* <http://www.gisdevelopment.net/technology/ip/ma03037abs.htm> [Accessed 24 Jan. 2006].

Taylor, C. 2003. *Geographic/UTM Coordinate Converter*. <http://home.hiwaay.net/~taylorc/toolbox/geography/geoutm.html> [Accessed 31 Jan. 2006].

TDWG. 2005). *Geospatial Element Definitions* v. 1.4. Taxonomic Databases Working Group – Extension to Darwin Core Version 2. http://darwincore.calacademy.org/Extensions/GeopatialExtension/GeospatialElementDefs [Accessed 13 Feb. 2006].

US Bureau of Budget. 1947. *United States National Map Accuracy Standards*. <http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/nmas/NMAS647.PDF> [Accessed 21 Jan. 2006].

US Navy. *n. dat.* Geographic Datums. PowerPoint from the *Navigator of the Navy* website. <https://www.navigator.navy.mil/navigator/geographic_datums.ppt> [Accessed 24 Jan. 2006].

Van Sickle, J. 1996. *GPS for Land Surveyors*: Ann Arbor Press, Inc: New York.

Wheeler, T.A., Huber, J.T. and Currie, D.C. (2001). *Label Data Standards for Terrestrial Arthropods.* Ottawa: Biological Survey of Canada (Terrestrial Arthropods). *Document Series No. 8* <http://www.biology.ualberta.ca/bsc/briefs/brlabelstandards.htm> [Accessed 24 April 2006].

Wieczorek, J. 2001. *MaNIS/HerpNet/ORNIS Georeferencing Guidelines*. University of California, Berkeley: Museum of Vertebrate Zoology. <http://manisnet.org/GeorefGuide.html> [Accessed 28 Jan. 2006].

Wieczorek, J. 2002. *Manual for Georeferencing Calculator*. MaNIS/HerpNet/ORNIS. University of California, Berkeley: Museum of Vertebrate Zoology. <http://manisnet.org/CoordCalcManual.html> [Accessed 28 Jan. 2006].

Wieczorek, J., Q. Guo, and R. Hijmans. 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*. 18: 745-767.

Further Reading

Beaman, R.S. and Conn, B.J. 2003. Automated geoparsing and georeferencing of Malesian collection locality data. *Telopea* 10(1): 43-52.

Beaman, R.S., Wieczorek, J. and Blum, S. 2004. Determing Space from Place for Natural History Collections. *D-Lib Magazine* 10: 5. <http://dlib.anu.edu.au/dlib/may04/beaman/05beaman.html> [Accessed 9 Jun 2006].

Chapman, A.D. 1999. Quality control and validation of point-sourced environmental resource data pp.409-418 **in** Lowell, K. and Jaton, A. (eds). *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources*. Chelsea, Michigan: Ann Arbor Press. 455pp

Chapman, A.D. 2002. Risk assessment and uncertainty in mapped and modelled distributions of threatened species in Australia pp.31-40 **in** Hunter, G. & Lowell, K. (eds) *Accuracy 2002 – Proceedings of the 5th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. Melbourne: Melbourne University 547 pp.

FGDC. 2005a. *Proposal for Geospatial Positioning Accuracy Standards, Part 3: National Standard for Spatial Data Accuracy*. Washington, DC: Federal Geographic Data Committee. <http://www.fgdc.gov/standards/projects/FGDC-standards-projects/accuracy/part3/progpas3> [Accessed 5 May 2006].

FGDC. 2005b. *Proposal for Geospatial Positioning Accuracy Standards*. Washington, DC: Federal Geographic Data Committee. <http://www.fgdc.gov/standards/projects/FGDC-standards-projects/accuracy/part1/progpas1> [Accessed 5 May 2006].

Foote, K.E. and Huebner, D.J. 2000. *The Geographer's Craft Project*. Department of Geography, The University of Colorado at Boulder. http://www.colorado.edu/geography/gcraft/notes/manerror/manerror_f.html [Accessed 17 Jul. 2006].

Moritz, T. 1999. *Geo-referencing the Natural and Cultural World, Past and Present: Towards Building a Distributed, Peer-Reviewed Gazetteer System*. Presented at the Digital Gazetteer Information Exchange Workshop, Oct 13-14, 1999. Transcribed and edited from audiotape. http://www.alexandria.ucsb.edu/~lhill/dgie/DGIE_website/session1/moritz.htm [Accessed 21 Jan. 2006].

Stein, B. and Wieczorek, J. 2004. Mammals of the World: MANIS as an example of data integration in a distributed network environment. *Journal of Biodiversity Informatics* 1: 14–22.

USGS. 2005. *National Mapping program Standards*. Washington: USGS <http://nationalmap.gov/gio/standards/> [Accessed 5 May 2006].

Software and on-line Tools

- ***BioGeomancer***
  http://www.biogeomancer.org/
- ***BioGeoMancer Classic***
  http://biogeomancer.org/
- ***DIVA-GIS***
  http://www.DIVA-GIS.org/
- ***GeoCalc***
  http://www.geocomp.com.au/geocalc/
- ***GeoLoc – CRIA***
  http://splink.cria.org.br/geoloc?&setlang=en
- ***GEOLocate***
  http://www.museum.tulane.edu/geolocate/
- ***MaNIS Georeferencing Calculator***
  http://manisnet.org/gc.html
- ***NGDC Magnetic Declination Calculator***
  http://www.ngdc.noaa.gov/seg/geomag/jsp/struts/calcDeclination

Gazetteer Look-up Services

- ***Alexander Digital LibraryGazetteer Server Client***
  http://middleware.alexandria.ucsb.edu/client/gaz/adl/index.jsp
- ***Fuzzyg – Fuzzy Gazetteer***
  http://tomcat-dmaweb1.jrc.it/fuzzyg/query/
- ***Global Gazetteer***
  http://www.fallingrain.com/world/
- ***GEOnet Names Server***
  http://gnswww.nga.mil/geonames/GNS/index.jsp

Appendix: Guidelines for Georeferencing Locality Types

FEATURE (NAMED PLACE)

Definition:

The simplest locality descriptions consist of only a named place, which is often a feature listed in a standard gazetteer and can probably be located on a map of the appropriate scale. Feature categories include:

- town, suburb, populated place, or homestead
- spring, bore, tank, well, or waterhole
- island, reef, or cay
- port, bay, gulf, or harbor
- airport, buoy, dock, or jetty
- point, cape, or peninsula
- cave
- dam, or lock
- hill, peak, pass, or mountain
- trig point
- park, reserve, or forestry zone
- junction of two paths (roads, rivers, contour lines, boundaries, etc.)

Despite how they might be presented in a gazetteer or on a map, features are not points; they are areas that have a spatial extent, though the extent may not always be obvious. In a very few cases (such as a trig point), the extent is very small as it is an accurately surveyed point. The important thing is to try to capture the information not only for where the feature is, but also how specific it is (i.e., how big is the extent of the feature).

Some features (e.g., river and road junctions, bridges) may not have gazetteer entries, while others (e.g., properties) may not appear on standard map series. These types of features can be a challenge to locate, and are therefore among the least efficient georeferences to produce. Nevertheless, additional resources, such as internet searches and field notes can often reveal these tricky places

Examples:

Example 1: "Bakersfield”
Example 2: “Point Lookout”
Example 3: “Bennetts Waterhole”
Example 4: “Isla Tiburon”
Example 5: “Lorne Reef”
Example 5: “Yosemite National Park”
Example 6: “Mt Hypipamee”
Example 7: “Junction of Dwight Avenue and Derby Street”
Example 8: “State Forest Reserve 607”
Example 9: “Where Dalby Road crosses Bunya Mountains National Park Boundary”
Example 10: “confluence of Labarge Creek and South Labarge Creek”
Example 11: “At 100 m contour line on Black street”
Example 12: “junction of Rio Claro and Rio La Hondura”
Example 13: “Victoria River Station” [Northern Territory, Australia]

Georeferencing Procedure:

<u>Features with an obvious spatial extent</u> — use the geographic center (i.e., the midpoint of the extremes of latitude and longitude) for the coordinates. If the geographic center does not fall inside the shape of the shaded area, then pick the nearest point to the center that lies within the shape (see Figure 4). Use the distance

from the coordinates to the furthest point within the named place as the extent. Some gazetteers give bounding boxes to describe the extents of large places and you can use these to determine the extent by measuring them from a map or by using a geographic distance calculator such as the Perpendicular Distance Calculator[16] from the Center for Biodiversity and Conservation (CBC).

Features without an obvious spatial extent — some features do not have a shaded boundary or a topographic symbol for buildings shown on the map (especially for non-USA locales). Some of these features may have large, but indistinct extents (mountains, trap lines). Other features may be relatively small (springs, junctions), with no apparent extent on a map. Use and document your judgment when placing the coordinates and estimating the extent of large features, and use a standard extent for small features based on the feature type. The extent of road junctions, for example, cannot be measured on maps, so use the following extent recommendations from Frazier *et al*. (2004):

- For 2-lane city streets and 2-lane highways, the extent is 10 m.
- For 4-lane highways, the extent is 20 m.
- For large highways with medians, the extent is 30 m.
- If unknown, use 15 m.

It is worthwhile to create a feature type extent table as part of your institutional best practices document so that there is consistency in extents for features whose size cannot be measured without ground-truthing.

'Exact' locations — if the locality appears to be 'exactly at' the locality cited (GPS reading) use the accuracy of the GPS as the extent.

In some cases – for example, an accurately recorded trig point – the extent and the uncertainty may be identical, however, collections are seldom made at the exact locality cited (e.g., right on top of the trig point), so the extent is usually much larger than a literal reading might suggest.

If you choose to use a gazetteer to obtain coordinates, keep in mind that they may not be at the geographic center of the feature. For example, the coordinates of a populated place may be at the main post office or the courthouse (if that place is a county seat). Coordinates for rivers and streams are usually at the mouth. For this reason, it is a good idea to use the gazetteer coordinates to find the feature on a map, and then use the map to find the geographic center of the feature.

When recording the method of determination of the coordinates and uncertainty in the remarks, use "measured from the main post office" or "measured from the geographic center of Bakersfield", etc.

Care! Some older gazetteers reference the bottom left hand corner of the position where the name is to appear on a printed map rather than the actual location of the feature. Most gazettteers have been fixed in recent years, but care should be taken when using an unfamiliar gazetteer. Always check the map, which you will need to do in any case to calculate the extent.

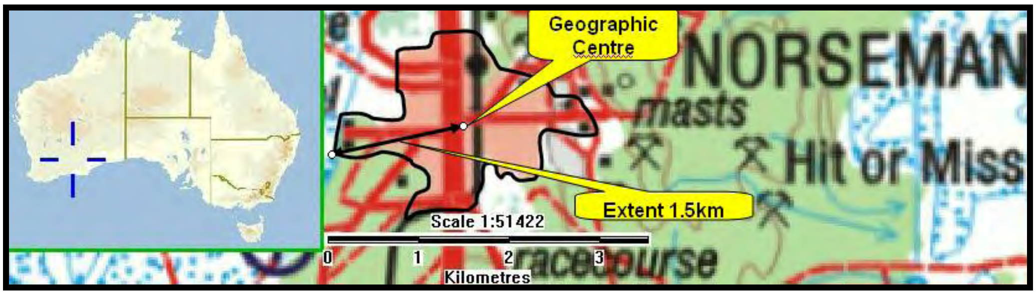


Fig. 4. Calculation of the geographic center and extent for Norseman, Western Australia. Background map from Geosciences Australia (2005).

---

[16] Perdendicular Distance Calculator <http://geospatial.amnh.org/open_source/pdc/index.html>.

Subdivisions of a feature — such as “N part of Mono Lake” calculate the extent based only on the subdivision and proceed as you would with a location consisting of a named place with a spatial extent.

Properties (ranches, farms, stations, etc.) — if you are unable to locate them in gazetteers or on regular maps, you may have to use a cadastral map, or carry out a search to see if you can locate them in relation to nearby cities or other geographic entities. If you are unable to find the boundaries, and thus determine the geographic center, then use the coordinates of the homestead or major property buildings and estimate the size of the property from the location of other buildings not on that property.

Caves — use the coordinates of the entry to the cave. This will usually be the location given in a gazetteer or on a regular map.

Uncertainty:

Use the ***MaNIS Georeferencing Calculator*** (http://manisnet.org/gc.html) to determine “Maximum Uncertainty Distance”.

- For **Calculation Type** use:
  “Error – enter Lat/Long for the actual locality”
- For **Locality Type** use:
  “Named place only”.

See Example 1, below

Example 1.

Locality: “Bakersfield”
Suppose the coordinates for Bakersfield came from the GNIS database (a gazetteer) and the distance from the center of Bakersfield to the furthest city limit is 3 km.

Coordinate System: degrees minutes seconds
Latitude: 35º 22' 24" N
Longitude: 119 º 1' 4" W
Datum: not recorded; 79 m uncertainty
Coordinate Precision: nearest second; 40 m uncertainty
Coordinate Source: gazetteer
Extent of Named Place: 3 km
Distance Units: km
Decimal Latitude: 35.37333
Decimal Longitude: -119.01778
Maximum Uncertainty Distance: 3.119 km

NEAR A FEATURE

Definition:

A locality given without an exact position, but with “near”, “in the vicinity of”, “adjacent to”, or some similar relation to a feature cited.

These locality descriptions imply an offset from a named place without definitive directions or distances.

Examples:

Example 1: “Near Las Vegas“
Example 2: “vicinity of Tumbarumba“
Example 3: “Big Bay vicinity”
Example 4: “near MS 117 on Dalton Hwy”
Example 5: “near Bend 43 on Great Western Hwy”
Example 6: “vicinity of bridge over Condamine River on Warrego Highway“
Example 7: “adjacent to railway underpass on Smith Street”
Example 8: “area of confluence of Black and Oshetna Rivers”

Georeferencing Procedure:

In these cases use the geographic center of the named place for the geographic coordinates.

If you are unable to determine the exact coordinates of the locality, then use the coordinates as near as possible to the referenced locality (and on the path if appropriate).

Extent:

The extent should be calculated as the greater of 2km or 200% of the extent of the named place. Clearly there is a measure of subjectivity involved here and you should use your judgement and evidence from other sources. Let common sense prevail and document the assumptions made.

Uncertainty:

Calculate the same as for ‘***Feature***’ but note the increase in extent.

BETWEEN TWO FEATURES

Definition:

A locality cited as ‘between’ two features or named places.

Examples:

Example 1: “between Point Reyes and Inverness“

Image

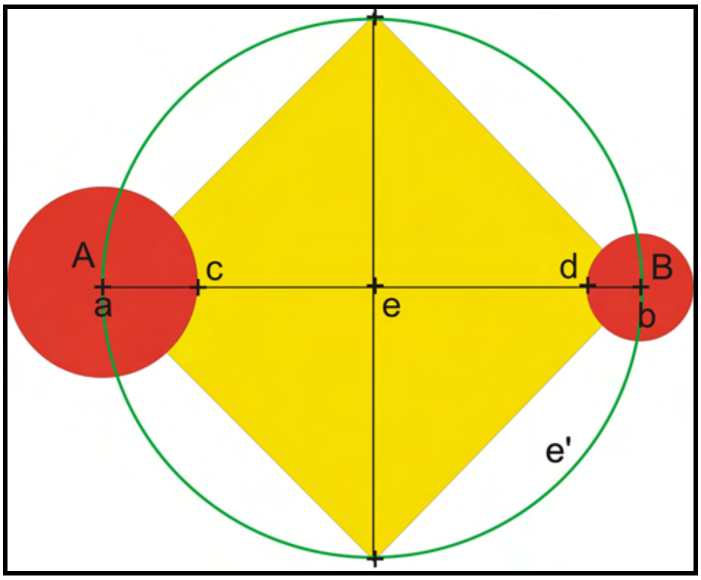


Fig. 5. The diagram above illustrates the general case of a locality description of the type "Between A and B.

Georeferencing Procedure:

Find the coordinates of the midpoint between the centers of the two named places (the point **e** in Figure 5).

Extent:

Use one-half the distance between the centers of A and B.

Uncertainty:

Calculate the same as for ‘***Feature***’.

STREET ADDRESS

Definition:

Locality is a street address – usually with a number, a street name, and a feature name.

In some places, street numbers in rural areas represent a metric distance from the start of the road.

Examples:

Example 1: “1 Orchard Lane, Berkeley, CA“
Example 2: “21054 Baldersleigh Road, Guyra, NSW“ (indicates that the locality is 21.054 km from the beginning of Baldersleigh Road).
Example 3: “Backyard of 593 West Street, Louisville, Boulder County, Colorado”
Example 4: “Greenhouse at 20th and Broadway, Boulder, Boulder County, Colorado”

Georeferencing Procedure:

Addresses are sometimes given when specimens are collected in cities or towns. When possible, plot the point at the indicated spot with the aid of a local road map or a mapping product such as Google Maps®, Maporama® or Mapquest®, White or Yellow Pages directory, or a GPS. If the exact address cannot be found, estimate the location as well as possible. Remember that many addresses reflect a grid system of labeling addresses. For instance, addresses between 12th Street and 13th Street would lie between 1200 and 1300. Be aware, however, that street names often change over time. To your best ability, locate this area on the map or map software you are using to get the coordinates, if the electronic gazetteer does not display them automatically.

Building names are often given to clarify the location within a town or city. Rarely are these buildings given coordinates in a gazetteer; however, they can sometimes be located using a Yellow Pages directory, which may be available via the Internet. Unlike natural features, most buildings change names or even disappear over time, so verify that the building named in the record existed in that location at that time.

Extent:

Use as the extent the smallest area that is identifiable and that cannot be any other address. If you cannot determine the location and size of an address within a block, use half the length of the city block for the extent and make note of this in the georeferencing remarks.

Uncertainty:

Calculate the same as for ‘***Feature***’.

PATH

Definition:

The locality is a linear feature such as a road, trail, boundary, river, or contour line. The locality may also refer to part (or subdivision) of the path (see Examples 5-7). Localities that are given without an exact position, but are cited as "near", "in the vicinity of", "adjacent to", a path such as a road, river, etc. (see Examples 8 and 9) are treated in the same manner as any other path, but with perhaps a wider footprint – see also under '***Near a Feature***', above.

Note! A path clause in a locality description is often meant to be read in combination with another clause. The relationship between the path clause and other clauses in the same description is important, because the resulting shape will be affected. For example, a description with a path followed by an offset from a feature at a heading ("Hwy 101, 2 mi N Santa Rosa) should actually be calculated as a clause of the type "offset from a feature at a heading along a path" rather than as the intersection of a path and a clause of the type "offset from a feature at a heading"

Examples:

Example 1: "Hwy 1"
Example 2: "Nepean River"
Example 3: "along 100 m contour line"
Example 4: "N. Boulder Creek, 1.3 miles above Boulder Falls"
Example 5: "mouth of Goodpaster River"
Example 6: "head of Mooney Creek"
Example 7: "Eastern part of Logan Motorway"
Example 8: "vicinity of Uyamitquaq Ck."
Example 9: "adjacent to eastern boundary of Foz do Iguaçu Park"

Georeferencing Procedure:

Roads — a path may be defined with reference to a named place (see Note above). This may influence where one places the coordinates. If there is no further refinement, then treat the road similarly to a river, as explained below. If there is reference to an offset or a position on the path, then treat the location as any other feature and refer to the appropriate sections, such as '***Offset***', or '***Feature***'.

Rivers — if you are unable to traverse the length of the river to find the geographic center, then make a straight line from the mouth of the river to the head of the river (or the extreme points within the county, state, etc. you are concerned with). Find the center of this line, and place your coordinate point closest to the center of the line on the river itself (see Figure 6). This method may lead to large errors in rivers that have large changes in direction. Use your common sense to determine the most appropriate point, bearing in mind the suggested methods above.

The mouth of a river is not always easy to determine, but is usually taken to be formed by a straight line across the river at the position where the river joins a larger body of water (sea, bay, lake, another river, etc.). In some rare cases, it may refer to the downstream end where the river changes its name. It is the position of lowest elevation of the river.

Similarly, the head of a river (where the river begins) can also be difficult to determine. Though the head is always the highest point on the part of the river bearing the same name, it may begin in a mountain, canyon, or lake, and may need to be estimated because it has become too fine or broken up into smaller streams to accurately identify on a map.

Sometimes the terms 'above', 'below', 'left bank', or 'right bank' are used with rivers instead of cardinal directions (see Example 4, above). 'Above' is used when referring to upstream of the feature while 'below' refers to downstream. The direction a river flows can be easily determined on a topographic map by looking at the contour lines and elevation. The contour lines will always point upstream as they cross the river. The terms left and right bank refer to the side of the river when facing downstream.

'Mouth of River' (Example 5) and 'head of River' (Example 6) are usually best treated as you would a '***Feature***'.

Care! When using older maps, be aware that rivers may have changed course and may have been in a different location at the time the collection was made, compared with the position drawn on the map at a different time. In addition, the apparent position of the mouth of a river can be strongly influenced by the scale of the map being used.

Note! Do not use the coordinates given by gazetteers, as these points usually correspond to the mouth of the river, not the geographic center.

Contour Lines — If the contour line has ends within the area of interest, treat it the same as you would a river. If the contour line is closed (i.e., forms a polygon around a hill or mountain, etc.), then treat the enclosed area the same as you would a '***Feature***' and use the geographic center of the polygon for the geographic coordinates.

Subdivisions of a Path — where a subdivision of a path still describes a path, continue to treat it as a road, river or contour line as above. In Example 7, for instance, you may take the midway point on the Logan Motorway as the western limit of the subdivision meant by the 'eastern part'. Use that limit as the basis to determine both the coordinates and the extent.

Image:

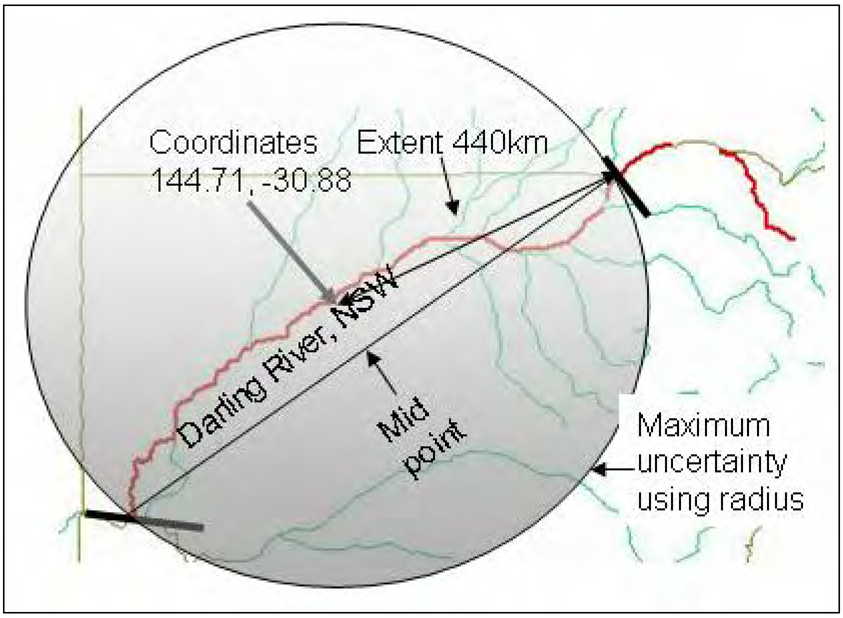


Fig. 6. An example of determining coordinates and extent for a path (in this case the Darling River in New South Wales, Australia). Use of an uncertainty polygon would give a more accurate representation for maximum uncertainty than the point-radius method.

Extent:

The extent is the distance from the point nearest the geographic center of the path to the point on the path furthest from that center point. Make sure to base the center point on only that portion of the path within the boundaries of interest.

Uncertainty:

Calculate the same as for '***Feature***'.

BETWEEN TWO PATHS

Definition:

A locality cited as being between two paths (two roads, two rivers, a road and a river, etc.).

Examples:

Example 1: “between Tanama R. and Clearwater Ck.”
Example 2: “between Aldersley and Bridge Streets” (i.e., two streets that don’t intersect)
Example 3: “on Hwy 14, between highway and adjacent fence”

Georeferencing Procedure:

Create a polygon from the two paths and the end points of each of the paths – for example, the state boundary, where the river joins another river or changes names, a road intersection, etc. (see Figure 7.)

Once the polygon is drawn – then the coordinates are determined in the same manner as for a ‘***Feature***’, above.

Image:

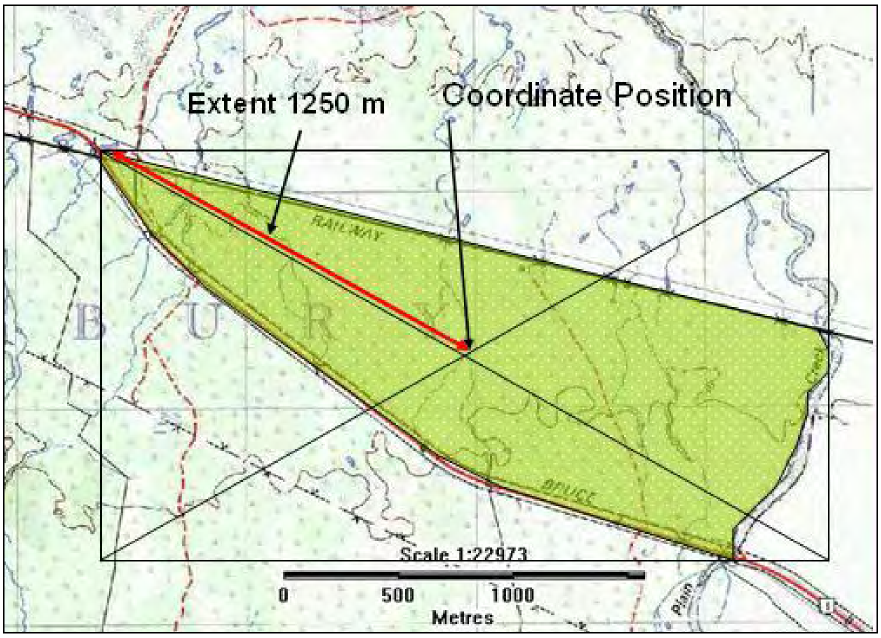


Fig. 7. An example of determining coordinates and extent for a location between two paths (in this case “between the Bruce Highway and the Railway line, West of Plain Creek and before the Railway crossing on the Highway”). Background map from Geosciences Australia (2005).

Extent:

Once the polygon has been drawn as above, then the extent is determined in the same way as for a ‘***Feature***’.

Uncertainty:

Calculate the same as for ‘***Feature***’.

OFFSET DISTANCE

Definition:

Locality consists of an offset from a named place without any direction specified.

Offsets without a direction are often the result of errors by the collector when recording the locality. Occasionally, these localities are data entry errors. Try to view the original collection catalogs or labels, as there may be more information in them.

Examples:

Example 1: “5 km outside Calgary“
Example 2: “15 km from Recife”

Image:

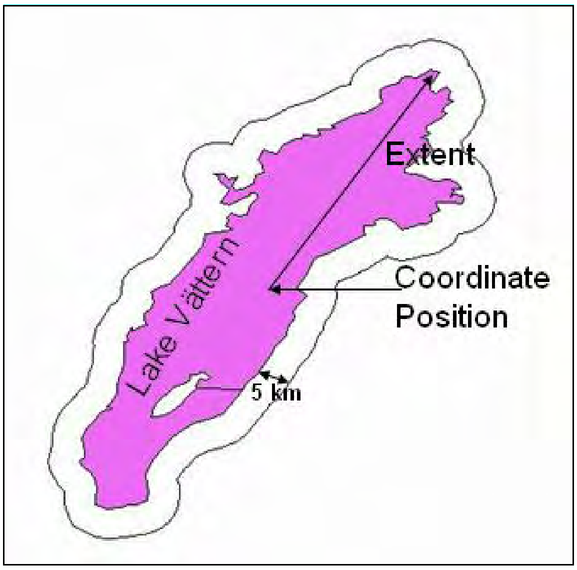


Fig. 8. An example of determining coordinates and extent for a location with offset distance only (in this case “5 km from Lake Vättern, Sweden”). The coordinates are 14.56º, 58.30º. Offset distance is 5 km, extent of the named place (Lake Vättern) is 61.2 km. These values are then used in the MaNIS Georeferencing Calculator to determine maximum uncertainty.

Georeferencing Procedure:

Record the geographic coordinates of the center of the named place, as you would for a normal ‘***Feature***’.

Sometimes offset information is vague either in its direction or in its distance. If the direction information is vague, record the geographic coordinates of the center of the named place and include the offset distance in the determination of the maximum uncertainty (see figure 8).

Extent:

Use the extent of the named place.

Uncertainty:

Use the ***MaNIS Georeferencing Calculator*** (http://manisnet.org/gc.html) to determine “Maximum Uncertainty Distance”.

- For **Calculation Type** use
  “Error – enter Lat/Long for the actual locality”

- For **Locality Type** use
  "Distance only (e.g., 5 mi from Bakersfield)".

Example 1.

Locality: "5 mi from Bakersfield"

Suppose the coordinates for Bakersfield came from Topozone® with the map coordinates reprojected in NAD27. Suppose also that the distance from the center of Bakersfield to the furthest city limit is 2 mi.

Coordinate System: decimal degrees
Latitude: 35.373
Longitude: -119.018
Datum: NAD27; no uncertainty
Coordinate Precision: 0.001 degrees; 0.089 mi uncertainty
Coordinate Source: gazetteer
Offset Distance: 5 mi
Extent of Named Place: 2 mi
Distance Units: mi
Decimal Latitude: 35.373
Decimal Longitude: -119.018
Maximum Uncertainty Distance: 8.089 mi

Example 2.

Locality: "5 km from Lake Vättern, Sweden" (see Figure 8).

Coordinate System: decimal degrees
Latitude: 58.30
Longitude: 14.56
Datum: unknown
Coordinate Precision: 0.001 degrees, 1520 m uncertainty
Coordinate Source: gazetteer
Offset Distance: 5 km
Extent of Named Place: 61.2 km
Distance Units: km
Decimal Latitude: 58.30
Decimal Longitude: 14.56
Maximum Uncertainty Distance: 68.559 km

NB! See discussion on "Estimating Uncertainty from Distance" earlier in this document. The Georeferencing Calculator uses the method for estimating uncertainty given in Wieczorek *et al.* (2004).

OFFSET DIRECTION

Definition:

Locality consists of a direction from a named place without any distance specified.

Examples:

Example 1: “N Palmetto“
Example 2: “W of Jondaryan"

Image:

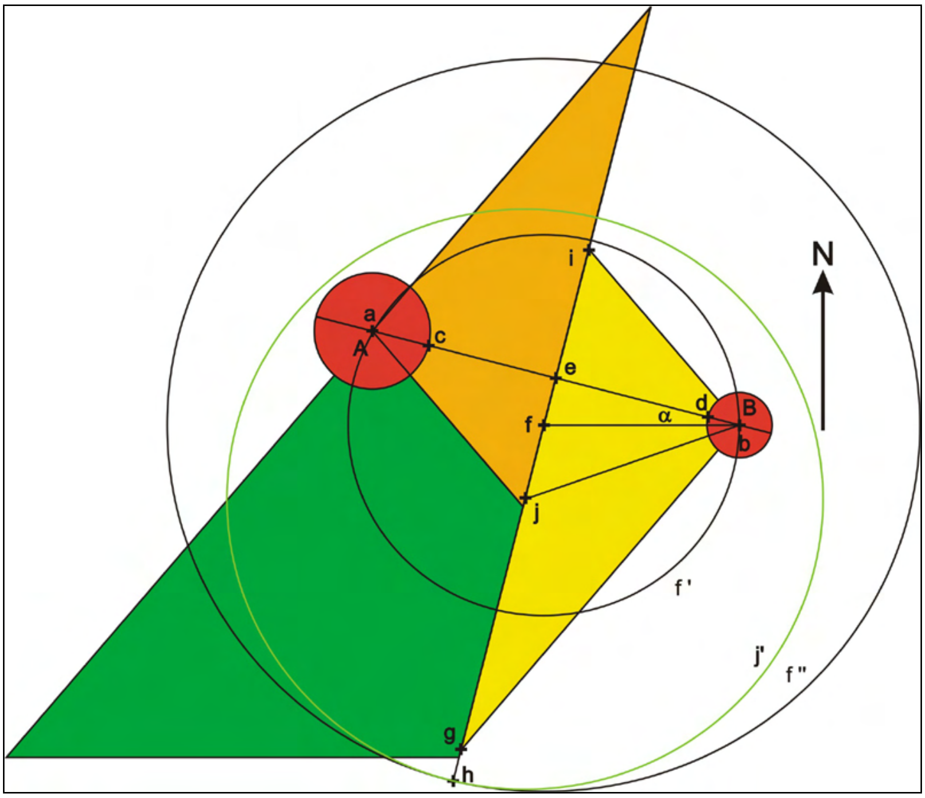


Fig. 9. An example of a locality description of the type "At a heading from B". In this diagram the specific example is "West of B". The area corresponding to "West of B" is encompassed by the bright yellow triangle connecting the three points b, i, and g. The orange triangle would be interpreted as "East of A" and the green triangle would be interpreted as "South of A”.

There are a number of ways one might calculate the coordinates, extent, and uncertainty for this complicated scenario, some of which are described below using the values shown in Figure 9.

Alternative 1:

Coordinates: Place the coordinates at the point **f**, which is at a distance r/cos(alpha) W from the center of B, where r is one half the distance between the centers of A and B and alpha is the angle between west and the direction from the center of B to the center of A.

Extent: The radius of **f'**, which is r/cos(alpha).

Disadvantage: This alternative leaves out some of the triangle (**big**).

Advantages:

1. The center of the uncertainty radius is the point due W furthest from the center of B within the triangle (**big**).
2. This is the simplest of the three alternatives to calculate.

Alternative 2:

Coordinates: Place the coordinates at the point **f**, as in Alternative 1.

Extent: The radius of **f"**, which is the extent of B plus r*sqrt(2)/(2*cos(alpha)*sin(theta-alpha)), where the angle theta is based on the direction uncertainty (45 degrees for West).

Disadvantage: The radius of uncertainty is larger than it needs to be to cover the area that might reasonably be called ‘West of B’.

Advantage:

1. This alternative leaves out none of the triangle (**big**).
2. It has its center on the point furthest due W of the center of B within the triangle (**big**).

Alternative 3:

Coordinates: Place the coordinates at the point **j**, which is half-way between the points **i** and **g**. The coordinates for this point is beyond the ability of a georeferencer to discern.

Extent: The radius of **j'**, which is the extent of B plus r*(tan^2(theta-alpha)+1)/(2*tan(theta-alpha)), where theta is the direction uncertainty (45 degrees for West).

Disadvantage:

1. This alternative leaves out none of the triangle (**big**).
2. The coordinates of the point **j** cannot be determined readily from a map; they have to be calculated.
3. The uncertainty for this alternative is the most complex to calculate.
4. The center of the uncertainty is not due W of the center of B.

Advantage:

1. This alternative leaves out none of the triangle (**big**).
2. The size of the uncertainty radius is as small as it can be and still encompass the whole triangle (**big**).

Alternative 4:

NoGeorefBecause: "no offset distance given".

Disadvantage:

1. No georeference is produced.

Advantage:

1. This alternative avoids all of the subjectivity required to interpret this vague description.

Georeferencing Procedure:

When only an offset is given with no distance, it is virtually impossible to georeference with certainty without additional information. For example, if we have a location 'East of Albuquerque' with no other information, there is no clear indication of how far one should go 'East' to find the location – to the next nearest named place; the next nearest named place of equivalent size, or just keep going? In reality, such a description could describe half of the Earth's surface. It is for this reason that we recommend using Alternative 4 above.

Seldom is such information given alone; there is usually some supporting information. For example, the locality may have higher-level geographic information such as 'East of Albuquerque, Bernalillo County, New Mexico'. This gives you a stopping point (the county), and should allow you to georeference the locality.

Uncertainty:

The ***MaNIS Georeferencing Calculator*** (http://manisnet.org/gc.html) does not explicitly calculate coordinates and uncertainty for this locality type. Nevertheless, the uncertainty can be calculated for any of the first three alternatives given above if one first determines the coordinates and extent. The Georeferencing Calculator can be used in two steps to georeference using Alternative 1, above. For the first step, determine the coordinates for the point **f**:

- For **Calculation Type** use
  "Coordinates and error"
- For **Locality Type** use
  "Distance at a heading".

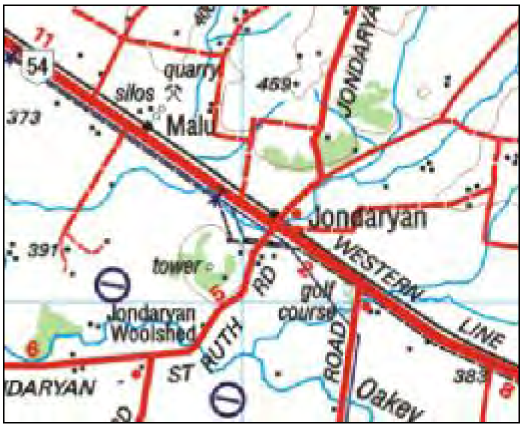


Fig. 10. Extract from TOPO250K digital map showing Jondaryan, Queensland, Australia. Map from Geosciences Australia (2005).

Example 1. (see Figure 10)

Locality: “W of Jondaryan”

Suppose the coordinates for Jondaryan came from a gazetteer using the Australian Geodetic Datum 1984 (AGD84). Malu is the next populated place in a westerly direction from Jondaryan at a distance of 3.65 and a heading of 305º. The scale of the map is 1:250,000 and the metadata for the map indicates an uncertainty of ~160 m (see Table 5).

Coordinate System: degrees, minutes, seconds
Latitude: 27º 21’ 50” S (for Jondaryan)
Longitude: 151º 34’ 59” E (for Jondaryan)
Datum: AGD 84; no uncertainty
Coordinate Precision: 1 second
Offset distance: 4.46 km (r/cos(alpha) where r = 3.65 km and alpha is the difference between the heading 305º and west 270º, or 35º).
Direction: W
Decimal Latitude: -27.36389 (for point **f**)
Decimal Longitude: 151.53797 (for point **f**)
Maximum Uncertainty Distance: 1.592 km

For the second step, determine the maximum uncertainty distance from the point **f:**

- For **Calculation Type** use
  “Error – enter Lat/Long for the actual locality”
- For **Locality Type** use
  “Named place only”.

Example 1. Step 2.

Coordinate System: degrees, minutes, seconds
Latitude: 27º 21’ 50” S
Longitude: 151º 32’ 16.69” E (based on the Decimal Longitude from Step 1)
Datum: AGD 84; no uncertainty
Coordinate Precision: 1 second
Coordinate Source: non-USGS map: 1:250,000; 0.16 km uncertainty.
Extent of Named Place: 4.46 km (radius of **f'** in Figure 9)
Distance Units: km
Distance Precision: 1/100 km
Decimal Latitude: -27.36389
Decimal Longitude: 151.53797
Maximum Uncertainty Distance: 4.751 km

OFFSET AT A HEADING

Definition:

The locality contains a distance in a given direction from a feature or named place. There are several variations on such localities.

Localities that have one linear offset measurement from a named place, but do not specify how that measurement was taken (see Example 1, below), are open for case-by-case judgment. The judgment itself must be documented in the remarks for the determination (e.g., ‘Assumed “by air” – no roads E out of Yuma’, or ‘Assumed “by road” on Hwy. 80’). In this case, the remark should be something like ‘Uncertainty encompasses both distance by air and distance by road on Hwy. 80’).

In Example 2, the locality is on the east side of the river, in Illinois, rather than on the west side, in Missouri. In this example, the 16 miles were assumed to be ‘by air’ – but see similar example under in the next Locality Type: ***Offset along a Path***.

The addition of an adverbial modifier to the distance part of a locality description, while an honest observation, should not affect the determination of the geographic coordinates or the maximum uncertainty. In Example 3, below, treat the locality as if it read "25 km WNW of Campinas"

Examples:

Example 1: “10.2 mi E of Yuma”
Example 2: “16 mi from St Louis on left bank of the Mississippi River – downstream“
Example 3: “about 25 km WNW of Campinas”
Example 4: “10 mi E (by air) Bakersfield“

Georeferencing Procedure:

Use the geographic coordinates of the named place (see ‘***Feature***’, above) as a starting point. Sometimes the locality description gives a method for determining the offset (e.g., ‘by road’, ‘by river’, ‘by air’, ‘up the valley’, etc.) For all cases except ‘by air’ (see Example 4), use the next Locality Type: ***‘Offset along a Path’***, below.

Where the method of determining the offset cannot be determined from the locality description or additional information and there is no obvious major path that can be followed in the rough direction and distance given, assume the collector measured the distance by air.

If there is no clear best choice between ‘by road’ or ‘by air’, you may wish to use the midpoint between the two possibilities as the geographic coordinate and assign an uncertainty large enough to encompass the coordinates and uncertainties of both methods.This choice is not recommended here for two reasons. First, the resulting coordinates will not match either of the two possible interpretations. Second, it will take about three times as long to calculate since the two interpretations have to be made, followed by the determination that encompasses them both. Since the offset at a heading “by air” will usually encompass the alternative by road anyway, this is the recommended option. You can increase the maximum uncertainty to encompass the other possible choice. Once again, this recommendation applies if you don’t have a compelling reason to use the offset along a path.

To calculate the coordinates, use the geographic coordinates of the center of the named place as a starting point (in the Example 1 above, use the center of Yuma) and enter its coordinates and extent in the ***MaNIS Georeferencing Calculator*** using the Calculation type: ‘***Coordinates and Error’***. Enter the distance and direction given – (make sure relevant parameters are filled in or selected, such as datum, direction, offset distance, distance units and precision, and coordinate source, system, and precision,) and push “Calculate.” The new coordinates that appear at the bottom of the calculator are the ones you can now enter in your database. They should be different from the coordinates you entered in the ‘Latitude’ and ‘Longitude’ spaces – if they are not, check to make sure you have chosen the correct Calculation Type. You should also check the resulting locality coordinates on a map (or for the USA, in Topozone.com) to make sure they make sense. Be sure to choose the same datum as the original coordinates when viewing the result.

Extent:

The extent is the extent of your starting point - usually a named place such as a city or a junction.

Uncertainty:

Use the ***MaNIS Georeferencing Calculator*** (http://manisnet.org/gc.html) to determine "Maximum Uncertainty Distance".

- For **Calculation Type** use
  "Coordinates and error"
- For **Locality Type** use
  "Distance at a Heading".

Example 1.

Locality: "10 mi E (by air) Bakersfield"
Suppose the coordinates for Bakersfield came from the GNIS database (a gazetteer), the coordinates of the locality were calculated to the nearest second, and the distance from the center of Bakersfield to the furthest city limit is 2 mi.

Coordinate System: degrees, minutes, seconds
Latitude: 35º 22' 24" N
Longitude: 118º 50' 56" W
Datum: not recorded; 0.049 mi uncertainty
Coordinate Precision: nearest second; 0.024 mi uncertainty
Coordinate Source: gazetteer (for which the datum is unknown)
Offset Distance: 10 mi
Extent of Named Place: 2 mi
Distance Units: mi
Distance Precision: 1 mi
Direction Precision: E (45 degrees precision – between NE and SE)
Decimal Latitude: 35.37333
Decimal Longitude: -118.67179
Maximum Uncertainty Distance: 16.588 mi

Example 2.

Locality: "10 mi ENE (by air) Bakersfield"
Suppose the coordinates for the locality were interpolated to the nearest second from the USGS Gosford 1:24,000 Quad map and the distance from the center of Bakersfield to the furthest city limit is 2 mi.

Coordinate System: degrees, minutes, seconds
Latitude: 35º 24' 21" N
Longitude: 118º 51' 25" W
Datum: NAD27; no uncertainty
Coordinate Precision: nearest second; 0.024 mi uncertainty
Coordinate Source: USGS map: 1:24,000; 0.008 mi uncertainty
Offset Distance: 10 mi
Extent of Named Place: 2 mi
Distance Units: mi
Distance Precision: 10 mi
Direction Precision: ENE (11.25 degrees either side of ENE)
Decimal Latitude: 35.46134
Decimal Longitude: -118.69326
Maximum Uncertainty Distance: 12.379 mi

OFFSET ALONG A PATH

Definition:

Locality describes a route from a named place.

If the distance was along a linear feature such as a road or river, measure along the feature for the distance and in the direction cited, rather than use a straight line. There is no uncertainty due to direction imprecision.

Examples:

Example 1: "7.9 mi N Beatty, on US 95”
Example 2: "13 mi E (by road) from Bakersfield”
Example 3: "18 km W of Guyra, on Baldersleigh Road"
Example 4: "2km downstream from Wallaman Falls”
Example 5: "3 km above Anita Grande on Rio Jimenez”
Example 6: "left bank of the Mississippi River, 16 mi downstream from St. Louis"

Georeferencing Procedure:

If ‘by road’ is specified in the locality description, or if there is an obvious major road that can be followed that complies to the direction and distance exactly, you can assume that the collector traveled by road. If there is a choice between multiple roads that fit the description, choose one of them as the basis for the georeference and increase the error to encompass the other possible choices.

Use the center of the starting point (in the Example 1, above, use the center of Beatty), and use the measuring tool found in Terrain Navigator®[17] (USA only), or your own appropriate application, to follow the road until you have gone the distance cited. Take the coordinates from this ending point. Be sure to make a note of the name of the road you followed in the Remarks if it is not already in the locality description.

Extent:

The extent is the extent of your starting point - usually a named place such as a city or a junction.

Uncertainty:

Use the ***MaNIS Georeferencing Calculator*** (http://manisnet.org/gc.html) to determine “Maximum Uncertainty Distance”.

- For **Calculation Type:** use
  “Error – enter Lat/Long for the actual locality”
- For **Locality Type:** use
  “Distance along a Path”.

---

[17] Terrain Navigator®, <http://www.maptech.com/land/index.cfm>.

Example 1.

Locality: “13 mi E (by road) Bakersfield”

Suppose the coordinates for this locality were interpolated to the nearest 1/10th minute from the USGS Taft 1:100,000 Quad map and the distance from the center of Bakersfield to the furthest city limit is 2 mi.

Coordinate System: degrees, decimal minutes
Latitude: 35º 26.1’ N
Longitude: 118º 48.1’ W
Datum: NAD27; no uncertainty
Coordinate Precision: 0.1 minutes; 0.148 mi uncertainty
Coordinate Source: USGS map: 1:100,000; 0.032 mi uncertainty
Extent of Named Place: 2 mi
Distance Units: mi
Distance Precision: 1 mi
Decimal Latitude: 35.43500
Decimal Longitude: -118.80167
Maximum Uncertainty Distance: 3.180 mi

OFFSET IN ORTHOGONAL DIRECTIONS

Definition:

Locality consists of a linear distance in each of two orthogonal directions from a named place (Figures 11 and 12).

Examples:

Example 1: “2 mi E and 1.5 mi N of Bakersfield“
Example 2: “6 km N and 4 km W of Welna“
Example 3: “2 miles north, 1 mile east of Boulder Falls, Boulder County, Colorado“

Georeferencing Procedure:

Where localities have two orthogonal measurements in them, it should always be assumed that the measurements are ‘by air’ unless there is a reference that indicates otherwise.

Use the center of the starting point (e.g., in the Example 2, above, use the center of Welna), and enter its coordinates and extent in the ***MaNIS Georeferencing Calculator*** using the Calculation Type: "Coordinates and Error" Enter the distances and directions given, and push "Calculate." The new coordinates that appear at the bottom of the calculator are the ones you can now enter in your database. They should be different from the coordinates you entered in the ‘Latitude’ and ‘Longitude’ spaces – if they are not, check to make sure you have chosen the correct Calculation Type.

Figures:

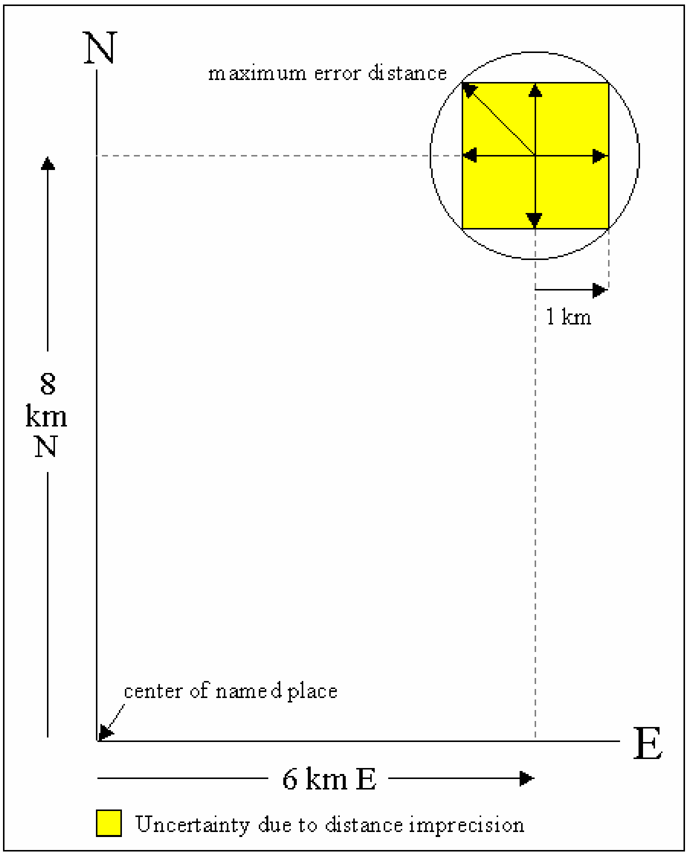


Fig. 11. Example of calculating maximum uncertainty using distance imprecision for two orthogonal offsets from the center of a named place. From Wieczorek (2001).

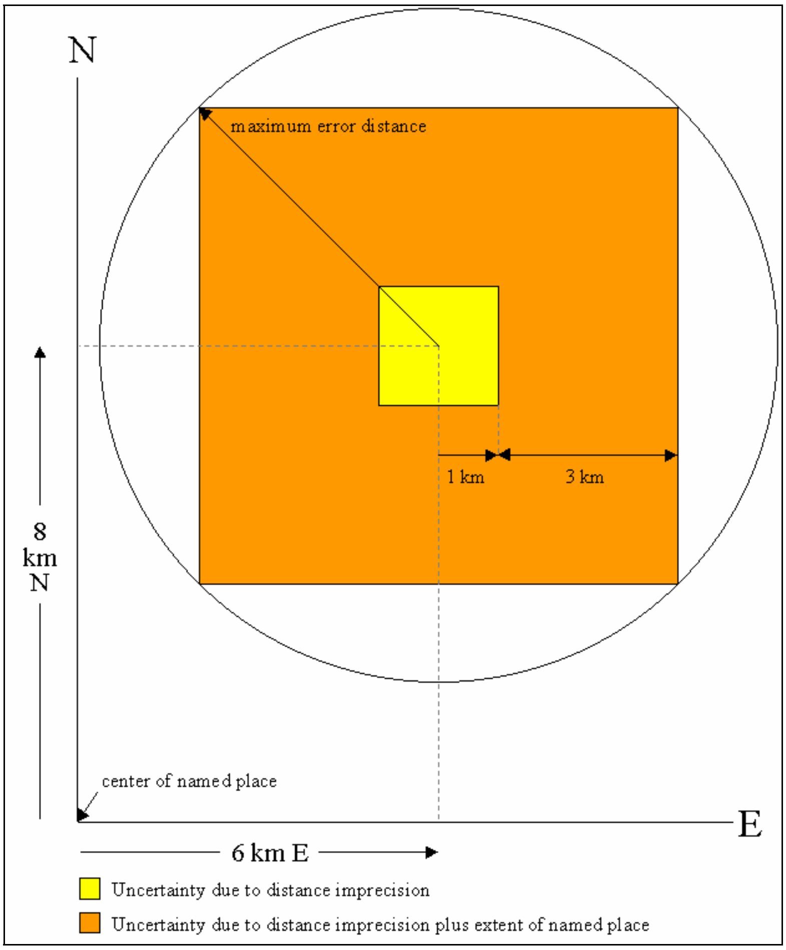


Fig. 12. Calculating maximum uncertainty from the combination of distance imprecision and extent. From Wieczorek (2001).

Extent:

Extent is the extent of your starting point - usually a city or a junction.

Uncertainty:

Use the ***MaNIS Georeferencing Calculator*** (http://manisnet.org/gc.html) to determine “Maximum Uncertainty Distance”.

- For **Calculation Type** use
  “Coordinates and error”
- For **Locality Type** use
  “Distance along Orthogonal Directions”.

Example 1.

Locality: "2 mi E and 3 mi N of Bakersfield"

> Suppose the coordinates for Bakersfield (the named place) came from the GNIS database (a gazetteer), the coordinates of the locality given to the nearest second, and the distance from the center of Bakersfield to the furthest city limit is 2 mi.

Coordinate System: degrees, minutes, seconds
Latitude: 35º 25' 4" N
Longitude: 118º 58' 54" W
Datum: not recorded; 0.049 mi uncertainty
Coordinate Precision: nearest second; 0.024 mi uncertainty
Coordinate Source: gazetteer
North or South Offset Distance: 3 mi
North or South Offset Direction: N
East or West Offset Distance: 2 mi
East or West Offset Direction: E
Extent of Named Place: 2 mi
Distance Units: mi
Distance Precision: 1 mi
Decimal Latitude: 35.4621
Decimal Longitude: -118.94623
Maximum Uncertainty Distance: 4.337 mi

OFFSET FROM TWO DISTINCT PATHS

Definition:

Locality consists of orthogonal offset distances, one from each of two different paths. This is a more unusual situation, but it does occur.

Examples:

Example 1: “1.5 mi E of LA Hwy. 1026 and 2 mi S of U.S. 190”

Georeferencing Procedure:

Locating the coordinates of a position like this is tricky. To do so, you have to draw a path parallel to, and at the appropriate distance and heading from, each of the reference paths. The place where they intersect (hopefully there is only one) is the coordinate.

Extent:

Use the extent of the wider of the two paths from which you are measuring. The uncertainty from the width of the wider highway will completely encompass the uncertainty from the narrower one. In Example 1, above, Interstate 190 is a big four-lane highway and LA 1026 is a two-lane highway. Since the measurements are orthogonal to each other and to the roads in this case, each extent would be half of the width of the respective highways. Since I 190 is the larger of the two, its extent would completely encompass the extent from LA 1026.

For standard extents of roads, use the values described under Locality Type: ***Feature***, above.

Uncertainty:

Use the ***MaNIS Georeferencing Calculator*** (http://manisnet.org/gc.html) to determine “Maximum Uncertainty Distance”.

- For **Calculation Type** use
  “Error – enter Lat/Long for the actual locality”
- For **Locality Type** use
  “Distance along Path”. Note: this isn’t actually the locality type, but it gives you all of the parameters you need to calculate the correct uncertainty.

LATITUDE AND LONGITUDE COORDINATES

Definition:

The locality consists of a point represented by coordinate information in the form of latitude and longitude. Information may be in the form of

- Degrees, Minutes and Seconds (DMS),
- Degrees and Decimal Minutes (DDM), or
- Decimal Degrees (DD).

Records should also contain a hemisphere (E or W and N or S) or, with Decimal Degrees, minus (–) signs to indicate western and/or southern hemispheres.

Examples:

Example 1: "36º 31' 21.4" N; 114º 09' 50.6" W"	(DMS)
Example 2: "36º 31.4566'N; 114º 09.8433'W"	(DDM)
Example 3: "36.524276º S; 114.164055º W"	(DD)
Example 4: "−36.524276; −114.164055"	(DD using minus signs to indicate southern and western hemispheres)

Georeferencing Procedure:

If a location has associated coordinates that are consistent with the rest of the locality description, there is generally little else to do except determine the maximum uncertainty.

Extent:

The extent of a locality should never really be zero. If a GPS was used to determine the coordinates, the accuracy of the GPS at the time (see the section ***Using a GPS***, above) should be used as the extent (or see estimates under '***UTM Coordinates***' below. If the coordinates were determined by some other or unknown means, use a reasonable minimum extent for the location, perhaps based on the rest of the locality description. For example, if the coordinates are associated with a point on a trap line, use the distance from the coordinates to the furthest end of the trap line as the extent.

Uncertainty:

Use the ***MaNIS Georeferencing Calculator*** (http://manisnet.org/gc.html) to determine "Maximum Uncertainty Distance".

- For **Calculation Type** use
  "Error – enter Lat/Long for the actual locality"

- For **Locality Type** use
  "Coordinates Only".

Example 1.

Locality: "35º 22' 24" N, 119º1' 4" W**"**

Coordinate System: degrees, minutes, seconds
Latitude: 35º 22' 24" N
Longitude: 119º 1' 4" W
Datum: not recorded; 79 m uncertainty
Coordinate Precision: nearest second; 40 m uncertainty
Coordinate Source: locality description
Distance Units: km, m, mi, yds, or ft
Decimal Latitude: 35.37333
Decimal Longitude: -119.01778
Maximum Uncertainty Distance: 0.119 km, 119 m, 0.074 mi, 130 yds, or 390 ft

Example 2.

Locality: "35.37,-119.02, NAD27, USGS Gosford Quad 1:24000"

Coordinate System: degrees, minutes, seconds
Latitude: 35.27
Longitude: -119.02
Datum: NAD27; no uncertainty
Coordinate Precision: .01 degrees; 1434 m uncertainty
Coordinate Source: USGS map: 1:24,000; 12 m uncertainty
Distance Units: km, m, mi, yds, or ft
Decimal Latitude: 35.37
Decimal Longitude: -119.02
Maximum Uncertainty Distance: 1.446 km, 1446 m, 0.899 mi, 1582 yds, or 4745 ft

UTM COORDINATES

Definition:

The locality consists of a point represented by coordinate information in the form of Universal Transverse Mercator (UTM) or related coordinate system (see Note below). When databasing using UTM or equivalent coordinates, a Zone should ALWAYS be included; otherwise the data are of little value when used outside that zone, and certainly of little use when combined with data from other zones. Zones are often not reported where a region (e.g., Tasmania) falls completely within one UTM zone. Be aware that UTM zones are valid only between 84ºN and 80ºS.

Note! There are many national and local grids derived from UTM and work in the same way – for example, the Australian Map Grid (AMG).

Examples:

Example 1: “UTM N 4291492; E 456156“ (Note: no zone cited).
Example 2: “AMG Zone 56, x: 301545 y: 7011991”
Example 3: “56: 301545.2; 7011991.4”

Georeferencing Procedure:

In Example 1, where no zone is cited, first find the UTM zone by using UTM Grid Zones of the World (Morton 2006) using any additional information in the locality description, such as country, state/province, county, etc.

Then fill in UTM data at Geographic/UTM Coordinate Converter (Taylor 2003). Remember that x is easting while y is northing.

Care! Care should be taken when determining UTM coordinates from a map as they are read in the opposite order to Latitude and Longitude, i.e., easting and then northing.

Extent:

See the recommendations under ***Latitude and Longitude Coordinates***, above.

Uncertainty:

Calculate in the same way as for ***Latitude and Longitude Coordinates***.

If unable to use the Georeferencing Calculator, a general rule of thumb is that the Uncertainty is of the order of

- 30 meters if determined by a GPS after UTC 00:00 2 May 2000, and the datum is recorded;
- 100 meters if determined by a GPS before UTC 00:00 2 May 2000 and the datum is recorded;
- 200 meters plus (depending on the location) if determined by a GPS and the datum is not recorded;
- Variable, depending on map scale if determined from a map (see Table 5, this document).

TOWNSHIP, RANGE, SECTION

Definition:

Township, Range and Section (TRS) or Public Land Survey System (PLSS) is a way of dividing land in the mid- and western USA. Sections are usually 1 mi on each side. Similar subdivisions are used in other countries, and should be calculated in a similar way, once the sizes of the rectangles have been determined. Map sheets are sometimes used and can also be calculates in this way.

A Township Range Section (TRS) description is essentially no different from that of any other named place. It is necessary to understand TRS descriptions and how they describe a place before trying to georeference. See the **References** at the end of this Locality Type, below, for links to further information on Township, Range and Sections and their meaning.

Note! Though TRS applies only to the USA, some countries may have equivalents and the principles elaborated here should be followed.

Examples:

Example 1: “T3S, R42E, SEC.2“
Example 2: “E of Bakersfield, T29S R29E Sec. 34 NE 1/4“

Georeferencing Procedure:

If there is no other usable locality data, or if TRS is the most specific information provided in the locality description, place the point at the center of the TRS or ¼ section. Otherwise, TRS is best only used as one factor in determining the final coordinates

To find the coordinates for the center of the TRS, use the TRS-data Website (Gustafson and Wefald 2003) and fill in the appropriate fields. Make sure to pick the correct state. The website will give you the geographic center of your section using the WGS84 datum.

If your locality includes something like "SW .25 Sec 15", then your work is not yet done. To georeference quarters of the section, use the coordinates from the TRS-data website and place them in the Topozone.com website, which shows section boundaries. Put your new point in the appropriate portion of the section and read the new coordinates from the top of the map. Be sure to record the datum used for the coordinates on Topozone®, since these are configurable while looking at the maps

Alternatively, to find the center of a quarter section, first find the center of the Section, then calculate the coordinates of the quarter section by using offsets of 0.25 mi in the appropriate directions from these coordinates. For example, the center of the NW 1/4 of Section 13 would be 0.25 mi N and 0.25 mi W of the center of Section 13.

Note! Not all TRS townships and sections are square. It is best to use a map to find the center of any subdivision of a section.

Extent:

For sections, the extent is half of the hypotenuse of the section, or 0.707 mi (the square root of 2 divided by 2). For quarter sections, the extent is half of that, or 0.354 mi. (Table 6).

Division	Example	Extent (mi)	Extent (m)
Township	T6S R14E	4.243	6828
Section	T6S R14E Sec. 23	0.707	1138
¼ Section	T6N R14E Sec. 23 NE ¼	0.354	570
¼ of ¼ Section	T6N R14E Sec. 23 NE ¼ SW ¼	0.177	285
¼ of ¼ of ¼ Section	T6N R14E Sec. 23 NW ¼ NE ¼ SW ¼	0.089	143

Table 6. Extents of Divisions of Townships in miles and meters. From Wieczorek (2001), Frazier *et al.* (2004).

Uncertainty:

Calculate the same as for '***Feature***'.

If unable to use the Georeferencing Calculator, see Table 6, above. The uncertainty estimate will be the extent plus the uncertainty due to the precision in the coordinates used – as long as the datum is recorded.

References:

Township, Range, Section Information:
http://www.esg.montana.edu/gl/trs-data.html
http://www.outfitters.com/genealogy/land/twprangemap.html
http://www.outfitters.com/genealogy/land/land.html

DUBIOUS

Definition:

At times, locality descriptions are fraught with vagueness. This may be due to any number of reasons, but in particular relates to historic collections in areas that at the time may have had no named features with which to reference.

The most important type of vagueness in a locality description is one in which the locality is in question. Such localities should not be georeferenced.

A cause of vagueness may be incorrect data entry and it is recommended that checking the original catalog books, field notes, specimen labels, etc. be the first step in removing the vaguery of a locality so that it can be georeferenced.

In Examples 1-3, below, the locality descriptions explicitly state that the information contained therein is in question. In Example 4, the location is not well enough bounded to identify a meaningful location. In the latter case, additional information (for example collector and date) may lead to a more specific location from diaries or other published information.

Examples:

Example 1: “possibly Isla Boca Brava“
Example 2: “presumably central Chile“
Example 3: “Bakersfield”
Example 4: “Nova Hollandia”

Georeferencing Procedure:

Do not georeference if the locality explicitly states that the information contained therein is in question.

Document in Remarks the reason for not georeferencing, e.g. “locality too vague to georeference”, “locality in question”, etc.

Note that subsidiary information may provide other information, which may help in determining a less dubious location.

Extent:

N/A

Uncertainty:

N/A

CANNOT BE LOCATED

Definition:

The cited locality cannot be located. This may be for any number of reasons, including:

- There is no locality information cited (Example 1),
- The locality fields contain other than locality information (Example 2),
- the locality cannot be distinguished from among multiple possible candidates (Examples 3 and 4), or
- the locality cannot be found with available references.

Examples:

Example 1: “locality not recorded“
Example 2: “Bob Jones“
Example 3: “"summit“
Example 4: “San Jose, Mexico“

Georeferencing Procedure:

Do not georeference.

Document in Remarks the reason for not georeferencing, e.g. “locality cannot be found with available references”, etc. Do still fill in the Georeferencing Resources field in your database so that the next researcher does not waste time using the same resources to track down the locality.

Extent:

N/A

Uncertainty:

N/A

DEMONSTRABLY INACCURATE

Definition:

The locality contains irreconcilable inconsistencies.

The worst type of locality description to georeference is one that is internally inconsistent. There are numerous possible causes for inconsistencies. Rather than determine coordinates for such localities, annotate the locality with the nature of the inconsistency and refer the locality to the source institution or collector for reconciliation. One common source of inconsistency in locality descriptions comes from trying to match elevation information with the rest of the description. In these cases (see Example 2), bear in mind that elevation data are notoriously inaccurate.

Another common source of inconsistency occurs when the locality description does not match the geopolitical subdivision of which it is supposed to be a part. At times, the locality can still be determined because the geopolitical subdivision is clearly at fault (see Example 3). In this case, georeference the locality and annotate it to describe the problem.

Often there is no way to know if the geopolitical subdivision or something in the locality description itself is at fault. In Example 4, the county may be wrong, the distance may be wrong, or the direction may be wrong. This locality cannot be disambiguated without going back to the originating institution, collection books, or by contacting the collector, etc.

Examples:

Example 1: "Sonoma County side of the Gualala River, Mendocino County"
Example 2: “10 mi W of Bakersfield, 6000 ft” (There is no place anywhere near 10 mi W of Bakersfield at an elevation of 6000 ft)
Example 3: "Delano, Tulare Co**.**" (Delano is in Kern Co.)
Example 4: “5 mi N of Delano, Kern Co." (5 mi N would put the locality in Tulare Co.)

Georeferencing Procedure:

Do not georeference. Record in remarks “locality contains irreconcilable inconsistencies”.

Extent:

N/A

Uncertainty:

N/A

CAPTIVE OR CULTIVATED

Definition:

Locality records the collection was made from a captive animal or cultivated plant, etc. The locality cited is often that of a zoo, aquarium, botanical garden, etc.

Examples:

Example 1: "lab born"
Example 2: "bait shop"
Example 3: “Cultivated in Botanic Gardens from seed obtained from Bourke, NSW.”

Georeferencing Procedure:

Do not georeference captive animals in standard georeference fields.

Do not georeference cultivated plant records in general georeference fields. If you must supply a georeference (for example for the location of the parent plant that supplied the seed), record it in a separate field, not in the general georeference fields.

Document in remarks “not georeferenced − captive/cultivated”, etc.

Extent:

N/A

Uncertainty:

N/A

Index

Index