Editing Methods Panel

David W. Rhind, Presiding
University of Durham, U.K.

Frank P. Edmondson
Central Intelligence Agency

George Steinke
CALMA Corp.

Walter M. Anderson
Applicon, Inc.

Barry K. Moritz
PRC Information Sciences Co.

Rhind: Let me introduce the panel members. The first to talk will be Frank Edmondson, a staff cartographer at CIA, with 3 yr of experience working on World Data Bank II tapes. This is an essential way to start; indeed I am going to emphasize the importance of the user end of editing methods. Following Frank, we will have George Steinke of the Calma group, and following him will be Walter Anderson; both of these speakers work for commercial firms and have had long and valuable experience in computer graphics. The afternoon will be closed with a short talk by Barry Moritz. Barry has an extensive background in mini- and maxicomputers and with a wide variety of data handling ranging from communications software to graphics. He will be talking about the PRC project for linear input system, but I asked him to stress the editing of raster data. Initially, he will cover user aspects, then a couple of applications of line editing, and finally, I hope, some consideration of editing raster data.

Anyone who has been involved in digitizing map data should, within a very short time, have their "hands covered with blood." It is almost impossible, in my experience, to digitize data without accepting, given our present technology, that there are going to be a lot of errors in that data. Data is not clean; the user must decide just how dirty he is prepared to have it. I would argue that, in fact, cleanliness of the data must be related to the task, and here, if I may take a stab at fellow cartographers, the standards of data cleanliness are often not dictated by the real requirements and the accuracy of the data, but by the historical connotations. In other words, we have always had nice clean data in graphical form. We want it the same way, even if it doesn't matter, even if this cleanliness is not justified in terms of data accuracy. I believe that we want to have our data only as clean as necessary because it costs more if you have it cleaner than that.

It is amazing how some people still do not see the need for an editing capability: One case in the U.K. recently came to light where someone who went to work for a large local authority opened the door and nearly 1,000 paper tapes fell on him. These had been digitized over the preceding 3 yr but were completely useless and subsequently burned because the people had no decent editing capability.

Another point is that I think that a real distinction has got to be made between finding errors and fixing them. You may very well not use the same techniques for both functions. In "finding" and in "fixing," you have to look at the two aspects of the data--its geometry (and/or the topology) and its attributes. How do we check data for errors? We can apply visual, logical, consistency, or sample checks. The first has always, to my knowledge, been carried out by simple "eyeball" comparison of plotted output and the original digitized document, although the possibility of comparing these in stroboscopic light might usefully be investigated. The greatest danger about visual checking is that you are dependent on the checker, and once he has checked a document, he will be convinced that it contains no errors. Certain logical tests, such as "is this point within the map bounds?," can easily be carried out by machine as can some types of consistency checks on redundant data, such as "is the sum of the depths to various geological strata equal to the supplied value for its total depth?" In current cartographic work, the onus of digitizing (and processing) the input data twice has meant that consistency checking to date has been done merely on attribute (code) data or on selected (e.g., line start and end) points from the geometry files.

Another interesting point is the frequency of editing: Do or can we do it all at one stage or must it be an interative process? A final point which might promote discussion is the cost aspect. Interactive digitizing using cathode ray tubes is seductive, very pretty, good fun, and exciting. But is it anywhere near as economic as offline working? There are a couple of commercial companies in U.K. at the present time who make no use of interactive digitizing because they find that a standard, very simple offline digitizing process, correcting the errors afterwards on a test plot, is still much cheaper than running a minicomputer linked to a standard digitizing table. In discussing costs, I suppose we are all aware how critical the technological aspect is: in addition to changing digitizing costs, the technology that we were hearing about this morning from i/o Metrics and others may present us with rather different editing problems from those we have had in the past.

Edmondson: I am one of those Government users that Dr. Boyle mentioned this morning. In the few minutes that I have, I will describe some of the editing techniques that have been used at CIA to produce a small-scale world data bank at an average input scale of 1:3,000,000. Coastlines, rivers, boundaries, and selected transportation have been input to the data bank for over 80 percent of the land area of the world. The bank now totals more than 5-million latitude/longitude points. Our objective is to create a world data bank that may be used to make small-scale maps of any area of the world, using 16 different projections at various scales and sizes. To achieve this flexibility, data are stored as one world data bank in latitude/longitude coordinates. Figures 1 and 2 are samples of maps made from the data bank. The line work is plotted, and the type and tones are added manually by cartographic technicians.

Since we do not have an interactive digitizer, all of the raw data are recorded, including any mistakes that may occur. Experience showed that most of the time the operator is aware of his digitizing

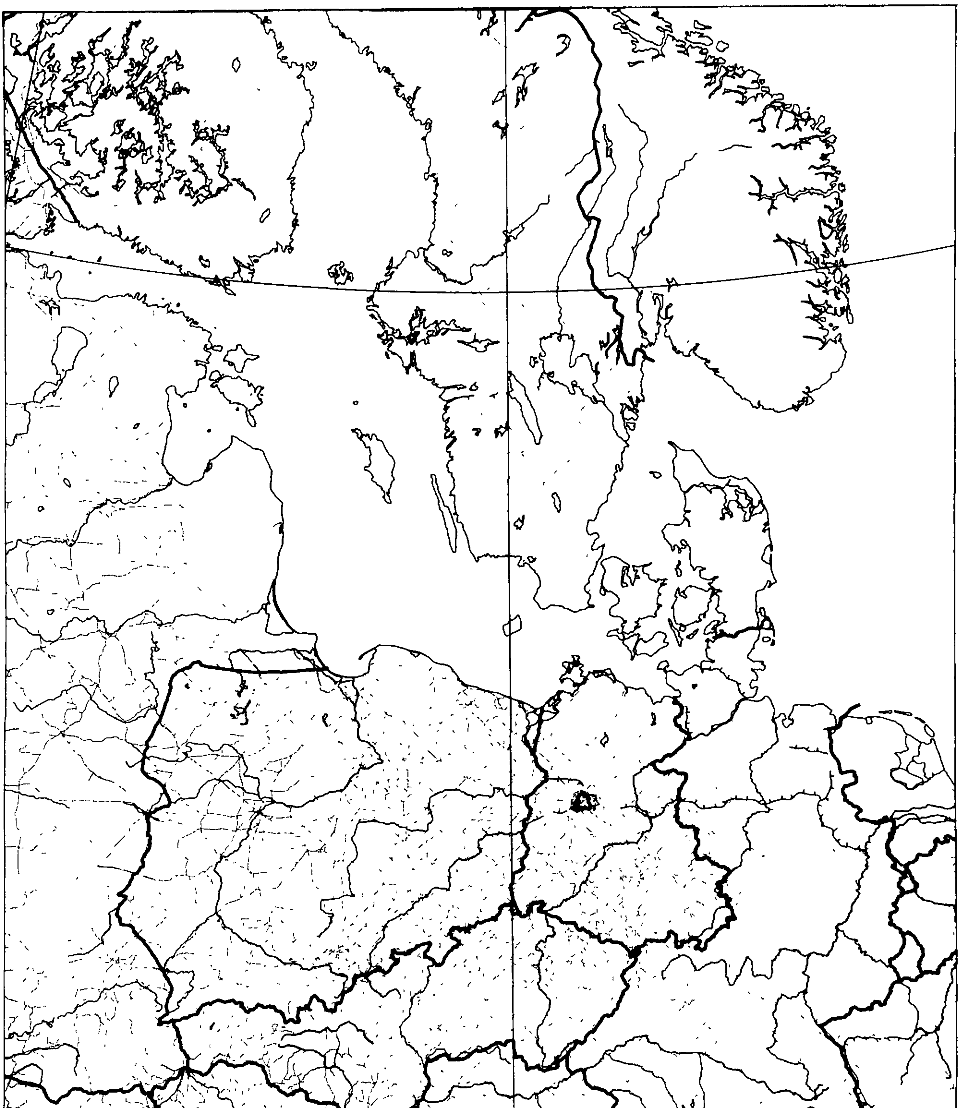


Figure 1

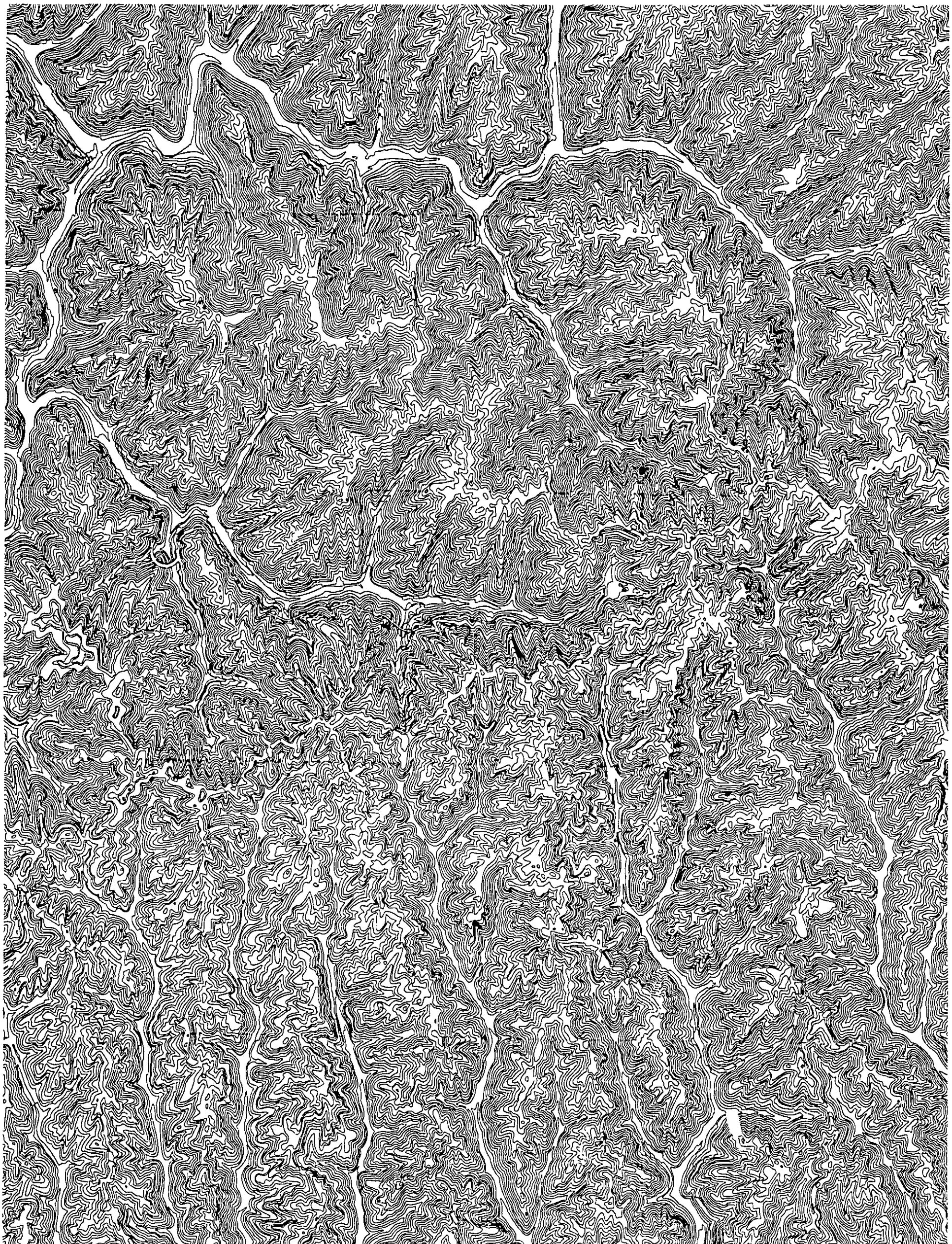


Figure 2

mistakes. For instance, when he follows the wrong line, gets too far off the line he is trying to follow, or does not join a river to a coastline correctly, he realizes his error, but the incorrect data have already been recorded. Thus options were added to our rotate-and-scale (R&S) program, which processes the raw digitizer data so that the operator can do some editing while digitizing. He may flag the incorrect line with the appropriate code so that it will be deleted completely, or he may flag for a partial deletion and redo only the portion of the line that is incorrect. The partial deletion and the automatic closing of the line save a lot of interactive editing time. The automatic closing of lakes and islands by this program also speeds up the digitizing since the operator does not have to try for a perfect closure. The R&S program also eliminates duplicate points and creates new points between 2 original points so that lines can be digitized point to point rather than in the stream mode.

After processing by the R&S program, the data are plotted on film and compared to the original map compilation. Corrections are then made interactively on the IBM 2250 which is online to a 360/65. For the most part, the digitized data are accurate enough but do require some cleaning, as was mentioned this morning, for the correct appearance. Gardner Hill of the Ordinance Survey refers to this as "cosmetic cartography". A point or two may need to be added or deleted from the end of a line to make a smooth intersection. Boundaries that follow rivers will not look correct unless the points in the rivers and the boundary are identical. We digitize them once and store them on separate files. All of these corrections are done interactively on the screen, so that you can see they are done correctly.

Since data are added to the bank one sheet at a time, it is absolutely essential that new lines join correctly with those already in the data bank. This is accomplished on the IBM 2250 by displaying the adjoining lines already in the bank and adding the new data to join them correctly. All of this is done in x-y coordinates on the IBM 2250. The final step is to convert this to geographic coordinates and file it on the disk.

In summary, I would say that with a skilled operator and the R&S program cleaning the data to some extent, the number of corrections that have to be made interactively on the IBM 2250 is usually quite small.

Wilkie (Dynamap Ltd.): What percentage of the actual digitizing time would be used in cleaning the data?

Edmondson: Approximately the same time is used--slightly less editing time than digitizing time.

Niedermair (National Ocean Survey): I find that these corrections tend to eliminate the useless points. How do you do that?

Edmondson: The R&S program throws out all duplicate points. This is the raw digitized data that are run through the program which will eliminate duplicate points, if there are any, within the same geographic line or feature. We just couldn't get off the ground with manual methods; I hope to never go back to graph paper and plotting points.

Falke (National Ocean Survey): I would like to know what the resolution of the information you are storing is, in terms of 0.001 degree or 0.0001 degree.

Edmondson: We digitize at approximately 1:3,000,000 scale, maybe a little larger or smaller, and sample a point about every 0.5 mi, if you are talking about distance on the ground.

Bonham (Chevron Oil Field Research): Do you have the graphic design feature on your IBM 2250 for the input of the edited data that allows you to screen data in on the tube; or is it point by point?

Edmondson: It is point by point, and the screen is limited to 1,000 points at any one time.

Steinke: Calma has recently introduced a new product in its line of cartographic equipment--the Calmagraphic Interactive System. The new system is based on a data base of 48-bit words and was designed to allow 3 or 4 main functions: (1) record input of cartographic data, (2) easy editing of this data online interactively, (3) a certain amount of computation, such as volume, area, and smoothing; and (4) the option to produce a tape of this data for processing on a large computer system. We will interface virtually any computer system in the world today.

The hardware is based around a Nova minicomputer. This is a completely offline system with a disk drive and a tape drive for the central section. We have input "stations" which consist of an interactive digitizer, a CRT, and a keyboard. An edit station provides 11- by 11-in interactive area that you can digitize and edit from, as well as a CRT and another keyboard for entering commands.

Before I get into the editing, I would like to give you a quick rundown of our data structure so you will be familiar with the terms. The first way that we identify anything is a drawing name that can be up to 15 characters long. Immediately after this, as a next classification of data in our tree-structured data base, we have a set of overlays which are numbered from 1 to 32. Below that in the tree structure, we have domains. We may have up to 32 levels of domains; at any one level, we have virtually unlimited domains available to us. For example, if you had a map, you could be talking about Overlay 1, and your drawing name could be Map 247; then at your first domain level you could have things like topography, improvements, and property lines. Under improvements in your first domain structure, you could have public, private, and commercial improvements. These can be broken down into a third domain level, such as streets, highways, buildings, and bridges under public. This can carry on for 32 levels. Attached to any group are overlays; attached to any overlay or domain you can have a group. A group is just a collection of related data items. You can have 1 to 65,000 groups, and a group of 1 to 65,000 could be attached to either a domain or an overlay.

Another item in the data base is an information block. The information block contains either numeric data, alphanumeric data, or computations such as area or volume. There can be one or more information blocks associated with any group, domain, or overlay. As another way of representing information, we also have associated a line-type table with an overlay. A line-type table contains from 1 to 32 different descriptors that tell you how pieces of information will be displayed. If you have a set of points that represent a line, they can be represented on your output device as a line that is 5 mil or 10 mil wide. You can also associate a different line-type table with each overlay in the system.

If you have an interactive system, first you need to be able to view the data that you have entered into the system. Several commands exist for this. We are able to define a window on our CRT, magnify the window, shift the window, and contract the window. We are also able to specify what data will be viewed. We can say we want Overlay 1 and Overlay 3; then we can also say we want to look at Overlay 1 and only a particular number of domains. We can also go down into the group level and specify which groups we would like to see.

I like to think of our edit commands as three types. In one type we can edit the data in the information block, and we can add, delete, rename, or add to the data blocks. We can also alter the information in the data blocks. At any time after you have entered the information--concurrent with the digitizing operation--you can stop and call the information from a previous drawing back into working storage. You can evoke the editing commands then, which at this point are usually limited to a particular overlay.

If you are going to modify some data, first thing is to identify it. One way of doing this is to define a window--again, we have a wide range of commands to pick out a particular window. After the window is picked out, we can specify that the item to be edited is inside, outside, or partially in the window. We also have the capability to search for a particular group by positions. Once we have identified a group, we are able to operate on that group in several ways. We can delete the group, copy the group to a new location, translate the group, or rotate the group.

As a further feature, we are able to operate on the individual items in the group. We can add items to a group; if you have missed a point, you can add a data point in. You can delete spurious data points, move points, and reshape lines. If you have a segment of a line that is badly digitized, you can erase and redigitize that segment. You are also able to do text editing on the items in the group. You can insert or delete symbols, which are items of the group. As a final feature, on the particular items in the group, you can change the line types for a different representation. Two other editing commands that I want to mention really quickly are: (1) you can always go back and rename your drawing, and (2) if you have several drawings that you want to merge into a large drawing, we have a merge feature for this particular data base.

We have an interactive data base that we are able to view in various ways, and we have a large number of commands that allow us to edit either the information block, the entire group, or the individual elements of the group.

Wilkie: When you erase data, do you in fact delete that data from your file, or do you tag it and retain the old data?

Steinke: The data are tagged as deleted, that is, the data have an identifier on them that says they are deleted. If at some point you want to go back and get that data, there is an unerased command in the system that allows you to go back and re-retrieve it if need be.

Anderson: Applicon has been in the interactive brackets business, producing minicomputer graphic systems for about 5 yr; it was just about 2 yr ago that we began to get inquiries about applying our system to mapping. Previously, most of our work was done with integrated circuit mass and printed circuit boards where it was clearly very cost effective and where we had quite a bit of background and expertise.

In talking to people about mapping applications, it became very clear to us that while we didn't have most of the software required, we did have many of the interactive techniques and graphic concepts that were clearly applicable. Layering in a printed circuit board is not much different than layering in a map. The mechanical line types are not too much different than the symbology necessary for railroads and rivers, etc. The concepts of windowing on a CRT over a map are ones that are probably more satisfying than windowing over a IC mask or a printed circuit board where it seems much less natural. Most importantly, we have developed some tools for multiprogram operation of a minicomputer and large disk systems.

(Slide) (Editor's note: Slides not available for publication.) The first slide shows a multiactivity system that was displayed at the ASP/ACSM Fall Convention in 1973. On the right is a free-cursor digitizer, a Bendix digitizer in this case. In the center is our Applicon editing console, and on the far left is a Xynetics plotter.

(Slide) This slide shows the central operating terminal with the stylus in the front. The operator works with the stylus, viewing the CRT. All of his motions on the tablet are mirrored in the CRT, which is quite simple to use. It does not take very long to pick up the editing techniques of selecting, moving, adding, windowing in, and blowing up small areas to very large drawings.

(Slide) The next few slides will show some of the possible graphics terminals that you might have. This one is called our tabletizer. It is a large-surface tablet which can be used with the two styluses that are shown at the top--one is nonmarking and one is marking. The display on the right is simply an alphanumeric display (no graphics there).

(Slide) This slide shows a similar large-surface tabletizer which is somewhat larger and backlit. On the right is a 19-in CRT, and on the left is a digitizer cursor which can also be used in lieu of the stylus.

(Slide) This slide shows an x-y cursor readout light in which the operator can read out the ground units for the internal Applicon control units or can read out the raw units as the digitizer is measuring them.

There are several possibilities for what one might want for a cursor in digitizing. The tabletizer stylus was shown earlier; in the next couple of slides we will show a couple of other possibilities. (Slide) This is a locked-arm type of thing where it is locked into a vertical or horizontal position. It is probably not too convenient for entering or editing map data. (Slide) This slide shows a free cursor style which is probably more convenient for editing.

(Slide) This slide shows another type of input terminal which is to input and edit from a coincident plotter on top of the digitizer. One can digitize a contour, for example, and immediately plot it back to see how faithfully that line has been captured as well as to mark off what has been digitized.

(Slide) The ETH system is probably the first one that we have sold that is being used exclusively for interactive mapping. Others of our systems and service bureaus have been used partially for mapping as well as for diagraming and similar applications. The ETH system consists of a deck PDP-1140 with 28 k rapport, a tablet editing console not shown here, the Graticon digitizer, the 19-in CRT, a Calcomp 936 for check plotting, and various periferals such as magnetic tape, card printer, line reader, and high-speed teletype supporting the graphics activities. Offline to the system is a Ferrante master plotter to produce the final artwork.

(Slide) This slide shows a closeup of the digitizer; notice a couple of other possibilities for digitizer cursors. The one being held is the simple crosshair type with a magnifier attached to it. To the right is a scribing cursor that can be used on scribecoat to mark exactly what is being captured. Towards the bottom of the picture is a function keyboard for quick re-entering commands to the system. These are user definable.

A number of special editing features in the software were developed to make the ETH system productive for map input and editing. First of all, a photogrammetric coordinate transformation was added to the software to relate up to 4 points on the digitizer surface to 4 points on the ground. Ground coordinates are accepted and displayed with up to 8 digits of precision relative to a user defined map origin. Retransformation and rescaling software are provided, so that over a limited range one can change the scale of the input or distort the data for plotting back on a distorted output medium. Continuous digitizing with time and distance sampling is provided along with precise clipping of map cells which the next slide will show.

As an operator prepares to do a large map by breaking it up into many smaller maps and digitizing them one by one, he may want to digitize over the edge of the map just to make sure that he has gotten to the edge of the map cell. The map clipping software provides the ability to clip the cells precisely at the edge and also provides data that can be carried forward to aid in the production of adjacent map cells which will meet when they are later joined together.

(Slide) This slide shows those selected areas, near the edges that are removed from the map. Segmentation software is provided to merge together separately digitized map cells into a continuum of maps. Again, to aid the operator in digitizing, there is close-looped snapping of contours and also snapping to existing lines in the drawing so that lines can be made to meet without the operator having to be very precise about it.

Moderate changes in scale either up or down may result in a data base with too many or too few points. Two programs are provided: one to filter the data, to thin out a data base with too many points when trying to generalize to a higher map scale; and the second to deal with map data that are too sparsely digitized (either I did not digitize it continuously or I set the parameters wrong when I was digitizing it). The next few slides demonstrate this spline fitting.

The lights that are provided are clustering software to allow points or contours that have been digitized at different times to be merged together in such a way that when they are plotted, they will plot one after the other in the same direction. The Ferrante formatting software is part of what is provided along with software for getting data into and out of the system on cards and magnetic tape. On magnetic tape, we can go off to a larger scale computer to do processing and other data base preparation. You may want to punch cards offline to enter known ground points into the data base with components of a certain type.

The ETH system will be installed next month in Zurich. We are looking forward to working with ETH and to determining how effective the software tools will be for interactive cartographic preparation.

Chrisman (Harvard Univ.): It seems to me that you spent a little bit more time on how you were going to add more points to a cartographic file. How do you proceed in deleting points?

Anderson: Points are deleted by means of an algorithm that simply takes a starting point on a contour and places a distance-window around that point. All the points that follow on the contour (that lie within the window) are deleted. Once a point is found outside that filtering window, it then becomes the center of a new filtering window, and you pass on from one polygon to another until the end of the polygon is reached. The first and last points of the continuously digitized polygon are always retained. This was a first approach that ETH felt would probably work. The program is modular, so that as the system is used, modifications will be a simple matter.

Moritz: The subjects I'm going to discuss are from the point of view of hardware related to editing and cover experience that the Information Sciences Company has had with the Rome Air Development Center (RADC) in Rome, New York. The company has been active in cartographic work at RADC for about 5 yr. I would like to describe a few of many different projects associated with the Experimental Cartographic Facility (ECF) at RADC.

The first project I will discuss is also the most recent--the Lineal Input System (LIS). It is clearly an automated cartographic process, and we accentuate, error free. The basic functional objectives of LIS are: (1) conversion of analog cartographic source material to error-free digital data cells (files) for subsequent automated cartographic processes, and (2) maintenance of a digital cartographic data base through timely updates based upon new materials. The second objective, we feel, is just as important as the first. The existing LIS is a manual digitizing entry system which we all know is time consuming, laborious, and prone to error. However, the existing system is also very cognizant of the fact that a human being is still the best representative for deciding what an error is and how to remove it. When building LIS, it was deemed extremely important to not only produce the data but also to be able to get it in any arbitrary coordinate frame (such as the transverse Mercator and photographic frame), to get pieces of it, to correct small pieces of it, and to put back into the geographic cartographic data base that information which would prove useful for a variety purposes. Therefore, LIS maintains data; it doesn't just create it.

The goals of LIS are:

- -Product-independent format
- -Use of a complete, flexible classification hierarachy
- -Rapid response to user requirements
- -Direct support to <u>production</u> (not R&D)
- -Enhancement--not elimination--of human participation
- -Reduction or elimination of duplication.

A product-independent format is one in which the data base information is not specific to any scale, projection, or other parameter of specific products, but which provides information for generating a whole range of products. The classification hierarchy is based on English cartographic classification terms. Since cartography is closely allied with the presentation of important information within human understandability, these goals are considered important as they enable the interaction of man with data in a variety of ways best suited to the data being presented. The cartographic data cell or data base is based on geographic coordinates. LIS has been installed at both Defense Mapping Agency Aerospace Center and Hydrographic Center and accepted for production.

LIS is depicted in figure 1. The goal was cost effectiveness which, compared to the dollar values considered by the small user, can still be rather high when the Federal Government is involved. LIS represents an initial investment of about $400,000 to $500,000 in hardware. Although the initial system only had 5 stations, it is expandable to 10 stations. Each station is represented by one IMLAC PDS-1G and one Instronics Gradicon digitizer; the cost of these two devices plus assorted small items is under $30,000. Each station is capable of both digitization and interactive editing, as well as interactive erasure and error correction during digitization.

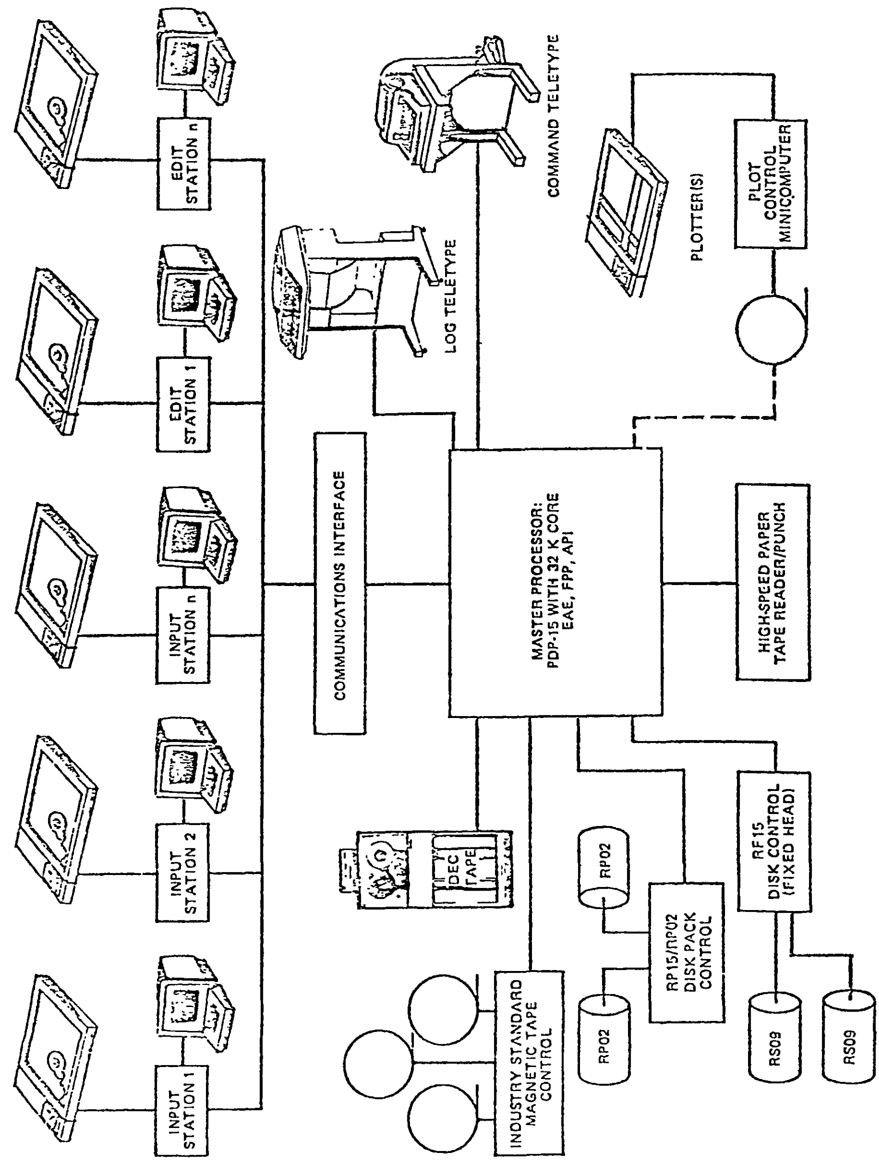

INPUT STATION 1
INPUT STATION 2
INPUT STATION n
EDIT STATION 1
EDIT STATION n
COMMUNICATIONS INTERFACE
LOG TELETYPE
COMMAND TELETYPE
PLOTTER(S)
PLOT CONTROL MINICOMPUTER
MASTER PROCESSOR: PDP-15 WITH 32 K CORE EAE, FPP, API
HIGH-SPEED PAPER TAPE READER/PUNCH
DEC TAPE
INDUSTRY STANDARD MAGNETIC TAPE CONTROL
RP02
RP02
RP15/RP02 DISK PACK CONTROL
RF15 DISK CONTROL (FIXED HEAD)
RS09
RS09

Figure 1

The major cost of modern minicomputer systems is, of course, in the peripherals, which are shared by all the LIS work stations; thus the system becomes even more cost effective as it grows. Every device depicted, with the exception of the communication interface, is an off-the-shelf item. All the interactive work is done at the work station in conjunction with a refresh-type CRT, keyboards, and the Gradicon digitizer. The CRT's are 15 to 17 in with 1,024-point resolution on each axis.

The work station configuration is shown in figure 2. The CRT is mounted on the digitizer controller, which is on wheels for moving with the operator. The digitizer cursor and the keyboards are free floating so that they can be placed anywhere on the table. The PDS-1G minicomputer could actually be installed under the table or in a variety of other places. The digitizing table is capable of resolutions on the order of 10 µm, although in practice the resolutions were never chosen exceeding 20 µm. With resolutions of 20 µm and sampling distances of 40 µm, smoothness of even the smallest radius curves was guaranteed, and class A cartographic standards were more than met.

However this is a session on editing, and this presentation is partly a discussion of how LIS provides the cartographer; in an interactive fashion with a highly effective, precision editing mechanism as good as the data being captured, i.e., 20-µm resolution and 40-µm sampling distance. These distances are smaller than the eye can resolve without optical aids. Thus if the high standards of data capture are to be maintained, the editing system must provide significant help to the cartographer.

Figure 3 illustrates the basic editing functions that are involved in correcting digitization errors, both as-you-go and later. Six types of edit are shown, although in reality only 5 basic operations are involved:

- -End-segment modify (encompasses extend-segment and delete-segment operations)
- -Mid-segment modify (or modify segment)
- -Joining features
- -Deleting features
- -Adding features

As an example, you will now go through the motions of a mid-segment modify operation. Figure 4 is a problem case where a coastline has had a landslide; the solid line is the coastline before the landslide, and the dotted line is the coastline after the landslide. The slide is labeled "Mid-Segment Modify" as it appears on the work-station CRT. The actual image size on the screen can be visualized by assuming a 17-in diagonal to the image.

First (fig. 5), the cartographer moves the digitizer cursor to a position in proximity to the point where he wishes to begin editing. He signals LIS his position by depressing one of the buttons on the

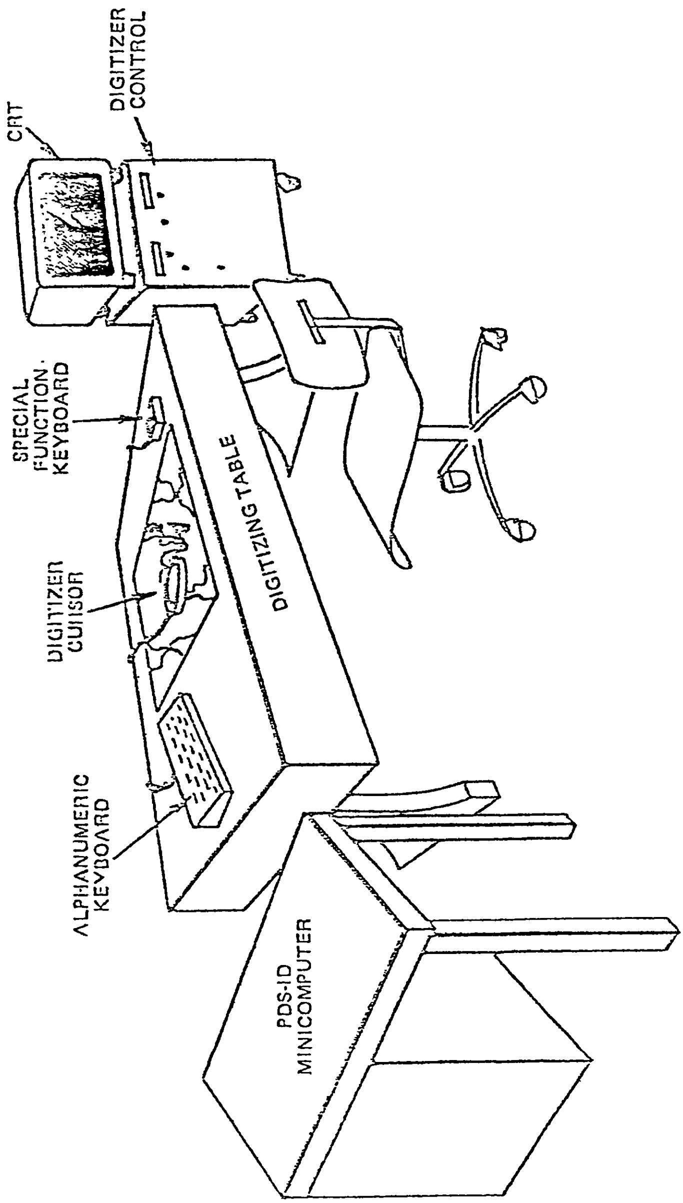

CRT
DIGITIZER CONTROL
SPECIAL FUNCTION KEYBOARD
DIGITIZER CURSOR
DIGITIZING TABLE
ALPHANUMERIC KEYBOARD
PDS-ID MINICOMPUTER

Figure 2

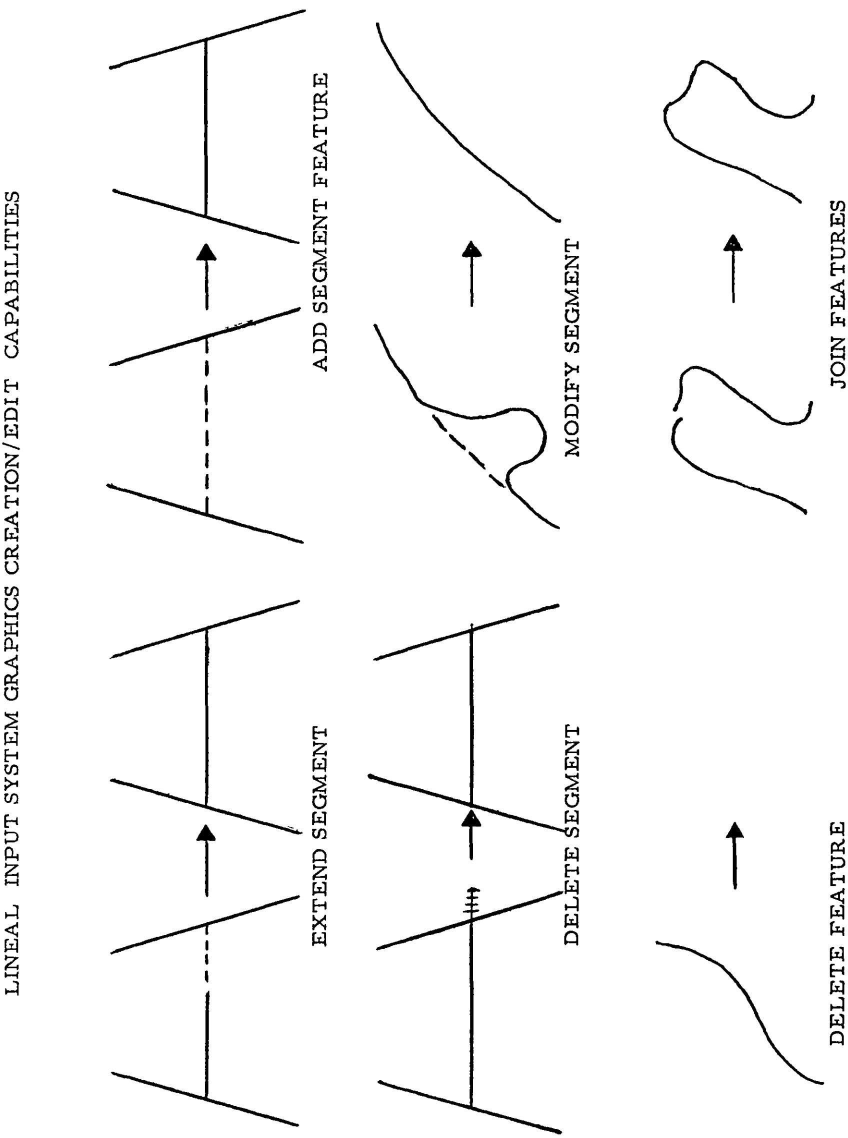
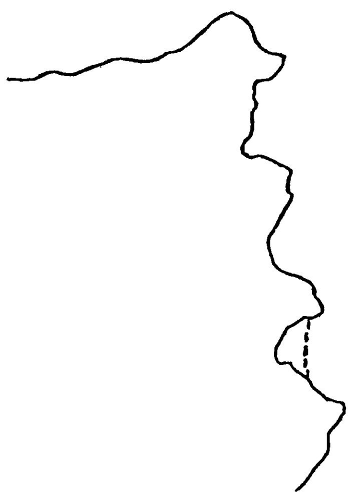
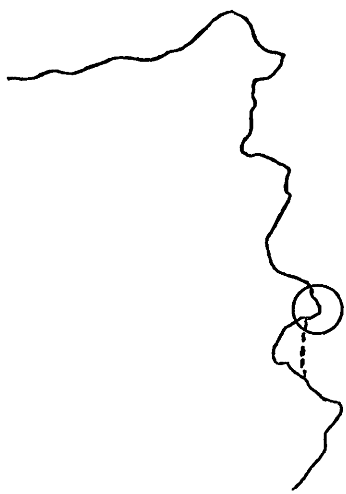

LINEAL INPUT SYSTEM GRAPHICS CREATION/EDIT CAPABILITIES
EXTEND SEGMENT
ADD SEGMENT FEATURE
DELETE SEGMENT
MODIFY SEGMENT
DELETE FEATURE
JOIN FEATURES

Figure 3



Mid Segment Modify

Figure 4



Mid Segment Modify

Figure 5

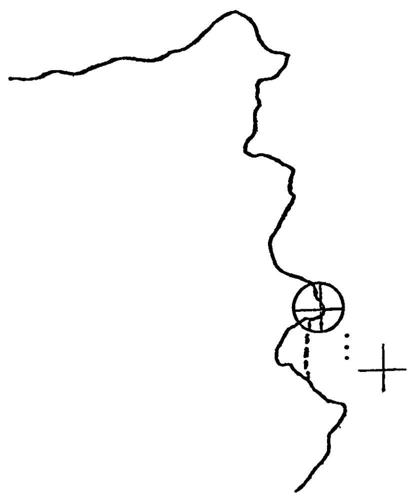
cursor, and the system responds by circling the specified point, or nearest feature point, on the screen. Thus the cursor does not have to be positioned exactly on the edit point, but need only come within eyeball precision. At this point, the cartographer accepts or rejects the circle. If rejected, the cartographer can reposition the digitizing cursor to a new location and get a new circle. Eventually the cartographer accepts the circle location and presses another button.

The system then responds by placing a crosshair on the screen (fig. 6) which is linked to the position of the digitizer stylus, that is, moving the stylus moves the crosshair on the screen. Now the critical point: How does the cartographer position the stylus so that the resulting edit will not have any slope discontinuity or jaggedness? In LIS, the cartographer places the crosshairs in the circle; gunsighting the stylus, a smooth, continuous line or join at the edit point is guaranteed. In practice, the cartographer can place the stylus within 2 resolution units of the actual feature position; with this closure, a special LIS software routine makes a smooth join with the old feature.

Figure 7 shows the result. The cartographer has locked onto the old feature and drawn what is now the solid line. The stylus is brought close to the other end of the edit and the CRT verifies that the smoothing algorithm can indeed be used. Upon seeing the results, the cartographer then either accepts or rejects the edit and repeats the process until satisfied.

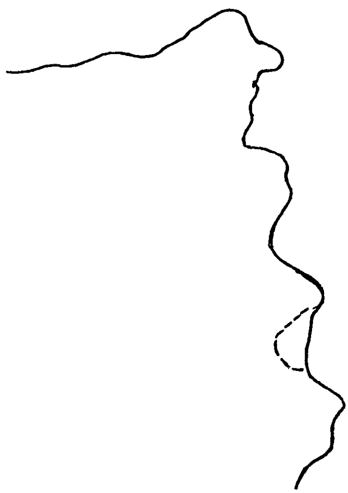
As important and time consuming as interactive editing of lineal features is, there is another type of editing that was developed for RADC with LIS and other systems. This is automated editing--those types of editing that don't require a human being because they don't require intelligence and judgment, just a lot of processing, smoothing, cleaning. One typical example of automated editing on the Remote Job Entry System is the clip/join illustrated in figure 8. We all know that it can take many minutes to try for an absolutely clean join of two lines and many more minutes to verify that all such conditions requiring joining have been identified by the cartographer and corrected. Wouldn't it be so much nicer if the cartographer, in digitizing and editing the features, only had to get within eyeball accuracy so that he could let the machinery clean it up for perfect clips or joins. This is exactly the technique that is employed at the Experimental Cartographic Facility and is incorporated into both the Batch Processing System and LIS.

I'd like to say a few words about a secondary but very important area in the field of cartographic editing. In existing systems to a large degree, lineal data are involved. Thus we are discussing the ability to enter lineal data either manually with a digitizing device or semiautomatically with things such as automatic line following equipment like that currently being developed at RADC. There is an alternative to lineal data and that is raster data. Raster data become especially important to people who have to rapidly develop a very large data base of cartographic information; classical lineal digitizing is just too slow. Enter--the automatic scan device.



Mid Segment Modify

Figure 6



Accept/Reject

Figure 7

Clipping/Joining Function

Description

The clipping/joining module functions to effect a clipping or joining action on adjacent digital features generated at the digitizing station. Digital clipping or joining is necessitated by inadvertant overruns or underruns, respectively, when the end point of one feature is in common with an intermediate point of another feature. Clipping involves the deletion of points produced as a result of a feature overrun. Joining involves the addition of points to a feature in order to close the digital gap produced by an underrun.

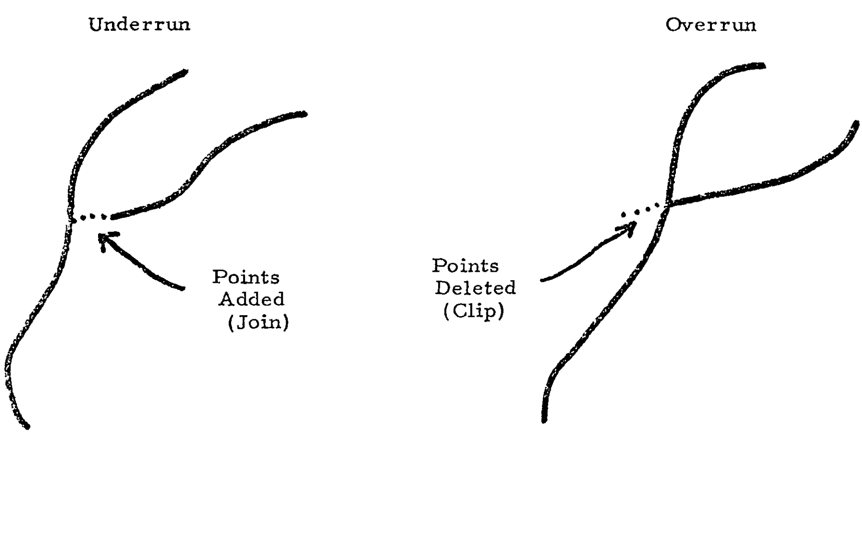


Figure 8

The automatic color scanning device (ACSD) has been available at RADC for several years. The output is raster data organized by x-y position rather than by features in the x-y field. It is easy to assume that the raster form is not readily usable for many cartographic processes, specifically those such as line symbolization and feature identification. These two processes are closely allied with the presentation of information on a chart in a form that is readily interpretable by the human user. For this type of operation, lineal organization now appears to be the best method. Thus we must resolve the problem of transferring the raster data obtained by the scanning device to a lineal format. This conversion really involves two steps: (1) a pre-edit function for cleaning up noise and machine-detection errors, and (2) the actual conversion. Keeping to edit problems, I will completely bypass the raster-to-lineal conversion problem.

The company recently installed an interactive CRT-oriented edit system for raster data--the Raster Edit System (RES)--at RADC. Here we work with the raw data from the scanning device and use a Tektronix storage tube display to handle the large density of line work. The RES design provides for looking at either a large part of the manuscript or for blowing up a specified window within the manuscript to look at smaller areas. Zooming can get down to individual points and lines where the human can identify noise from clean data and take editing steps. Noise can take the place of unwanted or missing points. Effectively a man using his associative eyeball can go in and find those dirty areas where he expects problems for any raster to lineal conversion software or where he expects problems will arise in later processing. Thus RES adds the intuitive, associative human brain to the interactive environment in providing an editing function on raw raster data.

Schuler (Control Data Corp.): Could you possibly give us a handle on the timing associated with your automatic editing and identify the computer?

Moritz: The biggest machine in figure 1 has 32-k word core. That machine is shared by up to 10 8-k word minicomputers. In these PDS-1G minicomputers all the graphics work is done right there. There is no need to call on the main or master processor for anything other than data through its data management system. All the editing operations are literally interactive, online, and real time or very nearly real time. From the time a man selects a feature to be edited by simply pointing the cursor on the graphic, our experience is that it never takes longer than 3 sec to retrieve the feature with 5 stations operational.

Schuler: I am referring more to the automatic editing where you have an overflow or an underflow problem at a join.

Moritz: We can take a typical file of several thousand inches, and it will run anywhere up to 5 or 6 min on a PDP-15 in a time-shared environment. It is fairly fast.

Rockwell (Dept. of Community Affairs): It is the conviction of the entire panel that the only way you can work with interactive editing is to create a new data base which is not the same as the data base you are ultimately going to have to be doing your analyses on? In other words, if you want to go back and correct something that existed before, there seems to be a need to pull it out of one data base format and put it into a new one. Is this something that we are never going to be able to get around? Can we ever work right on the data that we want to do our analyses on?

Moritz: I can speak in terms of how long it takes to do data transformations on individual points. It is theoretically possible, knowing all the connections between a variety of different coordinate systems, that you could always convert on the fly. Again, we are running into a cost situation. Given a parallel processor, perhaps, or a machine that operates at 10 to 100 times the current speed, you might be able to do such things. Take a look at CIA, for example, which requires a geographic global basis for a data base. This means geographic coordinates--latitude and longitude--or if you want to set up some form of Earth meter system, that's okay too. Here you are talking about taking a variety of scales--1:1,000, 1:50,0000, 1:100,000, 1:1,000,000, 1:3,000,000--and a variety of different projections--Mercator, transverse Mercator, Lambert conformal, and so on and being able to match or merge them together to give yourself a standard reference frame through which everything can be correlated. If you do not require it, there is no need to do this on a global basis; then there is no problem.

Rockwell: So the kinds of constraints that have been discussed so far in the hardware session are tied to the needs to meet National Map Accuracy Standards and that kind of thing. If you are working in a purely analytic framework, (as an analytic geographer would) and you were interested in 1-percent accuracy, the whole game becomes dramatically simpler.

Moritz: It could be; I think that it is too easy to answer yes or no. I would have to know more or less what the application is.

Peucker (Simon Fraser Univ.): I have a general comment. I don't think that we will ever have clean data. We might have accurate data or smooth data; we might have all kinds of other things, but clean is a dirty word. I will give you an example. The CIA has digitized the east coast of Australia. They have run it through their system, and for them it is clean. An Australian looks at it and doesn't find the Bay of Brisbane--for him it is not clean and of no use at all.

Moritz: Quickly, I want to point out that right now the existing systems are using coordinate frames which are better than .01 sec of arc. Granted, if you want to get less than a foot or two on the Earth's surface, then you might run into nonclean conditions. Everything that we have been able to work with to date on the systems at DMA-AC and HC is good to that type of resolution.

Chrisman: I am not here to speak in favor of the DIME system, but I think it should be brought up right now. We have been talking about people; we have been talking about interactive editing and some automatic editing. There has been a lot of U.S. money spent on systems for automatic clarification of what is effectively geographic representation--money spent by the Bureau of the Census on the DIME system. None of you mentioned any of that. It seems important to me that there is no interaction between the people who are trying to model urban systems at less-than-cartographic accuracy, and people who are trying to model cartographic information very accurately. There is a block somewhere between the two groups. I think that there should be more realization by cartographers that there is actually something to that form of editing. And I think that there is a need for the DIME people to be more aware of what is cartographically available to them. That is not something for this workshop. We do have people who are supposed to be here from Census sometime later; I am trying to present their case temporarily.

Aangeenbrug (Bureau of the Census): I am currently looking at the DIME file, and there will be a demonstration (twice) in seminar form where these topics will be discussed. I agree with the comment, having been a victim and a visitor at the USBC project, that there are problems in terms of dealing with absolute and relative accuracy. Going back to some of my lectures with Prof. Robinson, we are not shooting missles into urban areas, so the absolute accuracy is not that important. You get into a policy question here. I think that we might try to take advantage of the DIME seminar to address some of these points. Perhaps at a later session, we might want to discuss the needs for absolute and relative accuracy, because there is a big world of dirty cartography that gets most of the people to work daily, that caused presidents to make decisions, and so forth. Those are now becoming automated, so I urge us to discuss these later.

Rhind: I would just like to make a point that there appears to be a paradox in much of this work that cartography, which has high-accuracy requirements in some cases, needs much less carefully edited data than does certain geographical processing. If a small segment of line is missing in a map file, then it can be put in by pen--a semiautomated process. On the other hand, if a small segment of line is missing in card-fitting areas, then you have a real problem indeed if it is all being done inside the machine.

Boyle: When we were talking about drafting systems, I said that I believed it is in a professional state. If you want high-precision drafting, it is a professional job which should not be attacked by people on their own. As an exact opposite of that, I believe that everyone should have an interactive viewing and editing capability. I believe that this is their driving tool--to see what they have. If they are going to have digital data, they should be able to look at it and to play with it. The cost of these things is not very high relative to precision drafting and digitizing. You notice that it was very difficult for the last speakers to separate digitization--certainly interactive manual digitization--and interactive editing; one flows into the other.

Many people have tried to do their own digitization work over the last 10 yr or more. If they have bought any equipment at all, they have bought digitizing equipment. Many many people have ended up with rooms full of punched tape. Except for a few who have attacked it professionally, most of the people have been playing at digitization. Just about now it appears to me that we are moving into the professional stage of digitization, in which the work can be done reliably, predictably, and so on. We tried to bring some of these people together to tell you about the state of the art. I do not think that we are at the point where we just put out a contract, pay x thousands of dollars, and obtain the map data perfectly done. However, I think we are very near that.

You will want to update some of the digitized data without sending it out to some massive drum scanner, automatic line follower, or whatever. You will want to do this a little bit more accurately than you can eyeball on the display screen. A personal editing facility, in my view, is an interactive display with a part-time digitizing table connected to it. I am going to ask Olin Bockes to come up now and bring his panel with him. Bockes has been involved with the Forest Service in the Department of Agriculture for many years and has recently changed his allegiance to the Soil Conservation Service.