CARTOGRAPHICAL DATA AND DATA QUALITY ISSUES

Sindre Langaas
UNEP/GRID-Arendal
c/o Dept. of Systems Ecology, Stockholm University, S-106 91 Stockholm, Sweden.
Phone: +46-8-161737 Fax: +46-8-158417 E-mail: langaas@grida.no

Abstract

This paper reviews three readily available cartographical databases (DBs), the Digital Chart of the World (DCW), the World Vector Shoreline (WVS) and 30 Arc-Second DCW Digital Elevation Models (DEM) all originating from the US Defense Mapping Agency (DMA). These are presumed strong candidates as cartographical data sources for strategic needs at several Consultative Group of International Agricultural Research (CGIAR) centres. While most cartographical themes found within these DBs are of acceptable quality, some themes definitively needs improvement. No known substitutes to these are known to the author. An approach is being suggested to overcome these inadequacies. Furthermore, the data quality part of the US Spatial Data Transfer Standard (SDTS) is being briefly summarised with the aim to stress the importance of data quality reporting within the UNEP/CGIAR project when data sets are created.

INTRODUCTION, DEFINITIONS AND SCOPE

I will in this presentation briefly summarise information on some GIS databases (DBs) that originates from the US Defense Mapping Agency, are currently available, are potentially suited for strategic needs of several CGIAR centres and have been recommended to be used as such at the first CGIAR/UNEP workshop (**Arendal I**). *Strategic needs* was at Arendal I defined as requirements related to studies covering a large geographical area, for example in characterisation and classification to be used in strategic research planning (anon. 1992). I will limit this presentation to cartographical DBs. By *cartographical DB* is meant a structured collection of digital GIS datasets digitised from topographical (or equivalent) paper maps or created for the purpose of making topographic type paper maps. Commonly, the following themes are included in cartographical DBs: (i) Coastline, (ii) International boundaries, (iii) National administrative boundaries, (iv) Transport infrastructure, (v) Cities, (vi) Hypsography or Digital Elevation Models and (vii) Hydrography. In this presentation International and Sub-national Boundaries are excluded. These will be presented by Deichman and Fox later. Besides the value of this kind of GIS data for cartographical purposes, they are also of great value in GIS modelling. The use of administrative boundaries to derive statistics is a most prominent example.

I will not examine sources for GIS cartographical data for operational needs. *Operational needs* was defined at **Arendal I** to be related to much more detailed data requirements for studies applied to relatively small areas. The distinction between strategic and operational data are here set at a scale of 1:1,000,000 (or resolution 1 km^2).

Furthermore, I will briefly present the data quality part in the US Spatial Data Transfer Standard (SDTS). The SDTS is one out of several standards for geographical data (and information) that recently has been developed or are under development. The intent by presenting the data quality component of SDTS is to emphasise the need for this sort of meta-data information, increasingly important when GIS data are being combined in simple or complex models or being used for other purposes than the initial one(s). This presentation rely on experiences from several GRID-Arendal projects, Clark's (1992) presentation at Arendal I as well as information found in literature, analogue and digital on-line available from Internet

CARTOGRAPHICAL DATA FOR STRATEGIC NEEDS

Existing Datasets

Table *Global base layer data presented at Arendal I (from Clark 1992)*

	Database	**Size (MB)**	**'Ownership'**	**Scale**
1	Hershey	1.2	Public	> 1:40M
2	World Data Bank (WDB)-1	1.5	Public	> 1:12M
3	WDB-II	110	Public	> 1:3M
4	Micro-WDB-II	2.5	Public	> 1:10M
5	ARC/WORLD	?	Commercial	> 1:3M, 1:25M
6	Mundocarto	150	Commercial	> 1:1M
7	World Vector Shoreline (WVS)	150	Public	> 1:250K, 1:1M
8	Digital Chart of the World (DCW)	1,700	Public, Commercial	> 1:1M

In Arendal I the following cartographical DBs were described (Clark 1992):
Please refer to Clark (1992) for descriptions of most of these DBs. I will here focus upon the Digital Chart of the World (DCW) and the World Vector Shoreline (WVS). These DBs were recommended at **Arendal I** as the most important ones to be used by CGIAR Centres at the strategic level (anon. 1992). I will also introduce a digital DEM currently being derived from the DCW. GRID-Arendal and GRID-Nairobi has applied the DCW in several projects and have therefore accumulated quite considerable experience with its potentials and limitations, both regarding data qualities and practical problems. In particular, GRID-Arendal is, together with the Dept. of Surveying, Agricultural University of Norway, carrying out a project aimed at examining data quality issues of the DCW (Langaas and Tveite 1994, 1995).

Digital Chart of the World

Geographical coverage, description of content, spatial resolution

The DCW was made from two map series, the Operational Navigation Charts (ONC, 1:1 mill.) and the Jet Navigational Charts (JNC, 1:2 mill. Antarctica only) by the US Defense Mapping Agency, and collaborating partners in the UK, Canada and Australia. (ESRI 1992). The DCW is a digital representation of the global ONCs and JNCs and therefore a cartographical DB. Its development is documented in DMA (1992a). Both map series are made with a large number of mapping rules reflecting their intended purposes. For the ONCs the purposes area given in the product specification (DMA 1981):

> *"The 1:1,000,000 scale Operational Navigation Charts (ONC) Program provides aeronautical charts to support medium altitude enroute navigation by dead reckoning visual pilotage, celestial, radar, and other electronic techniques. In the absence of Tactical Pilot Charts (TPC's), these charts should also satisfy the enroute visual/radar navigation requirements of pilots/navigators flying low altitude operations (500 feet to 2000 feet above ground level). The ONC is also used for operational planning, intelligence briefings, and preparation of visual cock-pit displays/ film strips essential to aerospace navigation of high-performance weapon systems."*

The thematic content of the map series and their digital representation, the 1.7 GB DCW, reflects these purposes. Hypsography, Drainage, Roads, Populated Places, Political/Oceans, Land Cover, Railroads, Utilities, Cultural Landmarks, Transportation Structure, Physiography and Aeronautical are the major themes. For a complete description of themes (layers) and features, please refer to, e.g., ESRI (1992) or http://sun1.cr.usgs.gov/glis/hyper/guide/dcw.

Custodian, availability and format

The DCW was first released on four CD-ROMs by DMA as public domain data in the Vector Product Format (VPF) for a cost of US$200. Actually, the main purpose of making the DCW DB was to promote the use of the VPF, a recently developed military GIS format (DMA 1992a). Following this release, the DCW DB has now been released by a large number of GIS software vendors in their own proprietarian formats. Most of these DCW DBs are being sold at different price levels. The DCW DB now exist in ARC/INFO, MapInfo, Atlas and Intergraph formats, besides the initial VPF format. There exist also a number of more and less robust public domain conversion tools ftp'able on Internet.

Practical access

For the DCW in VPF and ARC/INFO formats, the relevant information for obtaining the CD-ROMs can be found in the GRID-Arendal Directory of Environmental CD-ROMs. For the DCW in various commercial GIS vendor formats, information should be available at the local dealer. UNEP/GRID also have an agreement with ESRI that allows the various GRID-centres to provide extracts upon request.

Shortcomings in content and quality

Users of the DCW, including ourselves, have after some years of usage, identified a large number of deficiencies in the various themes. These deficiencies are mainly related to the initial purposes of the ONCs and JNCs and the specific mapping rules guiding the compilation of the map sheets and the time of compilation (and updating) of ONC maps. Many of the features in several of the themes should only be included in the ONCs when they were of navigational value according to the specifications. To give an example from DMA (1981) related to populated places:

"702. Density and Selection
A. The following general rules are formulated to govern the selection of populated places.
1. In areas where populated places are very numerous, a selection of cities, towns and villages shall be shown to a density commensurate with scale.
2. In areas where populated places are generally sparse, cities, towns and villages shall be shown to a density comparable to the density on a standard 1:500,000 scale map of the area".

Similar kind of mapping rules related to the navigational significance of the various features .are found for most themes

The age of the original maps is another importance quality factor. For example most of the African tiles dates back to the 60ies and 70ies. It should further be kept in mind that the map sheet production year may deviate from the source material age. Therefore, several of the themes are not of a sufficiently high quality to be used for modelling purposes. Those that have been found by GRID-Nairobi to hold a quite high level in Africa are Hypsography, Drainage, Populated Places, and Political/Ocean boundaries (Goff 1994). Those found to be so-called problem coverages were the Roads, Railroads and Utilities. The rest where found somewhere in-between.

Processing requirements / problems

Due to the considerable data amounts, quite powerful HW and GIS SW are highly recommended. From own experiences we would strongly recommend UNIX type HW and SW, at least during DCW data preparation and editing phases. Further, it has been experienced that some themes, such as the Drainage layer, have had number of arcs per tile exceeding the software limitations of, e.g., UNIX Arc/Info. Although possible to bypass

Maintenance and update problems

The DMA is currently working on the second edition of DCW. We believe that this version will only remove errors related to transfer from paper maps to digital data, such as coding errors, duplication of lines, mislabelled edges, etc. and to a much lesser extent deviations from reality. Therefore, the actual content will predominantly remain the same. It is therefore quite obvious that many CGIAR centres will need to edit, update and correct DCW data. One technical solution to this can be to scan at high resolution another paper map at higher accuracy of the region of interest, to reproject the raster map and DCW to the same projection system and then to edit the DCW vector data superimposed upon the scanned (raster) map. Edit here means to fit to the positioning of the same geographical feature in the other map source.

World Vector Shoreline

Geographical coverage, description of content, spatial resolution

The World Vector Shoreline is a global dataset of shorelines created by the US Defense Mapping Agency (IOC et al.1994). It was developed by the DMA at a nominal scale of 1:250,000. Global coverage was complete in 1989. The sole feature is the shoreline. The primary data source for the WVS was DMA's Digital Landmass Blanking (DLMB). These were deduced primarily from the Joint Operations Graphics and coastal nautical charts also produced by the DMA. The DLMB data is a raster data set with 3 by 3 arc-second interval geographic grid, which explains the 3 arc-second stepping interval in the WVS when displayed at large scales. For parts of the world not covered by the DLMB, the shoreline was taken from the best available hardcopy sources at a preferred scale of 1:250,000.

Custodian, availability, practical access and format

The WVS is in the custody of the US DMA. A beta test version of the WVS was released on CD-ROM in the VPF format also used for the DCW described earlier. Another version of the WVS was released in 1994 by the International Oceanographic Commission (of UNESCO) and the International Hydrographical Organisation as part of the General Bathymetric Chart of the Oceans (GEBCO) on CD-ROM (IOC et al.1994). The latter version is stored in an internal binary format. Luckily, as a part of the GEBCO CD-ROM there a exist a conversion utility that enables conversion to DXF format, edible or importable by most GIS software.

Positional accuracy

The specification for positional accuracy, is that 90% of all identifiable shoreline features should be located within 500 meters (i.e. 2 mm at 1:250,000) of their true geographic position with respect to the World Geodetic System (WGS-84) datum. The precision as defined by the 3 by 3 arc-second steps imply at worst around 100 m at Equator and improving towards the Poles.

Comparison with digital coastline data from Norway in the same scale (1:250,000) made by the Norwegian Mapping Authority does not reveal differences of any significance worth mentioning (J.-A. Bordal, pers. comm.). Although far from most CGIAR centres' regions of interest, we tend to believe that the accuracy is good for most parts of the world. For strategic needs the WVS is of high quality.

30 Arc-Second DCW Digital Elevation Models (DEM)

Geographical coverage, description of content, spatial resolution

Terrain and hypsography information represented by digital contour lines or by Digital Elevation Models are becoming increasingly important for both analytical and visualisation purposes. Height information is provided in the DCW as contour lines at selected and unequally spaced intervals (ESRI 1992). For many purposes an equally spaced (in x and y direction) DEM is more suited. Therefore, USGS represented by the EROS Data Center, in association with UNEP/GRID-Sioux Falls, are currently developing a consistent 30 arc-second global DEM. The 30 arc-second resolution of course varies with latitude. At Equator this equals a raster cell size of approx. 930 m. The width decreases with latitude to the half at 60° N and S. The method used to create the DEM applies both the contour and point height information as well as the drainage network information from the DCW. As of 15. January 1994, Africa, Haiti, Madagascar and Japan were ready, and the South-America in advanced progress. For more up-to-date information about the global progress the Customer Services EROS Center or UNEP/GRID-Sioux Falls can be contacted.

Custodian, availability, practical access and format

The above mentioned already prepared DEMs can be obtained from the EROS Data Center by at least two ways, through Internet by ftp (for free) or on unlabeled CCTs (for a modest cost). It was also planned to be released on CD-ROM. The DEM data are provided as 16-bit straight raster (also termed unsigned 2-byte binary data) images. The height information is provided as feets. 4 ancillary files provide additional meta-information on issues such as file structure, world co-ordinates and position information. One of these files supports the ARC/INFO Image Integraton routine for image-to-world transformation.

Data quality

The absolute accuracy of the vector information in the DCW is 2000 meters circular error (horizontal) and ± 650 meters linear error (vertical) at 90 percent confidence according to the specifications (DMA 1992b). The DEM grid created obviously will be no more accurate than its sources. We are not aware of assessments that have been carried out to evaluate the accuracy of any of the sub-sets of this global DEM under preparation.

DATA GAPS

When one compares the cartographical themes offered by the three above-described DBs with those normally included by cartographical DBs, the situation looks quite good. The World Vector Shoreline is an obvious candidate for coastline data due to superior resolution and accuracy compared to the DCW. The DCW still has much to offer cartographically. However, as clearly illustrated in the initial aim of the ONC and JNC map series and by practical experiences, there are several themes that require improvement in data quality, in particular positional accuracy and completeness (see below) before fulfilling the needs of the CGIAR Centres and others. Transport infrastructure and Hydrography are two cartographical themes that seriously needs improvement. We are not aware of new data source at strategic scales that will resolve this need in the near future. An approach based on the existence of scanned higher quality paper maps and subsequent on-screen editing of co-registered DCW data is suggested as one feasible way of editing.

GIS DATA QUALITY ISSUES

With the quite revolutionary developments taking place in the field of digital geographical information technology and use the last decade, the need for meta-information to accompany the digital GIS data sets have become apparent. In particular, the possibilities to use GIS data sets for multiple purposes and several data sets to be combined in multi-layer model have enforced this. Several standardisation efforts have been initiated. One example is the US Spatial Data Transfer Standard developed by the joint US geodata community under the leadership of USGS (Fegeas *et al.* 1992). We here use the SDTS as an example, firstly because it is one standard already ready and implemented (NIST 1992), secondly because it is accessible over Internet.

While standards such as the SDTS encompass many parts, we will here specifically address the data quality part. We consider this of major importance within the joint UNEP/CGIAR project as a main aim of the project is to jointly compile, distribute and maintain high quality natural resource and socio-economic digital data sets. The data quality report of the SDTS consists of five portions being:

- lineage
- positional accuracy
- attribute accuracy
- logical consistency
- completeness

We here will briefly review these five parts, with the objective that the CGIAR centres compiling GIS data sets attempt to cover these quality parts in their data reporting. The following brief summary is taken from NIST (1992).

Lineage

This part shall include a description of the source material, methods of derivation, including all transformations involved. Appropriate dates should be included for relevant data sources and processing steps.

Positional Accuracy

This part shall include the degree of compliance to the spatial registration standard (another part of SDTS). Quality of control assessments shall be reported by using the procedures established in the geodetic standard. Descriptions of positional accuracy shall consider the quality of the final product after all transformations. The date of any positional test shall be included. Variations in positional accuracy shall be reported either as additional attributes of each spatial object or through a quality overlay (reliability diagram). Four optional methods for measuring positional accuracy are suggested.

Attribute Accuracy

Accuracy measurements for attributes on a continuous scale shall be performed using procedures similar to those used for positional accuracy. The report of a test of attribute accuracy shall include the date of the test and the dates of the materials used. In the case of different dates, actual changes in the phenomena shall be described. Spatial variations in attribute accuracy may be reported in a quality overlay. Three quantitative methods are suggested for attribute accuracy assessment

Logical Consistency

A report of logical consistency shall describe the fidelity of relationships encoded in the data structure of the digital spatial data. A number of tests can be carried out to assess various types of logical consistency, such as -

- Do the data contain permissible values only ?
- Do lines intersect only where intended ?
- Are any polygons too small, or any lines too close ?

The term "topologically clean" is allowed to be reported provided that
(a) All chains (arcs) intersects at nodes. Use of exact tolerance shall be reported.
(b) Cycles of chains and nodes are consistent around polygons. Or, alternatively, cycles of chains and polygons are consistent around nodes.
(c) Inner rings embed consistently in enclosing polygons.

The quality report shall report software (name and version) and dates of tests.

Completeness

Completeness in SDTS refers to information of selections criteria, definitions used and other relevant mapping rules such as minimum area or minimum width. Deviations from standard coding schemes, as well as definitions and interpretation shall also be reported.

CONCLUSIONS

The World Vector Shoreline and the Digital Chart of the World complemented by a DEM derived from the latter are believed to provide the CGIAR centres with most necessary cartographical GIS data at the strategic level (< 1:1,000,000). Some themes definitively need improvements. Which (DCW) themes and how much improvement is needed will depend upon the various centre needs in terms of data quality. To stress the significance of and need for proper geodata quality reporting, the quality part of the US Spatial Data Transfer Standard is briefly reviewed.

Acknowledgements

This paper has benefited considerably from the DCW Data Quality Project funded by the Geographic Information Technology Programme of the Norwegian Research Council and UNEP/GRID-Arendal. Critical comments by Christopher Smith, GRID-Arendal, improved this paper considerably.

References

Clark, D.M. 1992. Basic data layers for resource analysis and evaluation. Discussion paper prepared for the CGIAR/NORAGRIC/UNEP Meeting on Digital Data requirements

Defense Mapping Agency. 1981 - . Product specifications for Operational Navigation Charts (Code: ONC) Scale 1:1,000,000. First edition 1981 and changes and amendments thereto. Defense Mapping Agency, Washington., D.C.

Defense Mapping Agency. 1992a. Development of the Digital Chart of the World: Washington, D.C., U.S. Government Printing Office.

Defense Mapping Agency. 1992b. Military specification Digital Chart of the World (DCW). MIL-D-89009. 204 pages.

ESRI. 1992. The Digital Chart of the World for use with ARC/INFO® Data Dictionary. ESRI, Redlands, CA.

Fegeas, R.G., Cascio, J.L. and Lazar, R.A. 1992. An overview of FIPS 173, The Spatial Data Transfer Standard. *Cartography and Geographic Information Systems*, Vol. 19, No. 5., ??-??.

Goff, E. 1994. Evaluation of Digital Chart of the World for Africa. Internal Note, GRID-Nairobi.

IOC, IHO and BODC. 1994. Supporting volume to the GEBCO Digital Atlas. Published on behalf of the International Oceanographic Commission (of UNESCO) and the International Hydrographical Organisation as part of the General Bathymetric Chart of the Oceans (GEBCO); British Oceanographic Data Centre, Birkenhead, UK.

Langaas, S. and H. Tveite. 1994. Project proposal: Issues of error, quality and integrity of digital geographic data: The case of the Digital Chart of the World. URL http://ilm425.nlh.no/gis/dcw/dcw.html

Langaas, S. and H. Tveite. 1995. To Characterise and Measure Completeness of Spatial Data: A Discussion Based on the Digital Chart of the World (DCW). Paper to be presented at ScanGIS'95, the Fifth Scandinavian Research Conference on GIS, Trondheim, Norway, June 12th - 14th, 1995.

National Institute of Standards and Technology. 1992. Federal Information Processing Standard Publication 173 (Spatial Data Transfer Standard). US Dept. of Commerce, 199 pages.