W06.1 Castles built on sand – are GIS databases built on a solid foundation ?

Tom Timms, Director, Apic and Giles D'Souza, Consultancy Division, Infoterra Limited

1 Introduction

1.1 Background

One of the biggest advantages for map users in the United Kingdom is a common, continuous, universally accepted and unambiguous National Grid co-ordinate system (created in 1946). Such a widely used standard is not available in many other countries, including most of Europe.

Another advantage of UK mapping is the easy and relatively cheap access to very detailed large scale paper, and digital, map data from the Ordnance Survey (OS). This has driven a significant take-up of data derived from the National Topographic Database (NTD), and many users have used these data as the definitive backdrop against which to capture their own information. For example, many utility company surveyors have positioned their asset features by measuring offset distances to building corners shown on the OS large-scale map. To date, only the *local* consistency of the map has been important and maintained. Nowadays, however, surveyors may use differential GPS tools providing co-ordinates in a worldwide reference system to half-metre accuracy. Any *absolute* error in the Ordnance Survey map will therefore cause a discrepancy when compared with the GPS-derived data and make it difficult to create a true representation.

It is accepted that OS (and any other) mapping can never be perfect – a map is in fact a simplified model of actual physical reality, and any such model can only be an approximation. In order to emphasise this, OS large-scale mapping is supplied with a clearly defined positional accuracy specification which indicates the typical discrepancy that can be expected between large-scale map detail and precise positioning to National Grid co-ordinates (see tables below). However, during the reconciliation of the data to match the National Grid, in particular in the overhaul of the County series maps (which effectively existed on different local projections), a number of approximations were made, which mean that the *absolute accuracy* of the mapped data has become questionable; possibly outside the limits specified and, incompatible with data obtained from independent surveys.

The increasingly widespread adoption of GPS in surveying and navigation (along with the use of orthophotographs created from independently-collected GPS points) has highlighted such inaccuracies. The wider availability of GPS in surveying, is also making the *absolute positional accuracy* of map data with respect to a global reference system more important than ever.

With modern GIS applications increasingly making use of digital map data derived from OS mapping, and other digital data being more widely shared between organisations, these positional differences are becoming a significant issue.

1.2 Accuracy of Ordnance Survey map data.

The table below shows the expected absolute and relative accuracy values for well defined points within each accuracy category of mapping derived from the National Topographic Database. Those relevant to the problem of Positional Accuracy are shaded.

Table. 1. - Absolute Accuracy

Scale and method of original survey	Expected Absolute Accuracy at differing confidence levels		
	63%	95%	99%
1:1250 scale	<0.5m	<0.8m	<1.0m
1:2500 scale resurvey/ reformed	<1.1m	<1.9m	<2.4m
1:2500 scale overhaul	<2.8m	<4.8m	<6.0m
1:10,000 scale	<4.1m	<7.1m	<8.8m

Table. 2. - Relative Accuracy

Scale and method of original survey	Expected Relative Accuracy at differing confidence levels			Maximum measured distance
	68%	95%	99%	
1:1250 scale	±0.4m	±0.8m	±1.0m	60.0m
1:2500 scale resurvey/ reformed	±0.9m	±1.8m	±2.3m	100.0m
1:2500 scale overhaul	±1.2m	±2.3m	±3.0m	200.0m
1:10,000 scale	±3.5m	±6.7m	±8.8m	500.0m

1.2.1 Positional Accuracy Improvement (PAI) Programme

OS mapping in rural areas (1:2500 scale) was derived from County Series maps reconciled onto the National Grid by a manual 'overhaul' procedure. Potential problems with this mapping have been known for many years, and errors of up to 13 metres have been discovered. In order to address this problem, in April 2001, OS started the positional accuracy improvement (PAI) programme, with the aim of bringing all of the 1:2500 scale data within a uniform resurvey specification within five years.

This programme entails the "manual" movement of detail on OS maps – in some cases blocks of detail are shifted and rotated by several metres. In consultation papers, OS undertook to liaise closely with customers throughout the programme and advised that the overall timescale of the programme would be tailored to the needs of major users. The PAI programme was split into two parts - Rural Towns and Other Rural Areas.

Rural Towns

The Rural Towns programme covers some 2300 1:2500 map data tiles. The towns have been selected and prioritised according to their size (geographical extent), the level of building development planned in their vicinity, and the level of known positional accuracy discrepancies in that area. PAI in these towns will be achieved by full resurvey methods within 3 years. The resulting data will be to an absolute accuracy of ± 0.4 metres, ie. to the current 1:1250 scale specification, but will still be available as 1:2500 map tiles.

Other Rural Areas

The other rural areas account for some 155,000 1:2500 map data tiles. PAI in these areas will be achieved by methods that involve resurveying those features that are out of tolerance, or photogrammetric interpretation. The resulting data will have an absolute accuracy of ± 1.1 metres. Priority has been placed on those areas requiring most urgent rural revision update, but customers' own digitising programmes were also (meant to be) taken into account. The full programme is planned to be completed in five years *coinciding with the rural revision cycle that also began in April 2001.*

Rural Revision

The Rural Revision program aims to update 1:2500 scale mapping, accounting for the often significant changes that have occurred since the last revision cycle.

Potentially, the combination of both the PAI and Rural Revision programmes has major repercussions for users of rural maps, especially on any project which collects considerable amounts of data in relation to published OS data. Map changes as a result of rural revision combined with PAI effects mean that the overall differences between the map sheets are considerable. Figure 1 shows the combined effects of a PAI correction with Rural Revision (green lines are pre-PAI; blue lines are post-PAI).

Figure 1. Combined effects of PAI and Rural Revision (blue lines are post -PAI; green lines are pre-PAI)

In some areas the combined effects of Rural Revision and PAI, especially where the maps have not been revised for over five years, are very significant. For example, many field boundaries change and even river courses of can differ by several metres.

1.2.2 OS MasterMap

An additional complicating factor, has been the release of OS MasterMap in 2001. This new generation data set is derived from data held in the NTD. It suffers from exactly the same problems as any other large-scale digital data. However, the near-simultaneous release of PAI,. MasterMap and Rural Revision, has caused considerable confusion in the user community - with many struggling to develop a rational and structured approach to adapt to even one of these major changes. It should be noted that PAI will affect all new 1:2500 scale mapping, LandLine or Mastermap..

1.3. Link files

As a recognition of the problem, and to assist users to process their own data to fit the PAI data, OS have published a "link file" for each PAI map sheet. The link file is a set of "control points" that indicate the position of certain features before, and after the PAI. (see Fig 2).

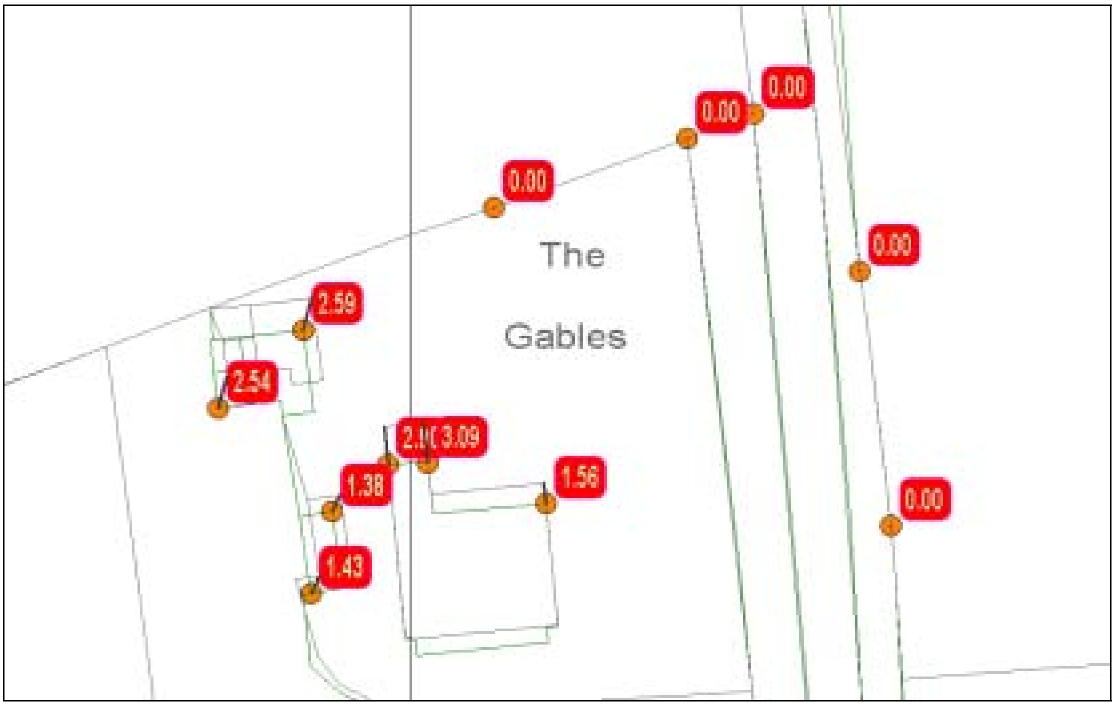

Figure 2. Example of link file data. The dot = position pre-PAI, and the end of the line = post-PAI (lengths of the links are labelled)

The first set of PAI data (with corresponding link files) was published in December 2001. Customers and software suppliers who received the data raised a number of issues mostly related to the quality, number and distribution of links, and edge effects.

Link file quality and distribution

The reliability of the links supplied with the first PAI tiles was questionable, with a significant number of "rogue" links within each sheet. Any mathematical transformation of data relies on the link files to provide details of the shift, and any erroneous control can have detrimental effects on the quality of the output. The spatial distribution of the links was also a major issue for many users with some rural tiles having only a handful of links.

Edge Effects

Whatever the block size, there will be an external edge to it and during the course of the PAI programme this may sometimes abut a 1:2500 tile that has yet to be improved. Where this occurs, there will inevitably be some discrepancies between detail across tile edges.

Ordnance Survey *smoothes out* these discrepancies on the unimproved tile edge to meet the corrected detail on the improved tile. This results in positional shifts on the unimproved tile. Figure 3 illustrates some of the adjustment made across a PAI-adjusted block. The original data on the lower tile are shown in green- the smoothed line is shown in red. Until the lower block is properly adjusted, the red line will form an artefact, which may or may not match the PAI-adjusted line. As PAI tiles become more available, OS plan to adjust all features within a certain "collar" of the PAI block. Thus users may also find PAI effects on the edges of map tiles which have not yet been PAI-adjusted.

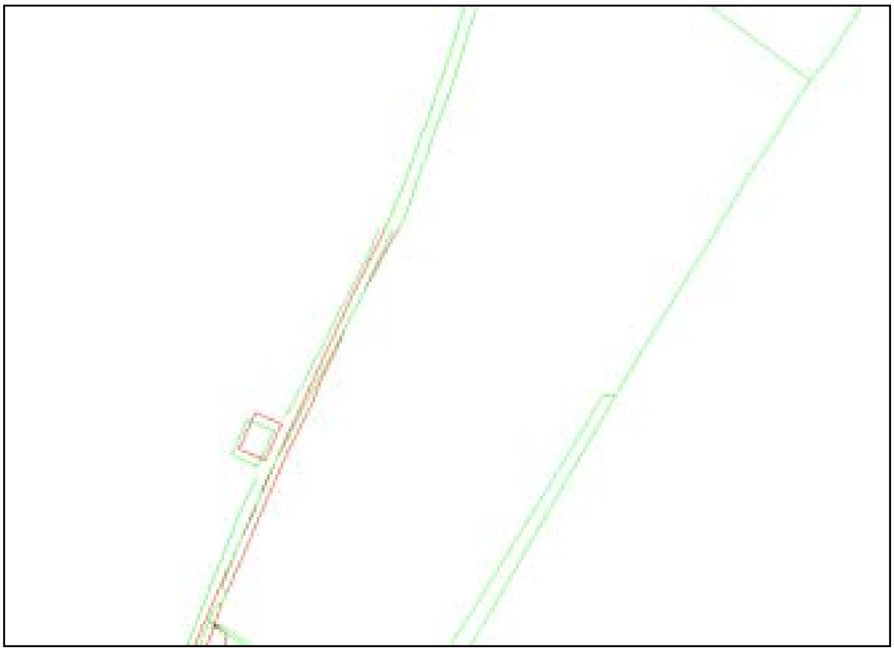

Figure 3. Example of an artefact of processing across a PAI-adjusted block. The red line is the manually-adjusted data to make it fit the PAI-adjusted tile above it. The horizontal discrepancy between the green and red line is nearly 5m at its extreme.

PAI programme postponement

The initial releases of PAI data and their associated link files caused some concern among the user community. Many discovered the scale of the problem facing them and potential problems in the quality and density of the link files provided. Following consultation with customers, OS decided to postpone the data banking of PAI data until further analysis.

OS have now revised the specification for link files - it is believed that the latest release will include significantly more points per tile (especially in rural areas) and fewer rogue links. The first set of improved data is set to be re-released shortly with a new programme timetable.

1.4 Modelling positional shift

Judging by link file data released to date, the accuracy variations caused by the overhaul of the County Series mapping to the national grid *are not systematic* - no large areas of constant shift can be identified. The situation is complex with several revisions having taken place on each map sheet, some selected revision by local survey (perahps using real-time kinematic transformations) and some by complete map overhaul (using a combination of survey and photogrammetric methods). Systematic shifts may be apparent in suburban areas where development has been surveyed using local control but it does not seem feasible to identify and model these in any practical way.

Therefore, it is highly likely that most users will have to use complex local transformations to adjust their data to match PAI data satisfactorily.

2 Issues, Techniques and Options for change

Given the positional accuracy problem with OS 1:2500 data (Landline or MasterMap), users need to examine what effect the PAI programme has on their data, how serious the problems are, and what options are available to help resolve them. The answers to these key questions depend, to some extent, on the nature and capture-history of the user’s data. For most organisations there may not be a single answer.

2.1 User data characteristics – absolute or relative data

One key characteristic that needs to be assessed is whether the users data have been captured relative to the OS large-scale map background, or fixed absolutely using some other measurement techniques (GPS or Ground Survey for example). Given the expense of ground survey and the only recent widespread availability of GPS, it is likely that most users have captured their data relative to the OS mapping. This is certainly the case where data have been manually digitised from old paper records or drawn on screen using heads-up digitising.

Data captured using 1:2500 Landline as a background are going to be significantly affected by the PAI programme. With median shifts of more than 1m within a tile and individual shifts of more than 10m in some places, there can be significant effects on the appearance, representation, and perceived “accuracy” of data.

The example below (Figures 4. and 5.) shows a utility network located in the pavement overlain on the pre-improved OS LandLine data, and the same information superimposed on the PAI version, which has shifted by an average of about 3m.

The utility data, originally shown in the pavement, now appears in the middle of a road, and crosses gardens - clearly an unacceptable situation for the utility company and it’s data users and customers.

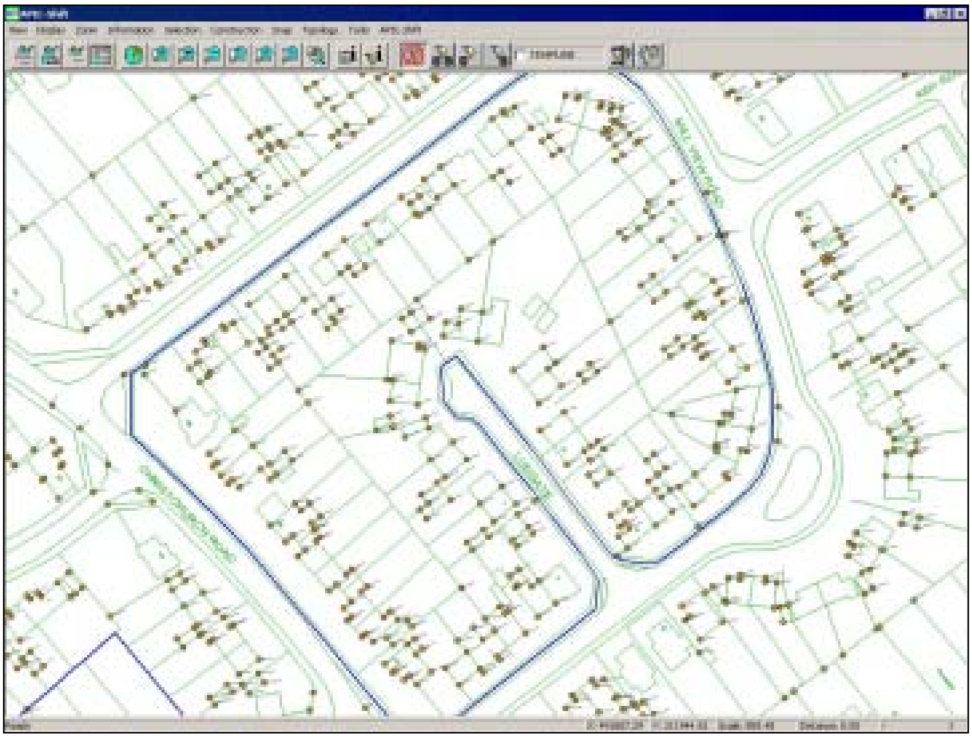

Figure 4. Pre PAI Landline with utility network overlain

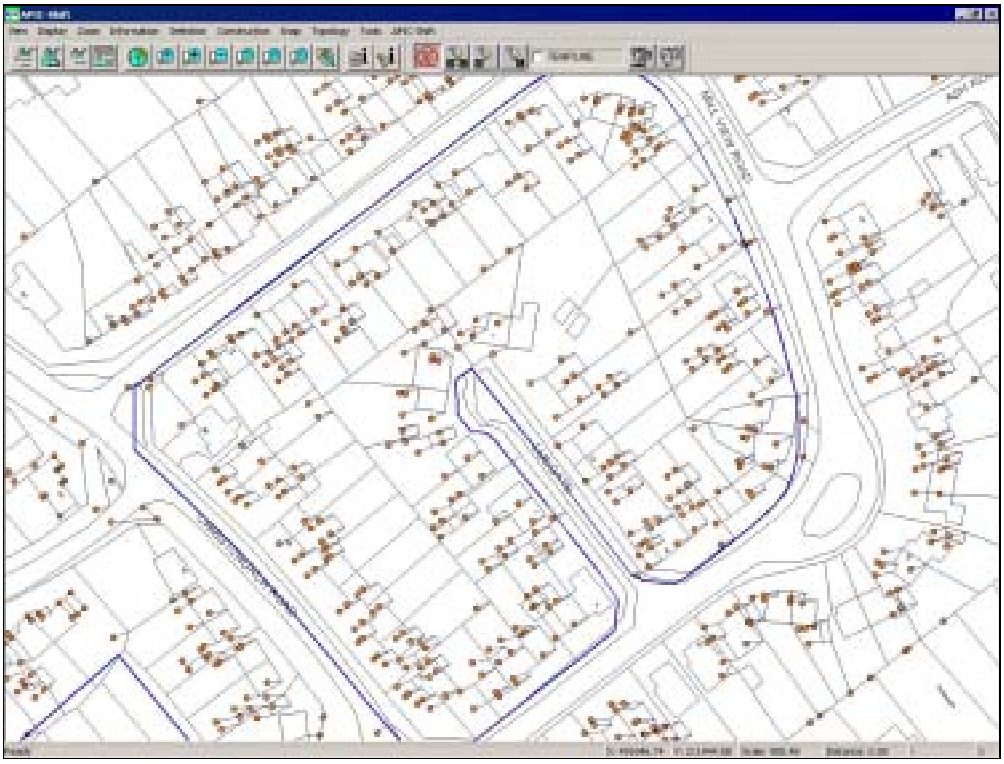

Figure 5. Post PAI Landline with utility network overlain

2.2 User data characteristics – relation to OS large-scale data.

Another key characteristic is the topological relation between the users data and the OS digital data. For example, where users data have been captured wholly by snapping to or "cloning", LandLine features, (as in the case of capturing buildings, for example) a strategy of "*shift and snap*" could be adopted. Cases where users data had been captured simply in relation to large-scale data with no actual snapping (as in the utility example above) could be treated differently. However, because of the nature of Landline, most users data are often only *partially* snapped to Landline, which means a mixed approach may be necessary. A classic example of this would be the digitisation of Basic Land and Property Units (BLPUs), or farm field boundaries, which may only have *part* of their boundaries indicated on the OS map.

Further, as the PAI data are often different to the pre-PAI versions because of real world change (see Figure 1 for example), there are likely to be complications to resolve where users data had been snapped to OS features which have changed or even ceased to exist on the revised map.

2.3 Possible solutions to the issues raised by the PAI programme

A range of possible solutions are available to those affected by the OS PAI programme with different costs and other implications. In organisations managing a number of datasets, no one option may be suitable for all datasets.

The options can be categorised into

- No action (do nothing)
- Re-capture or edit data manually
- Use the OS supplied link files to perform an automated transformation
- A combination of some or all of the above (hybrid solution).

No action (do nothing)
Although this is undoubtedly the cheapest option, it is unsuitable for many users unless absolute accuracy and good map representation is limited or unimportant, or where all data have been fixed by absolute methods only. A variation of this solution is to "*do nothing for a while*" and wait until the PAI programme is finished. However, this may mean "*freezing*" the background base mapping for five years and not introducing any customer data captured by absolute means. Most businesses need to update their base map data regularly and many businesses are starting to utilise GPS and ground survey more and more, so again, this may be an option only for a small number of users.

Re-capture or edit users data manually
This option is likely to be relatively expensive especially for organisations with large volumes of data that have been captured over a long period. Re-digitising or editing large volumes of data is expensive and time consuming, but may be the best way to guarantee the highest possible quality result. Indeed, it may provide a valuable opportunity to re-evaluate the accuracy, attribution, content and meta-data of existing digital data, and to use the PAI change as an opportunity to re-engineer or enhance information. However, it may be difficult to justify the costs to business managers.

Use link files to perform an automated transformation
The OS supplied link files provide the basis for a mathematical transformation of users data to match the PAI large-scale OS data. Theoretically a number of transformation options are available, though in practice these may be limited by those available within specific software packages. Common transformations available in a number of products include:

- Rubber sheeting
- Helmert transformation
- Affine transformation
- Triangular irregular network (TIN)

Whatever specific transformation algorithm is used, the quality of the result is directly related to the quality of the input points (the link file data). Any incorrect reference points can cause unexpected and potentially incorrect results. In addition, the distribution of the link file points is also important. Early releases of link files had only a handful of reference points in some rural tiles, which could result in only a few points having a disproportionately large influence on the transformation with potentially detrimental results.

Furthermore, a transformation is not guaranteed to provide a globally consistent match to the improved large-scale data, even if those data had been snapped to or derived from OS data in the first place. Whatever transformation is used, the transformed data will only match where there is link file information and may not necessarily do so in areas without reference points.

Hybrid solution
A common solution to the issues raised by the PAI programme may be to transform users data automatically using the link files and to then *post-process* the data (in automatic and manual ways) to resolve any residual problems.

Although some software packages offer the ability to snap users data automatically to OS PAI data after they have been transformed, in the authors' experience it is unlikely that any completely automated solution will be entirely successful. Some manual intervention and quality control is needed to ensure consistency and quality. The use of aerial orthophotographs may be useful as additional validation.

3. Lessons learnt

Many users of Landline and MasterMap data have a major issue as a result of the PAI programme. Given the advances in technology, it is possible that positional improvement of data will occur again in the future (for example, new datums and worldwide reference systems are already being discussed), although such a major step change is probably unlikely in the short-term. In many countries, much nationally-important data (for example cadastral information) is held on local projections. As these are digitised and harmonised across regional and country borders, similar problems will occur. Therefore, it is very important to assess what lessons can be learnt in order to minimise similar problems in the future.

3.1 Lessons for users data collection

The more information about any object in a GIS the better it can be managed. However, in the past, few GIS users have recorded information relating to the source of their data, for example their accuracy and quality (even at the dataset level let alone the feature level). Such information could be significant in assessing possible solutions to the issues described above. For example, knowing whether a feature was snapped or not to Landline in the first place can significantly simplify automated processing. With MasterMap data as a source it will now be possible to record the unique objects (TOIDs or part TOIDs) that make up an object and this may assist, to some extent, the management and possible automatic update of features when OS base maps change (from rural revision and PAI change).

3.2 Lessons for data providers

The PAI programme is an important and difficult issue for the OS. They would certainly be criticised for publishing data that does not match GPS surveys, but could also be criticised for changing their base maps in such a major way. Few users question the need for the programme, but few are happy about it and how it has been managed. Introducing major changes separately (Rural Revision, PAI and Mastermap), and allowing users to evaluate and manage the changes in a methodical, staged manner may have reduced a lot of confusion and consternation in the user community. A more systematic approach to providing data and differentiating sources of change would also simplify the problem.

Much better communication and information between the OS and its users is also necessary, perhaps with sample data sets and prototype link files being more widely distributed and discussed with all affected users.

4. Conclusions

The OS PAI programme will have a major impact on many users of OS large scale data, particularly those in rural areas. Users will have to develop strategy to suit their data and business. The programme is undoubtedly necessary to improve the quality of the OS data but the issues that it raises can teach us important lessons for the future.

The PAI program should also make us question our confidence in data and ask "How well do we know the foundations on which our GIS is built ?".

References:

Ordnance Survey PAI programme www.ordnancesurvey.co.uk/positional

Ordnance Survey MasterMap www.ordnancesurvey.co.uk/mastermap

APIC-Shift www.apic-eads.com/shift

Contacts:

tom.timms@apic-eads.co.uk

giles.dsouza@infoterra-global.com