<table>
<thead>
<tr>
<th>Mapsheet</th>
<th>Number of Monuments</th>
<th>Mean Difference</th>
<th>90-th Percentile Difference</th>
<th>Maximum Difference</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>94/A</td>
<td>53</td>
<td>0.47</td>
<td>1.13</td>
<td>2.08</td>
<td>0.18</td>
</tr>
<tr>
<td>94/B</td>
<td>83</td>
<td>0.59</td>
<td>1.18</td>
<td>2.41</td>
<td>0.20</td>
</tr>
<tr>
<td>94/C</td>
<td>64</td>
<td>0.72</td>
<td>1.46</td>
<td>2.50</td>
<td>0.29</td>
</tr>
<tr>
<td>94/D</td>
<td>69</td>
<td>0.58</td>
<td>1.24</td>
<td>3.34</td>
<td>0.31</td>
</tr>
<tr>
<td>94/E</td>
<td>80</td>
<td>0.48</td>
<td>0.81</td>
<td>1.82</td>
<td>0.11</td>
</tr>
<tr>
<td>94/F</td>
<td>72</td>
<td>0.66</td>
<td>1.44</td>
<td>2.43</td>
<td>0.28</td>
</tr>
<tr>
<td>94/G</td>
<td>68</td>
<td>0.58</td>
<td>1.28</td>
<td>2.72</td>
<td>0.27</td>
</tr>
<tr>
<td>94/H</td>
<td>47</td>
<td>1.53</td>
<td>1.66</td>
<td>16.45</td>
<td>3.61</td>
</tr>
<tr>
<td>94/I</td>
<td>181</td>
<td>0.69</td>
<td>1.43</td>
<td>10.30</td>
<td>0.89</td>
</tr>
<tr>
<td>94/J</td>
<td>145</td>
<td>0.67</td>
<td>1.40</td>
<td>4.32</td>
<td>0.41</td>
</tr>
<tr>
<td>94/K</td>
<td>55</td>
<td>0.24</td>
<td>0.60</td>
<td>0.91</td>
<td>0.05</td>
</tr>
<tr>
<td>94/L</td>
<td>139</td>
<td>0.46</td>
<td>0.92</td>
<td>3.17</td>
<td>0.19</td>
</tr>
<tr>
<td>94/M</td>
<td>86</td>
<td>0.57</td>
<td>1.21</td>
<td>3.12</td>
<td>0.34</td>
</tr>
<tr>
<td>94/N</td>
<td>51</td>
<td>1.37</td>
<td>3.37</td>
<td>8.45</td>
<td>2.80</td>
</tr>
<tr>
<td>94/O</td>
<td>177</td>
<td>1.37</td>
<td>2.75</td>
<td>8.52</td>
<td>1.99</td>
</tr>
<tr>
<td>94/P</td>
<td>226</td>
<td>0.84</td>
<td>2.13</td>
<td>7.90</td>
<td>1.43</td>
</tr>
<tr>
<td>103/A, H, &amp; E₁ G</td>
<td>34</td>
<td>0.40</td>
<td>0.78</td>
<td>2.82</td>
<td>0.25</td>
</tr>
<tr>
<td>103/B, C</td>
<td>32</td>
<td>0.38</td>
<td>0.86</td>
<td>1.55</td>
<td>0.13</td>
</tr>
<tr>
<td>103/F, K, &amp; W₁ G</td>
<td>124</td>
<td>0.61</td>
<td>1.37</td>
<td>3.03</td>
<td>0.35</td>
</tr>
<tr>
<td>103/I, J</td>
<td>221</td>
<td>0.52</td>
<td>1.14</td>
<td>4.34</td>
<td>0.29</td>
</tr>
<tr>
<td>103/P</td>
<td>68</td>
<td>0.67</td>
<td>1.43</td>
<td>3.74</td>
<td>0.44</td>
</tr>
<tr>
<td>104/A</td>
<td>171</td>
<td>0.50</td>
<td>0.95</td>
<td>5.93</td>
<td>0.37</td>
</tr>
<tr>
<td>104/B, C</td>
<td>80</td>
<td>0.52</td>
<td>0.68</td>
<td>20.92</td>
<td>5.41</td>
</tr>
<tr>
<td>104/F, G</td>
<td>234</td>
<td>0.43</td>
<td>0.91</td>
<td>1.95</td>
<td>0.13</td>
</tr>
<tr>
<td>104/H</td>
<td>96</td>
<td>0.84</td>
<td>1.93</td>
<td>3.75</td>
<td>0.63</td>
</tr>
<tr>
<td>104/I</td>
<td>70</td>
<td>0.51</td>
<td>1.31</td>
<td>3.72</td>
<td>0.45</td>
</tr>
<tr>
<td>104/J</td>
<td>136</td>
<td>0.80</td>
<td>1.39</td>
<td>2.64</td>
<td>0.26</td>
</tr>
<tr>
<td>104/K</td>
<td>62</td>
<td>0.50</td>
<td>0.85</td>
<td>4.00</td>
<td>0.38</td>
</tr>
<tr>
<td>104/M</td>
<td>87</td>
<td>0.22</td>
<td>0.55</td>
<td>1.55</td>
<td>0.08</td>
</tr>
<tr>
<td>104/N</td>
<td>179</td>
<td>0.50</td>
<td>1.11</td>
<td>2.68</td>
<td>0.19</td>
</tr>
<tr>
<td>104/O</td>
<td>98</td>
<td>0.32</td>
<td>0.62</td>
<td>2.39</td>
<td>0.12</td>
</tr>
<tr>
<td>104/P</td>
<td>143</td>
<td>0.38</td>
<td>0.90</td>
<td>1.74</td>
<td>0.13</td>
</tr>
</tbody>
</table>
"INTTAB" routine available in the National Transformation Software package should be used to get a hard copy listing of the grid shifts in the region of interest. If the grid shifts changes are reasonably smooth, the internal accuracy of the network will not be drastically affected. Large changes in the magnitude of the grid shifts will affect the internal accuracy of the network to be transformed.

**Independent evaluation of transformation**

To independently verify the precision and reliability of the National Transformation and to ensure its efficiency and effectiveness in a production environment, the following tests were performed.

### 5.1 4th order control monuments test

Table 2 lists the statistics derived from the Fourth Order Monuments for which coordinates are available in both systems. These monuments were excluded from the definition of the Version 1.1 Transformation due to the questionable accuracy implied by fourth order. As well, they were set aside as reserve points for checking the National Transformation. The data set includes all Boundary Monuments, MCE and NAT.DEF. and DTS monuments that were available. These 4th order monu-

<table>
<thead>
<tr>
<th>Quadsheet</th>
<th>Number of Monuments</th>
<th>Mean Difference (m)</th>
<th>90-th Percentile (m)</th>
<th>Maximum Difference (m)</th>
<th>Standard Deviation (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>82</td>
<td>274</td>
<td>3.09</td>
<td>10.07</td>
<td>46.51</td>
<td>5.42</td>
</tr>
<tr>
<td>83</td>
<td>38</td>
<td>12.44</td>
<td>25.49</td>
<td>77.29</td>
<td>14.88</td>
</tr>
<tr>
<td>92</td>
<td>548</td>
<td>2.02</td>
<td>4.32</td>
<td>68.93</td>
<td>4.54</td>
</tr>
<tr>
<td>93</td>
<td>375</td>
<td>2.08</td>
<td>4.37</td>
<td>35.26</td>
<td>3.91</td>
</tr>
<tr>
<td>94</td>
<td>40</td>
<td>5.23</td>
<td>13.37</td>
<td>22.55</td>
<td>5.67</td>
</tr>
<tr>
<td>103</td>
<td>52</td>
<td>2.70</td>
<td>7.27</td>
<td>9.56</td>
<td>2.73</td>
</tr>
<tr>
<td>104</td>
<td>112</td>
<td>3.15</td>
<td>8.65</td>
<td>22.60</td>
<td>3.65</td>
</tr>
</tbody>
</table>

Table 2: Coordinate difference statistics for 4th order monuments, listed by quadsheet

![Figure 11: Bar chart of coordinate difference statistics for 4th order monuments](image)
ments extend across the province. Figure 11 is a provincial summary of Table 2 in bar chart form.

The statistics are higher for this data set partly due to the lower accuracy of the NAD27 coordinates. The mean and 90-th percentile for the small corner of quad sheet 83 that falls within the province of British Columbia are high, thus special care should be exercised in this area. Nevertheless, on a provincial scale the results are very good. 95% of the differences between transformed and rigorously adjusted coordinates are within 10 metres, while 68% of these differences are within 2 metres (Figure 11).

5.2 Comparison with the NADCON transformation from USNGS

The software put out by the United States National Geodetic Survey (USNGS) for transforming coordinates from NAD27 to NAD83 (NADCON) within the continental United States has a limited region of overlap with that of the National Transformation. This section describes a comparison of the accuracy of the two transformation packages.

A test set consisting of 2558 monuments within the region of B.C. below 50° Latitude and east of 126° longitude was used. Monuments excluded from the definition of the transformation surface for the National Transformation were not rejected from this data set. Two monuments were excluded as obvious outliers with coordinate differences greater than 50 metres. Statistics of the coordinate differences between the two predicted sets and the NAD83 coordinates from the July 1990 readjustment are shown in Table 3, below.

There appears to be a systematic difference between the two solutions of 1 metre in latitude and 1.5 metres in longitude. Overall, the National transformation is providing transformations with better accuracy in the region of interest. Users working in this area with coordinates obtained from both transformation packages should be aware of this incompatibility.

5.3 Cadastral map data conversion pilot

When using the National Transformation to convert digital map data, usually an interface specific to the map system environment will also have to be built. Thus, an Intergraph based system will require reformatting of the Intergraph Design (IGDS) files so as to accept the National Transformation software. As well, the final file converted onto NAD83 will again have to be reformatted into the proper IGDS file format.

Currently, a pilot project within the Ministry of Lands and Parks, Surveyor General Branch, is examining the issues associated with converting the 1:20 000 scale cadastral map data using the National Transformation. Interface software was first written to accommodate the Intergraph IGDS file structure. Preliminary data within the four selected sites in the province (Saanich peninsula, Kootenays, Skeena and Peace River) show excellent results. The areas selected attempted to reflect a cross-section of data densities, age of data compilation and area of province. The quality control (map comparison) was provided by the NAD83 TRIM mapsheets, agreement with which so far is extremely good. The information converted was land tenure, planimetry and limited forestry information. The largest file conversion in the Saanich peninsula (including downtown Victoria) entailed 138 071 coordinate conversions and took fifteen minutes in total elapsed time on a DIGITAL VAX 3500 computer.

Final editing - entailing removal of old neat lines and replacing with new NAD83 surround; merging of surrounding affected files; continuity check at former map neat lines; and substitution of old planimetry with

<table>
<thead>
<tr>
<th></th>
<th>National Transformation</th>
<th>NADCON</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Δ φ</td>
<td>Δ λ</td>
</tr>
<tr>
<td>Max (m)</td>
<td>7.564</td>
<td>7.609</td>
</tr>
<tr>
<td>Min (m)</td>
<td>-12.673</td>
<td>-11.612</td>
</tr>
<tr>
<td>Mean (m)</td>
<td>-0.012</td>
<td>-0.073</td>
</tr>
<tr>
<td>σ (m)</td>
<td>0.716</td>
<td>0.671</td>
</tr>
</tbody>
</table>
newer TRIM planimetry - amounted to the most time consuming part of the conversion exercise. For the mapsheets in this project, the final editing averaged 3 hours per map sheet.

5.4 Thin plate spline function alternative

Quite apart from the tests reported here, an investigation by Barrodale Computing into thin plate spline functions for defining the transformation surface has also proved valuable in verifying the integrity of the National Transformation. The thin plate spline function is different from a polynomial fitting process, such as that used in the National Transformation, in that it fits the transformation surface exactly at the defining control monuments. While still under investigation by Dr. Ian Barrodale, the methodology may be well suited for the urban transformation surface definition later on since these(urban) areas are thought to have fairly homogeneous and consistent control networks. That is, the least squares best fitting process of the polynomial at the control points, needed to reject non-conforming control networks/monuments, may be less accurate than an exact fitting thin plate spline in those situations.

Currently, the Ministry of Forests is evaluating the feasibility of the thin plate spline approach for the conversion of 1:20 000 forest cover maps. Since these maps represent a large investment it is desirable to convert them to NAD83 and overlay them against the newer TRIM (non-forest cover) base maps. However, the forestry maps were compiled using approximate methods for georeferencing so that their relationship with the geodetic reference system is considered tenuous. Hence the consideration of the thin plate spline function to deal with heterogeneous data sets.

Transformations in urban areas

Further to section 3.8, the current plans are to provide a new version of the National Transformation on April 1, 1993, to accommodate the transformation needs within ISAs. The high density control networks, with their monuments spacings of 300 - 800 m., should permit a transformation surface definition that will yield 5 - 20 cm. accuracy, depending on density and networks consistency and accuracy. This should be generally sufficient to transform map bases of 1:1 000 and smaller.

Users requiring NAD83 ISA coordinates immediately may approach Surveys and Resource Mapping Branch, Geodetic Control Unit, to arrange for preliminary information. Preliminary coordinates will not change by more than 0.5 m. when the final values are published in 1993. As well, the relative accuracies between control monuments will not change significantly (say, within 3rd Order accuracy). Therefore, municipalities wishing to proceed on NAD83 now may be able to easily perform up to 0.5 m. shifts of their NAD83 data holdings in 1993, without significant loss of accuracy. Of course each organization should study its requirements carefully in order to decide whether these small errors (few cm.) after the shift affect their application(s).

National transformation software and users guide

The actual shifts from NAD27 to NAD83 for any given point(s) in B.C. is computed using the National Transformation software. The Users Guide for the National Transformation software is published within the CISM (1990) publication 'Proceedings of the CISM seminar; Moving to NAD83, the new address for georeferenced data in Canada'. It is also available with the National Transformation software as distributed by MAPS-BC at Surveys and Resource Mapping Branch.

The diskettes supplied from MAPS-BC contain the source and executable codes for programs GRIDPT, DIRFIL, INTGRID, INTTAP, READDA and GSRUG. The source code is compatible with DIGITAL VAX, Macintosh, and IBM/PC computer platforms. There are two versions of the executables supplied. The ones with a 7 at the end of the name require a math co-processor, whereas the other ones do not require a math co-processor. As well, the diskettes contain the actual grid data required by the software execution. The data only covers the British Columbia land mass and an arbitrary buffer around it.

For assistance with the National Transformation in B.C., contact the Geodetic Control Unit at Surveys and Resource Mapping Branch, Ministry of Lands and Parks, in Victoria.

Summary and conclusions

The analysis of the 9272 First to Third Order monuments used to define the transformation surface indicates a mean difference between the transformed and adjusted NAD83 coordinates of half a metre. Ninety percent of the monuments exhibited coordinate differences of less than 1.14 metres. Actual values can vary significantly throughout the province. The transformation should provide coordinate conversions with an accuracy acceptable to many applications such as GIS, cartography and thematic mapping. Users with much higher accuracy requirements may need to look at other alternatives.

The statistics presented in this paper are meant as a guideline only. They give a general idea of the accuracy of the transformation from NAD27 to NAD83 for the monuments utilized for this analysis. Positional data maintained by other agencies may be of far differing accuracy and internal reliability and will thus produce widely varying results. In addition, the statistics presented here cover broad regions. Smaller, more localized regions can have significant differences from the mean values.
To offer practical guidance, conversion to NAD83 of existing NAD27 1:10 000 scale mapping (and smaller scale) will be supported by this National Transformation version 1.1. Nevertheless, it is recommended that any agency contemplating transforming their data holdings complete further studies, if possible, to identify areas of concern — especially when transforming larger scale mapping. Of particular interest will be positional data sets tied to monuments excluded from the National Transformation surface definition. Entire networks which were found to be inconsistent with the surrounding control were excluded from the definition of the transformation surface. Many of these networks can be viewed as localized datums. A transformation specific to these monuments can be derived that will provide far superior accuracy in the transformed coordinates but the compatibility with the national standard will be lost. A good source of quality control in checking a conversion is the NAD83 based TRIM 1:20 000 mapping. However, this mapping is incomplete at present, covering 25% of the province (100% by 1997).

Finally, users should exercise extreme caution in using version 1.1 in developed (urban) areas of the province since the rigorous readjustment of the geodetic control networks onto NAD83 is as yet incomplete. A new version of the transformation based on the new provincial readjustment and incorporating the Integrated Survey Areas (i.e. many of the urban municipalities in the province) will be made available April 1, 1993. It is expected that this new version will support 1:1 000 scale (and smaller) mapping conversions in these ISAs, while continuing to support 1:10 000 scale (and smaller) mapping conversions in other areas of the province. Note that the new version will not change significantly outside of the ISAs, i.e. the changes will be well within the tolerable limits of the largest map scales (1:10 000).

The cadastral data conversion pilot conducted by the Surveyor General has shown the viability of the National Transformation in a production setting, in this case conversion of 1:20 000 cadastral map data residing in Intergraph’s IGDS file format. Thus, entire map files are converted in two to fifteen minutes, although final editing requires about 3 hours per map.

The NAD83 technical challenge now lies largely with the mapping and LIS / GIS communities (as opposed to the geodesists). While the conversion of point coordinates from NAD27 to NAD83 is easily achieved with the National Transformation, the treatment of map topology and integrity will vary from one application to another, due to both the hardware / software system environment and the information content characteristics. The greater challenge of course is to have sharable and seamless data structures built, while accommodating NAD27 to NAD83 transformed data.

Acknowledgments

My thanks go to the CCSC, the Canadian Council on Geomatics (CCOG) and especially EMR Canada’s Canada Centre for Surveying (CCS) for the national direction and coordination of the NAD83 project, and for providing excellent support in developing a workable version 1.1 of the National Transformation in British Columbia. Particular thanks go to Mike Pinch, Paul Henderson, Don Junkins, Steve Farley, Dan O’Kane and Gord Garrard of CCS in this regard. Don Junkins was instrumental in formulating the National Transformation methodology.

I would like to take this opportunity to acknowledge the efforts of staff within my Unit for the dedicated and thoughtful work that continues to be applied within the NAD83 project. Ted Rowlandson and his group are responsible for the data preparation and the actual readjustment exercise, whereas Kent Pointon (no longer with the Branch) provided significant contribution to the National Transformation analysis, the results of which are presented in this paper.

* Kent Pointon is now President of Manthanein Geomatics Ltd. in Victoria, BC

References


Pinch, M.C. (1990) *Differences Between NAD27 and NAD83*, Proceedings of the CISM Seminar; Moving to NAD’83, the New Address for Georeferenced Data in Canada.

University of Maine at Orono (1984) *The use and value of a Geodetic Reference System*, Research study by The University of Maine at Orono (1984), Orono, Maine 04469

Putting the records in order: GIS database construction

C. Warren Ferguson
Cantex
1425 Pearson Place
Kamloops, BC
Canada, V1S 1J9

Abstract
The construction of an accurate, intelligent, and above all, usable database is of utmost importance in the development and implementation of a Geographic Information System. This paper will begin by looking at the essential elements of the GIS database construction process. Next, it will explore the requirements for an accurate database construction process, comment about new technology and upcoming improvements and, lastly, share a few ways to achieve success in geographic database construction. Some specific project experiences will be shared.

Introduction
After a brief introductory history of GIS conversion and why it is implemented, I'd like to touch on the essential elements of the GIS database construction process. Next, we'll look at the requirements for an accurate database construction process, add a few comments about new technology and upcoming improvements, and lastly, share what I think are some ways to achieve success in geographic database construction. As we proceed, I will share some specific project experiences.

The conversion process as we described it in 1980 had seven steps to it. Let me review those. First, landbases had to be created. Second, the data we wanted to add had to be corrected because it was rarely accurate in the form that we found it. One of the things that we did then and still have to do today is standardize the records in preparation for entering them into the computer database. We had to establish data relationships because in many cases they were not evident from the manual records and the maps we had. Then the facilities had to be digitized.

After all the information was in the computer, there was considerable time spent doing edits to insure that we had the accuracy and the correctness we wanted. Most often we found that there were several correction cycles required in order to get the data in the format that we finally wanted. Even back in the early '80s it became obvious to me that the trademarks of excellence in conversion were accuracy, low cost and timeliness. These are items that were easy to identify with success and they still are the result of excellence in conversion.

In the early '80s conversion was a mystery and many people were looking for that magic box where they could stuff their original in on one side and get a digital tape out the other side. Even then, there were those who wanted it yesterday. There was no way to satisfy all of their many demands.

Conversion, or as you will hear me suggest, the activity better described as geographic database construction, is a very unique process and one that John Antenucci has recently called "the worst of both worlds because it is a
process that is both capital intensive and labor intensive”. I don’t particularly agree with John’s description that it is the worst of both worlds. But I will tell you that it is capital intensive and it is my conviction that people are what make success in conversion.

Let’s back up a minute and look at why you have chosen to implement a GIS system. Most often it’s a combination of several factors. There probably are problems with your existing manual system. Perhaps more efficient management of your information is the need in your organization. Surely all of us can be more cost effective. In today's times and with the need for infrastructure renewal, we find that we must have an accurate inventory of our facilities.

When we look at putting the records in order, one of the things we must remember is that the problems with a manual system is one of the most compelling reasons to develop a GIS project. Very often there are several things that are found lacking in existing manual systems. First is the physical condition of the documents. They’re simply old, they’re tattered, they’re torn, they have coffee stains on them, they may be very difficult to read, and in some cases they may have even been lost. A proper examination often reveals that there is a very high cost to maintain these paper documents. Sometimes there is confusion over which set of records are the “official” records. Often there is duplication of records, and obviously duplication is a needless waste of time, energy, and money. Very often, the important records are in the hands of those old-time employees who know exactly where everything is and what keeps the system working.

These dilemmas usually don’t show on any of the records. Added to this is the difficulty of keeping records current. To some extent this is the description of the records that we have to work with when we try to put the records in order and do data conversion. But the manual record is not the only thing we work with these days when we begin conversion. In the mid-to-late ‘80s we recognized that conversion was the process for building a digital facilities model on a geographic base from mechanized records as well as existing manual records.

Again, conversion is a process, it is not an event. It will not take place suddenly, quickly, or all at one time. It’s a lot more than most first imagine.

The conversion process involves much more than just converting the existing data to digital format. It involves an accurate landbase created to meet present needs as well as needs anticipated for the future. It represents a standardizing of symbology throughout our record set or database. The data sources for input have to be compiled and methods found to combine the variety of sources that exist within most organizations. Some data may already be in a form of mechanized records. Finally, it involves converting, changing, modifying, adding and verifying all existing and acquired data to build the geographic database.

Requirements for an accurate process

The requirements for quality information in a geographic database are paramount. The data converted must have the following characteristics: It must be accurate, and that means free from mistakes and errors; whatever you see is correct. It must be complete, and that means that all necessary elements are displayed appropriately. It must be timely, current and up to date. It must be reliable if users are to depend on the information. It certainly must be credible, which means that different output products from the common database must agree when viewed or used by a variety of users. In other words, different sources must agree. The information must pass validity checks because these are necessary if the impact of changes is to be predictable (and it must be). Finally, the information must be convenient and readable in that it is clean, legible, easy to use, and easy to read.

At this point let me say that to achieve the excellence in conversion that I believe is appropriate, each conversion process must be tailored to fit the user organization perfectly. That’s why each conversion process is different unto itself, driven by the needs of the specific organization, the users themselves and the existing or available data that can be included in the geographic information system database.

The database construction process

We talk about database construction as a process with four components. Four essential components. The first, information sources, prompts us to ask the question, what are the inputs to the process? Second, the technology causes us to define what tools are going to be used to perform the process. Third, methodology makes us ask how the technology is applied to perform the process. The fourth essential component is the output or end product. How sophisticated must it be? How much is expected of the outputs from the process?
In recent months, I have come to feel that geographic database construction is a more appropriate description of the process than is conversion. Clearly, to build a geographic database requires a lot more than simply conversion as data entry. It involves the design of real world data models so that the applications we have today, as well as those that will step forward in the future, can be enabled by our database. Very often we have to inventory or acquire data. Augmenting the existing data is nearly always a requirement of the geographic database construction process. Digitizing may be either conventional in nature at a digitizing table or scan digitizing and vectorizing those records into the digital database. A key activity in nearly all of the work done today is to translate and merge data from other digital files already existent within the user organization. Often there is a requirement to blend multiple databases to achieve the best of the variety of source data that is available.

Finally there is an absolute requirement to create connectivity throughout the geographic system database. This connectivity is a mandatory requirement of both linear and polygonal networks that form the topological basis of our geographic information systems. As you know, topology is generated by line/node structures. Nodes can be formed at the intersection between lines. They define the connectivity between adjacent line segments. Nodes may have associated user defined attributes as well. In linear networks lines are used to represent features such as roads, pipes or cables. Nodes represent the point at which these features connect to each other, for example a valve or street intersection.

In polygonal networks lines define the boundaries of areas and nodes represent the points where shared boundary lines meet. Usually an identifier for the enclosed areas is included to define topological relationships more thoroughly. An example of a polygonal network might be the service area boundary of a particular district office.

The skill sets required for GIS database construction are certainly far more than just data entry. Most projects today involve perhaps geodesy and/or photogrammetry. For sure, however, they require computer science skills, digital cartography, and a variety of engineering skills. In addition, database design and network understanding are required skills for the construction of a good, useful GIS database. The age old skills of project implementation and quality assurance are certainly required to build the complex GIS databases that we use today.

If we look at it in light of today’s technologies, today’s GIS systems and the GIS database construction requirements, we see that what we’re talking about is making a “real world database of geographic information”. The information that we put in this database tends to be landbase information which is geographic and is related through specific land data elements as well as the coordinate system. We include in our database a great deal of boundary information which is polygonal in format. And finally we put in asset or facility information, generally speaking the location, the relational features of that asset, as well as descriptive attribute information.

Why all the talk about conversion? What’s the big deal? Money! If you recall the typical system development cycle for a data processing system you remember how little conversion costs — maybe 15%. But in the GIS system development cycle conversion can cost 50% to 60% of the total. And some people say it can cost as much as 75%, but that’s only if you already own the hardware and software.

When you compare GIS system development to a conventional data processing system, you see how much different the conversion cost is (as a percentage of the total system development). However, you should remember that while hardware has a 5-8 year life cycle and software a 3-5 year cycle, the data has a 25-40 year life cycle.

What makes up the cost of GIS database construction? To many it’s a puzzle, but to those of us that do it for a living there are a wide variety of factors that influence each specific project. The base maps that are to be used for the landbase, the geographic accuracy, the facility density, the geographic size of the organization, as well as the scope and complexity of the project are those that most often impact cost on a large scale.

There are several development and conversion alternatives. Do it yourself, hire a consultant, purchase software, hire a conversion company or combinations of all these. Most organizations use some combination, or series of combinations, to gain the greatest benefit.

Let’s be sure we know what are the functional requirements. Conversion costs a lot and we must do it right the first time. Functional requirements are answers to the what, when, and why questions (not the how question - that comes later). We must determine the project scope, because scope affects the requirements. Is this a corporate project or only a district project — do we want full functionality from the start or must we more quickly solve some onerous problem?

We must also look at our present system and ask what should be Changed? Deleted? Streamlined? Added? “What if” and “if then” analysis must be evaluated vs benefits. And the result stated as a functional requirement. When we look at all the functional requirements we see the magnitude of the risk and the challenges. GIS cuts across departmental lines. We must solicit requirements from the right levels within the organization. Don’t assume the current procedure/organization is correct. Try to identify new opportunities.

Nearly all users decide to use a conversion service company because the cost is reduced, the accuracy is improved by the conversion specialist’s verification or QA software, and the data is converted much more quickly.
A clear scope of work is simply a must. It must include:

- Geographic Coverage
- Feature types and quantities
- Number/type of customers
- Existing records - types, quantities
- Project subdivisions
- New records - types, quantities
- Pilot project
- Backlog handling

The schedule must be defined and resources allocated.

Existing data specifications such as types of data, drawings, file records, computerized data, graphic specifications, methods and procedures documentation, computer record formats, and field inventory notes are required.

Deliverables must be clearly defined. Physical format — computer media or hard copy — plots, aperture cards, database reports, edit reports, or all of these must be specified.

The host-dependent GIS data structure — logical/physical file structures, working units, naming conventions, symbology, feature definitions, mapping standards, positional accuracy, and drafting standards must be clearly defined.

Database specifications — data structure (host/dependent), entity/attribute definition, data relationship and data sources, plus priorities of multiple sources, as well as valid attribute values/default values and critical/non-critical attributes must be defined.

With this we get to the point of decision making — are we going to be in the action, or are we just going to tell stories about the action? It takes a knowledgeable and a committed executive to decide to proceed.

**Technology innovations**

Innovations in conversion technology can improve not only the cost of converting to a GIS system, but the accuracy and timeliness of the conversion effort as well. Innovations come about in two ways. There are some that have great benefit within an existing functional area of the conversion process, and others that use new technology to redefine the conversion process itself.

An example of the first type of innovation (one that benefits an existing functional area) might be the use of artificial intelligence (AI) in automated editors to perform data purification and edits. An example of a new technology that essentially redefines the conversion process might be the use of scanning with pattern recognition and intelligent object association combined with record class analysis and clean-up. This could create an automated conversion process.

---

**Table 1:** Identifies various technologies that may either improve the typical conversion process, redefine the conversion process, or both.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Improve typical Conversion process</th>
<th>Redefine the Conversion process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial intelligence</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Scanning &amp; vectorization</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Voice data input</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Pattern recognition</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Neural networks</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Automated edgematching</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Statistical character</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>recognition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faster workstation</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>New CPU's</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Since conversion is usually a large part of an GIS project’s cost, there have been, and are now, several innovations being touted as “major breakthroughs”. It is important to recognize that, as a new technology proceeds from introduction through design, expectations rise greatly. Then, as applications are implemented, and as the new technology is integrated into the process and production use occurs, expectations are reduced to a practical value or proven reality (that may or may not be an improvement).

The conversion industry has seen this pattern occur frequently in the recent past. Innovations to date have been limited in their ability to reduce conversion cost. Most of the time, results have been lower than expectations. Nonetheless, the state-of-the-art in GIS data conversion has advanced dramatically in the past five years. Scanning and vectorizing is cost effective if the maps to be converted are up-to-date, complete, and of reasonable quality. There now exists a great deal of specialized conversion software that enables conversion operations to take place much faster than can be accomplished with system vendor software. Translation programs now exist that enable data to be converted on one system and translated to the user’s system.

While these are all welcome improvements in conversion technology, they have reduced the cost of conversion only enough to maintain its level at about 50% of the total cost of GIS implementation. This is because CPU, workstation, and GIS software costs have also decreased substantially in the past five years.

Conversion technology is clearly improvable. It is taken for granted that tomorrow’s technology will be far more powerful than today’s. But people must grow and improve as well, or the innovative technology will not mean much. Conventional wisdom tends to blame “the computer” for the foul up, but results rarely have much to do with the mechanism that executes the underlying math or logic. Instead, the credit or blame should go to the people who understand, or fail to understand, the problem they assign to a particular information processing system.

Let’s return to a perspective of where we are today in the world of computing. In the ’50s computing was number crunching. Data, and data management, was the focus from the early ’60s until the late ’70s. The Information Age began in the late ’70s. At precisely the same time GIS emerged as a unique information technology.

Looking forward we can see that very quickly our corporate information will become knowledge based. It will also be geography based. The application potentials are truly unlimited. The application of geographic knowledge based technology is an area we should all learn more about. I urge you to find out about it.

Let me take a moment to summarize.

GIS database construction - a summary
- GIS Database Construction is a construction process, not a conversion event.
- The process must be tailored to fit the user perfectly.
- It is a large, complex, detail-oriented job.
- It is both capital and people intensive.
- Accuracy, low cost and timeliness are the trademarks of success.
- GIS Database Construction enables GIS benefits to begin.
- With the records in order you have a solution that turns your data into profits.
Technology trends for GIS applications

Michael J. Fisher
Data General Corporation International
GIS Marketing Consultant
3400 Computer Drive
Westboro, MA USA 01580
Tel: 508-366-8911

Abstract

Both hardware and software development is under way, focused on improving performance and security issues realized when running database-critical applications, such as Geographic Information Systems (GIS). The marriage of multi-threaded relational databases and multi-processor based systems appears to be a trend in meeting the needs of the large GIS environment.

As a GIS database grows more sophisticated and increases in size, access to information and performing complex analysis create a burden on both CPUs and databases. When combined with the importance of protecting the most valuable aspect of a geographic information system - the data - these represent the most common problems for advanced GIS users.

This presentation offers insights into these developments by computer manufacturers and relational database vendors while outlining associated benefits. Industry standard benchmarks and other related specifications will be presented in support of the paper's findings.
D6 Scanning and conversion

Automated map data conversion
Wayne Coleman, Laser-Scan, Inc., Sterling, VA, USA

Impacts of scanning resolution on GIS data conversion
Lisa Klapwyk Crouser et al., Infotec Development, Inc., Portland, OR, USA

Scanning technology: considerations in the preparation of scanner-ready manuscripts
Daniel Haskell et al., Infotec Development, Inc., Portland, OR, USA

Data collection for GIS applications using scanning in conjunction with automated raster to vector conversion technology
Jeff Marcei, Scan Conversion Services, Burnaby, BC, Canada
Automated map data conversion

Wayne C. Coleman
President
Laser-Scan
45635 Willow Pond Plaza
Sterling, VA, USA 22170-4457

Abstract

GIS are dependent on, as well as hampered by, the availability of digital data. Historically, GIS applications development has been dependent on the time-consuming and expensive process of manual digitizing. However, today there are numerous companies providing a wide range of cost-effective scanning systems that convert hard copy maps to raster data. Additionally, there are automated systems that efficiently convert raster data to intelligent vector map data.

It is generally agreed within the GIS community, that the cost of converting hard-copy source maps into intelligent attribute-coded vector data constitutes 60 to 80% of the total GIS expenditure. The requirement of converting data for mapping and GIS can vary in complexity from simple “spaghetti” linework to multi-attribute object level topologically structured data.

The success of emerging automated data conversion technology is measured in the final quality assurance phase. If the converted data is not tested and proven to meet all specifications and requirements, the conversion is a failed exercise.

Incorrectly converted and misrepresented data is worthless. For this reason the evaluation and procurement of automated conversion systems should be as well planned and as thorough as the procurement of the GIS system itself.
Impacts of scanning resolution on GIS data conversion

Lisa Klapwyk Crouser
Infotec Development, Inc.

Dan Haskell
Infotec Development, Inc.

John Dabritz
U.S. Forest Service, Pacific Northwest Region
Infotec Development, Inc.
500 N.E. Multnomah, Ste. 329
Portland, Oregon 97232, USA
Tel 503-238-0034
Fax 503-238-1224

Abstract

Scanning technology can provide the opportunity for efficiencies in GIS data conversion. However, the extent of these efficiencies is directly affected by the resolution of the scanned image. Map scale, data accuracy, line work quality, and data intensity influence choosing a scanning resolution. Resolution, in turn, impacts machine sizing, processing, and storage requirements. These cause-and-effect relationships have been evaluated using Line Trace Plus (LTPlus) data capture software at Infotec Development’s GIS/LIS facility in Portland, OR. LTPlus was developed by the U.S. Forest Service in cooperation with the Soil Conservation Service. It operates under UNIX on a PC 386+ workstation. The “average” RAM and hard disk size of the workstation is in excess of 8Mb and 300 Mb respectively.

Introduction

The cost of converting data for use in a GIS is a concern for most users at some point in time. Scanning data for use in a GIS has always been an alternative but has become more affordable due in part to the increased processing power available in 386/486 PCs. These systems can provide users with the capability to convert the large data maps that natural resource analysis requires. Still, converting scan data requires a good size machine, and a budget conscious GIS user looking at scanning as a solution for data conversion needs to know what variables will impact system size, speed, and storage.

Scanning resolution is one of these variables. It is anticipated that scanning resolution impacts all phases of the data conversion process and carries over into the GIS analytical process. Validating its impacts on conversion processing and the final GIS output is the focus of this paper.

Infotec Development Inc. is a software engineering and integration firm that has been providing GIS/LIS services since 1981. The GIS/LIS facility in Portland, Oregon specializes in implementing GIS for natural resource management and has been providing GIS scan conversion services using Unix-based 386 systems since October, 1990.
Determining scanning resolution

The choice of a scanning resolution is influenced by data accuracy, line work quality, data intensity, and map scale. The first objective in selecting a scanning resolution is to choose a high enough resolution that it will not degrade the positional accuracy of the data. When a raster representation of a line is vectorized, the resulting vector can deviate slightly to either side of the original line. Therefore it is important that the total distance a line “drifts” as the result of vectorization is not more than the on-the-ground-accuracy. Since the deviation is usually one pixel in width there is a general starting rule for determining a scanning resolution. The pixel size should be, at most, half of the on-the-ground-accuracy of the original data.

Resolution can be calculated as follows

1. Determine the on-the-ground accuracy of the data
   Example: on-the-ground accuracy is 50’ (15.24m).
2. Divide the on-the-ground accuracy in half to determine the maximum pixel size.
   Example: 50’ / 2 = 25’ maximum pixel size required (15.24m / 2 = 7.62m).
3. Determine what distance is represented by a map inch (mm).
   Example: 124 000 scale is equivalent to 1” = 2000’ (1mm = 24m).
4. Select a dots per inch (dpi) scanning resolution
   Example 100 dpi (3.94 dpmm).
5. Divide the map inch distance (#3) by the scanning resolution to determine the size of the pixel.
   Example: 2000’ / 100 dpi = 20’ per pixel (24m / 3.94 dpmm = 6.08m).
6. Compare pixel size with the quotient in #2. Pixel size should be less than the quotient.
   Example: 20’ pixel < 25’ quotient from step #2 (6.08m < 7.62m).

Once this baseline is determined, the following factors may require increasing the scanning resolution (decreasing pixel size)

- Density of data on the manuscript
- Quality of linework
- Number and tightness of curves
- Number of low angle intersections

Setting up the analysis

In order to examine how scanning resolution impacts file sizes, editing times, and end product quality it was necessary to define the project parameters. Infotech's GIS/LIS facility expertise is in the conversion of natural resource data. This data rarely has on-the-ground accuracy that exceeds USGS National Map Accuracy Standards. The accuracy standard for a USGS 124 000, 7.5’ quadrangle is 40’.

Condition Class data was selected for the project. Condition Class maps contain productivity and size/age class information for major timber species and is representative of thematic information required in natural resource management. The Condition Class map was scan-ready and contained the following characteristics:
- On-the-ground accuracy is 60’ (18.29m)
- Drafted on a stable base media (mylar)
- Drafted using opaque black ink.
- 124 000 Map scale, USGS 7.5’ quadrangle
- Approximately 210 map objects (1144 segments)
- Attributes on a separate overlay
- Scan input format is a binary Run Length Compression file (.RLC)
- Scan is monochrome
- Output format is an ASCII file, USGS Digital Line Graph 3 (DLG-3)

The analysis was performed on a 386/20 LTPlus Workstation. The workstation components are listed in Appendix A.

First the Condition Class data was scanned at four resolutions, 75, 100, 150, and 200 dpi, which corresponds to 2.95, 3.94, 5.90, and 7.87 dots per mm. Given that the on-the-ground accuracy of the data was estimated at 60’, the four scan resolutions easily fit our primary criteria for determining a resolution. All four provide a pixel size that is less than half the on-the-ground accuracy (30’). The relationship of resolution to pixel size and on-the-ground accuracy and scale is represented in Tables 1-1 and 1-2.

Once the map had been scanned, it was processed using LTPlus software. LTPlus was developed by the U.S. Forest Service in cooperation with the Soil Conservation Service. LTPlus was not designed to be a GIS, rather it was designed as a GIS “Front-end” capable of converting data for use in a number of GIS packages.

In simplest terms, LTPlus begins with raster input and ends with topologically structured vector output. This output can then be imported into a GIS for database linking and analytical processing. The basic LTPlus conversion process used for the project is outlined below.

1. Set up the LTPlus map file
2. Gather the input data. LTPlus recognizes data from three main places
   - Importing a scan file
   - Manually digitizing (onto raster)
   - Adding data from an external vector source (onto raster)
3. Edit the raster
   - Pass one - separate touches
Table 1-1 Relationship of resolution to pixel size and On-The-Ground Accuracy and Scale - U.S. equivalent conversion

<table>
<thead>
<tr>
<th>Scale 1:24,000</th>
<th>Dots per Inch</th>
<th>Feet per Pixel</th>
<th>Maximum on-the-ground accuracy represented</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&quot;=2000'</td>
<td>75</td>
<td>26.67</td>
<td>54'</td>
</tr>
<tr>
<td>1&quot;=2000'</td>
<td>100</td>
<td>20.00</td>
<td>40'</td>
</tr>
<tr>
<td>1&quot;=2000'</td>
<td>150</td>
<td>13.33</td>
<td>27'</td>
</tr>
<tr>
<td>1&quot;=2000'</td>
<td>200</td>
<td>10.00</td>
<td>20'</td>
</tr>
</tbody>
</table>

Table 1-2 Relationship of resolution to pixel size and On-The-Ground Accuracy and Scale - Metric System Conversion

<table>
<thead>
<tr>
<th>Scale 1:24,000</th>
<th>Dots per mm</th>
<th>Meters per Pixel</th>
<th>Maximum on-the-ground accuracy represented</th>
</tr>
</thead>
<tbody>
<tr>
<td>1mm=24m</td>
<td>2.95</td>
<td>8.14</td>
<td>16.3 m</td>
</tr>
<tr>
<td>1mm=24m</td>
<td>3.94</td>
<td>6.08</td>
<td>12.2 m</td>
</tr>
<tr>
<td>1mm=24m</td>
<td>5.90</td>
<td>4.07</td>
<td>8.2 m</td>
</tr>
<tr>
<td>1mm=24m</td>
<td>7.87</td>
<td>3.05</td>
<td>6.1 m</td>
</tr>
</tbody>
</table>

4. Assemble to create vectors (topology) and quality control
   • may return to Step 3
5. Edgematch and snap data to coincident lines in external data
   • may return to Step 3 or 4
6. Attribute, attribute a second time, then compare for accuracy
7. Output

Integration of vector data from external sources, edgematching and attributing were not evaluated as they are primarily vector processes and their relationship to scan resolution is minimal.

The effect of resolution on file size

To begin this project, the map data was scanned at standard size (24"x30") even though the actual data only covered approximately 18"x24". This was done to establish the same scanning area at all four resolutions. The data was scanned into Run Length Compression (RLC) format which compresses the data and reduces file size. Still, the high resolution 200 dpi scan file is over 2.5 times larger than the 75 dpi scan file. When the scan files were imported into LT Plus, and the data is decompressed for use, the 200 dpi LT Plus raster file is over 7 times larger than the 75 dpi LT Plus raster file. Table 2 shows the original scan and LT Plus raster sizes and LT Plus import times.

Table 2. Impacts of resolution on .RLC Scan and LT Plus raster Files

<table>
<thead>
<tr>
<th>Scanning Resolution</th>
<th>75 dpi (2.95 dpmm)</th>
<th>100 dpi (3.94 dpmm)</th>
<th>150 dpi (5.90 dpmm)</th>
<th>200 dpi (7.87 dpmm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resulting Pixel Size</td>
<td>26.67'</td>
<td>20.0'</td>
<td>13.33'</td>
<td>10.00'</td>
</tr>
<tr>
<td>.RLC Scan File size of an E size drawing (24&quot;x30&quot;)</td>
<td>0.165 mb</td>
<td>0.219 mb</td>
<td>0.329 mb</td>
<td>0.440 mb</td>
</tr>
<tr>
<td>Comparison of .RLC Scan File Size (Ratio to 75 dpi)</td>
<td>1 : 1</td>
<td>1.33 : 1</td>
<td>1.99 : 1</td>
<td>2.66 : 1</td>
</tr>
<tr>
<td>LT Plus Import Time (Ratio to 75 dpi)</td>
<td>1:1</td>
<td>1:2</td>
<td>1:4</td>
<td>1:6</td>
</tr>
<tr>
<td>LT Plus Raster Size</td>
<td>0.504 mb</td>
<td>0.900 mb</td>
<td>2.034 mb</td>
<td>3.600 mb</td>
</tr>
<tr>
<td>Comparison of LT Plus Raster File Size (Ratio to 75 dpi)</td>
<td>1 : 1</td>
<td>1.78 : 1</td>
<td>4.04 : 1</td>
<td>7.14 : 1</td>
</tr>
</tbody>
</table>
It is possible to reduce file size without altering the resolution of the scanned data. Some scanners allow the user to define or crop the map area to be scanned. Cropping a scan minimizes “null space” outside the data area and the result is a smaller scan file.

GIS scan conversion software, such as LTPPlus, can further reduce raster file size. Once the data is in LTPPlus, the user registers the raster using the grid ticks of the scanned manuscript. After the map geodetics have been entered and the registration verified, the user can then “margin from the registered raster”. Margining will remove any or all data outside of the registration. For example margin 20 erases all data that lies greater than 20 pixels outside the registration (Figure 1).

If a margin 20 is used and the pixel size is 20', a 400' wide border remains on all sides of the map that can be used for edgematching. For this project a 400' margin was used and the resulting LTPPlus raster files sizes were reduced over 40% (Table 3).

**The effect of resolution on edit and processing times**

Following the generalized procedures previously outlined, the next step is editing the raster data in LTPPlus. On the first editing pass, low angle intersections are opened and false touches are separated. Lines that do not touch on the manuscript but do touch in the raster are referred to as false touches. False touches and low angle intersections are separated to prevent the creation of erroneous polygons. Then the data is thinned to a single pixel width. On the second pass, if the data is polygon, data gaps are closed and spurious data is deleted.

Separating false touches and low angle intersections on the first pass and repairing identified data gaps on the

---

**Figure 1** Example of the effect of margining to reduce file size

---

**Table 3. Results of retaining only the registered map data and a 400' margin**

<table>
<thead>
<tr>
<th>Scanning Resolution</th>
<th>75 dpi (2.95 dpmm)</th>
<th>100 dpi (3.94 dpmm)</th>
<th>150 dpi (5.90 dpmm)</th>
<th>200 dpi (7.87 dpmm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTPlus Raster Size</td>
<td>0.504 mb</td>
<td>0.900 mb</td>
<td>2.034 mb</td>
<td>3.600 mb</td>
</tr>
<tr>
<td>Pixel Size</td>
<td>26.67'</td>
<td>20.0'</td>
<td>13.33'</td>
<td>10.00'</td>
</tr>
<tr>
<td><strong>margin from registered raster (400')</strong></td>
<td>16 pixels</td>
<td>20 pixels</td>
<td>30 pixels</td>
<td>40 pixels</td>
</tr>
<tr>
<td>Resulting LTPlus Raster Size</td>
<td>0.294 mb</td>
<td>0.512 mb</td>
<td>1.143 mb</td>
<td>2.025 mb</td>
</tr>
<tr>
<td>Reduction in File Size</td>
<td>42%</td>
<td>43%</td>
<td>44%</td>
<td>44%</td>
</tr>
</tbody>
</table>
second are done interactively by the user. The number of false touches and data gaps increased at lower resolutions as shown in Table 4. This was expected and it was believed that although there would be more errors the editing time would be offset by faster graphics regeneration after editing zooms. By and large this was true, but not to the extent expected. LTPlus has four pre-defined view levels and at 75 dpi, the views were either too far out to get a good feel for the error or too close to see the data trend. As a result it was necessary to increase the number of zooms and some efficiency was lost.

While the number of false touches and data gaps were less at the higher resolutions editing was slower. This is because more pixels represent a line and more editing decisions can be made. For example, a line one pixel wide at 75 dpi will be three pixels wide at 200 dpi. If the line creates a false touch, turning off one pixel and turning on another to separate the false touch presents no major editing choice to the user. On the other hand at 200 dpi most users begin to “cartographically” edit turning off and on pixels to not only separate the false touch but also create a line that is more visually pleasing. In addition the graphics took more time to regenerate after a zoom, but having to zoom in and out in order to get a proper editing view was unnecessary.

After false touches are separated the data is thinned to a single pixel width in an automated process. The amount of machine time required to thin the data is dependent on the resolution. For the 75 dpi map three passes were required to thin the data to a single pixel width. At 200 dpi, six passes were required. Table 4 compares overall time for completing the editing process at 100, 150, and 200 dpi to 75 dpi.

Once the raster data is edited and thinned it is vectorized and topologically structured. Table 5 compares the time for topological structuring and saving the data. Times are presented as a ratio comparing 100, 150 and 200 dpi to 75 dpi. Again, the higher the resolution the longer the time to create the topology. The time required to save a map was also compared. This is important because it is common practice to save on at least three different occasions in the conversion process. For the 200 dpi map, each save takes over five times longer to complete than the 75 dpi map. This means that completing a map at 200 dpi would take at least 5 times longer if all other factors remained the same.

<table>
<thead>
<tr>
<th>Table 4. Results of the Edit Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scanning Resolution</td>
</tr>
<tr>
<td>Number of false touches</td>
</tr>
<tr>
<td>Number of thinning passes</td>
</tr>
<tr>
<td>Number of data gaps found</td>
</tr>
<tr>
<td>Raster Editing Time, man &amp; machine (Ratio to 75 dpi)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5. Results of Topological Structuring and Map Saves</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scanning Resolution</td>
</tr>
<tr>
<td>Topological Structuring Time (Ratio to 75 dpi)</td>
</tr>
<tr>
<td>Map Save Time (Ratio to 75 dpi)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6. Results of Exporting GIS Vector File</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scanning Resolution</td>
</tr>
<tr>
<td>Output File Size - DLG-3</td>
</tr>
<tr>
<td>Comparison of DLG-3 File Size (Ratio to 75 dpi)</td>
</tr>
</tbody>
</table>
The effect of resolution on vector output

The vector data was exported in an ASCII format, USGS Digital Line Graph-3 (DLG-3). USGS DLG-3 file is composed of three separate but related elements: nodes, lines, and areas. Table 6 shows the size of the DLG-3 files and compares the 75 dpi file to the higher resolution files. Predictably, the size of the final vector output file increases with resolution. This relationship will continue once the data is imported into a GIS and increased analytical processing and graphics times generation can be expected.

The vector data is displayed in Figures 2-1 through 2-4. These figures show a section of the map plotted at the original drafted scale of 1:24,000. To the naked eye at the original map scale, the vectors appear similar. It is possible to see the effects of the resolution at tight corners and low angling lines. Also when the data is presented at larger scales, the smoothing differences become more pronounced. Still, all four plots maintain the on-the-ground accuracy of the data. When the plots are compared with the original drafted data, the 100, 150, and 200 dpi resolution maps consistently lie within one line width of the original data. The 75 dpi map achieves this on over 85% of the map. See Figure 2-1 through 2-4.

Conclusion

Scanning resolution impacts all phases of the data conversion process and carries into the GIS analysis process. The project attempted to evaluate the general effect of scanning resolution on LTPlus data conversion by eliminating activities that could have minor influence at certain resolutions. Varying scanning thresholds and thinning vertices could have optimized the higher resolution results.

For this project, machine processes not manual editing were most impacted by changes in resolution. Primary machine processes include importing a scan file, thinning raster data to a single pixel width, vectorizing, creating topology, thinning vertices, exporting the vector data, and saving. Carrying a map through from import to export may require executing a machine process multiple times. For example, the user may save after redefining the map margin, before deleting all remaining spurs on a polygon map, and at the end of the session. Or the user may find an error after vectorizing and creating topology which can only be repaired by editing the raster map and vectorizing and creating topology again. Saving three times and assembling vectors twice would take over 4.5 times longer at 200 dpi than at 75 dpi. The cumulative effects of these repetitive processes should not be underestimated when selecting a scan resolution.

For this project editing the 200 dpi map took twice as long as the 75 dpi map while creating topology and saving took over 4.5 times as long. A key reason why the resolution impacts to manual editing time were so low was because the data was scan ready. Scanning reproduces the original manuscript. If a map is well manuscripted, lower resolution scans will have a few more false touches and data gaps than higher resolutions scans. For this project, the 75 dpi map had 8 false touches and 14 data gaps while the 200 dpi map had 5 false touches and 8 data gaps. The total number of number of errors that were the result of manuscripting and/or scanning was less than 2% of the total number of data segments.

File sizes cannot be overlooked, as resolution increases so does file size. This is true throughout the conversion process and knowing how the software handles the raster and associated files is important. When a map is acquired in a program like LTPlus, LTPlus makes a duplicate of all associated files (21). It also copies a number of files into RAM including the raster image, vector image, edgematching vectors, registration, and attribute information. While it is highly advantageous to compress files when they are not in use, the raster map and its associated files will not be compressed during conversion.

Finally, all resolutions produced data that fell within the positional on-the-ground accuracy of the original map. There is a visual difference between resolutions with the 200 dpi map producing the most flowing linework. The 75 dpi map produced the most raster-like linework with 100 and 150 dpi maps falling somewhere between. It is the GIS users decision as to whether the vector ends justify the resolution means.
Appendix A

LTPlus system requirements

Minimum LTPlus requirements

- Hardware
  PC
  80386 (or better)
  Hard Disk _300 Mb
  Memory _8 Mb
  80387 Math CoProcessor
  BIOS & register-compatible
  VGA graphics card
  One parallel port, two serial ports
  Floppy disk &/or tape drive

Monitor/Graphics
  Color multisync
  (400x600 resolution)

Mouse for scan editing
  3 button, serial

Scanner
  Producing standard scan formats (e.g. .RLC, )
  or

Digitizer
  Varied

Plotter
  Varied (HPGL)

- Software
  Unix or Xenix Operating System
  Text Editor (vi)
  LTPlus

Infotec Analysis Components

- Hardware
  PC
  Compaq 386/20e
  Hard Disk - 320 Mb
  (Xenix Partition = 288 Mb)
  (DOS Partition = 32 Mb)
  Memory 10 Mb
  80387 Math CoProcessor
  BIOS & register-compatible
  VGA graphics card
  1 parallel port, 8 serial ports
  Floppy disk & tape drive

Monitor/Graphics
  NEC 20" multisync color
  (1024x768 addressable)

Mouse for scan editing
  Logitec 3 button, serial

Scanner
  Houston Instruments LDS 4000
  (Monochrome)

Plotter
  Hewlett Packard 7585B

- Software
  Xenix Operating System v. 2.3.2
  Text Editor (vi)
  LTPlus version 2.30
  MS-DOS version 3.2 (for scanning)
Scanning technology: considerations in the preparation of scanner-ready manuscripts

Scott Oerding, GIS Specialist
Daniel Haskell, Technical Leader
Infotec Development Inc., GIS/LIS Facility
500 N.E. Multnomah, Suite 329
Portland, Oregon. 97232 USA
Tel: 503-238-0034
Fax: 503-238-1224

Abstract

With the increased availability of scanning technology, more organizations are recognizing it to be a "quick and easy" means of GIS data capture. However, the speed and accuracy of this approach is strongly tied to the quality of the source manuscripts. Not all manuscripts are scanner ready. This paper first examines the considerations necessary for the creation of scanner ready source maps. Secondly, it discusses criteria for evaluating the suitability of existing manuscripts for this type of capture.

Introduction

Scanning technology provides a high tech solution to the problem of GIS data capture. In a short period of time a manuscript can be run through a scanner, loaded in the editing software, cleaned up, structured topologically, attributed and added to a database. In fact, given that much of this processing can be performed by the computer with minimal user input it could be said that scanning is also a "smart" technology. Unfortunately its speed and high tech smarts depend heavily on a low tech issue: manuscript quality. Poor quality or inappropriately drafted manuscripts can negate all of its advantages. To receive the full benefit of scanning technology it is necessary to start with scanner ready manuscripts.

What makes a manuscript scanner ready? In the last 14 months Infotec's GIS/LIS Center has used scanning to capture over two thousand maps for the U.S. Forest Service and, based on that experience, would like to offer a few suggestions. These suggestions have been culled from our experiences scanning with a Houston Instruments LDS4000 Plus scanner and editing with the Line Trace Plus (LTPlus) software.

Defining the problem

Poor quality or inappropriately manuscripted maps require much more time to capture. In many cases the capture time can even equal what could be achieved through manual digitizing. This happens because scans of poor manuscripts require a large amount of manual editing before they can be processed: A human must go through the scan, separating and/or completing the object images to make them discernible for the computer. While scanning is usually computer intensive, poor manuscripts make the process labor intensive. The goal in creating scanner ready manuscripts, then, is to draft images that can be processed with as little manual editing as possible.
There are two occasions when it is important to consider scanner readiness: (1) when preparing new maps and (2) when evaluating pre-existing maps for capture. The first section of this paper addresses how to draft scanner ready manuscripts. The second section addresses how to evaluate existing manuscripts.

**Drafting a solution**

The best way to ensure that your manuscripts are scanner ready is to draft them that way. To do this it is necessary to “think digital”. This is accomplished by imagining how the manuscript will look to the scanner. The following is a list of areas of concern in creating a manuscript that will be easily scannable:

**Stable base media** Most GIS professionals know that paper and vellum are not appropriate media for manuscripts. When the humidity and temperature in the room changes so does the size of paper or vellum maps in that room. They are also more vulnerable to stretching and tearing. However, when manuscripting for a one time use, like GIS data capture, it is tempting to cut corners. Don’t. Mylar may be more expensive, but it will help safeguard the accuracy of the data. An opaque, single matte mylar works very well.

**Ink vs pencil** Pencil usually works rather poorly for scanning and should be avoided. Penciled lines often smear, creating an undesirably dark background which could obscure the linework. Also, even though penciled lines look dark to the human eye, the scanner often “sees” them as inconsistently opaque causing gaps and splotches. A bold black pen seems to be the best choice.

**Linework** It is important that linework be drawn as evenly and smoothly as possible. Overly thick sections may clump together in the scan, obscuring the linework. Overly thin lines may be lost or only partially captured. Dashed lines are scannable but not recommended as they often require additional editing time to connect. Smears and smudges on the media can make sections of the linework look like a muddy mess to the scanner. In all it is best to use a pen that produces an opaque black line and try to maintain a uniform level of pressure while drafting. The size of the pen point would vary depending on the estimated accuracy of the data.

**Registration points** It is important the map contain four accurately placed registration points, preferably outside the data. If these points are not placed in their correct geodetic locations the data will be distorted. The further off the placement the worse the distortion. Ideally each map would have four simple “plus sign” tics inked onto each corner of the map. Stick-on tics are not recommended because they can be moved by the traction wheels of the scanner.

**Colored lines** When using a monochrome scanner, it is best if all the linework be drafted with the same color pen. Almost all colors of ink will scan, but some scan better than others. When multiple colors are used it may be impossible to make the scan light enough to capture one color without making the others indiscernible. Yellow comes the closest to being non-scannable and works well for lines which are included for reference only (such as the perimeter of a quadrangle).

**Object attributes on a separate sheet** It works best to draft attributes onto a separate sheet. Attributes included on the manuscripts will require editing time to remove. The separate sheet could be a registered clear acetate overlay or a blue line copy of the manuscript. The attributes should be drafted in a dark ink for maximum legibility.

**Map name/number** Whenever possible draft the map name/number somewhere within the tics. This allows the worker to verify the identity of a scan file at edit time.

**Gaps in linework** The elimination of gaps in the linework of both polygon and line data is important. All lines should extend to meet boundaries and/or the map's geodetic or planar edge. This practice can eliminate the need for interpretation of the data by workers during the editing process and, thereby, speed editing time.

**Edgematch all mylars** During the manuscripting process carefully edgematch each mylar with its neighbor. Taking the time to ensure that the data matches across each map sheet will greatly reduce the need for interpretation, editing and re-work at the time of capture. Adding the name or number of each edgemap to the corresponding edge of the manuscript will speed the edgematching process as well as make future updates easier.

**One theme per manuscript** Composite themes generally do not lend themselves well to digital data conversion through scanning. They tend to be very complex, confusing and usually require a great deal of time to interpret and edit. This is especially true if the manuscripts were drafted with a single color pen and/or they are being scanned with a monochrome scanner. Each theme must be carefully separated from other background information. These maps also require more processing time and storage capacity.

**Evaluating pre-existing manuscripts**

Unfortunately, there are many organizations that have large catalogs of pre-existing manuscripts they need converted to a GIS digital format. More often than not these manuscripts were not drafted with GIS, let alone scanning in mind. While the ideal solution might be to custom manuscript data explicitly for scanning, it may not always be feasible. In these cases it is important to be able to determine whether or not scanning is feasible at all.

When evaluating pre-existing manuscripts for scanning it is useful to use the following approach:
(1) Evaluate the maps against the suggestions listed in the "Drafting a solution" section (see appendix A for a checklist).

(2) Use the suggestion list to determine ways in which the maps could be made more scanner ready.

(3) Use the list in the "Handling Special Cases" section (below) to determine if your maps may be captured using a special procedure.

(4) Beta test the capture process with a representative sample of the maps. Determine whether the time and resources that are required will be acceptable. Consider both human and computer resources.

(5) Consider re-manuscripting. If it appears that capturing your existing maps would require more time or resources than are acceptable, it may be preferable to trace the maps onto new sheets and scan the new sheets. In practice we have found it to be more cost effective to trace and scan than to hand digitize. This may sound outrageous, but digitizing is a slow tedious and inaccurate process compared to scanning.

Handling special cases

Below is a list of situations that can make otherwise perfect manuscripts un-scannable or poor manuscripts scannable. It addresses a few of the special cases that can make or break the chances of capture through scanning.

Multi-color manuscripts Multi-color manuscripts can be readily scanned if you have a color scanner. Some color scanners come bundled with color image processing software which allows users to improve the image before taking it into the GIS capture software.

Multi-color composites When the manuscript contains a number of themes which are coded by color linework, it may be possible to separate the themes using a color scanner. If the roads, for example, are in red and the rivers are in blue the two sets of linework could be separated through use of the image processing software bundled with the scanner.

Smudged, smeared, or dirty manuscripts Pay careful attention to the condition of the manuscripts. An excessively smudged, smeared or dirty map may be useless for scanning. It may be possible to salvage such maps using variable thresholding on a monochrome scanner. A better option might be to perform some editing on a grey scale or color image of the map before taking it in to the GIS capture software.

Paper and pencil From an accuracy standpoint paper maps and pencil linework are horrendous. We have seen variation in paper manuscripts of greater than a quarter inch (which equates to a greater than 500 foot discrepancy on a 1:24000 scale theme). As for pencil, it smears worse every time it goes through the scanner. The more the linework smears the less accurate it becomes. The smearing may eventually make the background so dark it obscures the data.

Monochrome composites If the manuscripts in question are monochrome composites consider either digitizing or tracing the data on to a separate sheet. Such maps are labor intensive in the extreme.

Orthopho-quad and aerial photos Some organizations have drawn data directly onto orthophoto-quad or other types of aerial photos. Orthoquad photo have dark backgrounds and aerial photos have cluttered backgrounds that make them unsuitable candidates for scanning.

Conclusion

Scanning of geographic data for use in a GIS can be a quick easy process where the computer does most of the work; or it can be a slow and tedious one, requiring extensive human input. The nature of the experience hangs on manuscript quality. Properly manuscripted maps allow the user to take full advantage of their computer’s intelligence. Poorly manuscripted maps require much user time and input.

Appendix A

A brief checklist of points to consider when preparing manuscripts or evaluating existing manuscripts for scanner readiness:

___ Stable base media
___ Ink linework
___ Consistent linework
___ Four accurately place registration points
___ Single colored lines
___ Object attributes on a separate sheet
___ Map name/number included inside grid tics
___ No gaps in linework
___ Data edgematched with adjacent maps
___ One theme per manuscript
Data collection for GIS applications, using scanning in conjunction with automated raster to vector conversion technology.

Jeff W. Marcel (presenting author)
Scan Conversion Services Inc.
4228 Manor Street
Burnaby, BC
V5G 1B2
Tel: (604) 439-0054
Fax: (604) 439-1459

Abstract

With the 'coming of age' of scanning technology, considerable efficiencies in data collection for GIS applications can now be realized. Perhaps even more significant than the scanning component of the process (although the quality and accuracy of a scanned image is critical) are the developments in raster to vector conversion technology. This technology utilizes the end product of a scanner (raster or bit map data) and produces an intelligent vector based file. It can then be cleaned and built to produce a fully functional, accurate GIS product.

Proper utilization of the automatic capabilities, found in both raster to vector and GIS software, can result in the conversion of source maps to fully structured GIS files with very little operator time required. Consequently, the cost of data collection can be substantially less than other methods of conversion. To ensure the most cost effective execution of such a project, it is imperative that several critical steps be followed to maintain conversion efficiencies.

Introduction

In recent years it would appear that the level of knowledge obtained by many digital graphic users on the subject of scanning has increased substantially. This is likely due to their exposure to the technology either directly or through the increased amount of literature and articles available on the subject. As a member of a scanning service bureau, namely, SCAN Conversion Services Inc., the writer respectfully makes the following observation. Many digital mapping users, feel they already know all they need to, on the subject of scanning. However, many still have some misconceptions about the technology, and its capabilities. Case in point: often managers of a conversion project have been heard to say, "But, with scanning, I'm going to end up with everything on one layer (level, feature class, coverage, etc...)" or, "all the line work in a scanned file is made up of small broken line segments...", or "but, a scanned file is in a bit map format, what can I do with that?" If approached correctly, the end product produced through a scanning system should satisfy the end user, both cartographically and functionally.

With the objective of clearing up these misconceptions in mind, this paper is not intended to impress the
reader with a lot of technical jargon, but to examine the issues of scanning. Hopefully, some constructive suggestions can also be made, to help the technology produce it’s full capabilities in data collection.

**Raster to vector technology**

The principal by which most scanning systems operate is fairly simple in concept.

1) Using a high resolution scanner (available in all shapes and sizes) which can physically accommodate the desired map, the document is optically read and that image is converted to a bit map or raster image. This raster image is much like a photocopy in a digital format.

2) The raster image is then converted to a vector image. This is accomplished in various different ways depending upon what system is used. Some systems require considerable interactive work by an operator, such as, on screen raster tracing/digitizing. Other systems have automatic vectorization, in which case, the raster to vector conversion is done with little or no operator intervention. Such systems utilize artificial intelligence and parallel processing to perform the vectorization. Yet, other systems incorporate techniques from both of the above procedures to carry out the conversion.

3) In systems with automatic vectorization, the vector file is then cleaned up interactively by an operator, using highly specialized software. Any areas in the map which the vectorization system could not resolve, are repaired. This produces a cartographically correct digital image.

4) Using a compatible exchange format, the digital file can then be, imported, transformed, cleaned, built, edgematched and further processed as required, in a GIS environment. These functions can also be carried out using utility software, working in conjunction with CAD or Digital Mapping software.

**Project analysis for optimum scanning results**

When approaching a digital mapping project for AM/FM or GIS applications, there is one question that certainly now enters the minds of most project managers. “Is scanning a feasible method to accomplish the desired end result?” The answer to this question is a resounding, “yes!” In this regard, there are some key questions which must be considered, when evaluating the place of scanning technology, in a conversion project.

Perhaps the most important aspect, of evaluating a conversion project using scanning technology, is to approach the exercise with an open mind. Of course there will be specific goals or objectives that are desired upon completion of the project. However, be reasonably flexible regarding new methods to reach that end.

Rumours you may have heard might also sway your opinion on scanning. A colleague’s experience with scanning, may be an isolated incident. It could be the result of several variables which were not completely explored. Consequently, the end product was not completely satisfactory. Of course this can happen regardless of what conversion process is used.

As with any situation, when dealing with scanning as a means of conversion, there are always many options open to the user. Specifically, with regards to how a desired end product will be accomplished.

**Accuracy of scanning**

The first question a digital map user must ask himself/herself, is regarding the purpose of the data, and how accurate it needs to be. More specifically, will the data be representational or positional. Also, will the data be related to an existing base map or will it be the base map itself? We all want to have “perfect accuracy” in our mapping. However, what usually brings us back to reality, is the cost to achieve this kind of accuracy. Especially, when quotes come in and budget day comes around. In other words, the more accurate the desired end product, the larger the price tag. Perhaps this is scanning’s biggest advantage. If handled properly, scanning can accommodate the accuracy issues while coming in at a reasonable cost.

There is one key factor to remember when dealing with scanning technology. Scanning will do a wonderful job of reproducing a hard copy map with precisely the same accuracy that was present in that original document.

Many times digital mapping users are convinced that their source material is perfectly accurate. This is until it has been scanned and errors reveal themselves. For instance, when a mylar plot of the converted file is produced, it lines up perfectly with the original map. However, when distances are measured within the file, they do not correspond perfectly with the dimensional annotation in the original. Usually those distances are very close... but not perfect. If the source material was produced using manual systems incorporating human intervention this error can only be expected. Consequently, without additional processing, scanning technology cannot be expected to produce a file more accurate than the original.

**Augmenting the accuracy of scanned files**

As stated earlier, there are always several options opened to the user, to achieve a desired result. One technique which can greatly increase the accuracy of a scanned file,
is to utilize the "Rubber Sheeting" or transformation capabilities found in most GIS systems. As mentioned earlier, this type of capability is also available from several vendors as a separate utility program to run in conjunction with existing application software.

The procedure to increase the accuracy in a map is, to tie "control", such as UTM or Latitude and Longitude coordinates to geographic features in a digital file. Upon inputting these coordinates into the transformation software, it will massage the map relative to the supplied coordinates. If a tic or grid system is used as control, those features will be moved to their respective, precise coordinates. All other detail in the file is then transformed relative to that grid or tic system. Many transformation software packages can calculate an unlimited number of control points. The more control points used in a map sheet, the better the result.

In the final analysis, prior to using any special techniques and software to resolve accuracy issues, the determining factor of a successful scanning project, is the quality and accuracy of the source material that was used.

Source document preparation

The preparation of source maps plays an important role in determining what level of success may be achieved in a scanning project. This preparation can facilitate the best production efficiencies, resulting in cost benefits for the overall project.

The first rule in scanning is, not to try and shove something down the scanners throat, that it doesn't like. The first source material chosen for a project, may not be suitable for scanning conversion, in its initial state. This may be due to the quality of the original, or how the features in the map where compiled. Some streamlining of the system used in source preparation may be needed. Or, by using a different source map, which lends itself better to scanning, the conversion costs may be affected substantially.

The objective of source preparation, is to result in material which allows for the full utilization of the automatic vectorization features in the scanning process. The efficient conversion of entities in a map, decreases the need for operator time. Such time, is required to clean up areas of the map, which the automatic vectorization system left unresolved.

The following are general guidelines, for choosing scanning source material. It is best to look for maps processing these characteristics.

a. Clean line work — The less background noise, or fuzziness of linework present in a source map, the easier it is for the scanner to interpret the information. Mylar or some other stable media is preferred to produce the most accurate results.

b. Intersecting lines — The more lines that intersect in a map, the higher the chance that the automatic vectorization may break the line or continue along the wrong line. If this is widespread, more cleanup will be required, which affects the price of the conversion. One method which has been found to be effective, is by separating certain features onto different source maps. i.e.) Mylar Separations

1. Forest polygons
2. Planimetry
3. Hydrography
4. Contours

Separating features in this way, eliminates the number of intersecting lines on a map when a compilation of features are required.

Energy Mines and Resources, Mapping and Remote Sensing Sector in Ottawa has utilized this in the scanning of 1:250 000 and 1:50 000 NTS maps with considerable cost benefits.

c. Line types — A clean solid line is the preference of most scanners. Generally, any dashed or dotted line will either require tracing or repair. The clean up software used in conjunction with scanning can quickly repair these lines to a continuous solid line very efficiently. The end product will always be excellent (depending upon the operator), but the price will be affected somewhat.

Pilot projects

Enough cannot be said for the importance of pilot projects. There are many areas where a conversion project manager, or the subcontractor could misinterpret the requirements of the project. Of course, this holds true no matter what the project, or method of conversion.

Carrying out a small test, or pilot project, allows the scope and details pertaining to the total project, to become much more defined. The methodology to be followed for the duration of the project, can be streamlined and set in place. A pilot project simply allows for some good constructive communication at all levels of the project hierarchy. It can also determine, if in fact a contractor has the resources or expertise to effectively carry out the total project.

Meshing of scanning and GIS technologies

Certainly scanning has come along way, and the capabilities of the technology are far beyond what they were just a few years ago. However, perhaps equally as important, is the evolution in understanding of the technology by it's users.
Case study

A fairly important realization at SCAN Conversion Services Inc. has been brought to light. It was determined that we were cleaning up scanned files too much for GIS applications. We had always been very careful to produce cartographically correct digital maps as an end product. Consequently, we had cleaned up all vector files until they were a thing of beauty.

While thinking through a proposal for a client, regarding the conversion of several timber and soil type maps, we asked ourselves the following. The client wanted the end result to be in ArcINFO format. Therefore, did we need to clean up the vector file so meticulously if it was to be topologically cleaned later? In other words, could we leave breaks in lines, and intersections which the scanning system left unresolved, if the cleaning function in ArcINFO could do it for us?

We ran some tests, and determined that this was in fact the case. By setting both the dangle, and fuzzy tolerances at appropriate settings, the errors left purposely in the file where repaired automatically. This resulted, in very little time spent cleaning up the files at the scanning level. The same amount of time was spent at the GIS level of the conversion process. This is because, that cleaning function would have needed to be undertaken in the GIS, regardless.

The result of this process refinement, has been a savings in time and money. This has been passed onto that client, and all of our other GIS clients. Using this method of conversion, SCAN Conversion Services Inc. has completed numerous map sheets for B.C. Ministry of Forests, Re-Inventory Contractors. Both cost effectiveness and a high level of quality have been found to be the results.

Conclusion

Scanning has proven itself to be an exiting and effective form of AM/FM and GIS data collection. As opposed to being an alternative form of conversion, as it has been perceived in the past, scanning is now becoming the preferred method of conversion for many applications. This is due to the considerable time and cost savings that can be realized from using scanning. Highly accurate digital mapping can be achieved, if scanning technology is utilized properly.

It is also worth noting that, at last years GIS’91 conference, there were only a few papers on scanning technology. For the most part they were lumped in with other unrelated presentations. This year GIS’92 feels the technology is important enough to represent it’s own exclusive scanning forum. It looks like scanning is here to stay.
D7  Integrating images with GIS

Utilizing satellite data and GIS to map land cover change
Paul Maus et al., Pacific Meridian Resources, Emeryville, CA, USA

Replacing aerial photo interpretation in resource inventories with integrated digital images and GIS
Frank Hegyi et al., Ferribill Technologies Ltd., Victoria, BC, Canada

Managing forest resources with raster tools
Andrew Bury, ERDAS, Inc., Atlanta, GA, USA

Using ancillary data in post-classification modelling to increase the accuracy of conifer species classifications from LANDSAT data
Mike Golden et al., Pacific Meridian Resources, Portland, OR, USA
Utilizing satellite data and GIS to map land cover change

Paul Maus
Vaughan Landrum
Jan Johnson

Pacific Meridian Resources,
In Residence at Nationwide Forestry Applications Program,
USDA Forest Service, Salt Lake City, UT 84116

Henry Lachowski

Nationwide Forestry Applications Program,
USDA Forest Service, Salt Lake City, UT 84116

Bruce Platt
Mike Schanta

Mark Twain National Forest,
USDA Forest Service, Rolla MO

Abstract

Monitoring vegetation changes in the landscape over time is an important concern that land use managers have been confronted with for many years. This problem has increased in importance as the land base available for natural resource extraction has decreased and the demand for the resources has increased. Landsat Thematic Mapper (TM) digital imagery has been available since 1982 and continues to be an excellent source of information about the Earth’s land cover. Two TM images were acquired, seven years apart, over an area of land managed by the Mark Twain National Forest in south-central Missouri. Band 7’s from both images were combined and an unsupervised classification was run to classify five categories of change in the landscape. The classification was integrated into a GIS database which contained land ownership and management unit boundaries. A prototype map and table were produced for selected management units to determine how well the Forest Service is meeting the objectives of the Forest Plan.

Introduction

The use of remote sensing systems for mapping land cover change has been well documented and continues to be one of the most promising emerging technologies (Colwell et al., 1980, 1981; Nelson, 1982; Lee, 1980). Satellite imagery has many advantages over other remote sensing systems for mapping land cover change. The benefits, including repeatable coverage, high spectral and spatial resolution, and highly accurate image registration methods have helped make satellite imagery an excellent vehicle for mapping land cover change. These advantages are augmented by the fact that satellite data exists in digital form, making it readily available for integration into a geographic information system (GIS).

Early problems associated with mapping land cover change from digital satellite imagery have included poor registration, lack of established change
detection procedures, and poor methods for determining the boundaries and acreages involved in change (Malilis, 1980). Improvements have been made in all of these areas due to advances in computer hardware and software, combined with new methodologies for manipulating digital data, including classification and positional accuracy calculation procedures.

**Background and study area**

This paper summarizes the cooperative project conducted by the Mark Twain National Forest and the Nationwide Forestry Applications Program (NFAP) under the “Integration of Remote Sensing” (IRS) program. Contract support to IRS is provided by Pacific Meridian Resources of Emeryville CA. Analysis was completed on a SUN workstation with ERDAS and ARC/INFO software.

The Forest Land and Resource Management Plan for the Mark Twain National Forest was begun in 1980 and approved in 1986. The Plan’s mid-point review was begun in 1991. The project developed from a need to gather and analyze information to monitor Plan accomplishments in conjunction with the mid-point review. The Forest Plan sets standards for vegetation change by management prescription areas in which data is needed for private and federal lands. The project integrates the use of satellite image analysis in conjunction with a GIS to determine the type and amount of vegetation change during the Plan period.

The Mark Twain National Forest is located in the area of southern Missouri known as the Ozarks. The Forest encompasses an area about 285 miles east to west and 100 miles north to south. There are nine districts located on nine separate units. Net ownership of National Forest System lands in 1990 was approximately 1.5 million acres, an average of about 50% within the proclaimed National Forest boundary. Ownership within the individual units ranges from about 30 percent to 70 percent. The Forest is located in the transition zone of several major ecological plant associations. The oak-hickory, oak-hickory-pine, cedar glades, southern floodplain forest and bluestem prairie all converge within the Forest.

The project described in this paper was conducted on one of the nine units on the Forest called the Fristoe Unit. Figure 2 shows the boundaries of the Fristoe Unit and the three shaded management unit boundaries used in the sample analysis described in this paper.

**Analysis methods**

**Image classification and polygon creation**

Landsat TM data from August 1982 and August 1989 was acquired for a portion of the Forest. The August 1982 image was registered to the August 1989 image which had been precision geoded and terrain corrected. The image registration was tested by performing a visual inspection of individual bands from the different dates in separate color gons on an RGB (red, green, blue) display. Image registration of the two scenes was excellent because of the careful preprocessing procedures used.

Clouded areas create problems for change detection analysis and will classify incorrectly as land cover changes if not removed. Clouds were confined to a small portion of the images but were most persistent in the southern third of the project area. A “cloud mask” was created from an initial unsupervised classification and areas classified as clouds were removed. This method creates holes in the data but prevents misclassifications and results in fewer errors in the analysis.

Landsat TM data has seven spectral bands or channels in which it gathers information about earth surface features. In this project only band seven’s were used from each of the images. Band seven’s from each of the masked images were combined into a single file containing the two bands. Band seven was used because the digital values in this portion of the spectrum (middle infrared) were consistent for earth features that showed little or no change between the two image dates.

An unsupervised classification of the combined band seven image was completed with 35 classes. The unsupervised clustering program and classifier separated out various degrees of change between the two bands and created an output classification with 35 information classes. The classification was further aggregated into five categories which represented various classes of land cover change. The land cover change classes are shown in Figure 1.

The evaluation and labeling of the unsupervised classes was based on visual and statistical analysis and was guided by the expertise of Forest Service personnel who were familiar with the forest. A forest history database, containing a history of activity in individual stands, was also available and proved valuable for assigning the unsupervised classes to land cover change categories. A ground-based accuracy assessment was initially completed for a portion of the project area and is currently being completed for the entire project area.

The resulting land cover change map was subjected to a series of scanning filters which served to create homogeneous pixel areas of land cover change. An assortment of majority scans were used to firm boundaries between classes and eliminate outlier pixels in the classification. Vector polygons were created from the scanned classification and polygons less than twu acres were eliminated from the analysis. Additional vector GIS layers of land ownership and forest management unit boundaries were created from the Forest Service’s Cartographic Feature Files (CFF’s).
Figure 1. Land cover change classes derived from the Landsat TM classification.

<table>
<thead>
<tr>
<th>Land Cover Change Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Open Areas</td>
<td>Areas that were classed as open in 1982 and in 1989.</td>
</tr>
<tr>
<td>2. Change</td>
<td>Areas that have not had recognizable change in forest canopy density from 1982 to 1989.</td>
</tr>
<tr>
<td>3. Regrowth Areas</td>
<td>Areas that show an increase in woody vegetation since 1982. These would include recent and older clearcuts.</td>
</tr>
<tr>
<td>4. Moderate Change</td>
<td>Areas that show a moderate decrease in forest canopy density since 1982. These include commercial thinnings, moderate fire killings, etc.</td>
</tr>
<tr>
<td>5. Major Change</td>
<td>Areas that show a major decrease in forest canopy density since 1982. These would include clearcuts, conversion of forest to pasture, fire kill, etc.</td>
</tr>
</tbody>
</table>

Future Condition (DFC) for each of the management prescriptions and areas. Each management prescription has a specific goal, or limitation, for the amount of permitted vegetation manipulation. Questions that may be answered using this technology are: Where and what types of change are occurring? How much change is occurring? Which has more impact — Forest Service activities or private activities? Is the Forest following Standards and Guidelines? Achieving its goals?

Comparison of the amount of change or disturbance during the period versus the amount desired under the Plan is one measure of how well the Forest is progressing toward attaining the desired vegetation conditions. In addition, the Forest is charged with monitoring the effects of these changes on adjacent lands as well as the effects upon National Forest lands by other agency and private land activities.

Monitoring of changes to vegetation was tracked using the process developed in this project. Even-aged management areas show clearcuts as major change and thinnings as moderate change. The shapes of the polygons are variable in size and shape. Size limits of the clearcuts can easily be determined and compared to the maximum size allowed. Total acres in the 0 to 9 age class and percent of the area, can be determined and compared to the goals of that management unit. Distribution (grouped versus scattered) of these activities is also very important and can be monitored with the methods developed here.

Need for resource information

Basic resource data for the Forest is obtained from a detailed field inventory. It is updated based on management activities and new inventories and forms the basis for district management activity planning and scheduling. Data is only available for National Forest lands. Approximately 3 to 7 percent of the Forest is reinventoried each year.

With government ownership ranging from 30 to 70 percent on the units, data is often needed on private as well as the National Forest lands to assess total impacts. The Mark Twain National Forest received over 90 appeals of proposed activities in the past two years. With its proximity to large population centers and the large number of people living within the Forest boundaries, it is inevitable that people will disagree with some management decisions. There is a real need to show these proposals and their effects in a visual format (e.g. images and maps) in order to make easily understandable presentations to land managers and the public. Analysis has to be on larger areas and on a landscape basis rather than simply the affected acres and immediate surroundings.

The objective of this project was to apply change detection procedures in monitoring Forest Plan. Results of the image analysis are in the form of a digital data layer of “change”. This layer, in addition to the other data
TABLE 1. SAMPLE PRIMARY ANALYSIS:
The following table is an example of the primary analysis product of this project. It was created by overlaying a GIS layer containing the boundaries of three selected management units over the change detection GIS, and producing a summary of change by prescription and land status (public vs. private). This type of summary can be generated by prescription, by management unit, by ranger district, etc., limited only by the layers of data input into the Forest GIS:

<table>
<thead>
<tr>
<th>MANAGEMENT PRESCRIPTION</th>
<th>LAND STATUS</th>
<th>NO CHANGE</th>
<th>AG/Open</th>
<th>Regrowth</th>
<th>MOD. CHANGE</th>
<th>MAJOR CHANGE</th>
<th>TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acres %</td>
<td>1982 to 1989</td>
<td>CHANGE CLASSIFICATIONS</td>
<td>Acros %</td>
<td>Acros %</td>
<td>Acros %</td>
<td>Av. Size</td>
</tr>
<tr>
<td>3.1: Natural Vegetation</td>
<td>PVT</td>
<td>931</td>
<td>57</td>
<td>256</td>
<td>16</td>
<td>216</td>
<td>13</td>
</tr>
<tr>
<td>Communities</td>
<td>Total</td>
<td>3986</td>
<td>78</td>
<td>286</td>
<td>6</td>
<td>369</td>
<td>7</td>
</tr>
<tr>
<td>3.4: Wildlife Habitat</td>
<td>PVT</td>
<td>1449</td>
<td>42</td>
<td>948</td>
<td>28</td>
<td>781</td>
<td>23</td>
</tr>
<tr>
<td>Emphasis</td>
<td>Total</td>
<td>7315</td>
<td>70</td>
<td>1112</td>
<td>11</td>
<td>1488</td>
<td>14</td>
</tr>
<tr>
<td>6.1: Semi-Primitive</td>
<td>PVT</td>
<td>210</td>
<td>84</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Recreational</td>
<td>Total</td>
<td>3099</td>
<td>89</td>
<td>48</td>
<td>1</td>
<td>265</td>
<td>8</td>
</tr>
<tr>
<td>TOTALS</td>
<td></td>
<td>14400</td>
<td>76</td>
<td>1446</td>
<td>8</td>
<td>2122</td>
<td>11</td>
</tr>
</tbody>
</table>

Technical note: The table above was derived from a vector polygon coverage that did not have a minimum mapping unit applied. Hence, the GIS contains many very small polygons carried over from the raster classification. Thus, the average size of major change polygons as listed above is much smaller than if a minimum mapping unit of 1-2 acres had been applied to the analysis. Percentages are based on total acreage by ownership.

TABLE 2. SAMPLE SECONDARY ANALYSIS:
The full value of the type of data collected by this project becomes apparent when the above table is compared to goals and objectives presented in the Forest Plan. For example, the following table compares selected Forest Plan goals with the change detection results for three selected management units in the pilot area:

<table>
<thead>
<tr>
<th>MANAGEMENT PRESCRIPTION</th>
<th>SELECTED PLAN OBJECTIVES</th>
<th>CHANGE DETECTION RESULTS</th>
<th>WERE GOALS MET?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Timber Manage. System</td>
<td>Retention of Open/Semi-Open Areas</td>
<td>Major + Moderate Change per Year</td>
</tr>
<tr>
<td>3.1 Natural Vegetation</td>
<td>Even-age .7 to 1.5% of area per year</td>
<td>4-15% of area</td>
<td>NF: 1.0%/yr</td>
</tr>
<tr>
<td>Communities</td>
<td></td>
<td></td>
<td>PVT: 2.1%/yr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ALL: 1.3%/yr</td>
</tr>
<tr>
<td>3.4 Wildlife Habitat</td>
<td>Even-age .8 to 1.5% of area per year</td>
<td>7-12% of area</td>
<td>NF: 0.6%/yr</td>
</tr>
<tr>
<td>Emphasis</td>
<td></td>
<td></td>
<td>PVT: 1.2%/yr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ALL: 0.8%/yr</td>
</tr>
<tr>
<td>6.1 Semi-Prim. Non-Motor Recreational</td>
<td>Uneven-age 0 to 1.5% of area per year</td>
<td>1-5% of area</td>
<td>NF: 0.1%/yr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PVT: 1.5%/yr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ALL: 0.2%/yr</td>
</tr>
</tbody>
</table>
Figure 2. Map showing the boundaries of the project area and location of the boundaries of the three management units (shaded) used in the primary and secondary analysis.

Management Units Over Change Detection Classification

Figure 3. Map of the three management unit boundaries illustrating the results of the Landsat TM classification with overlay of land status.
layers of ownership and management unit prescription, serve as input to a number of other analysis and modelling processes. Tables 1 and 2 illustrate these ideas.

**Analysis of the overlay process**

Table 1 provides a sample analysis of three management units with different management prescriptions. The number of acres and percent of total land area for each type of activity were created by overlaying a GIS layer containing the management unit boundaries for three selected units over the change detection GIS. The additional layer of land status allowed for summary of activity on private versus public lands. Table 2 provides a secondary sample analysis that builds upon the results of Table 1. This table compares the goals and objectives presented in the Forest Plan to the results of the change detection summary. The full value of the data becomes apparent in this type of analysis.

Figures 2 and 3 show the extent of the project area and the three management unit boundaries used in the sample primary and secondary overlay analysis. Figure 3 illustrates the change detection map created from the digital Landsat TM classification for the three management unit boundaries. The numbers inside of the management unit boundaries refer to the management prescription in Table 1 and Table 2.

**Conclusions**

Based upon the results of the Landsat TM classification and the GIS overlay analysis, the following conclusions were reached:

- Forest canopy changes can be accurately detected and classified in a straightforward procedure using multitemporal Landsat TM images.
- The derived TM classifications can greatly aid in the monitoring of forest related activities over time and the results can be readily integrated into a GIS.
- The results of the change detection become most effective when compared to the goals and objectives presented in the Forest Plan. Results can be tabular or maps.

**References**

Replacing aerial photo interpretation in resource inventories with integrated digital images and GIS

Frank Hegyi
(presenting author)
Paul Pilon
Penny Walker
Ferihill Technologies Ltd.
1025 Pandora Avenue
Victoria, B.C. V8V 3P6

Abstract

Information acquired for resource management through current operational methods lacks some of the critical data needed for the protection and enhancement of the environment. In particular, conventional aerial photographic techniques are limited mainly to visual observations of the earth's surface. Even when satellite imagery is used, the resolution of the data does not accommodate the monitoring of individual trees and stands, especially during their early renewal development. Also, the existing technology has limited capabilities in demonstrating the impacts of resource management practices on the environment. For example, the use of video images for the evaluation of Visual Quality Objectives (VQO) may not be cost-effective when dealing with large areas.

Ferihill Technologies Ltd. is developing, in cooperation with Canada Centre for Remote Sensing, National Research Council of Canada and Forestry Canada, techniques to replace conventional aerial photography and video imaging in resource management with airborne digital data. Information collected with sensors capable of examining the earth's surface from a wider range of the electromagnetic spectrum than that of the visible bands, is integrated with 3-D thematic GIS information to provide unique cost-effective analysis techniques. This new approach is especially designed to collect and process information which is both sensitive and responsive to environmental issues in forest management. In addition, the system contains a wide range of tools to present the results of alternative management strategies including the generation of diagrams, graphs and images with their descriptive text.

In this paper, an overview of the new Airborne Digital Data Information System (ADDIS) is presented and its unique capabilities are illustrated through a slide presentation.
Managing forest resources with raster tools

Andrew S. Bury
ERDAS, Inc.
2801 Buford Highway NE, Suite 300
Atlanta, Georgia 30329 USA
Tel: 404/248-9000
Fax: 404/248-9400

Abstract

Image processing techniques, in conjunction with raster GIS modeling and vector data overlays, are being used to develop successful forest management practices throughout the United States. Ecological issues concerning timber resources, wildlife habitats, and watershed protection are seriously affected by forest management practices. Remote sensing and GIS offer an integrated approach to forest management for both publicly and privately held lands. This paper focuses on the process of managing timber sale compartments using satellite imagery and scanned aerial photography. Satellite data are compared to aerial photography for updating timber stand boundaries. This process was used to define impact-free stands needed to protect Red-Cockaded Woodpecker habitats and to improve tree stand updating and harvesting procedures.

Introduction

Image processing and geographic information system (GIS) modeling tools provide an integrated approach to forest management. Together, they can be used to support a variety of mapping and resource management activities. The highly accurate information generated through image processing and GIS techniques not only allows foresters to more efficiently manage their resources, but provides information that acts as a common ground between forestry companies and environmentalists. This information allows opposing groups to assess and resolve issues regarding the preservation of critical forest stands or wildlife habitats. Several recent projects are described in this paper to illustrate how image processing, raster GIS modeling, and vector data analysis can be combined for tree stand mapping and wildlife habitat protection.

Updating forest stand maps using satellite imagery

One project, recently conducted by ERDAS, Inc., for a large forest management company, used satellite imagery to update out-of-date forest stand maps. In the past, the company had delineated forest stands on unrectified aerial photographs, then digitized them into their vector GIS. This technique generated a GIS data base that was not accurate, either in terms of the actual location of stands or in stand size. The inaccuracies caused errors in assessing the company's resources, especially for tax and harvest purposes. In addition, map updating was difficult because the updates took the form of hand-drawn maps done by foresters in the field. The GIS attribute data base, however, contained valuable information on tree stand characteristics. An important goal of the project was to preserve the attribute data from the company's
original stand database, while increasing the overall positional accuracy of the mapped information. The company also needed an efficient technique for updating old stands as well as adding new stands.

ERDAS acquired 10 meter SPOT panchromatic imagery, which was rectified and georeferenced. The tree stand maps previously digitized from uncorrected aerial photographs were overlayed on top of the imagery, and warped to fit it. New stands that were readily apparent on the imagery were digitized directly from the computer screen. Others that were more difficult to detect on the imagery were annotated on image plots and digitized from a digitizing tablet. Changes in the boundaries of previously mapped stands were also rapidly updated using the same process.

Changes in the attributes of tree stands were made from field notes typed into on-site computers by foresters while conducting their forest management activities. These changes were uploaded to a central system where map updating took place. When a particular forest stand was updated, the system automatically searched for the corresponding attribute information and recorded the update in the stand data base.

Map updating with satellite imagery brought the company's inventory of forest resources up-to-date quickly and efficiently. It also presented a means for maintaining the data more effectively. Satellite imagery provided the most cost-effective and accurate base for updating maps of the company's entire land holdings at one time. Following the initial update, aerial photography was used to record changes in timber stands occurring in areas of high logging activity. Paper copies of the imagery and aerial photographs were provided to the field personnel on which to record changes as they occurred.

Protecting Red-Cockaded Woodpecker habitats

A project recently conducted by graduate students in Landscape Architecture at Louisiana State University identified individual tree stands that should not be harvested because of their proximity to the habitat of the Red-Cockaded Woodpecker. The woodpecker, which is protected by law, favors a certain combination of forest species. This project focused on determining the impact the woodpecker habitats would have on USFS forestry management plan.

First, U.S. Forest Service (USFS) compartment and stand maps of a section of the Kisatchie National Forest near Alexandria, Louisiana, were digitized. These maps delineated homogeneous forest vegetation species. Because of the diversity of the area, the maps were highly complex, showing 382 distinct areas of forest vegetation in an area less than 23 square miles in size. USFS attribute data on tree stands, based on the U.S. Department of Agriculture Silvicultural Examination and Prescription Field Book, were imported into an ERDAS system. The attributes included forest type (e.g., Hemlock-Hardwood), and wildlife codes, among many others. The wildlife codes were used to create a file of stands that would typically include the Red-Cockaded Woodpecker.

Proximity searches of 0.25, 0.5, and 0.75 mile ranges were generated around the woodpecker habitats. The 0.25 mile buffer encompassed woodpecker foraging areas, whereas the 0.75 mile buffer represented its flight radius. The buffer zones were intersected with the USFS forest stand maps to identify any tree stands that at least partially crossed woodpecker habitats.

A software program was developed to read the histograms of the stand-buffer intersection files and create three new attribute fields. These fields indicated which stands intersected with which buffers. Acreage statistics were also generated, which indicated how many acres of each stand intersected with each buffer.

The new attribute fields were used to generate three images which showed the stands that intersected with each buffer zone. The maps and the accompanying acreage attributes provided USFS personnel with useful information to conduct forest harvesting studies, determine which stands and how many acres of timber could be harvested safely without threat to the woodpecker habitats.

Mapping old growth stands in the Pacific Northwest

Pacific Meridian Resources completed a project for the USFS to produce GIS data on old growth forests in all national parks and national forests in western Washington and Oregon. Old growth forests are critical because of their species diversity and their role in providing a habitat for wildlife. Unfortunately, old growth is defined differently from person to person. In general, however, the definition centers on the following four characteristics:

Crown Closure - the percentage of ground area that is covered by the tree canopy.

Tree Size Class - measured according to the tree diameter at breast height.

Tree Species - identified when the canopy of a particular species occupies more than 25 percent of cover.

Stand Structure - either even storiéd (all trees approximately the same size and age), two-storied (two distinct generations), or multi-storied. Most old growth forests are multi-storied, supporting trees of many sizes and ages.

By mapping each of these characteristics separately, the USFS would have the flexibility to change the definition of old growth as needed and recalculate the extent and location of old growth forests. Additional information that
was mapped and added to the USFS GIS data base included slope, aspect, elevation, hydrology, flight line maps, plant associations, location of research and inventory plots, and training site locations.

Pacific Meridian used a unique combination of GIS, image processing, and statistical analysis procedures on Landsat TM imagery to analyze variations in spectral reflectance and vegetation. Training sites were captured from the imagery and passed to the statistical analysis package to identify variables (e.g., aspect, elevation) which significantly affected vegetation distribution. Both supervised and unsupervised image classification were performed, and compared. The analyses resulted in four classification maps, one for each old growth vegetation characteristic (i.e., species, size class, stand structure and crown closure).

A set of spectrally and informationally unique training signatures were also generated, along with an identification of training signatures with confusing spectral reflectance. For example, spectral confusion typically occurred between similar species (e.g., Sitka Spruce and Pacific Silver Fir). These areas were further analyzed to determine if ancillary data (such as aspect, elevation, or latitude) could be used to stratify the imagery. Fortunately, the distribution of species was usually found to be a function of these ancillary data.

The four old growth vegetation characteristic maps were combined to create a vegetation type polygon map. Polygons were delineated around areas of homogeneous vegetation using algorithms which replicated the decision rules used by photo interpreters delineating vegetation types. First, the crown closure pixel classification map was generalized to produce crown closure polygons. Crown closure was then held constant, and size class was generalized within each crown closure polygon. Next, both crown closure and size class were held constant and stand structure was generalized. Finally, species were summarized within each crown closure-size class-stand structure polygon.

The vegetation polygon map was overlaid with each of the four pixel classification maps to produce an attribute file. The attribute file described the distribution of species, size class, structure, and crown closure within each vegetation polygon, and provided an indication of the vegetation diversity of each polygon.

The old growth mapping projects required a classification accuracy of at least 80 percent. Pacific Meridian and Dr. Russ Congalton, a consultant from the University of California at Berkeley, developed procedures and a sample design for map accuracy assessment. Guided and random procedures were generated to compare manual photo interpretation with results obtained through the above procedures. At least 20 percent of the areas assessed for classification accuracy were ground-checked in the field. An error matrix was generated for each vegetation characteristic.

The use of the mapping procedures described above lowered the cost per acre of obtaining accurate information by one-third, and produced fourteen layers of detailed GIS data. Traditional air photo interpretation techniques would have yielded only one layer of data. The pixel classification maps, and vegetation polygon map with its attribute file provided the means by which users could query and model the data for forest management practices. These enabled the USFS to address issues such as the fragmentation of old growth and its implications to wildlife habitat, and to develop estimates of the biological diversity of forest vegetation.
Using ancillary data in post-classification modelling to increase the accuracy of conifer species classification from Landsat Thematic Mapper (TM) data

Michael Golden
Lisa Lackey
Pacific Meridian Resources
Suite 570,
5200 S.W. MacAdam Avenue
Portland, Oregon 97201

Abstract

A post-classification model was used to identify and correct mis-classifications of forest tree species from Landsat TM data of US National Forests in Western Oregon and Washington. The model used elevation and aspect data, and Forest Service ecologists' expertise and plot data, to identify pixels with species labels outside their "allowable" elevation/aspect ranges. These pixels were relabeled with a species label from historical Forest Service maps of cover type. Separate models were developed for each National Forest. Manual editing after the model was run was used to further improve the map. The Mt. Hood National Forest model is presented as an example.

Introduction

Pacific Meridian Resources, under contract to Region 6, USDA Forest Service, produced a Geographic Information System (GIS) data base for old growth on national forests in western Oregon and Washington (Tepl y and Green, 1991). The old growth project was part of a larger project to produce a GIS data base for all the National Forest lands in Washington and Oregon. This data base includes 15 layers of information and represents more than 30 million acres. The old growth study involved 12.5 million acres.

Vegetation layers of tree crown closure, tree size class, forest stand structure, and tree species were produced using Landsat Thematic Mapper (TM) satellite data. This paper presents an example of using ancillary data in a GIS modeling process to improve the computer classification of TM data for tree species. The Mount Hood National Forest is used as an example. Similar models were developed for each of the nine national forests in the old growth mapping project.

Background

The success of image classification is based on the assumption that there is a correlation between the thematic classes of interest and the spectral response of those classes. Unfortunately, this correlation is not strong with respect to conifer tree species. While some tree species in the Pacific Northwest (most
notably lodgepole pine, *Pinus contorta*) have a distinct spectral response, most species and species mixes have very similar spectral responses. In order to accurately classify tree species, additional data was needed to differentiate species and species mixes to the level desired for this project. Fortunately, species distribution in the Pacific Northwest is highly correlated with elevation and aspect. We used this relationship to improve the classification of the imagery.

Ancillary data can be defined as "secondary data pertaining to the area or class of interest, such as topographic, demographic, or climatological data" (Swain and Davis, 1978). In a remote sensing context, ancillary data refers to any data not obtained by the sensor. Ancillary data, such as topographic data, can be used either before, during, or after classification, through stratification, classifier operations, or post-classification sorting and post-classification modeling (Hutchinson, 1982).

Stratification is a technique to divide a large data set into smaller homogeneous subsets. Separate classifications can then be run on each subset. The objective is to reduce the variability within each data set. However, stratification divides the data into subsets with distinct boundaries. No gradations, such as those found in nature, exist between mapping classes (Hutchinson 1982), and "seam lines" are often apparent when the subsets are "stitched" back together.

A second use of topographic data is during classification. Digital terrain data can be used as an additional "band" of data, called a "logical channel" (Strahler et al. 1978). This use of terrain data has been disappointing. The "simple addition" of non-spectral data appears to add little to classification accuracy (Hutchinson 1982). Topographic data can be used to modify prior probability in the maximum likelihood classifier (Strahler 1980). If the distributions of cover types across the topography are known, the probability of occurrence can be used to "weight" the maximum likelihood classifier. Strahler et al (1978) reported that classification accuracy was improved 13 to 19 percent using topographic data to modify the maximum likelihood classifier. Hutchinson (1982) discusses two problems with using modified probabilities. The assumption of a normal distribution is often not met for ancillary data. Secondly, determining prior probabilities requires more intensive ground sampling and therefore increases costs.

A third major use of topographic data in image classification is for post-classification sorting, usually with unsupervised classification. In a study in California, Woodcock et al. (1982) found that "due to varying slope angle and aspect, poorly-stocked open stands of conifers under low angles of illumination exhibited spectral signatures which were very similar to those of dense, well-stocked stands under more direct illumination". They used digital terrain data to produce a map of illumination. Overlaying this map on the clusters divided all the classes that were spectrally similar but differed in illumination. This allowed the analyst to separate poorly-stocked stands from well-stocked stands. Post-classification sorting suffers from its deterministic nature (Hutchinson 1982). It uses data that is essentially continuous, but produces classes based on abrupt criteria boundaries.

<table>
<thead>
<tr>
<th>Elev (feet)</th>
<th>E</th>
<th>NE</th>
<th>N</th>
<th>NW</th>
<th>W</th>
<th>SW</th>
<th>S</th>
<th>SE</th>
<th>flat</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Aspect/Elevation ranges for Mt. Hood tree species.
In this study we used a fourth method of incorporating ancillary data, post-classification modeling. Topographic data and knowledge about their relationship to tree species distributions were used to identify problem areas in the classification. Historical cover type maps and manual editing were used to correct these areas. This approach avoided the "seam lines" inherent with pre-classification sorting, and we felt we did not have adequate information for estimating prior probabilities. This was not a research project and so there was no hypothesis to test. Rather, it was an attempt to incorporate available information into the classification to improve the accuracy of the map.

Data used

The primary data source for species classification was Landsat TM imagery, geocoded and terrain corrected to UTM coordinates. Digital Elevation Model (DEM) data, ortho photo quads, and resource aerial photography were also available. Field data included representative training sites established by Pacific Meridian Resources staff and Forest Service ecology plot and inventory data. All training site and plot data included information on species composition, aspect, and elevation. Pacific Meridian Resources staff also recorded extensive field notes on ortho photo quads while driving between training site locations. Forest Service historical maps dating from the early 1950's were also available for each national forest. These maps were digitized and converted to raster format using ESRI ArcInfo (Redlands, California) software.

Methods

A combined supervised/unsupervised approach was used to classify the TM imagery. All classifications were performed using ERDAS (Atlanta, Georgia) image processing software. For the old growth study, only those forested areas meeting the tree density, size, and stand structure criteria for old growth were mapped for species. Forested areas not meeting those criteria were mapped as "young conifer". A particular species was included in the species label if it represented at least 25 percent of the crown closure of the stand, up to 3 species.

Based on Forest Service plot data and training site data, aspect/elevation ranges for each major tree species were developed. After consulting with Forest Service ecologists, these ranges were combined and generalized into one aspect/elevation matrix which delineated ranges where certain species were not allowed (see Figure 1).

An ERDAS gis model was written that compared, pixel by pixel, the species label in the original classification with the aspect and elevation and the allowable range for that species. In this way, "unreasonable" species calls (i.e., species calls outside their allowable ranges) were identified. The model did not insure accurate species labels, but rather identified labels which were not allowed. "Unreasonable calls" then had to be replaced with an acceptable species label. Both Forest Service ecologists and Pacific Meridian Resources staff agreed that the historical species map would be a reasonable substitute for the mis-classified labels. It was assumed that species composition in the old growth stands would not have changed appreciably over time. However, the historical species labels were less detailed than the classification label and so were not desired for the entire coverage.

The final step in producing the species coverage was to manually edit the coverage based on plot data, training site data, and field notes from the ortho quads. The editing was completed for all the forests on a quad by quad basis.

Results

Results of classifying, modeling, and editing species on the Mt. Hood National Forest are summarized in Table 1. The Mt. Hood model identified 77841 pixels (1.5%) labelled with species outside their allowable ranges. These pixels were relabelled with the historical species call. Manual editing relabelled an additional 790910 pixels (14.8%). These changes are assumed to be a more accurate species label. Similar results were obtained with the models for the other National Forests.

<table>
<thead>
<tr>
<th></th>
<th>% of pixels classed by original classification</th>
<th>% of pixels classed by model</th>
<th>% of pixels classed by editing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Classification</td>
<td>100.0 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>After Modeling</td>
<td>98.5 %</td>
<td>1.5 %</td>
<td></td>
</tr>
<tr>
<td>After Editing</td>
<td>83.7 %</td>
<td>1.5 %</td>
<td>14.8 %</td>
</tr>
</tbody>
</table>

Table 1. Modeling and Editing Summary.
Quantitative accuracy assessment is currently being implemented on each of the individual vegetation layers in the old growth data base. At this writing, no results are available for the species layer.

**Conclusion**

A post-classification model was used to improve the image classification of tree species on national forests in Western Oregon and Washington. The model used DEM data, field data, and personnel expertise on species distribution. The model identified species labels which were "unreasonable" with respect to aspect and elevation. Unreasonable calls were replaced with a label derived from historical maps of species distribution. After modeling, the species maps were manually edited.

The Mt. Hood model relabeled only 1.5% of the pixels. While this is a small percentage of the total number of pixels, we feel that this is an important improvement. Geographically, they were found in both areas where we had field notes and data, and in areas where we did not. Those pixels that changed were clearly misclassified. From a remote sensing perspective, those errors are easily understood. However, correcting the pixel labels which should not occur is an important step in insuring acceptance of a map by clients with little or no remote sensing background. The absolute percentage of corrections may be less important than the type of corrections made.

Manual editing relabeled 14.8% of the pixels. Manual editing is a critical step in commercial mapping applications of satellite imagery. We must understand, and be willing to admit, the limitations of imagery for mapping purposes. We can then use ancillary data to improve the map. Manual editing allows us to use site specific information that would be difficult, if not impossible, to model.

The Mt. Hood model was based on the data available and was relatively simple to implement. While it did not guarantee correct species labeling, it eliminated labels which were clearly incorrect. Manual editing further improved the accuracy of the map. Obviously, the model is limited by the data used to derive it. The model did not consider small micro-climate influences such as cold air pockets, or precipitation/moisture gradients. However, the model was a useful tool in improving the classification of forest species.

**References**


Human Resources
Working Group

Hardware, software and flesh: a human resources architecture for geographic information systems
Robert Menes, Ministry of Environment, Lands and Parks, Victoria, BC, Canada

New GIS/LIS education and training initiatives in Australia and southeastern Asia
Stephen R. Kessell, School of Computing, Curtain University of Technology, Perth, Western Australia
Hardware, software and flesh:
A human resources architecture for
geographic information systems

Robert Menes, P.Ag.
Senior Land Information Analyst
Surveys and Resource Mapping Branch
Ministry of Environment, Lands and Parks
4th Floor, 1802 Douglas St.
Victoria, B.C. V8V 1X4, Canada
Tel: 604-387-9353
Fax: 604-356-7831

Abstract

The Land Information Infrastructure (LII) of British Columbia is being designed to produce a system for data access and interchange by the year 1998. The system architecture, precursor to the high level design, identified the logical components of the system from a hardware and software point of view. However, the architecture and design must also include the definition and management of the human resources necessary for system operation. Identification and management of the human resources is complicated by the fact that the organizational structures for managing people do not map onto the system structures for the Land Information Infrastructure. To do so, an additional architecture must be defined which deals with the human resources necessary to assume responsibility for the computer/communications system. This paper presents a human resource architecture which can be used for the operations and management of the LII and presents a possible implementation of the architecture. The architecture permits the development of plans for realizing the operations team and facilitates the estimation of resource needs for the system.

In the paper, responsibility categories, assigned to people, are defined which are necessary for the operation of the system, such as: spatial data management, internal user management, policy and human resource coordination. Every responsibility category inherits the operations of a management class: planning, resource allocation, reporting, and monitoring. The definition of the classes, and their interrelationships, form a model for human resource responsibilities. Once the model is defined, it is applied to British Columbia's Land Information Infrastructure to illustrate the methodology for identifying the number of people needed to manage a system and their responsibilities.

Introduction

Motivation
Geographic Information Systems are becoming essential tools in the management of land-related information throughout North America. People responsible for making land-use decisions at all levels are aware of the new technology and are eager to make it part of the standard set of management and analysis tools. However, the implementation of GIS is far from simple and, regardless of the scale of the implementation, key issues arise which cannot be solved by the application of technology alone. GIS requires the management of multiple sets of data, spatial and nonspatial,
linked together in various ways. The underlying software and hardware platforms must be matched and the communication with the "outside" world must be defined. In short, GIS places serious management requirements on the use of the support tools.

GIS is not the only technology which impacts the modern planner and policy maker. Communications networks enabling distributed processing and distributed data storage are making disparate data sources accessible. Technologically, they are accessible, but managing the information across regional and administrative boundaries, across systems, is an unresolved nightmare.

Historically, systems design has been performed for the "technical" components of a system: the hardware and software and communications networks. After the system is designed, the human aspects of the system are considered: how it will be run, maintained and redeveloped. To date, there is no clear methodology for identifying the human resources necessary. Usually, standard job titles such as "data administrator" or "systems operator" are used post hoc to assign the human resources. The management of the human resources and the planning for system growth is often not defined.

British Columbia is constructing a system for sharing land-related data within all of the ministries in the government. The system spans ministries, geographic regions, data types, hardware and software platforms, and user levels. The extensive system will be created over a seven year period in four phases. At this time, Phase I is nearly complete. A set of requirements has been identified, a target system and migration strategy developed, and a logical model of the working environment has been created. The system, known as the Land Information Infrastructure, is already defined by a System Architecture and a High Level Design. Phase II entails the design and creation of a prototype system. A fundamental piece of the design is missing: the human resources necessary to manage the system. Is the Government of British Columbia creating a system which cannot be operated and maintained? A means of identifying and assessing the human component of the system is needed. This paper attempts to address this requirement.

Objective

The objective of this paper is to provide a means of defining the human resource needs of the British Columbia Land Information Infrastructure (LII). The LII is more than a Geographic Information System: it is a large scale network of services, provided by hardware, software, and data, to facilitate the sharing and exchange of data across the province. A system architecture has been defined, but the full scale system has not been designed in detail. This paper adds a human resource architecture to the systems architecture to provide a basis for the design of a system which includes, explicitly, the people needed to run it.

The objective of the architecture is to define the human resources for the B.C. Land Information Infrastructure such that increases in the volume or size of the system require only a scaling of the human resources allocated to manage the system.

A secondary objective of the paper is to provide a model for human resource systems design for a) general systems development and b) geographic information systems development.

Methodology

The methodology used to define the human resources architecture relies on identifying a model for the responsibilities of the people required to manage a computer-based data exchange system. Common terms for job responsibilities will not be used to avoid confusion with conflicting definitions.

An architecture is a plan for the design of a system which defines the types of system components and the relationships between the components. It is general in nature and usually relates to logical entities rather than physical objects. For example, a human resource architecture could include an entity that manages all data in the system without identifying whether that entity is one person or a group of people. Use of the architecture involves defining how these logical entities map onto real physical objects.

The human resource objects — those entities which perform the functions usually assigned to people — are identified first. While broadly defined, these objects provide a way of making distinctions between types of responsibilities to be assigned. One person may be assigned more than one of these human resource objects. The objects are then further refined, and additional objects will be created which are subclasses of the more general objects. Objects can inherit behaviour from their parent class. The result of the human resource object definition is a class hierarchy of objects which spans all of the human activities required to manage the system.

The human resource objects must be defined in terms of their functions (responsibilities), protocols (communication with other objects), and structure (parts or attributes). Once this has been done, we have a complete set of entities which can be used to form a model. A model will be formed which identifies how the human resource objects relate to each other and the world external to the system.

The last phase of the methodology is the application of the model to a specific system architecture. This entails mapping the components of the system architecture onto the human resource objects in this model. One additional step is required to implement the model: the human resource objects must be assigned physical values (levels of effort in man-years) to define the actual human resources needed. The process is depicted in Fig. 1.
The human resource management model

The System

The system which must be managed is composed of hardware, software, and information (data). The hardware runs the full range from personal computers to mainframes. The software includes both operating systems and data-specific applications. The data includes internal system information (accounting, usage, etc.) as well as the spatial and nonspatial information which is the main commodity of the system.

Model requirements

The model used to describe the human resources of a geographic information system must meet the following requirements:

- simple: It can be understood by a wide range of planners and systems architects
- general: It can be applied to systems with complex components of varying types
- decomposable: It can be broken down into its parts and analyzed in discrete pieces
- specifiable: It can be described unambiguously and built into a computer-based system
- implementable: It can be applied in a practical situation
- organized: It can be applied and described in linear, logical steps

Furthermore, it must be able to address the following key issues:

- Common management functions must be able to be described and recognized by system planners
- Specific lines of communication must be able to be described to eliminate confusion in the assignment of responsibilities
- Communication with both man and machine must be specified to acknowledge the fact that management must be performed for both the human and computer systems
- Implementations will vary in complexity and the model must be able to respond accordingly
- Widely disparate responsibilities may be combined in an implementation. One person may assume responsibility for many aspects of the system management
- External system interfaces must be identified to recognize the impact of the world outside of the system

Basic model components

There are five basic components to the model which, when broken down further, are associated with human resource functions. These five components are:

- DataManagement
- TechnologyManagement
- ApplicationsManagement
- UserManagement
- MetaManagement

The first three management components relate directly to the three basic aspects of any computer-based system: data, technology and applications. The last two components add to the system the management of the users and the management of the resources for management, i.e. human resources themselves and the policy implemented by the system.
Every model component serves a management function within a particular domain. As such, there are functions - tasks - which must be done by each of these model components. These functions are the responsibilities of any management construct and include a) planning, b) resource allocation, c) reporting, and d) monitoring. There can be management of other people or of the hardware/software resources.

Data management

Data is the primary currency of the system and requires rigorous management to ensure a common understanding, efficient access, optimal physical organization and storage, and integrity. Data Management for spatial data systems is particularly burdened by the need to manage the link between baseline geographic data and a host of nonspatial, or attribute data.

The Data Management aspect of the architecture encompasses all of the traditional database management functions as well as the data management functions [Martin, 1983]. Roles which perform these functions have been defined as Data Strategist (planning and co-ordination), Data Administrator (data analysis, definition of views and schemas, data modelling, validity, consistency and accuracy of data), Database Designer (physically structuring data and planning access methods, database backup and recovery, physical schemas, database performance, security, searching techniques, concurrency), and Data Operations Supervisor (data errors, system restarts and recovery, reorganization of databases and indexing, audit trails, computer scheduling, file transfer for export). In the ANSI/NDL model for database systems [Gorman, 1991], these roles fulfill the management of the dictionary and logical database, physical database and system control, and interrogation and system control (respectively). The physical database and system control aspects of database management are only partially performed under the Data Management component. These aspects are part of the functions of the Applications Management and Technology-Management components.

Data Management requires human resources which deal with the semantics, access, quality, and physical storage of the data.

Applications management

Applications are the hardware/software systems which provide the specific services to the user. They are the business functions integrated into automated systems. An example of an application for the Land Information Infrastructure is the Repository, which holds and manages all of the metadata (data dictionary and directory information) for the system. Applications can be of two types, Data General and Data Specific. Data General applications, like the Repository, are not dependent upon the domain of the data which they utilize. Data Specific applications are designed to perform a set of tasks on data from a particular domain, e.g. forestry timber supply. For the LIH, most applications identified in its High Level Design are Data General BC, 1991).

Applications Management involves the operation, maintenance, and redevelopment of the computer-based systems. The responsibility for the physical database and system control functions of database management systems would fall under this component. In addition, the management of all of the input/output services of the system would be the responsibility of Applications Management.

Clearly, there is a close interaction between Data Management and Applications Management. However, Applications Management is more concerned with the "how" while Data Management is focused on the "what".

Technology management

Applications are built on top of technology platforms. These are the specific products and standards which can be used to create the applications which, ultimately, satisfy the user needs. Technology includes the computer systems, operating systems, input/output devices, communications networks, database management systems, geographic information systems etc. which are generic in nature and which must be combined with context-specific data to meet the needs of the system. Management of the technology involves the application of agency-wide policies to the purchase and implementation of the products, optimization and review of their performance, maintenance and upgrading, and migration to newer, more reliable systems.

User management

Users interact with the system in terms of the data, applications, and technology. However, the interface to users and the management of their concerns is of such a fundamental nature that this set of functions should be extracted from those of the other components. User management entails the communication of user needs to the system management as well as the communication of system management changes to users. System training, information, and marketing fall under user management.

There are two types of users who must be managed as part of the system: internal and external. External users are users who request the data management and applications services of the system. This category can also be identified with "end-users". Internal users are the system operators and maintainers themselves: the human resources of the system management. These users require a different set of services and yet still require all of the management services provided for External users.
<table>
<thead>
<tr>
<th>Data Management</th>
<th>Applications Management</th>
<th>Technology Management</th>
<th>User Management</th>
<th>MetaManagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantics</td>
<td>Operations</td>
<td>HW/SW Purchase</td>
<td>Training</td>
<td>Policy</td>
</tr>
<tr>
<td>Access</td>
<td>Maintenance</td>
<td>Implementation</td>
<td>Information</td>
<td>Human Resource Coordination</td>
</tr>
<tr>
<td>Quality</td>
<td>Redevelopment</td>
<td>Optimization</td>
<td>Marketing</td>
<td></td>
</tr>
<tr>
<td>Physical Storage</td>
<td>Migration</td>
<td>Feedback</td>
<td></td>
<td>Accounts</td>
</tr>
</tbody>
</table>

**Meta management**

Meta Management is the management of the human resources who manage and operate the system. There are two aspects to the metamanagement: human resource coordination, and policy. The first aspect is the traditional management function of personnel assignment, allocation, task definition, etc. The second aspect is the identification of procedures, standards, policies, etc., which must be implemented by the management of the system.

The five components are summarized in Table 1 above.

In addition to the five basic model components, there are also external components which require some interface. These components are also functional groupings identified with human resources:

- Users
- ApplicationsDevelopers
- DataProviders
- Vendors
- AgencyManagement

Users are the "end-users" of the system (discussed previously in the context of External users). ApplicationsDevelopers are those people or agencies who are contracted to provide specific applications for the system. DataProviders are those groups who contribute data to the system and, perhaps, manage that data independent of the system itself. Vendors are businesses providing commercial hardware and software for the technology platforms. AgencyManagement is the larger human resource planning authority to which the Human Resource Coordination must report.

Given the model components described in this section, we can present a simple model depicting the interaction of the these management entities.

The diagram shows the interactions between the main model components and external components as well as the interactions between each of the main components.

These interactions will be described in greater detail in the following section. At this point, a complete set of objects which represent human resource responsibilities can be specified.

![Diagram of Human resource management model components](image)

**Fig. 2. Human resource management model components**

**Human resource object hierarchy**

The entities that represent the human resource responsibilities are complex: they share behaviour, they have compound structure, and they have subgroups which involve an additional level of specific behaviour. For these reasons, a hierarchy of objects which are divided into classes is an appropriate way of describing the components of the model. Once all of the classes of objects have been described, applying the model to a real system only requires that we instantiate the classes as needed.

The object hierarchy could be devised around the responsibility groupings or around the human management groups (e.g. people, or teams). The approach here utilizes the responsibility groupings. When the model is implemented, these responsibility groupings — the objects — are assigned to people. A number of objects could be assigned to a single person.