Cartographic Data Bases Panel

Frederick R. Broome, Presiding
Bureau of the Census

Donald J. Orth
U.S. Geological Survey

Nevaire M. Serrajian
Defense Intelligence Agency

E. Llewellyn Robe
Central Intellignece Agency

Charles Meyers
Department of the Interior

Broom: A data base may take any form, and a cartographic data base is no exception. It can be a table in a book, a photo, a digital file, or a map. The average USGS 7.5-min topographic map has been estimated to have over 100 million bits of information--quite a good-sized data base. The form of the data base is usually dictated by the equipment available and the functions anticipated. The product may be as simple as the answer to "What's the center of Washington, D. C., in latitude and longitude?" or as complex as a map of a city's housing quality.

A later panel will discuss data structures for performing operations on data bases. This panel, however, is an introductory panel with representatives from agencies that have data bases which can be used for cartographic purposes. Regardless of the number of people we have on this panel, some excellent examples of data bases for cartographic activities will undoubtedly be omitted.

Orth: My interest in cartographic data bases is geographic names. Names are essential in making our lives intelligible. Actually nothing really exists for man until he has discovered it and named it; futhermore, names are necessary for communication. Though we tend to take them for granted, names are as important in cartography as in our everyday lives. Features of the landscape have common names, but here we're particularly interested in proper names given to specific features. Names are the language of the map; in fact, I consider names to be part of map symbology.

USGS is presently developing a data base on geographic name information to assist the mapping program and the standardization of geographic names throughout the Federal Government. A small staff assists the U.S. Board on Geographic Names (BGN) in standardizing names for Federal usage. The data base now has only about 13,000 names and attendant data in the system. All of the geographic names for Massachusetts (about 12,000) which appear on maps are in the computer. About 1,000 out of 60,000 BGN decisions made since 1890 have been computerized. We are also in the process of negotiating to computerize about 45,000 Alaska names.

We use the GYPSY system stored in the 360/65 IBM computer on mag disks. The name records are somewhat complex, and although this particular system may be slow by some standards, it certainly has proved useful. Each record includes the name and the kind and size of feature; the location in terms of geographical coordinates, State, county, town, section, township, and range; the map where the name appears; any variant names or BGN decisions; and other descriptors such as elevation and population. Information on the origin of the name can be included. Sometimes the record includes geologic names, although they cannot be retrieved separately. The GYPSY system can be queried for any of these elements or any combinations of elements. Of course the system would be a tremendous tool for standardization of geographic names--the goal of BGN.

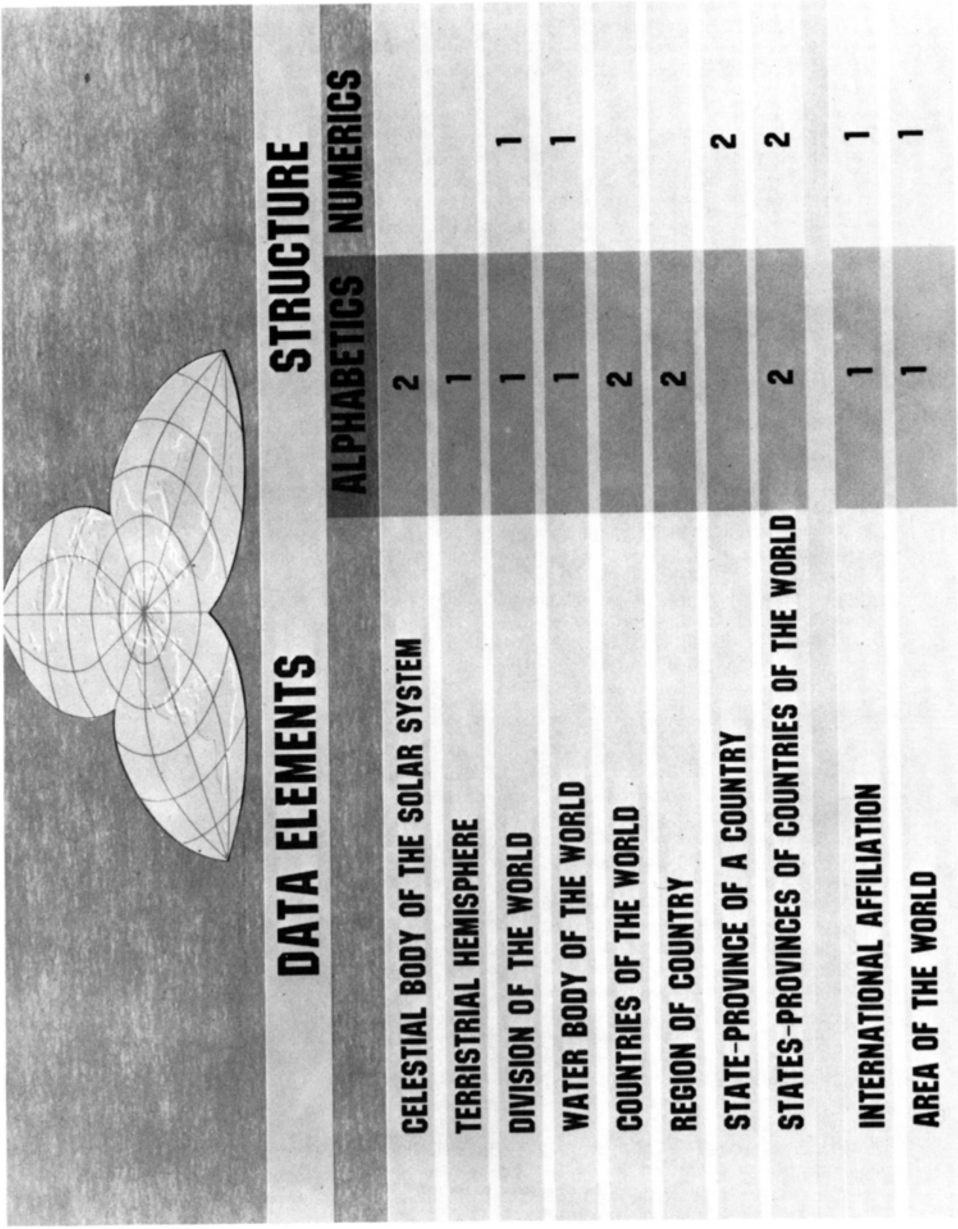
Moritz (PRC Information Sciences Co.): What will be the form of the output--hard-copy lists of names or special digital products? Are you considering the possibility of telecommunications for interfacing with other data bases.

Orth: There is the potential for all of these options. For the time being, hard-copy printout is the only output available. I hope that someday this particular data base can be queried by remote equipment. There is no end to the sophistication that could be achieved.

Serrajian: The Defense Intelligence Agency (DIA) has developed numerous standard data elements, largely in the geopolitical area, for the interchange of information. A data element is the primary building block of information in the data system, whether manual or automated, and DIA develops and maintains geopolitical data elements for all the components of the Department of Defense (DoD). An example of a geopolitical data element is "Countries of the World," which includes items such as United States, United Kingdom, and Canada.

Figure 1 shows examples of geopolitical data elements: First, DoD-approved data elements and second, those currently under consideration. A future project, not pictured, is a hierarchial coding system for the rivers of the world. When the geopolitical data elements were developed, a requirement to identify parts of the solar system surfaced; thus the data element "Celestial Body of the Solar System" is included. "Division of the World" represents continents and major water aggregations, while "Waterbody of the World" identifies regions that fit into major water aggregations. The Federal standard, "Countries, Dependencies, and Areas of Special Sovereignty," is the basis of the data element "Countries of the World." "Region of Country" includes geographic regions below the country level which do not have political boundary significance and is designed to be used in the same field as "Countries of the World." "State-Province of a Country" encompasses the entire political land surface of the Earth below the country level; the first-order political subdivision must be linked to its applicable country code to be unique. "International Affiliations" includes such items as NATO, SEATO, and CENTO. "Areas of the World" represents aggregations of countries, but presently is difficult to standardize as everyone categorizes the world differently.

All geographic names used are those approved by the U.S. Board on Geographic Names. When researching each data element, a technical expert is consulted. For information on countries, states, and provinces, the U.S. Department of State is consulted; for waterbodies of the world, the U.S. Navy Oceanographic Office. Although data elements are developed separately, the system design must be considered. Each data element, except for the data chain "States-Provinces of



DATA ELEMENTS	STRUCTURE	
	ALPHABETICS	NUMERICS
CELESTIAL BODY OF THE SOLAR SYSTEM	2	
TERRISTRIAL HEMISPHERE	1	
DIVISION OF THE WORLD	1	1
WATER BODY OF THE WORLD	1	1
COUNTRIES OF THE WORLD	2	
REGION OF COUNTRY	2	
STATE-PROVINCE OF A COUNTRY		2
STATES-PROVINCES OF COUNTRIES OF THE WORLD	2	2
INTERNATIONAL AFFILIATION	1	1
AREA OF THE WORLD	1	1

Figure 1

Countries of the World," is identified within two characters. The items of each element are uniquely coded, not only within the parent data elements but within the whole group of geopolitical data elements. If necesssary, all elements can be placed in one field, called a shared field.

A hierarchy of geopolitical data elements (fig. 2) is an alternative to the shared-field concept. The five columns represent five fields in a file--A, B, C, D, and E. In hierarchial order the AA column represents two characters within the field. The data element "State-Province of a Country" normally must be linked to a country code for uniqueness; however, if only one country is being dealt with in a file, the country code is unnecessary and can be omitted. When a file deals with more than one country, the country code is also unnecessary if it already appears in another field within the same record. A minor programing effort could link the two separate fields together.

The DIA geopolitical data elements are documented and available through the National Technical Information Service (NTIS) of the U.S. Department of Commerce (fig. 3). A copy of the DIA manual is available for review after this panel. DIA geopolitical standards are presently being proposed as American National Standards. Three standards appeared on the agenda of the International Organization for Standardization which met in Tokyo, Japan, last October. The DIA geopolitical data element manual is revised continually, and as the changes are published, they too are filed with NTIS.

Robe: Unlike my predecessors on this panel, I will discuss a data base that is mainly composed of lines and is plotted. We actually have two such bases: World Data Bank I, simple and quickly generated in 1966, and World Data Bank II, complex but more flexible and more than half completed. Figure 1 illustrates the refinements being incorporated in the newer system.

World Data Bank II, started in August 1971, was conceived as a base for thematic maps. Originally, the plan was to digitize at a larger scale than the desired output scale. Experimenting with some new equipment, however, we digitized one of our more elaborate maps made for another purpose and found that we had the capability to work at 1:1 scale. Since then we have compiled data for 65 percent of the Earth's land area. Very shortly we hope to have digitized all of the world except North America. The data bank now contains about 5 million points; when completed, it will have about 8.5 million points, as compared with 100,000 points in World Data Bank I. The coverage of World Data Bank II was developed as a patchwork (fig. 2), that is, no sheet touching a sheet in progress was begun until the latter was in the bank.

Both of our data banks, as is true of most such banks, were designed to meet our specific needs--primarily bases for thematic overlay data which could be automatically plotted. We wanted flexibility in choosing the output projection and format, independent of the input form. Also we wanted World Data Bank II to have the capability of choosing among several categories of data and of controlling the density of base information. Both data banks can be updated or changed, though not easily.

This portion of a coded worksheet (fig. 3) illustrates the method used to identify the lines in the data bank. (In the original, the information categories were identified by color.) The space problem on the coding sheet reveals one reason why a variety of scales were used.

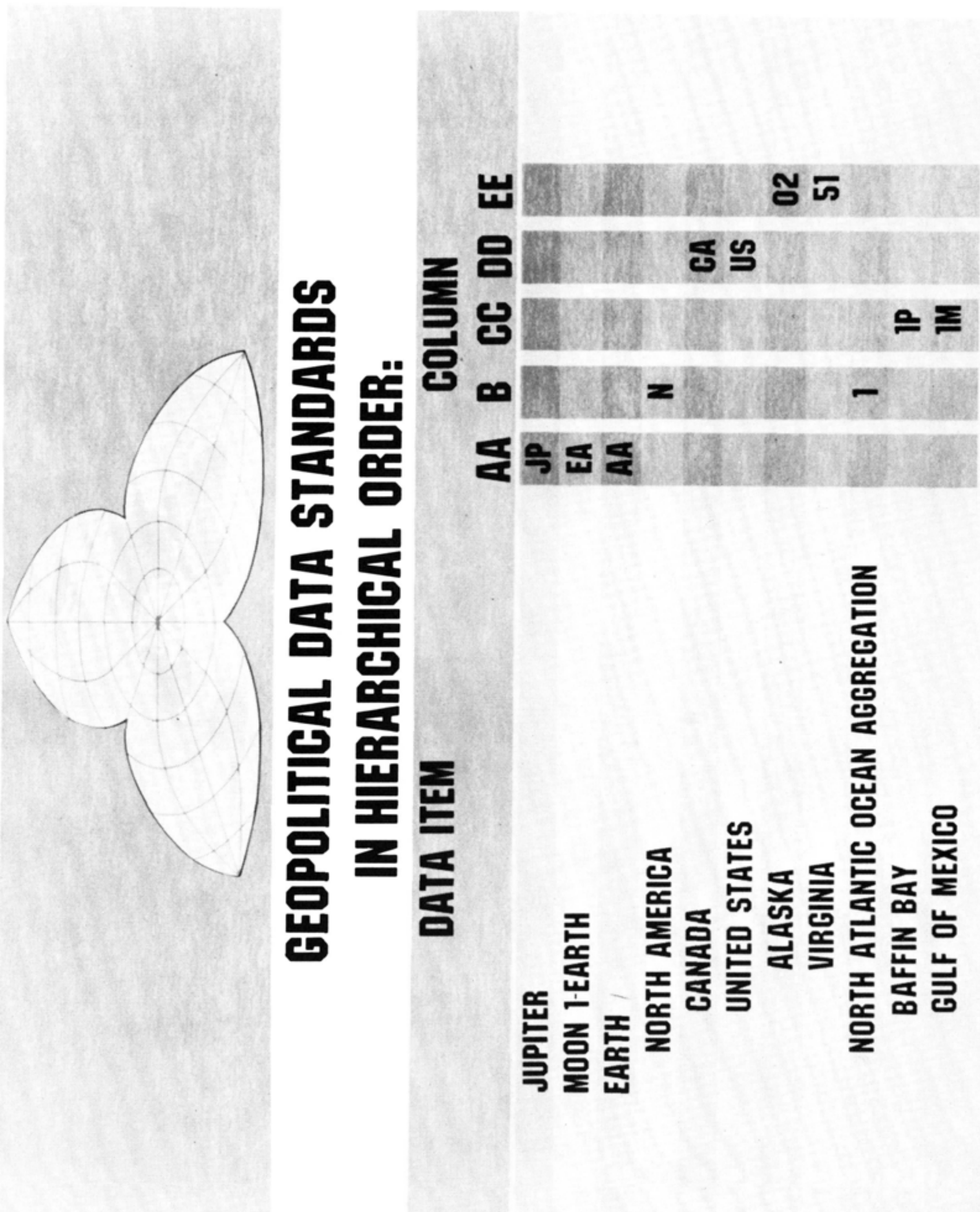


Figure 2
(Serrajian)

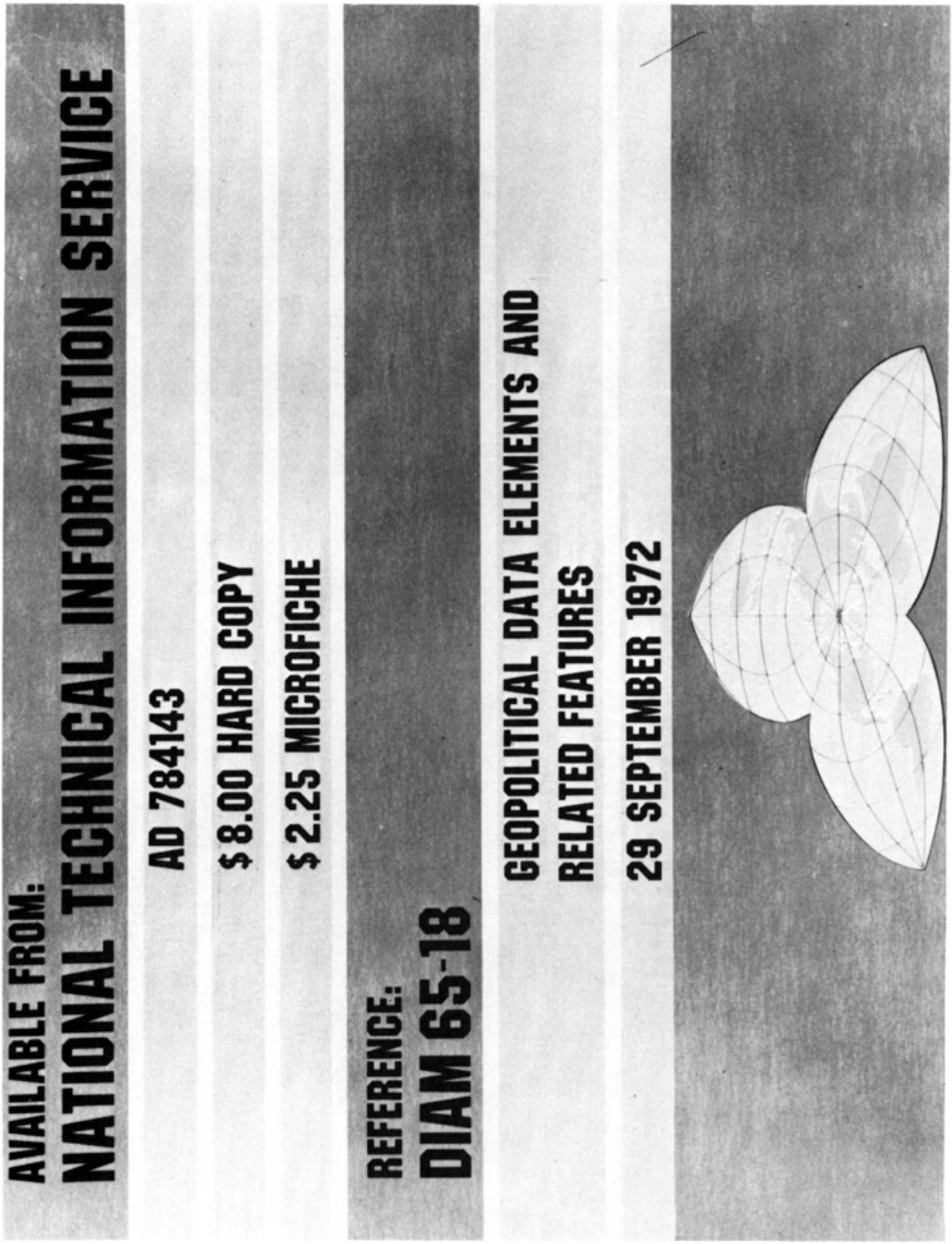
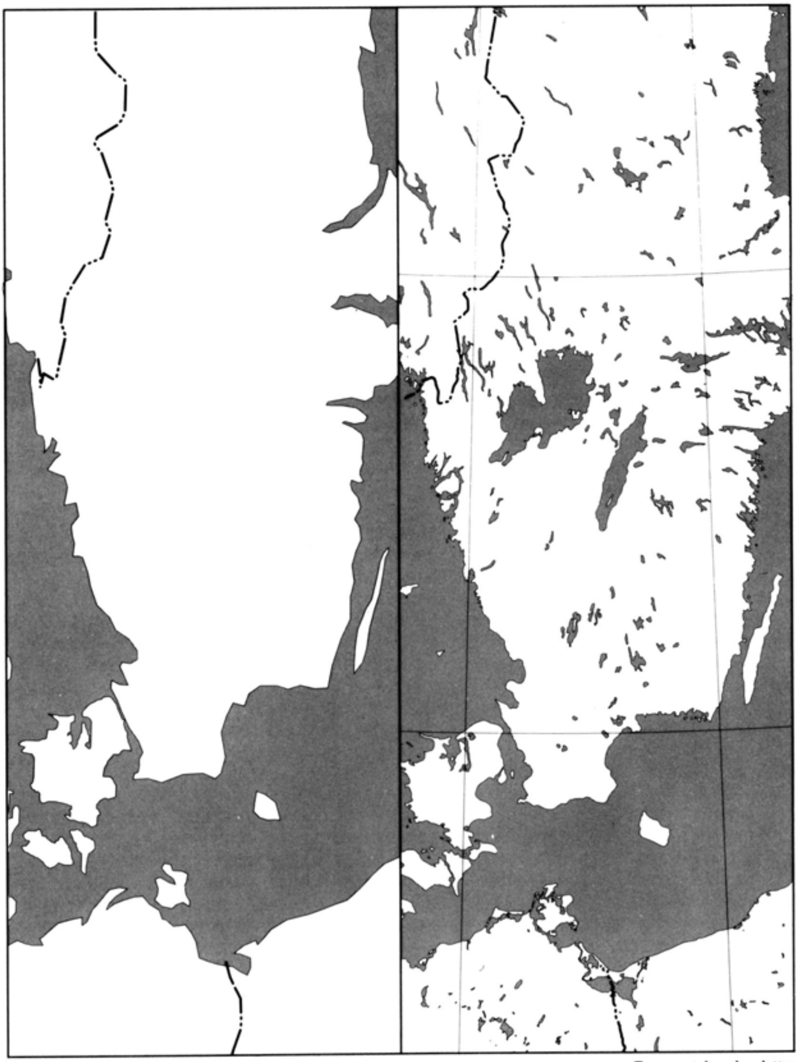


Figure 3
(Serrajian)

WORLD DATA BANKS I & II

Line Character



Tones not done by plotter

Figure 1

WORLD DATA BANK II

Coverage

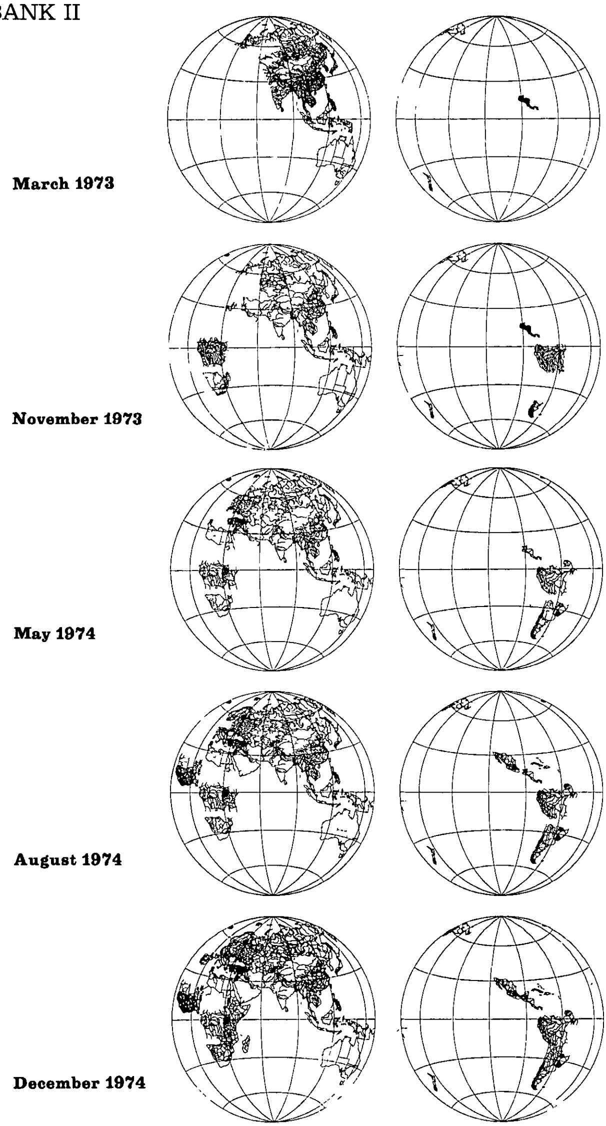


Figure 2

WORLD DATA BANK II

Coded Worksheet

Figure 3

Since each line must have a unique address in the computer, it is coded with a 9 digit number--7 digits for continent, country, and feature and 2 for rank. (On this worksheet some numbers are abbreviated.) Rank governs the amount of detail shown and enables the compiler to regulate the types of base data to suit the thematic overlays.

Figure 4 shows the distribution of input scales over the world. Most of our map has been digitized at 1:3,000,000. A few very small island groups were digitized at a larger scale. Areas such as Antarctica, Australia, New Zealand, and Siberia, where data are very limited, were processed at 1:4,000,000. This scale speeds plotter time and reduces the amount of storage, but if we were beginning again, we might digitize such areas at 1:3,000,000. Incidentally, the omission of Port Phillip Bay, Australia, has been remedied.

Figure 5 shows an area in which four different input scales were used. Although the computer filters out excess points to make the various sections compatible, the area of 1:4,000,000, plotted at scale, is noticeably less intricate. (The slide reproductions shown are approximately 1:5,000,000.) We discovered very early that no plotted base is absolutely flexible. In the slide of Data Bank I the lines became very angular at 1:4,000,000, 3X the input scale. In contrast Data Bank II holds line quality even beyond 6X although anomalies do occur. For instance, double line streams broaden as scale increases while single line streams do not, giving the map a strange appearance.

The various features of the Data Bank are divided into six different files, five of which are shown on this map (fig. 6), roads are not shown. Each file may be used alone or in combination with the others. Not all categories have been compiled for all areas. The coastline, islands, and lakes are grouped on one file; the rivers and canals are on another. In other files reefs are grouped with islands and glaciers with lakes. This grouping was largely a matter of convenience. To illustrate the variety of possibilities ranking provides, take rivers as an example. This file has 11 possible ranks: 5 for permanent streams, 2 for intermittent streams, and 4 for canals.

As well as providing a choice of scales, projections, and features, these data banks may provide a flexibility of orientation. They may enable cartographers to overcome the "north at the top" syndrome that has been with us since the days of Mercator. Although people seem to accept a computer-plotted map with an unusual orientation, they still expect a hand-drawn map to have north at the top.

Rockwell (Department of Community Affairs): You have said a great deal about the content of the data bank, but very little about its structure. Is the data structure in chains of coordinate points or single points that are somehow identified as connected?

Robe: Since I don't deal with the technical part of the bank, could either Edmondson or Angel answer that question?

Angel (CIA): The data structure is in line chains, identified by a seven digit number. With that number we can pull the line chain out, replace it, or replace part of it.

Rockwell: I believe that those seven digits were described as content identifiers. What is the geocoding mechanism? How do we tie these

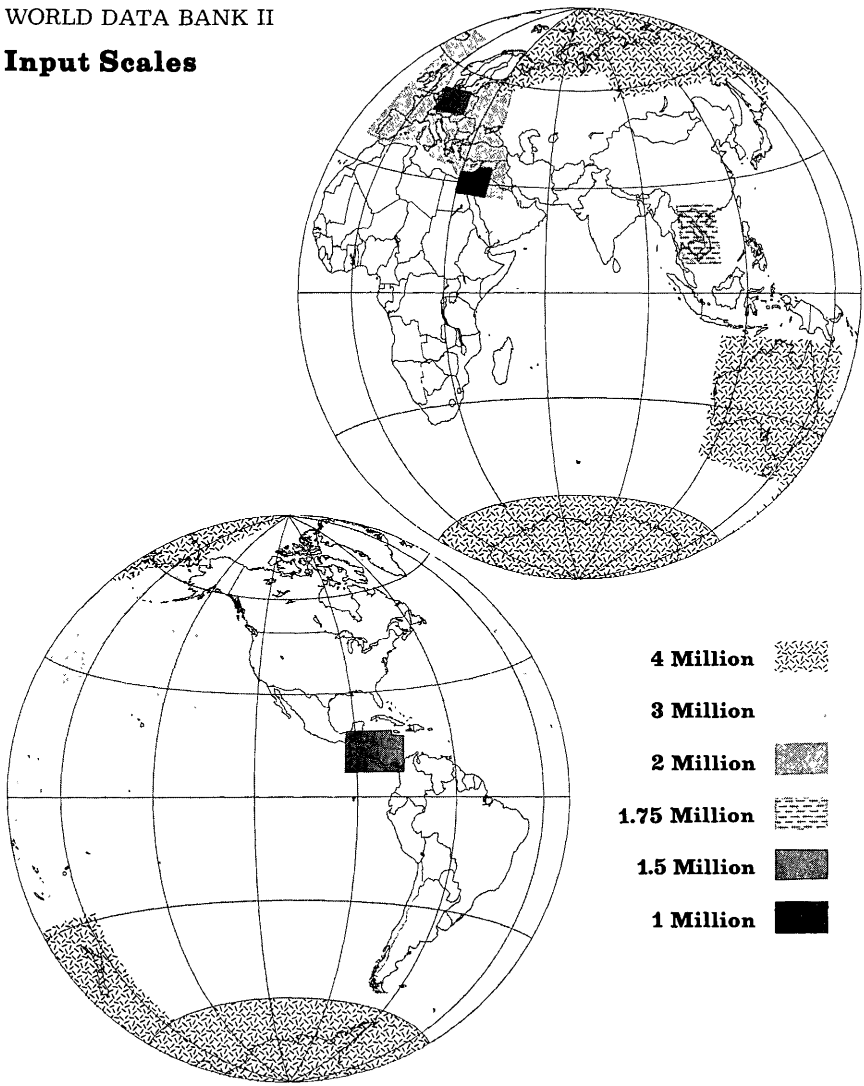
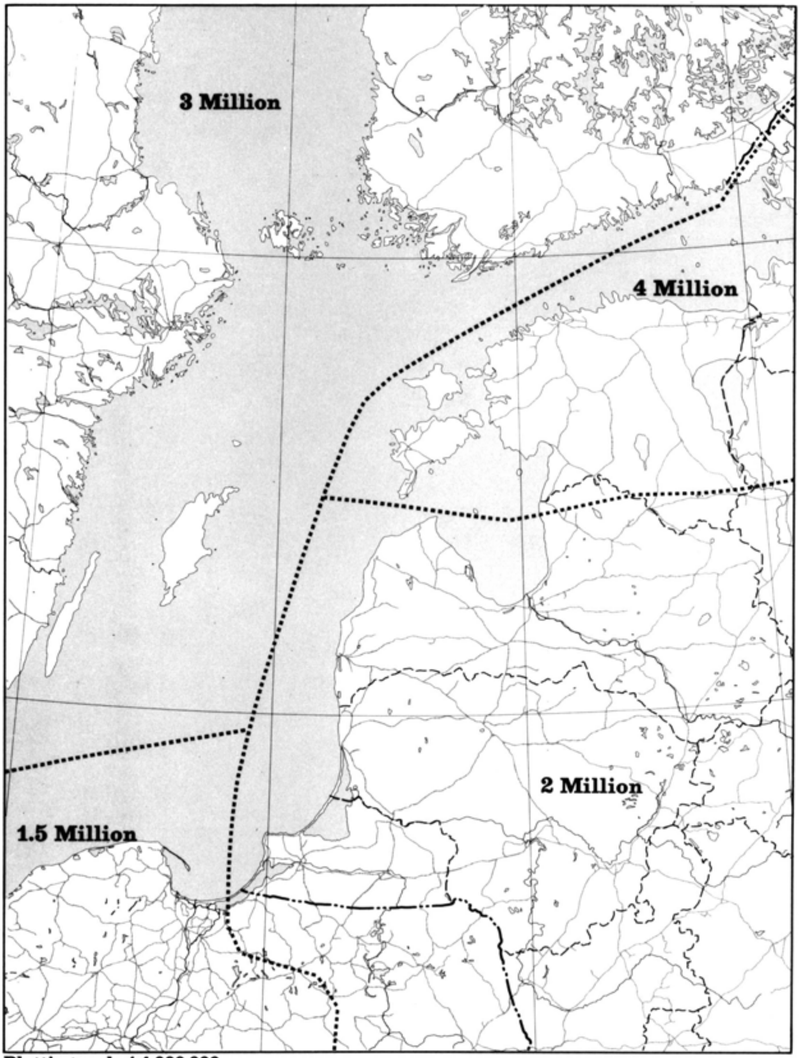

WORLD DATA BANK II
Input Scales
4 Million
3 Million
2 Million
1.75 Million
1.5 Million
1 Million

Figure 4

WORLD DATA BANK II

Input Scale Merge

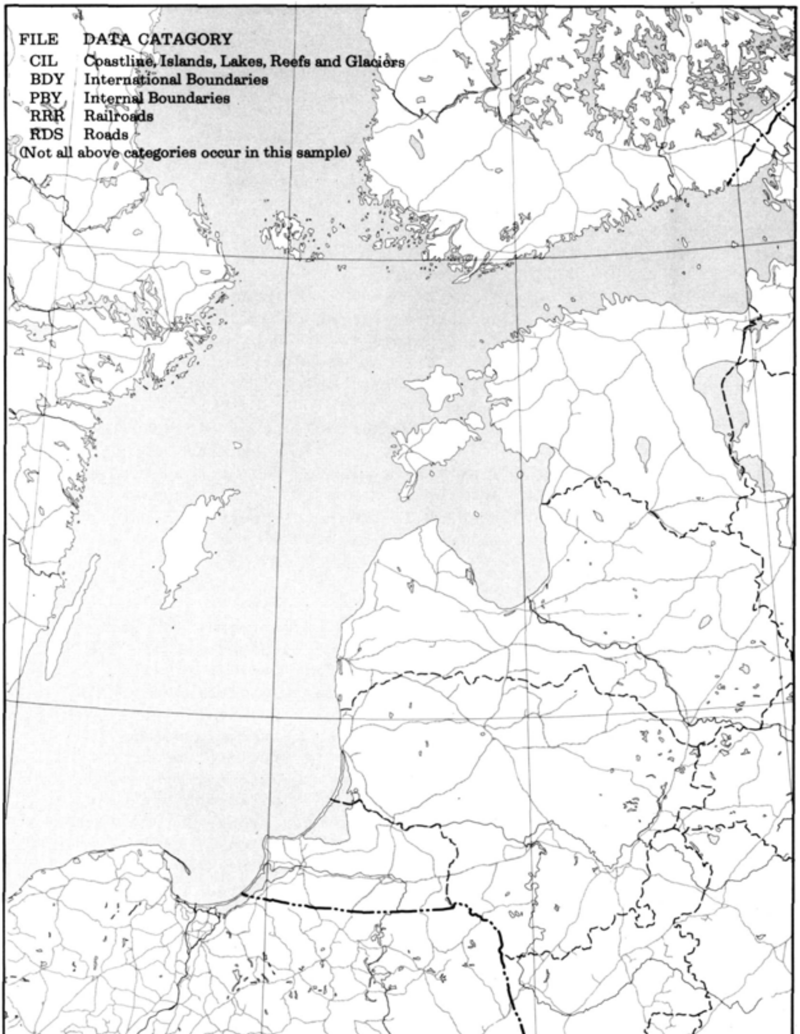


Plotting scale 1:4,000,000

Figure 5

WORLD DATA BANK II

Information Categories



Plotting scale 1:4,000,000

Figure 6

back to UTM?

Angel: Since each line chain is a series of individual points which are in geographic coordinates, they are not tied back in until the actual CAM produces grid. Then the line chains are tied back into the grid at the scale required.

Rhind (Univ. of Durham): World Data Bank II seems very useful. Will it be available through NTIS, and if so when?

Robe: I knew somebody was going to ask that question, and I really can't answer it. Right now World Data Bank II occupies an enormous amount of storage.

Schmidt (U.S. Geological Survey): I believe World Data Bank II will be available in 1975, either through NTIS or NCIC.

Peucker (Simon Fraser Univ.): Working on eastern Australia, we generalized 18,000 points and with an offset of 0.02 in, which is our plotting density, we reduced the points to 1,000 or so. Don't you think that you should reduce the data before distribution? You could reduce your bank to about 500,000 points.

Robe: Undoubtedly we will perform many functions as time goes on because a data bank of this complexity has never existed before. But we are learning as we go. I am sure that much can be done to simplify the use of this data bank; however, I doubt if we will because of the time involved. Once the data bank is complete though, I'm sure there will be people who can and will modify it.

Meyers: Enactment of various Federal and State legislation pertaining to land-use planning will require States to acquire data, conduct land-use inventories, and develop data-handling techniques or processes. Considerable time, effort, and resources are required for any one State to assess its data and inventory needs and to identify feasible alternatives for meeting those needs. Some States have developed highly sophisticated data techniques or centralized processes; others have only single-purpose or decentralized systems or techniques of limited applicability to land-use planning. There is little or no compatibility between States or between State and Federal techniques, processes, or systems. Similarly, there is inadequate and inefficient transfer of knowledge and State experience regarding needs, problems, and feasible solutions. This situation is leading to duplication of effort, massive compilations of unnecessary data, deficiencies in exchange of data due to access problems, and development of data and data-handling techniques that are inadequate for land-use planning and regulation.

The U.S. Department of Interior's Land-Use and Water Planning, in cooperation with the States and under an agreement with the Resources and Land Information program of the Department, has undertaken a study to provide a comprehensive framework from which individual States may assess their information needs and the alternatives for data acquisition and handling. The study has been organized into five tasks:

1.--To survey, evaluate, and report on Federal, State, and key local statutory and U.S. case law related to land-use data and inventory requirements for land-use planning and regulation.

2.--To review data acquisition/handling and land-use inventory systems potentially useful for State land-use planning and regulation; determine the effectiveness of such existing systems; and develop a number of case studies to demonstrate a variety of State experiences.

3.--To identify and demonstrate, through case studies or other means, alternative processes by which States could identify and evaluate their needs for land-use data and inventories.

4.--To identify and evaluate alternative systems for data acquisition/handling and land-use inventories; compare them with existing or pending State and Federal legislative requirements for land-use planning and regulation; and demonstrate, through case studies or other means, how the alternatives may be applied to satisfy State land-use data and inventory needs.

5.--To prepare a handbook for States to use to identify, assess, and develop systems for land-use date acquisition/handling and inventorying.

Workshops are being held at various stages of the project, and a technical steering committee is assisting and reviewing progress.

Results of the study are being documented in several forms: a primer on data handling to support decisionmaking, a handbook on data handling and inventorying primarily for State program personnel, and a series of technical papers ranging from an information systems review to a collection of papers by experts in the field. Publication and distribution of these reports is planned for early fiscal year 1976.