Biodiversity GIS

postal	Private Bag X7 Kirstenbosch 7735
physical	Kirstenbosch Research Centre
website	http://bgis.sanbi.org
enquiries	**Selwyn Willoughby**
telephone	+27 21 799 8698 **fax** +27 21 797 6903
email	willoughby@sanbi.org
reference date	**18 October 2004**

DATA & META-DATA STANDARDS, INCLUDING EXPLANATORY TEXT

Please contact the Biodiversity Geographical Information Systems (BGIS) for assistance on the use of the data and meta-data standards

GIS DATA STANDARDS

VECTOR DATA

- All vector data (points, lines and polygons) must be provided in ArcView/ ArcMap Shapefile or ArcInfo Coverage format as this facilitates upload onto the BGIS website.
- An *AVL file (ArcView) or *LYR file (ArcMap) must accompany the Shapefile.
- All vector data should be provided in Geographic projection (Latitude and Longitude) and WGS84 Datum.
- All text in the attribute table must be in CAPITALS.
- Attribute feature names and descriptions should be in line with current standards such as SAGDAD (The South African Geospatial Data Dictionary).
- All data must be accompanied by metadata

RASTER DATA

Before acquiring remotely sensed imagery (Satellite images and aerial photography), a detailed terms of reference must be drawn up. The BGIS must be consulted in the drafting of these terms of reference.

The terms of reference must include but may not be limited to the following:

- The projection and datum type and parameters:

  The recommended projection to be used is the Universal Transverse Mercator Zone 34 South (UTM34S). This decision is based on the fact that the 1:10 000 and much of the urban colour orthophotography, as well as the Consolidated Municipal Infrastructure Programme (CMIP) database and disaster management database for the province is in this projection. The parameters are as follows:

  UTM 34 S parameters:
  WGS_1984_UTM_Zone_34S
  Transverse_Mercator

False_Easting: 500000.000000
False_Northing: 10000000.000000
Central_Meridian: 21.000000
Scale_Factor: 0.999600
Latitude_Of_Origin: 0.000000

Datum: WGS84

- The setting of the sensor (A calibration certificate should be requested)
- The resolution
- The rectification technique (ortho-rectification is preferable to rubber-sheeting techniques. Nearest Neighbour/ Bilinear)
- The DEM to be used in the rectification
- The acceptable RMSE (Root Mean Square Error) and maximum error (in meters). All error reports should be provided.
- The provision of control points (Will they be provided or captured by the consultant)
- Acceptable levels of cloud cover
- Colour-balancing and mosaicing requirements
- Standard naming convention (E.g. 3341ADCA)
- Standard Format (E.g. Mr SID)
- Copyright issues (single or multi-user license)
- Metadata

GPS

A detailed set of GPS standards will be defined at a later stage. It is *vital* to use the appropriate instrument for the scale and accuracy of data required. In general the *reliable* accuracy of the cheaper hand-held units is between 15 and 30m. For most purposes a machine capable of sub-meter accuracy (differentially corrected either in real time or post-capture) should be specified. The operator should have proven experience.

Data quality ("clean" data)

All intersecting lines must be processed in GIS to remove overshoots and undershoots, (dangling arcs), "bowties" (or "fish tails"), and sliver polygons, resulting from incorrect closing of polygons. This is also a common problem where importing from GPS data.

Lines between adjacent polygons must be captured once only.

If data are to be used for specific application/models they may require specific ArcGIS attributes in terms of linear networks or topology.

Base data

Administrative boundaries, infrastructure and where possible, all other base data must be used from a common source – to be made available on the Provincial IT server.

Digitising Accuracy

Many factors can impact upon the accuracy of a map. Any number of factors can cause error. Note that these sources can have a cumulative effect, such that it is imperative that error for each factor be kept to a minimum, where possible. Error, is calculated as:

$E = f(f) + f(l) + f(c) + f(d) + f(a) + f(m) + f(p) + f(rms) + f(mp) + u$

where,

f = flattening the round Earth onto a two-dimensional surface (transformation from spherical to planar geometry)
l = accurately measuring location on Earth (correct projection and datum information)
c = cartographic interpretation (correct interpretation of features)
d = drafting error (accuracy in tracing of features and width of drafting pen)
a = analogue to digital conversion (digitizing board calibration)
m = media stability (warping and stretching, folding, wrinkling of map)
p= digitizing processor error (accuracy of cursor placement)
rms = Root Mean Square (registration accuracy of tics)
mp = machine precision (coordinate rounding by computer in storing and transforming)
u = additional unexplained sources error
(Extracted from ArcInfo Help File, ESRI, 2001)

To minimise error, consider the causes of each error factor presented above. A certain level of error is unavoidable, but it is critical that both parties agree to all methodologies used, and that quality checking and ground truthing is carried out.

Scale

Map Scale	One mm on map =	A 0.5mm pencil line =
1:10 000	10m on the ground	5m
1:50 000	50m on the ground	25m
1:250 000	250m on the ground	125m

The required scale must be specified by the client to the data provider. The above table indicates the level of potential drawing/digitizing error at each common map scale, which can be compounded with other errors as in the previous section.

Base data used must be appropriate to the scale of operation.

META-DATA STANDARDS AND EXPLANATORY TEXT

IDENTIFICATION INFORMATION

Title: Name or title of the dataset

Brief Description: Very short description of the dataset

Date: Date that the dataset was published

Abstract Description: More detailed description of what is depicted in the dataset. What does the dataset represent?

Description of Purpose: A detailed description of the purpose for which the dataset was created.

Lineage Description: A detailed description of how the dataset was created. (What was used as source data? For raster imagery, what was the RMSE? What rectification technique was used?)

Product Classification: Standard/ Value-added. Has additional work been done on the dataset to make it more valuable?

Data Capture Source: Who (individual and organisation) has captured the dataset?

Temporal Extent: Actual survey duration details. When did the dataset reflect the on-the-ground reality?

Descriptive Keywords: Keywords used to describe the dataset.

Supplemental Information: Are there any additional reports/ documents giving further information about any component of the project.

Project Name: Was the dataset created as part of a larger project or initiative (such as C.A.P.E. or STEP)? If 'yes', what is the name of the project?

Bounding Polygon: North, South, East, West: The coordinates of the polygon that bounds/ encloses the geographic area of the dataset. These numerical values are generated automatically in ArcCatalog.

Scale: The scale of the dataset (or resolution of raster/ image data).

Language: The language used in the all components of the dataset (such as the attribute fields).

Theme Type: Feature (Vector)/ Image (Raster).

Content Type: For feature data, is the theme depicted using points, lines or polygons? For image data, is the data remotely sensed imagery, grid or TIN?

Minimum Zoom: The minimum zoom level of the theme.

Maximum Zoom: The maximum zoom level of the theme.

Thumbnail: A small J-Peg map image of the theme.

DISTRIBUTION INFORMATION

File Name: The actual name of the file (E.g. cedb_infrastructure.shp)

File Type: (E.g. ArcView Shapefile)

List of formats: Alternative formats in which the dataset may be requested (E.g. GeoTiff)

Dataset Size: The digital size of the dataset in megabytes.

Decompression Technique: Has the dataset been decompressed? If 'yes', what decompression technique was used?

Online Resource URL: Is the dataset available online? If 'yes' what is the URL?

Distributor Organization Name: Name of the organization responsible for distributing the dataset

Distributor Name: Name of individual responsible for distribution

Distributor Address: The distributor's postal address.

Distributor Telephone Number: The distributor's telephone number.

Distributor e-mail address: The distributor's e-mail address.

Ordering Instruction: How does one go about ordering the data?

Fees and Terms: How much will it cost to obtain the dataset? What are the terms or conditions for obtaining the dataset?

Turnaround Time: How long will it take to obtain the dataset once it has been formally requested/ ordered?

DATA OWNER AND METADATA INFORMATION

Owner Organisation: Name of the organization that owns (normally also created) the dataset

Contact Person: Person responsible for the creation (or ownership) of the dataset.

Position of Contact Person: Position of the contact person in the organisation.

Contact Address: Postal **a**ddress of the owner organisation**.**

Contact telephone number: Telephone number of the contact person

Contact e-mail address: E-mail address of the contact person.

Metadata Creator Organisation: Organization of the person completing the metadata form.

Metadata Creator Name: Name of the person completing this form.

Metadata Creator Position: Position of the person completing this form.

Metadata Creator Telephone Number: Telephone number of the person completing the metadata form

Metadata Creator E-mail Address: E-mail address of the person completing the metadata form

Metadata Date Stamp: Date that the metadata form is being filled in

CONSTRAINTS AND PARAMETERS

Use Constraints: Constraints to using the dataset

Access Constraints: Constraints to accessing the dataset (Can the dataset be distributed to the public?)

Data Copyright: Does the data have a copyright? Yes/No

Projection Parameters

A list of standard projects is available through the BGIS (Geographic-WGS84, Albers-Clarke 1880 unmodified, Geographic-Cape and Lambert WGS84). If the projection used for the dataset is one of these projections, only the projection name must be provided. If the projection is not one of these, then the following metadata fields must be filled in:

Projection Name: Name of the projection.

Reference System Name: The reference system information (name, code) can be obtained from the Chief Directorate Surveys & Mapping (Mowbray).

Reference System Code: The reference system information (name, code) can be obtained from the Chief Directorate Surveys & Mapping (Mowbray).

Projection Type: Type of projection (Conic, Azimuthal, Cylindrical)

Central Meridian: The north-south meridian or longitude of origin (central longitude) of the projection. It runs between the poles and perpendicular to the equator.

Upper Parallel: Vertical grid lines in the UTM system are oriented parallel to the central meridian. The upper parallel is the top vertical grid line for the particular zone.

Lower Parallel: Vertical grid lines in the UTM system are oriented parallel to the central meridian. The lower parallel is the bottom vertical grid line for the particular zone.

Latitude of Projection Origin: The latitudinal value of the projection centre.

Projection Units: Decimal degrees, Meters

False Easting: By having separate origins for the Northern and Southern hemispheres, UTM uses no negative values. 'False Easting' is the value or the 'easting' assigned to the projection origin. Easting values increase to the east.

False Northing: The value of the 'northing' assigned to the UTM projection origin. Northing values increase to the North.

UTM Zone: The UTM coordinate system represents small regions of the earth's surface. (Sixty UTM zones, each 6° of longitude wide, cover the globe). It is important to know which zone has been used, as for each UTM zone the projection parameters are unique.

Scale Factor at Equator: The scale factor at the equator is a multiplier used for reducing a distance obtained from a map by computing or scaling to the actual distance along the equator.

Datum Parameters

A list of standard datums is available through the BGIS (WGS84, Clarke 1880 unmodified and Cape). If the datum used for the dataset is one of these, only the datum name must be provided. If the datum is not one of these, then the following metadata fields must be filled in:

Datum Name: Name of the datum.

Datum Semi Major Axis: Half the longer axis of the ellipsoid.

Datum Semi Minor Axis: Half the shorter axis of the ellipsoid.

Datum Inverse Flattening: The inverse of the flattening value of the ellipsoid.

MAINTENANCE INFORMATION

Maintenance and Update Frequency: How often is the dataset revised or updated?

Revision and Update History

For each recorded updated or revision the following fields must be filled in:

Revision Date: Date of the revision or update.

Revised by: Person and organisation responsible for the revision/ update.

Revision Reason: Why was the dataset revised/ updated?

Revision Notes: Additional notes pertaining to the revision/ update.

CONTENT INFORMATION

Has SAGDAD been used: Yes/ No. SAGDAD is the South African Geospatial Data Dictionary. It is a standard feature catalogue that provides standard attributes names and descriptions.

If SAGDAD has not been used:

Has another Feature Catalogue been used: Yes/ No

If another Feature Catalogue has been used:

Catalogue Title: Name of the catalogue that was used

Catalogue Date: Date that the catalogue was published.

Attribute Codes and Descriptions: A comprehensive list of all attribute fields, alias and descriptions must be provided.