THE USER PROFILE FOR DIGITAL CARTOGRAPHIC DATA

A.R. Boyle
University of Saskatchewan
Saskatoon, Saskatchewan, Canada, S7N 0W0

ABSTRACT

The paper reports on a year long examination of the general user profile for digital cartographic data. The investigation casts doubt on the needs for this data to be in vector form as far as topographic overlays and cadastral boundary lines are concerned. The only actual or proposed used for such data appeared to be as background imagery, and this could equally be met by the much lower cost scan data format. The main question raised for discussion is relative to the policies of the larger digital base map producers. Are the large front-end expenditures of time and money for the production of digital vector data really warranted, or could they more usefully serve the public in a much shorter time period and at a lower cost? The writer believes that the user profile has not been given the attention it deserves. It is hoped that this paper will raise discussion and comments from present and prospective users.

The controversy over raster versus vector digital map data has raged to the writer's knowledge for at least 25 years. Changing situations, user needs and technology require it to be continually reviewed. The writer has just completed a sabbatical year discussing problems with spatial data users and producers in many parts of the world. Some new, although perhaps with hindsight, obvious aspects have appeared and these seem worth discussing at this time.

The most important change is the fact that the 'users' are now starting to make their comments felt; until recently little data except self-generated has been available to make general needs formalized. The previously unstated 'wants and non-wants' now present unexpected viewpoints to data producers who previously thought they knew the user. Unfortunately the question to many data producers of 'what is your user profile?' is met with blank looks or even hostility. Have data producers been spending enormous sums on equipment and hours of work on a product that is not wanted? The large costs involved make continual re-assessment critical.

While the object of this paper is to raise questions and create lively discussion, nevertheless this should be limited to the real points at issue and periodically certain exclusions must be made. The first point of importance is that the main part of this paper is limited to cartography; imagery and GIA will only be discussed later and in no great depth.

Cartographic production departments, when moving to digital mapping, have traditionally preferred to think in terms of the more easily relatable vector format. There is a natural connection with manual digitization where points are pairs of XY coordinates and lines are streams of coordinates pairs. A coordinate resolution adequate to recreate the line graphically at good quality, is used; typical resolutions are 0.001"-0.004".

Many cartographers were disappointed when Automatic Line Followers turned out to be technical and costly failures and they were forced to turn to the less comprehensible precision scanners; however, they still felt the need to convert the scan data to vector format. While

scanners of high quality are expensive, nevertheless they are reliable and efficient, making a raster digital copy of a separation sheet in about 20 minutes. Recent discussions with data producers have shown production costs of the vector conversion process to be extremely high; the format conversion is easy, but the problem is the amount of interactive edit required to make precision and program-usable vector data. Times of over 100 hours per sheet are reported; in fact, one producer commented that in comparison with manual digitization times for the same sheet the latter proved to be faster and the equipment costs to be appreciably lower. Digital map data producers have tended to pour money and effort into the creation of vector format data, usually now following the route of scanning and subsequent interactive edit, believing that this meets the demands of the user. However, the front end costs and times have been enormous and it is only recently that the user has been in a position to commment on the usefulness of the produce which has only been gradually provided.

Discussion carried out in 1985 with many digital cartographic data users in the topographic area showed that, at that time and into the foreseeable future, they had little use for the digital data except as a graphic background; in fact no general users of those met could cite any example of general cartographic vector data usage in their computational operations. Until this time they have been using drawn sheets as background but this causes difficulty and they certainly do have a need for a digital graphic presentation so that they can manipulate it for display and plotting; their needs would be adequately and perhaps preferentially met by the cartographic data in raster format. Scan data is still digital and can be manipulated by computers in all the normal ways; it can as easily or even more easily be changed in scale, projection, than vector. It can thus be made to overlay or underlay any other data in vector or raster form and many display systems can handle these two formats simultaneously.

Arguments against the data being in raster format are frequently raised and these must be looked at carefully. The first concern is that automatic data selection can only be done easily on a full separation sheet basis rather than on individual points, lines or polygons; each separation raster image is in a different file. It is only on rare occasions, however, that this is of real importance.

Secondly data is often assumed to be far less compact in raster format than in vector. However, when we examine this we have to appreciate that we are only talking about a raster pixel being black or white (1 or 0) and thus run-length encoding can be used. We can assume that the resolution of the pixel is the same as the coordinate resolution and then only in the case of long straight lines do vectors show an advantage. As soon as the line is curved or irregular then the two formats are relatively similar in bulk. The present rapidly reducing cost of storage makes any difference unimportant in most cases; it is interesting to look at the enormously uncompact storage of vector data to realize that neither producers nor vendors worry about this and even seem to prefer to use excessively large disk and tape storage. The argument of compactness thus has little or no validity.

Cartographers sometimes say that raster data is not as visually 'smooth' as vector, but when the same resolution is used this comment cannot be warranted. In many cases it arises from the fact that for many years coarse grid cells (similar to a large raster pixel) produced this type of presentation.

If scan data were to be agreed as the desired output then two possible forms are possible. Data from scanning may be 'as is' with lines being as wide as they appear on the original separation sheet. The

second is to reduce all lines to a standard single pixel width by a line thinning program. This latter software operation takes appreciable cpu time and can lead to a need for interactive edit, because lines can thin so much that they disappear. If high costs and edit times are to be avoided the direct scan data should be produced and used if proved to be acceptable.

It should be remembered that the proposals under consideration are only that data should be supplied to the users directly after scanning and prior to vectorization and intensive interactive edit. If this were acceptable to the user then data could be provided at a much earlier date and at a much lower cost. This is because we now have operational scanners in wide use and do not still depend on manual digitization.

A note on costs gleaned from many agencies over a period of time might be useful. Taking as a basis a reaonably complex small scale separation sheet, eg. for hydrology, the digitization cost range can be anything between $1000 and $10,000. With multiple map series such a difference can have a large financial impact. An efficient manual digitization can often be done for $1000, equipment costs being relatively low. Automatic scanning to produce a raster image need only cost $200; however, to edit this to high quality vector data can cost at least $2000 and frequently much more because of combined operator and equipment costs. If the supply of raster image data alone meets a majority of user requirements ($200) why go to the tediousness of manual digitization or the high costs of editing the raster/vector image data?

It is, of course, not possible to state that all needs could be met by raster data as background imagery. Vector data can be very useful in some cases and essential in others. Technically this does mean that all user interactive stations must be capable of at least visually overlaying raster and vector data. There would seem to be very good arguments for digitizing the culture separation sheets of small scale maps in vector form and storing them in that mode; some boundary sheets may also be treated in this way. However, the vector digitization and storage of parcel boundaries on cadastral maps seems to be best done as raster, because the main use is very much that of imagery only; few people seem to have enough confidence in the boundaries themeslves for exacting computer manipulations. One cadastral data user reported that 94% of his needs were met by the alphanumeric data only, another 5% by the addition of a centroid point and only 1% from the boundary image. Utility mapping is probably best treated as vector digitization and storage, but it is to be hoped that within a few years no digitization of drawn utility maps would be required, the first input being on a display in computer form.

In addition to the fact that input of cartographic data is moving to raster techniques, the user now can have powerful cpu capabilities. This makes it possible to suggest that the usually small amounts of raster data required in vector form (eg. along a transmission line) could be converted by the user. It seems very wasteful to convert all data from raster to vector at high cost because it then might be useful to someone at some time. A number of interactive display systems have a capability of pointing to a pixel and creating a vector line data stream from adjacent pixels; moreover, this can be done at very high speed, the data usually stopping as soon as a junction or line end is reached. In the future a preferable method would be to indicate a vectorization rectangle or polygon and proceed by a batch program.

The question of labelling points and lines must be discussed. Using

the raster data image form, any labels on the original will of course be passed through the system to the viewer. If the user carries out vectorization, that user also has to label such lines from the viewed data. Methods have been proposed for automatically labelling node points on a separate overlay and this procedure might be useful in the future.

At this point in the paper we must move from pure cartography to more complex structures such as imagery or Geographic Information Systems. In raster terms this means moving from a simple 2 bit pixel or fine grid cell to one where there may be 256 states or even reams of alphanumeric descriptive information. As would be expected, the data storage increases proportionally and we have to consider the advantages of polygon or raster methodology in a new light. Again we must not be misled by the fact that for many years coarse grid cells were the only method possible, with crude manual digitization and slow manipulation software on the slow computers of that time. We must now consider grid cells as fine as the resolution of the coordinates used in the polygon work, perhaps equivalent to a few centimeters or meters on the ground.

Some years ago the new availability of efficient polygon manipulation changed the capabilities of systems to the extent that they had economic applications in such areas as forestry management. However, not everyone changed to that procedure and the proponents of the fine grid cell as an alternative have a good case. They believe that storage is not appreciably greater and that the grid cell overlay process is superior, particularly when historical, remote sensed or DTM data are concerned. It may be that the polygon methodology has been stretched to its limits and that the next advances will be in fine grid cell work. The phrase 'fine grid cell' must be stressed as many arguments are erroneously based on the noisy visual appearance of coarse grid cells.

Can the complete use of raster data be far distant with the increasing application of scan technology in digitization, edit and plotting, together with the rapidly increasing application of DTMS, DEMs and remote-sensed scan imagery? With the extraordinary increase in memory capacity and cpu speeds both tied to lower costs, the tendency to fine grid cells seems to be there; the economics seem to be in place, the user appears to be pushing in that direction and rapidly improving software is helping.

A major factor in the proposal for more scan data is that it enables both producer and user to proceed towards the future in a step by step way without the initial commitment of enormous funds. Vector digitization may even be the last straw that breaks the back of the camel of traditional cartographic presentation. However, on the other hand it may be that raster digitization is the start of a new type of precision storage cartography because the base material could well be a drawn document, updated manually, and rescanned as needed to produce new data products. This is reminiscent of the RADC developments in the late sixties which were overtaken in the seventies by the explosion of vectorization; perhaps the wheel is turning full circle.

Many of you will know me as a 'vector dedicated' man for many years, even while carrying out appreciable developments in the raster area. I believe the time has now come for me to change my viewpoint to one where data is basically in raster form, but nevertheless not forgetting the great advantages that can be obtained in selected cases from vector formats and being prepared for rapid interchange between the two as the need arises.