W5.2

DNF local holdings: pros and cons

Jamie Justham, Dotted Eyes
Peter Roberts, Powys CC

Introduction

As part of the OS DNF Early Adopter Program Powys County Council and Dotted Eyes have been working together on the issues surrounding the management of DNF data. The aims of this collaboration are to identify clear data management regimes, benefits and migration paths for existing Landline users. The paper in its current form reflects the aims of the initial phase of the work that was to identify viable migration paths for existing Landline users.

This work remains ongoing and has not yet reached any final conclusions. As such this represents a record of the areas that the organisations are exploring and the presentation at the workshop will offer some conclusions on the models that this paper proposes as possible migrations paths.

In addition it will also offer a clearer perspective on the additional benefits that the improved data structures in DNF can be used for within an organisation, in particular focussing on benefits for users who are maintaining their own data rather than just using Land-Line DNF as a mapping backdrop.

Developing a model for your organisation

Determining your needs

Determining the model an organisation will use will come as a result of considering a number of key factors. These will include, but not be limited to, the size of data holding, the frequency of change in the data, the uses that the data is to be put to and the level of dependency within the organisation on the data being available. The cost of implementing the desired data model is also likely to be a determining factor.

The start point will however be considering how your organisation is currently using Land-Line and what limitations individual users are finding with it. In this area a key question is: are users just reliant on Land-Line as a backdrop mapping product or is it the basis for data capture and exchange?

If the answer is the former then in the short term a migration and use of one of the simpler models outlined may be appropriate. However if you are going beyond just simple mapping one of the more complex management routines will be required in order to expose the real benefits that can be gained. There is of course nothing to say you can't start small and then expand and change models; in many cases this may be an appropriate mechanism to grow your use of DNF over the medium term.

Land-Line and DNF – the key differences

While there are clear similarities between Land-Line and DNF in that they are a delivery mechanism for the OS survey scale data there are also a number of significant changes that mean there is a lot more to DNF than just Land-Line polygons. Many of these differences impact directly on how you will need to approach the management of DNF data.

The differences have been explained in detail elsewhere but can be sumarised as follows:

- Unique identifiers, (TOIDs), that have a defined life cycle in relation to a given object.
- A tile-free polygon based delivery mechanism based on a user-defined area.
- Inclusion of polygon data in the dataset.
- Change only update delivered on demand via the internet.
- A clear movement away from feature coding to a structured data description.

Taken in isolation these changes would not have major implications for data management. The fact that they are combined into a single product opens up a wide range of data management options. This also makes possible the opportunity to link data into common boundaries that can then allow for the movement of data at a later date if required due to positional accuracy improvements.

Data management models

Taking all these factors together it is possible to recognise a number of different data management models that an organisation might adopt. The potential benefits and pitfalls within these models will be explored in more detail within the presentation and subsequent supporting documents.

1 Local holdings using proprietary GIS systems

The Land-Line duplication model

This is the simplest model and would apply to organisations working on a themed approach to their existing Land-Line holding. Under this model organisations would identify the themes and feature category combinations required to match their existing Land-Line holding. Their data conversion software would then identify only the non-polygon objects and convert them into the same table structure. On supply of a change only update the same principles would apply to the data with items being altered on a delete/insert model.

Analysis: The only advantages that this model offers are that the data is maintained in a more timely fashion relative to survey date and the end user will appear still to be using Land-Line. This may also be the only option available to users on low budgets and limited data storage facilities. None of the major potential benefits of DNF will be realised. In addition unless the updates were spaced several months apart and data was archived before an update the ability to maintain a history would be significantly compromised.

The Land-Line extension model

This model would work based on the same principles as above. The linework that is present in DNF would be converted to mirror Land-Line as closely as possible. However in this instance the polygon data would also be converted and stored in a parallel set of tables. Supply of change only data would then be imposed over the top of existing data most likely on a delete/insert model

Analysis: This model offers some significant improvements. In this instance the polygon data is now available to users, which should have significant presentational advantages. On the fly linkage of point and polygon datasets is now possible but in the absence of historic data the use of data association is not advisable.

The polygonised Land-Line model

This model builds on the fact that the DNF polygons are actually reworkings of the polyline geometry. Whereas in the earlier models the polygons are discarded in this model it is the polygons that have primacy and the linework that is discarded. Again a lookup table would be required to mirror an existing Land-Line data model and fill styles would be left blank

Analysis: This model appears to be an improvement but needs testing to prove that it is possible to use only the polygons to duplicate Land-Line. In particular it will require careful consideration of the ordering of tables within a map view so as to ensure that for example the water river border is clearly visible and not

the field boundary. The theoretical model also fails to identify clearly how features such as rivers, which may be a polyline or a polygon, would be handled. This model needs formal validation with the use of sample data before it can be recommended as being viable.

The flat history model

This model assumes that a spatial database is unavailable to the end user and they will be maintaining the data within a proprietary GIS file format. The data will then be stored in either a single flat file or a number of files divided by category. The store will be used as a master with the periodic spinning off of current data files for more general usage.

Under this model on initial load the data will be placed into its storage environment with the addition of a new locally derived and empty end date field as an additional attribute. On supply of a change file the system would either insert a new record (insert), place an end date on a record and create a new entry with identical TOID but new version number and attributed data (modify) or place and end date on a record (delete).

Analysis: This is the first option to start to explore the full potential of DNF. The maintenance of all data within a single database means that as change occurs over time it will be possible to track back into an historic record. This facility will be essential if we are successfully to create a model for the use of TOIDs in the transfer of data and also to maintain them over time in order to cope with shifts that result from positional accuracy movement. The main problem of this approach will be the eventual size of the data holding within a single file. Separating the data into themes could of course reduce this, however the potential to move the data between themes will in itself create management problems. The need to spin off data and the system's flat nature would also mean that this model is intensive on disk space. To some extent it also reduces the flexibility of the system to cope with daily information updates.

2 Local holdings using spatial database

The historic spatial database model

This model is a variant on the flat history model. In this instance all the data would be stored within a spatial database. Attribute information would be rationalised by use of a relational structure and start and end dates would again be stored. Data would be supplied to the end user either through direct access to the server by use of spatial and SQL queries including in them a temporal element. While additional routines would allow the database to output a flat file environment for use in other systems where appropriate.

Analysis: With careful configuration this model offers the greatest potential to DNF users. By placing the data in a database environment there is the potential substantially to reduce attribute data volumes by virtue of lookup tables. At the same time the introduction of a spatial element would make it more realistic for users to access the core data and to do so both with current and historic snapshots. The most significant advantage to the data manager is that it removes the need for a duplicate set of data. The problems with this approach are largely to do with finance and infrastructure. The system is reliant on the introduction of a spatially enabled database complete with its own new cost and possible incompatibility with existing infrastructure. For many small to medium sized organisations this is likely to be a significant barrier to the use of this model of data management.

3 Remote data holdings

The supply partner model

This model could utilise any of the above models of data storage but is most likely to use a spatial database. In this instance all data conversion and management would be undertaken at a remote site most likely by an OS licensed partner. Organisations would then leverage Internet technologies to deliver this data to specific user interfaces within their departments.

Analysis: For small organisations with specific needs e.g. map enabling a call centre there are distinct advantages to this approach. It would clearly facilitate the benefits of a spatial database approach without incurring the initial costs. It is however limited by the need to specify specific applications and is reliant on the still emerging Internet technologies and infrastructure.

Streamed GML model

In this model no data would be stored, instead GML would be requested on the fly and downloaded and translated en route to a browser or GIS application. Data would be likely to remain current, as it would possible be sourced directly from the OS.

Analysis: GML is proving to be a very verbose format and questions remain as to how fast data can be converted and rendered. For large areas download times are likely to be significant and if your application needs to link to historic data it is not going to be available. Version 1 of DNF via the OS will not be supporting direct streamed supply of data.

Implications for change

All these models will necessitate a degree of change within an organisation. In some cases this change may be as limited as the retraining of the corporate data manager to enable them to use a new software translation solution, while others will require a substantial investment in IT hardware infrastructures. The next stages of this work will be to examine some of the claimed benefits of DNF in relation to the infrastructure that will be required to deliver them.

Of specific interest to the project are benefits to the user community in terms of data supply in areas of limited feature change, the benefits for users responsible for linked datasets and the implications for the management of change due to positional accuracy work.

Conclusions

With real world sample data now available to users it is becoming apparent that while simple data management techniques can be transferred from the management of NTF holdings, to fully explore the potential of DNF data will require a more rigorous approach to data management procedures than has been put in place in the past. While the potential benefits are clearly greater than Land-Line so are the pitfalls that await an unwary data custodian.

References

Roberts, P D “Land-Line and DNF Data Management based on a local store of data” OS DNF early adopter discussion paper. 2001