

## 連載「誰も書かなかった GIS」第 10 回

### 地図データフォーマット変換の憂鬱（その 2）

**（株）エヌ・シー・エム 代表取締役社長 柳田聡**（やなぎだ さとし）

1982年東京大学工学部土木工学科卒業。同大学院修士，博士課程を経て1985年より現職。工学博士。専門は画像処理及び地理情報システム。

前回到引き続き、地図データフォーマット変換のお話しをします。最初に軽く前回の復習をします。前回のお話しの概要は以下の通りでした。

#### 1 はじめに：なぜこの文章を書きたくなかったか

地図データフォーマットあるいはフォーマット変換に対する世の中の誤解を嘆き、もって話しの糸口としました。

#### 2 フォーマット変換の難しさの感覚的な説明

フォーマット変換がなぜ難しいのかを、お饅頭の箱詰めを例として、感覚的に説明しました。

#### 3 異なる地図データフォーマット間の高い壁

##### 3 - 1 何を記述するか、記述の範囲は何か

一口に地図データフォーマットと言っても、記述範囲が千差万別で違いが大きいことを解説しました。

##### 3 - 2 どう記述するか

記述範囲のみならず、データ構造、モデル化技法も地図データフォーマットの間で著しく異なっていることを、ポリゴンを例として説明しました。

## 4 フォーマット変換の曖昧さ

ここまでのお話でフォーマット変換って何だろうと悩み始めて頂けるならば、言いたいことが伝わったと言う点で、嬉しいです。フォーマット変換を考える際には以下がキーポイントとして、挙げられます。

情報損失を許すか、逆に言うと可逆性を求めるのか：一般に情報損失は望ましくありませんが、フォーマット毎に記述範囲の違いがある以上、情報損失は避けられません。それでも良いのでしょうか。あるいは情報損失してまで行うフォーマット変換って一体何でしょうか。

不足情報を無理に作って良いのか：例えば、本来記述されていない位相構造を無理に座標情報から作ったとして、その妥当性は誰がどうやって保証するのでしょうか。そもそも、これってフォーマット変換なのでしょうか。

目的はある特定データを処理することか、フォーマットそのものが対象なのか：この両者の間では明確にプログラムの作り方が違う筈です。前者の場合は、プログラムの汎用性など気にせず兎に角ある特定のデータを救うことだけに集中しますし、後者の場合は、出来るだけ一般的な事象を想定して、それに対する処理を考えます。

## 5 データの強い主張、影響

### 5-1 データの強い個性

今までの説明で、フォーマットのみならず、(実際の世の中に存在する)データもフォーマット変換(のアルゴリズム)に大いに影響を与えることが御理解頂けたと思います。例えば ArcView-Shape の Arc は PolyLine Group、つまり複数のポリラインのグループという、余り聞きなれない概念を含みます。もし実際にグループ化された Arc データに出会ったら、それをどう処理すれば良いかプログラマーは悩むでしょう。しかし、実際のフォーマット変換の目的であるデータにこれが含まれなかったら、この心配は消えます。これは楽になったケースです。逆に大変になったのが、前述の DXF データの穴あきポリゴンの記述法です。またこの様な例もあります。某役所(日本)の仕事で独立点データを DLG フォーマットで納品することになりました。しかし DLG には独立点の概念は含まれていません。ノード(結節点)の記述は可能なのですが、ノードとはラインの端点、交点であり、独立点ではありません。そこでどうしたか。なんと起終点一致のラインを DLG 内で記述し、それを独立点と解釈すると決めたのです。

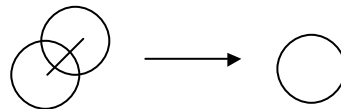


図8 起終点一致のライン

このことを知った私は、怒り心頭に発したのですが(なんでこんな不自然な運用をするんだ!)、ところがあにはからんや、たまたま別件の仕事で入手した文献(USGeoData Digital Line Graphs from 1:100,000-Scale Maps Data Users Guide 2。どうもアメリカ

の USGS、つまり地質調査所が作成している地図データに関するガイドみたいです。)を読むとこの様な技法が運用手法として認められているのです。そこでは、この様な起終点一致の長さ 0 のラインを degenerate line (縮退化ライン) と命名して、独立した点とみなしているのです。GIS の本場、本家本元アメリカの文献に明記されてしまうと、振り上げたこぶしを下ろさざるを得なくなるのですが、それでも反論させて下さい。まずこれは、常識で考えるとおかしな、と言うか極めて不自然な運用です。この DLG データを、別のフォーマットに変換したい人は恐らく、独立点は独立点として読み込みたいでしょう。でも DLG のフォーマット書を見ているだけでは、起終点一致ラインを独立点として解釈するなんて、お釈迦様でも思い付かないと思います。現に私が最初に出会った DLG のフォーマット書では縮退化ラインについては記載されていません。くだんの文献はデータガイドであり、その意味で縮退化ラインはこのデータのみで成立する特殊規約ではないかと思えてなりません。もし仮にそうならば、そしてこの様に拡張されたデータも DLG フォーマットであるならば、それはデータ運用がフォーマットを拡張出来るという点で、フォーマット変換プログラマーにとっては、またまた悩ましい事態が出現したわけです。

この様に実データ(の様子、状態)はフォーマット変換に大きな影響を与えます。このことは特に汎用フォーマットの場合、顕著です。と言うのはフォーマットが厳密なデータの作成規則、様態を縛らないので、実際のデータを見ないと、その実態が分からないからです。逆に言うと、一般にフォーマット(と言う言葉で括られる規約)はフォーマット変換ロジックを書くことを可能にするに十分な情報を規定していないとも言えます。それに対して特定フォーマット、例えば前述の「(財)日本デジタル道路地図協会 - 道路データフォーマット」では完全にデータの内容さえ予測出来ます。従って対策を立て易いのです。

## 5-2 新規の問題：データの質の重要性

前節で述べたケースは、「参ったなあ。なんてヘンテコリンなデータなんだ。」とブーブー文句を言いつつもまだ我慢出来る範囲です。ここから先は我慢出来ないケースです。

某役所の仕事で、某フォーマットの地図データを読み込んだ時の話です。一番驚いたのは、このデータフォーマットではポリゴンが一種の二重記述をされているのですが、この二重記述間で完全な矛盾が存在したことです。例えば一方ではある領域は属性 A だと言いながら、反面他方では属性 B になっているみたいな矛盾です。弊社はそのデータを読んで図化する立場ですので、当然困ってデータ作成担当の某会社に疑問を投げかけたのですが、その時の対応が余り好意的でなかった。まあ私も厳格・緻密に物事を進める性格ですので、これ以外のちょっとした問題(それでも未定義のコードがデータ内に含まれたなど、明らかな論理エラーだったのですが。)を全てその会社に投げ返しました。その数は実に夥しくなりましたが、そうしたらどうも気分を害したのか、しまいには、まともに相手にしてくれなくなってしまいました。なんともお粗末な話しです。

結局、その時達した結論は、「結局データ交換の際に最後に大事なものは、フォーマットでもソフトでもメタデータ（データのカatalog）でもキャッチフレーズ／掛け声でも営業の『何でも出来ます。問題ありません。』式発言でもなく、実はそのデータ作成を担当した技術者なり会社なりの責任感・誠意・実力なのだ。」と言う至極当たり前のことでした。「ポイントは案外泥臭い、人間臭いところにあるのではないか。」と言うことです。考えてみればこれは当たり前のことです。フォーマット変換の目的はデータの共通利用であり、それゆえデータの質が問われることは当然です。でもどうもその当然なことが忘れ去られ、単に「フォーマット変換」という用語だけが一人歩きしている様な気がします。

「あっそうそう。」ついでに指摘しておきますと、フォーマット変換の際に手渡されるフォーマットに関する文書・資料もいい加減ですね。と言うか余りちゃんと世の中では、ドキュメンテーションされていないみたいです。「N・C・Mならその位持っているよね。」ってな感じでマトモなドキュメントが来ないのが普通です。

ちなみにこの様なデータ読み込み作業は果たして「フォーマット変換」と呼ぶのでしょうか。寧ろ「データ調査」、「データチェック」、「データの洗練化」、「データ修正」と言った方が近いと思います。にも関わらず世間の大方の理解はこれらの作業を全てひっくるめて「フォーマット変換」です。「ワカンナイナア、この発想。」

## 6 まとめ

本書で述べた通り、フォーマットは一つの哲学・文化です。よって互いの間の壁を乗り越えるのは難しくなります。そのせいか、フォーマット変換の捉え方が難しく、曖昧な気がします。そして実際のデータがフォーマットと必ずしも等価ではないので、混乱に更に拍車をかけている。無理にまとめるとこんな感じでしょうか。

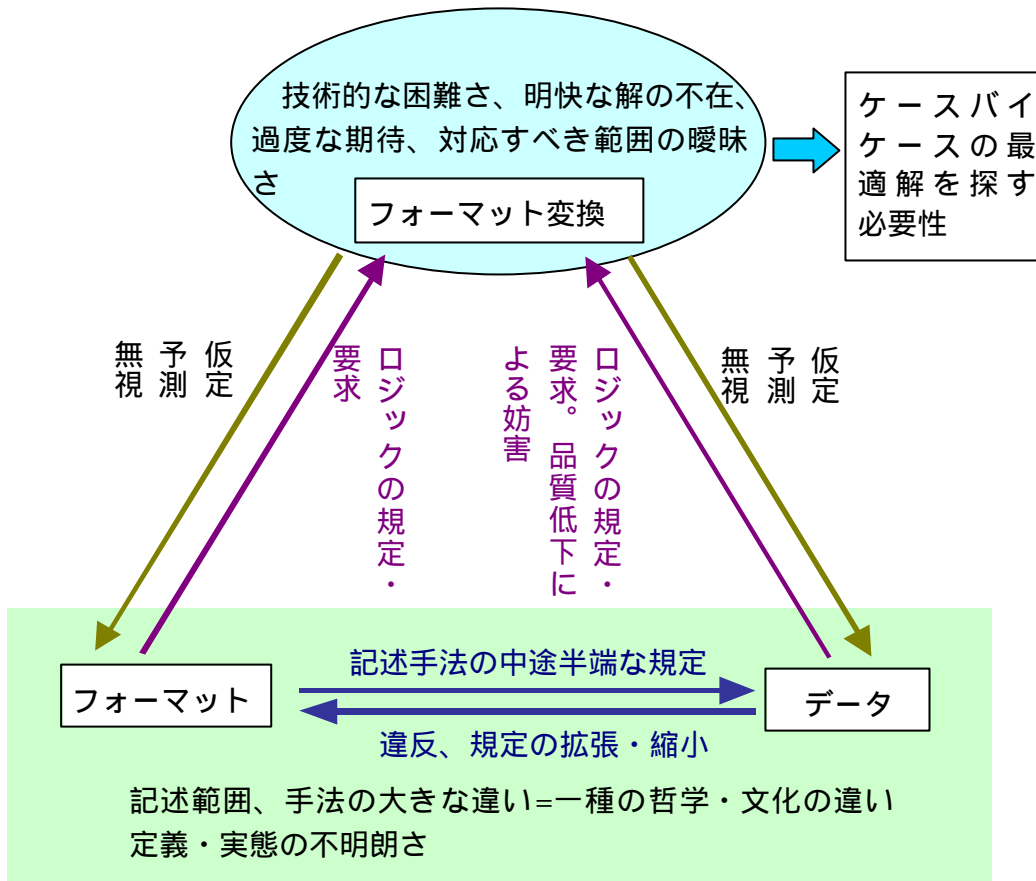


図9 地図フォーマット変換をめぐる曼荼羅模様

## 7 今後どうなるか

今、GIS業界のキーワードの一つは「Open」です。要は「閉じた世界に安住しては駄目である。GIS同士がデータ交換または会話が出来なければいけない。」という考えです。アメリカでは「OpenGIS」というコンソーシアムが出来、データの交換を真剣に議論&実現しています。

いつまでも本編の様に、ブツブツ不平不満、屁理屈を並べる時代でもなさそうです。そうです。私が言っていることはもしかして、単に極端な場合を拡大視して組み立てたロジックなのかもしれません。今までの苦勞の余りヒネクレ、イジケ、愚痴っているだけなのかもしれません。自分でもそれは分かっているのです。グスン。でもそれでもですよ、私の経験上では迷走しているデータが多すぎます。たまたまそう言うデータばかりに巡り合う幸運?の持ち主なのでしょうか。想像ですが、例えば「数値地図(地図センター販売)」など商品化されているデータはある程度はまともなのでしょう。問題は受託業務の一環として作成された地図データです。外部チェックを受けていない、言わば密室の閉じた世界のデータなので、いい加減なものが多いのではないのでしょうか。その意味では、外圧にさらされずにぬるま湯

につかっていた時代の銀行に似ています。特に某社(あー、許されるならばその名前をここに書きたい!!)にはこの状況からの脱却を強く望みます。

「Open」と言う外圧にさらされた時、地図データがどう変化して行くか。興味は尽きない所です。