UNEP/GRID Documentation Summary for Data Set:

World Cities Population Database

The World Cities Population Database (WCPD) has been produced by the Birkbeck College of the University of London in 1990 at the request of the United Nations Environment Programme (UNEP) in Nairobi, Kenya. The database contains head counts for 3000 cities world-wide which either have a population of over 100,000 or are capitals of their nation state. Listed are the geographical location of the cities and their population sizes. The material is primarily derived from the UN Demographic Yearbook 1987, published in 1989. For the United Kingdom, the UNDY data were inadequate and thus were replaced with the OPCS 1981 Urban Areas volume, giving cause for concern about other countries.

WCPD is an Arc/Info point coverage. It has no projection, as the cities are located on the basis of their latitude and longitude. The coordinates were assigned on the basis of gazetteers produced by The Times, by the Permanent Committee on Geographical Names, and by other sources or by direct reference to maps.

Each record in the data base contains details of the city name, country name, latitude and longitude of the city, its population at a defined time, the population of the urban agglomeration and the data compilation method. The Arc/Info attribute table contains the following fields:

- AREA — Arc/Info item
- PERIMETER — Arc/Info item
- WCDP# — Arc/Info item
- WCPD-ID — Arc/Info item
- ID-NUM — Unique number for each city
- CITY — City name
- COUNTRY — Country name
- CITY-POP — Population of city proper
- POP-AREA — Population of urban agglomeration
- YEAR — Latest available year of collection
- CEN-EST — Census/estimate in the above year, 1=census, 2=estimate
- CAP — Capital city or not; 1=capital, 2=not capital

WCPD comes as an Arc/Info EXPORT file called "WCPD.E00" and contains 410 Kb of data. The file has a record length of 80 and a block size of 8000 (blocking factor = 100). The file can be read from tape using Arc/Info's TAPEREAD command or any other generic copy utility. If distributed on a diskette it can be read using the ordinary DOS 'COPY' command. The file has to be converted to Arc/Info

internal format using its IMPORT command.

References to the WCPD data set can be found in:

- SERLL News, Issue No. 1, January 1991, Birkbeck College, London, UK.
- D. Rhind. "Cartographically-related research at Birkbeck College 1987-91" in: The Cartographic Journal, Vol. 28, June 1991, pp. 63- 66.

The source of the WCPD data set as held by GRID is Birkbeck College, University of London, Department of Geography, London, UK.

The Birkbeck family of global databases of population

Enormous attention has been given in recent years to study of the global environment, mostly by remote sensing means. Models of what might happen to the earth's climate, to world sea levels, etc have been built and discussed. Unhappily, it has not proved possible to assess the impact of different scenarios on the human race since we have no comparable data files describing the distribution of people around the world. This Birkbeck project set out to improve this situation.

The approach taken is to create a family of data bases. This has been adopted for three reasons:

- different projects require data at different resolution;
- low resolution data sets can be provided quickly whilst higher resolution data sets assembled from multiple sources require considerable effort to assemble and may be impossible to complete for all of the earth's surface;
- the funding of this work will inevitably be from multiple sources, each prepared to support one particular product.

A policy decision taken at an early stage was that we would work in parallel on different members of the family, thereby ensuring that detailed data sets were not too delayed after the less detailed ones. Another was that the population material sought would be limited - at least so far as the detailed data sets were concerned - to 'head count' data. And a third policy decision was that we would use existing and, if necessary proprietary, files of boundaries as the spatial framework - thereby saving us enormous amounts of work.

The data sets completed to date (March 1992) or in active preparation are:

1. National-level Population Database (NPD), also referred to as 'Level 0'

This includes population figures derived primarily from the Statesman's Yearbook 1989-90, together with the date to which the information relates, the country name and area and other details for 196 countries in the world. It is held in an ARC/INFO attribute file which can be mapped using the PC version of ESRI's revised derivative of World Data Bank II. This is conveniently run on the IBM PS2/80 and - concidently - the work was funded by IBM UK Ltd.

2. World Cities Population Database (WCPD)

This contains head counts for the 3,000 cities world-wide which either have a population of over 100,000 or are capitals of their nation state. The material is primarily derived from the UN Demographic Yearbook 1987, published in 1989. For Britain, the UNDY population data were inadequate since the 'cities' used were a mixture of parliamentary constituencies, local government districts and towns; these were replaced by urban area statistics taken from the OPCS 1981 Urban Areas volume for all towns with more than 100,000 people. Latitude and longitude was assigned on the basis of gazetteers produced by The *Times,* by the Permanent Committee on Geographical Names and by other sources or by direct reference to maps. Each record contains details of the city name, country name, latitude and longitude of the city, its population at a defined time, the population of the area (in some cases, no city population is available) and the method by which the data have been compiled (census or estimate). It should be noted that the definition of cities differs around the world. Moreover, this research revealed serious flaws in the UNDY as a source: checking of city names and the process of allocating latitude and longitude revealed ambiguity in which city was intended, largely through use of place names which do not conform to accepted international standards. Moreover, whilst the population data for other countries which are readily checked (e.g. France) seem of reasonable quality, that for Britain is totally unacceptable and gives cause for concern about other countries. That said, the UNDY is the single, up-to-date and supposedly consistent data source at this level of detail so its use was unavoidable.

The data were originally entered in Statview on a Mac SE and then converted into id,x,y and id, attributes file form for ARC/INFO purposes. The final versions of the files are held in an ARC/INFO attribute file and can be mapped against either the PC or the Workstation version of the ESRI World Data Bank II file. This work was funded by the UN Environment Programme.

3. African Cities Population Database (ACPD)

This closely resembles the WCPD except that the data source is a 1988 report of the

Economic Commission for Africa or ECA (kindly made available by Prof. J. I. Clarke), the population threshold is lower (all towns above 20,000 are said to be included) and that fewer details are available for each town (only its name, latitude and longitude, country and population are stored). The coordinates were derived from gazetteers or from African maps (largely held in SOAS and the RGS). Significant problems were found with several countries such as Togo (where the town names provided by the ECA did not match the great bulk of names on the maps) and South Africa (where the population figures were entirely absent from the ECA report and were derived from the WCPD after comparison with others derived from South Africa House).

The data compilation procedure was as for WCPD and the final files are held in ARC/INFO attribute file form. They may be mapped against either of the two ESRI World Data Bank II files.

This work was funded by the UN Environment Programme.

4. Medium Resolution Population Database (MRPD) (work in progress)

This is designed to provide population data at much higher resolution than NPD. It is being compiled on the basis of administrative district (or equivalent) units. Thus in the UK the 452 local government districts are being used as the base and in India the similar number of units is being used. For the USA, the county level data file is being used.

Where these boundaries are already in the ESRI Workstation version of the World Data Bank II these are being used. In future, those boundaries in the US Defense Mapping Agency's Digital Chart of the World may be much better solutions but, for the moment, whatever other boundaries are available and match minimal quality requirements are pressed into use and converted to form part of a global map.

This work is funded by IBM UK Ltd.

5. High Resolution Population Database (HRPD) (work in progress)

For some areas of the world, quite high resolution data is already available or will shortly become available. Thus in Europe, commune-level (ED in UK) already exists and is crudely referenced in geographical terms (in France, for example, the boundaries for the 30,000 communes have been digitized and are sold by IGN Paris). In the USA, the county or sub-county files are also available and TIGER has facilitated the creation of such databases. It is our plan to obtain such data on every possible opportunity, subject to availability of human resources, and gradually build up a 'quilt' of national datasets. It seems unlikely that this will provide true global coverage in the foreseeable future.

Populations of cities.

Project started 10 September 1990 **Projected completed** 23 October 1990
Time period :- 5 weeks.

Updated 7 November 1990 **Update completed** 26 November 1990
Updated to the 1988 Demographic Yearbook.

Work completed by: A. Wilkes, S. Rhind, M. Abbott and staff at the RGS.

Sources

Cities update: '1988 Demographic Yearbook' 40th edition.
Table 8 'Population of capital cities and cities of 100,000 and more inhabitants: latest available year.'

Cities world: '1987 Demographic Yearbook' 39th edition.
Table 8 'Population of capital cities and cities of 100,000 and more inhabitants: latest available year.'

Cities Africa: 'Demographic Yearbook' 1973,1976,1978,1979 and 1981.
Table 11 'Latest available population estimates of capital cities and towns of 20,000 and more inhabitants since 1970.'

Cities UK : Office of Population Censuses and Surveys Registrar General, Scotland, Census 1981, Key Statistics for Urban Areas, Great Britain, Cities and Towns. Table 1 'Population, age and birthplaces.'

Method

The data was copied onto a Macintosh from the source documents with the longitude and latitude added. The longitude and latitude were collected from a variety of sources including the Times Atlas, and gazetteers at the Royal Geographic Society. All the data was typed onto the Macintosh by Samantha Rhind.

Two files were produced with the data; one had the id, longitude and latitude, the other had all the other attributes; including id, city, country, city population, urban agglomeration, year and census or estimation.

The data was imputed as one file as follows:-

City	Country,	City pop,	Urban agglom,	Year,	Latitude,	Longitude
Alger,	Algeria,	1523000,	1740461,	1977, C,	36, 50, N,	3, 0, E
Annaba,	Algeria,	239975,	246049,	1977, C,	36, 55, N,	3, 47, E
Benta,	Algeria,	102756,	102756,	1977, C,		
Blida,	Algeria,	138240,	158047,	1977, C,	36, 30, N,	2, 50, E
Constantine,	Algeria,	344454,	378668,	1977, C,	36, 22, N,	6, 40, E
Oran,	Algeria,	490788,	543485,	1977, C,	35, 45, N,	0, 36,W
Setif,	Algeria,	129754,	129754,	1977, C,	36, 11, N,	5, 24, E
Sidi-bal-Abbes,	Algeria,	112998,	112998,	1977, C,	35, 15, N,	0, 39,W

The file was divided into two files. The longitude and latitude was recalculated so that the values west and south become negative values and all the minutes were divided by 60 and added to the degree value to give the following files.

File 1 looks like this:-

1, 3, 36.8333
2, 7.7833, 36.9167
3, 6.1667, 35.5667
4, 2.8333, 36.5
5, 6.6667, 36.3667
6, -.6, 35.75
7, 5.4, 36.1833
8, -.65, 35.25
9, 13.25,-8.8333
10, 2.5167, 6.4

File 2 looks like this:-

1, Alger, Algeria, 1523000, 1740461, 1977, 1
2, Annaba, Algeria, 239975, 246049, 1977, 1
3, Banta, Algeria, 102756, 102756, 1977, 1
4, Blida, Algeria, 138240, 158047, 1977, 1
5, Constantine, Algeria, 344454, 378668, 1977, 1
6, Orari, Algeria, 490788, 543485, 1977, 1
7, Setif, Algeria, 129754, 129754, 1977, 1
8, Sidi-bel-Abbes, Algeria, 112998, 112998, 1977, 1
9, Luanda, Angola, -9999, 475328, 1970, 1
10, Cotonou, Benin, 383250, -9999, 1981, 2

Both files were then copied over to the vax 11/750 from the macintosh using the Netcopy in the Tssnet file. A simple copy command was used once the Macintosh was logged on to the Gvax1.

On the Vax, file 1 was edited to remove any values with -9999 (no data) and any duplicates in the file. The file was then ARC GENERATED to produce the coverage CITIES.

```
$ ARC GENERATE CITIES
generate: INPUT CITIES.DAT
generate: POINTS
generate: QUIT
```

The coverage was built as a points coverage. Plots of each area were produced for checking against the Times Atlas, as some of the locations were in the ocean and not in the correct place. Any incorrect cities were noted and these were deleted and a new data file was edited to replaced the points. These were generated and checked before being added to the main coverage. Any cities with the values of -9999 were located they were added in the same way to the main file.

In ARC INFO the attribute data file was defined as follows:-

CITY-ID	4	5	B	0
CITY	30	31	C	
COUNTRY	32	33	C	
CITY-POP	10	11	I	
POP-AREA	10	11	I	
YEAR	4	4	I	
CEN-EST	1	2	I	

The data was added to the define file C2.dat in INFO.
An ARC JOINITEM was used to append the attributes to the main file.
$ ARC JOINITEM CITIES.PAT C2.DAT CITIES.PAT CITIES-ID CITIES-ID

The coverage was built with point attributes before ID-NUM, 4, 5, B was added to the main file. In ARC INFO, ID-NUM was calculated to equal the CITY-ID value, thus retaining all the unique ids for each city. Also using ADDITEM, CAP, 2, 1, 1 was added to the file, this was calculated to 1 if the city was a capital city of the country. This was mainly calculated in ARCEDIT. The final coverage was built.

Coverage data.

World Cities Population Database - WCPD

Labels	2832		108 bytes of attribute data	
Tics	4			
Fuzzy	0.1249334E-01			
Dangle	0.0000000			
Boundary	Xmin	-179.0008	Ymin	-53.16670
	Xmax	178.4167	Ymax	71.76670
Size	772 blocks			
Export size	402858			

African Cities Population Database - ACPD

Labels	479		96 bytes of attribute data	
Tics	4			
Fuzzy	0.2000000E-02			
Dangle	0.0000000			
Boundary	Xmin	-23.66667	Ymin	-33.98300
	Xmax	57.53228	Ymax	36.91700
Size	134 blocks			
Export size	65464			

UK Cities

Labels	63	108 bytes of attribute data
Tics	4	

Fuzzy	0.002			
Dangle	0.000			
Boundary	Xmin	-4.483	Ymin	-50.383
	Xmax	1.300	Ymax	57.167
Size	28 blocks			

Attributes

For WCPD.PAT

AREA	}	4, 12, F, 3
PERIMETER	} ARC/INFO items	4, 12, F, 3
WCPD	}	4, 5, B
WCPD-ID	}	4, 5, B
ID-NUM	Unique number for each city.	4, 5, B
CITY	City name.	30, 31, C,
COUNTRY	Country.	30, 31, C,
CITY-POP	Population of city proper.	10, 11, I,
POP-AREA	Population of urban agglomeration.	10, 11, I,
YEAR	Latest available year of collection.	4, 4, I,
CEN-EST	Census or estimation in the above year. 1 = Census 2 = Estimation	1, 2, I,
CAP	Capital city or not. 1 = Capital 2 = non capital	1, 2, I,

For ACPD.PAT

AREA	}	4, 12, F, 3
PERIMETER	} ARC/INFO items	4, 12, F, 3
ACPD	}	4, 5, B
ACPD-ID	}	4, 5, B
ID-NUM	Unique number for each city.	4, 5, B
CITY	City name.	30, 31, C,
COUNTRY	Country.	30, 31, C,
CITY-POP	Population of city proper.	10, 11, I,
YEAR	Latest available year of collection.	4, 4, I,

For UK .PAT

AREA	}	4, 12, F, 3
PERIMETER	} ARC/INFO items	4, 12, F, 3
UK	}	4, 5, B
UK-ID	}	4, 5, B
ID-NUM	Unique number for each city.	4, 5, B
CITY	City name.	30, 31, C,
COUNTRY	Country.	30, 31, C,
CITY-POP	Population of city proper.	10, 11, I,

POP-AREA	Population of urban agglomeration.	10, 11, I,
YEAR	Latest available year of collection.	4, 4, I,
CEN-EST	Census or estimation in the above year. 1 = Census 2 = Estimation	1, 2, I,
CAP	Capital city or not. 1 = Capital 2 = non capital	1, 2, I,

Requests

1. M. Norton-Griffiths - United Nations, GEMS/UNITAR Africa Programme P.O. Box 30552, Nairobi, Kenya.

   Wanted the population of cities for the world and greater detail for Africa. Received the coverages CIT4 and AFRICA as PC export coverages. Cit4 was sent 22 October and Africa sent 26 October 1990.

2. Ms Carol Owen - MARC, Kings College London, Old Coach House, Campden Hill, London, W8 7 AD.

   Wanted a listing of cities with a population of 3 million or over for the world. A listing of 63 cities was sent on the 26th October 1990.

Missing Cities

For Africa:-

140	Ashianan	Ghana
143	Effia-Kuma	Ghana
147	Nkawkan	Ghana
234	Gnayna	Togo
235	Ikponou	Togo
244	Owue	Togo
245	Ouma	Togo
250	Woudou	Togo
252	Yeque-Adele	Togo

For the World:-

314	Ciudad Lopex Mateos	Mexico
2540	Tankoping	Sweden
	Parana	Chile
	Neyagawa	Japan
	Nabeul	Tunisia
	Qaeim shahr	Iran

Parana	Chile	178000	- 9999	1985	2	0
Neyagawa	Japan	257355	257355	1986	2	0
Nabeul	Tunisia	- 9999	334702	1984	1	0
Qaeim shahr	Iran	110137	- 9999	1986	1	0