WO2 Positional Accuracy – A Moving Story

Colin Smith, Computer Mapping Centre Manager, HM Land Registry

1 Land Registry – Business Background and GI Assets

Land Registry (LR) is a self-financing government agency. It gives legal effect and title to all conveyancing business in England and Wales and provides a range of services to organisations and individuals concerned with interests in property. It employs over 8000 people and deals with about 20 million applications and enquiries annually.

Guaranteeing title to land is our business and this depends upon the effective management of Geographic Information Assets held by LR. 650,000 Ordnance Survey (OS) maps form the backdrop of the LR Index Map (IM) on which are overlaid the positions of every registered title in England and Wales. Searching the IM reveals the unique title number essential for the Conveyancer to progress the property transaction.

A 3 year project to convert paper IM records into vector form is at an advanced stage and is on target for completion Spring 2004. This will have created vector polygon indexings and seedpoints for 18 million titles and is a key asset development to support LR business. The IM vectorisation project also lays the foundation for future electronic access and e-conveyancing initiatives.

2 Impact of Positional Accuracy Improvement on Land Registry Business

It is widely understood that the OS PAI programme for rural areas has significantly “shifted” boundary, building and other features to more accurate positions. What perhaps has not been publicised is the practical affect this has on GI Assets within individual organisations and how the problem has been addressed. This paper describes the affect on, and the actions taken by, LR.

Within LR, PAI has had a major impact and presented a business critical problem that could not be ignored. In essence, when a PAI map is substituted for a ‘pre PAI’ map the positions of the IM polygons (indexings) are no longer synchronised with the new geometry.

In many cases the amount of shift is significant such that the polygon for a property, e.g. 10 High Street, may now straddle or completely overlay the geometry for 11 High Street. This compromises the quality and function of the IM and, if left unresolved, will risk giving misleading search results to Conveyancer's as to the registered status and title number of land. Such errors would generate service complaints and could expose LR to claims for compensation.

Given that the IM is in constant use for searching and under continuous update as new titles are added or changed, consistency in the underlying OS map is vitally important.

To compound matters, the IM vectorisation project had to progress and complete indexings in areas that will be updated later as part the OS PAI delivery programme.

Accordingly, LR business representatives required a solution that would support ‘business as usual’ and also ‘move’ polygons and seedpoints to coincide with updated PAI geometry.

3 LR Strategic Business Decision and Initial PAI Investigations

In maintaining 'business as usual', an early decision was made that it would be best to 'freeze' the IM in its 'pre PAI' state and record indexings thereon. 'Polygon shift' would have to be done but it could be deferred until a later date under a fully managed solution. This had the advantage of ensuring that existing geometry, the AddressPoint data and the indexed polygons remained synchronised and thus protected the business from issuing erroneous results to Conveyancer's. Preparation of new title plans did however continue to be based on the latest PAI maps but the indexed position was re-adjusted to fit the 'pre PAI' IM thus maintaining pictorial relationships with existing polygons.

This was an important business decision as the strategy also enabled the IM vectorisation project to proceed and gave the time needed to pursue solutions to the 'polygon shift' problem.

APIC, Dotted Eyes and Tenet IT Mapping (Tenet) all offered possible solutions and each were helpful and supportive in understanding the LR polygon shift problem. LR acknowledge their efforts and record that the decision to explore the Tenet solution was driven principally by existing Tenet data formats and system integration issues and not on the credibility of the other suppliers or their offerings.

4 LR Business Requirements and PAI Feasibility Project

The business specified an exacting target – deliver a fully automated solution that will correctly reposition any LR polygon without any operator intervention. A feasibility project was launched and detailed analysis of LR polygon types commenced. Tenet were commissioned to build a prototype Positional Accuracy Control Module (PACM) which, driven by a rule set and using 'pre PAI' and PAI tiles, LR IM polygons and OS link files produced new polygon positions. Initial results were poor and it was very quickly established that the defined business target was unachievable.

The complexity of polygon types, link file weaknesses, real world changes intermixed with PAI change, a rule set that needed far more refinement and a much improved PACM were identified as the key issues. These were referred back to the Project Board and authorisation was given for further development work. The business also abandoned the requirement for a fully automated solution and agreed a more realistic target that allowed for a low percentage of polygon shift failures. The business recognised that user intervention to manual re-alignment 'failures' was unavoidable but for resource and costs reasons had to be minimised.

5. Revised LR Business Requirements and New PAI Project Development

The development of a revised User Requirement and investigative work with Tenet began. Immediately the most difficult issue was determining what percentage success rate the business would deem as acceptable, could it be achieved with a software solution and how would success be measured for system acceptance purposes. The business required a controlled environment to be built which enabled at least 90% or better of polygons to be realigned without user intervention. The controlled environment and interface would be designed and built by LR's software development team in Plymouth and Tenet would build a new polygon shift module (PACM).

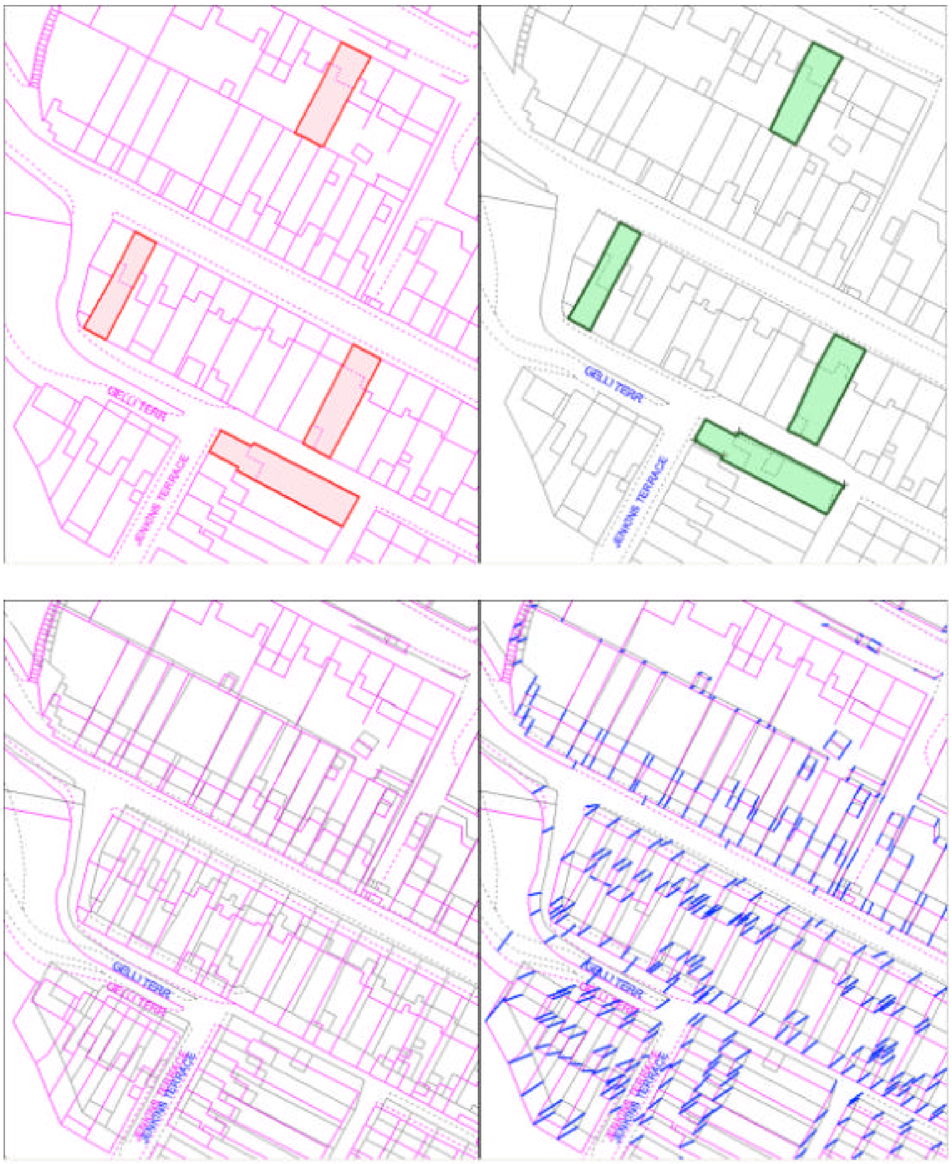
Given the huge variety of LR polygons it was agreed that Tenet would need a 1000 test case polygons defining both the 'before and after' shift results that had to be delivered. Against each polygon, the business rules had to be precisely defined in order that the logic for each polygon shift decision was understood and agreed. All parties acknowledged that without agreed rules it would not be possible to achieve consistency or accept a delivered solution.

It was also agreed that an iterative approach to the PACM software development would be essential as queries would inevitably arise on some of the polygon shift results and the interpretation and priority of the rules. This would be progressed through delivery of interim beta versions of the PACM; strong technical communication links with the Tenet team and detailed technical meetings at which problems and solutions were moved forward.

6 Preparation of 1000 Test Cases and Rules.

Preparing this data set and defining the rules was a complex and time-consuming task. It quickly highlighted a number of data issues concerning the polygons, seedpoints and the PAI data. It also revealed that user X did not always concur with the results sought by user Y and common agreement had to be negotiated. The data-set preparation involved creating a vector polygon overlaid on the pre PAI map and the equivalent vector polygon to be created on the PAI map.

The test case examples were selected from 1/2500 towns and villages and included the widest selection of property types and boundary situations that could be found. Old and new residential houses, industrial estates, farms, fields and rural properties, complex blocks of flats, undefined and partly defined land with or without buildings, overlapping interests (rooms over passageways and maisonettes), roads, passageways and garages were all included.



Top left: Pre-PAI map with overlaid LR polygon. Top right: PAI map with shifted polygons
Bottom left: Pre-PAI and PAI map differences Bottom right: Pre-PAI and PAI map with link data

These illustrate dramatic differences that can occur in small areas and generalisation of property divisions which appeared in greater detail on the Pre-PAI maps.

LR developed a set of 12 main rules and recorded the affecting rule(s) against each polygon and also the precedence that was to be applied to each rule.

As the project progressed, it proved necessary to re-visit and refine the rules and test cases to cater for new circumstances and precedence conflicts. This largely arose as the Tenet solution evolved and greater understanding of the data and the PACM operation and algorithms developed.

7 The Tenet Solution

Tenet's approach was based on a detailed analysis of the input data and establishing their relationship to the original mapping.

In essence, the shape of an input polygon is abstracted to derive key waypoints that are expected when creating the PAI-corrected polygon. Coincidence with the map data and adjacency with neighbouring polygons are assessed, bearing in mind that correspondence does not have to be exact to be intended.

The input data are then checked against the PAI-corrected map data, as it is quite possible for them to have been captured against several generations of mapping, up to and including PAI-corrected data.

Even with this level of analysis and the OS link files for guidance recreating the polygons is far from just "joining the dots". Rather, the information plays the role of a treasure map. Just as Jim Hawkins relied upon Captain Flint's famous map in Treasure Island for hints and general directions, so the shape of the input polygon and details of its relationship with the original mapping controls the manner in which the path that forms the edge of the new polygon edge is followed.

Unlike its human counterpart, however, the software is able to evaluate thousands of possible options instantaneously. At any given step, there may be a choice between following the map data, jumping to the expected location of the next waypoint, stopping at a nearby edge or vertex, following an alignment implied by the map data, and many other possibilities. And, of course, each is the starting point for decisions on the next potential step.

Balancing the likelihood of the various options gives the approach its characteristic fuzzy nature. Minor deviations from the expected route are tolerated if no better alternative can be found. As each successive waypoint is reached, the overall confidence in the solution increases. However, backtracking is allowed should one encounter a blind alley, even after a waypoint has apparently been reached.

Link data also permit the application of soft constraints within the treasure map algorithm: if the original data pass through or close to a link point, then deviations from the corresponding point in the PAI-corrected map will be penalised.

Material in Section 7 was provided by Tenet – see acknowledgements

8 OS Map Data Issues Arising during the Investigation and Development

On the PAI OS maps, real world changes are intermixed and not readily distinguishable from true PAI changes. No attribute information is held against features and many test case decisions relied heavily upon years of accrued LR 'plans experience' to enable interpretation and acceptance of 'common' polygon boundary features.

Detailed building divisions on some pre PAI maps had been 'improved' by generalisation and straightening on PAI. This loss of granularity then presented common feature interpretation difficulties and led to questions being raised with the OS in respect of the quality of translation of aerial photography. OS responded positively and have instigated data quality improvement initiatives.

Top left: Pre-PAI map with overlaid LR polygon. Top right: PAI map with shifted polygons
Bottom left: Pre-PAI and PAI map differences Bottom right: Pre-PAI and PAI map with link data

These also illustrate major change and the 'loss' of property forecourts and front gardens which appear to be neither in the road, nor the individual properties.

Initially, link points were found to be scant, non-existent or of questionable value in rural areas and in areas of complex change. The position, pattern or degree of shift reflected by links was not always consistent it was considered that reliance on links alone would not be a safe basis for polygon shift decisions. Incidence of 'rogue links' and the absence of links in rural areas were raised with the OS. OS responded and later increased link coverage and released new link files. Whilst an improvement is noted experience has shown that the additional links have given only a marginal benefit and that some rogue links, inevitably, persist.

On areas of mixed residential development where extreme shifts – up to 6.0m – were found visual identification of the 'same' building and its associated boundaries proved difficult but was helped with the use of dual colour overlays of pre PAI and PAI maps. In preparing the test cases where streets of similar properties exist special care was needed to ensure the correct properties were selected particularly if link points were absent.

Further, extreme shifts also had dramatic effects at PAI tile edges. Consideration of 'before and after' positions can have the effect of 'moving' a complete property from one tile to the next. In all cases, the affect on boundaries and polygons at edges of the eight tiles which abut any PAI tile have to be accounted for. The PACM was designed to include all peripheral tiles (collar or otherwise) into the process.

Delivery of incomplete PAI blocks and collar tiles, unexpected delivery of PAI tiles in non PAI areas and a lack of certainty as to which PAI tile was the very first PAI version in any area have all presented development problems. The processes in the LR designed solution required rigid and timely management of large volumes of maps, polygons, seedpoints and link files. Supplementary systems outside of the scope of the original project had to be developed to quality check the completeness of OS data and establish other control data unavailable at the OS. Defect reports were supplied to the OS to support improvement to their internal processes.

Seedpoints to which AddressPoint data and LR title numbers are linked also needed to be re-aligned to accord with the PAI geometry. Initially, this did not appear to present any new problems and only required an equivalent seedpoint shift to retain compatibility with the new map detail. However, whilst the UPRN value would be retained, the shifted co-ordinates would no longer accord with AddressPoint values. This approach would then run contrary to the OS decision not to move seedpoints that remained within the perimeter of its building. As updates to PAI maps can be received daily by LR but AddressPoint updates are only received quarterly a data synchronisation issue was raised with OS. This has however been a data management problem that LR has had to resolve internally by adding change history to seedpoints.

9 LR Polygon Data Issues Arising during the Investigation and Development

Polygons with fully or partly defined boundaries on pre PAI maps but now undefined or differently defined on PAI. Undefined situations are common place in areas of demolition/re-development or road widening schemes. Partially defined boundaries regularly arise on open plan housing estates and in the division of passageways, roads or parking areas. Comparing and achieving consistency of shift where map detail does not exist required special consideration. The general shift (assessed from the links), and maintaining the pictorial relationships of polygons shapes to the topography of the area were deciding factors in the rules and algorithm development.

Small inconsistencies introduced by operators during data capture and IM vectorisation were found. Minor over-laps and under-laps between polygons therefore presented a data quality issue that needed to be resolved. Other inconsistencies not previously found had been introduced through minor changes to map detail on successive editions of the underlying IM. Careful investigation established that a tolerance test could be applied in most instances to readjust these discrepancies. This considered the ‘gap’ distance, adjacency and polygon relationship to the pre PAI and PAI map detail. This also drove the rules.

The adjacency – or otherwise – of polygons was also important. Where pre PAI polygons adjoined then that adjacency had to be maintained on PAI shift. Equally, any unregistered separating strip had also to be maintained.

Top left: Pre-PAI map with overlaid LR polygon. Top right: PAI map with shifted polygons
Bottom left: Pre-PAI and PAI map differences Bottom right: Pre-PAI and PAI map with link data

These illustrate polygon variety in rural areas. These include very narrow strips of land and undefined boundaries on both Pre-PAI and PAI maps.

10 LR PAI System and Process Flow

With the exception of the Tenet PACM component, all software used in the LR PAI process has been developed in house. The PACM is a component of the Tenet PAI Correction Studio and performs the shift of seedpoints and polygons. The interface to the PACM has been developed entirely in-house specifically to meet LR business requirements. This system manages all stages of the PAI process from delivery and validation of PAI data; extraction of existing polygons, historic and adjacent maps, seedpoints and link files; submitting the data for PACM processing; reviewing, accepting and amending the results sets; committing the updated data to the LR database and, finally, archiving the history.

The following describes the system stages.

a Delivery of PAI Data

To avoid the PAI data delivery problems experienced in the past, OS will send a pre-delivery list before commencing on-line despatch to LR Information Systems Directorate at Plymouth. This is checked and agreed using in-house developed software and any discrepancies are resolved before the delivery starts. Any discrepancies are reported back to OS, and a new pre-delivery list is supplied for re-checking.

Data is received daily electronically, and PAI data is only accepted if it is on a confirmed delivery list. When the delivery is complete, a final check is run to confirm all the data is present.

b Planning and Allocation of PAI work

At LR Head Quarters, a senior user then uses the in-house system to plan and allocate work to the LR Local Offices responsible for specific geographic areas. This ensures that Local Offices work on their own maps and also to manage edge-match problems between PAI blocks.

This planning process releases individual blocks to a Local LR's 'queue'. The queue is processed by an overnight application, which copies the PAI maps, the pre-PAI maps and the link data to the Local LR's server. It also copies polygon and seedpoint data from the live environment to work tables so the shift can be done without risk to the live database tables whilst still enabling business as usual. Data for two blocks is made available to each Local LR, which is enough data to keep the PAI staff actively employed the following day.

c Executing the Shift

Local LR staff are presented with a pictorial representation of the PAI tiles in the block, and fire the shift process for each in turn. The software finds all the polygons and seedpoints that fall in or partly in the tile, and passes these together with the relevant maps and link data to the Tenet PACM module to be shifted. The results are colour coded on the confidence value returned from Tenet's PACM module.

Green	Good result
Amber	Possible result
Red	Doubtful result
Blue	Already PAI or no map shift

The Local LR staff review the results and accept or amend them as required using the standard Tenet toolset. Finally, the reviewed and amended results are committed to the work tables in the database.

d Committing the Updated Data to the Live Environment

The following evening, any changes to the data in the live environment are reflected in the work tables, making the changes available for shifting the following day.

This process is repeated until there are no new changes to the data and all the data in the work tables has been shifted.

At this point, the overnight process:

1. copies the polygons and seedpoints from the work tables to the live environment
2. copies the new PAI map as the new IM backdrop
3. archives the original polygons and seedpoints

e AddressPoint Updates

Finally, where the PAI AddressPoint data is available, a load program is run for the PAI areas that have been completed (data previously held in abeyance awaiting polygons shift and new maps) thus synchronising UPRN co-ordinates with the new IM detail.

f Cyclic Process

The processes then repeat as new PAI blocks are progressively released by the OS as the PAI programme continues in the years ahead.

11 LR Project Development Summary and Implementation Plans

This project entirely concerns the problem of maintaining the quality and updatedness of LR's Geographic Information Assets. It has presented great technical challenges and has required new and innovative approaches to be explored. Our experiences and the work undertaken by the OS and Tenet in supporting this project will, it is hoped, enable others who are presented with similar data issues to tackle their own PAI projects with greater confidence and benefit from the work that LR has done.

At the time of writing, 18 July 2003, the system build is close to completion and final stages of system testing and user acceptance are planned in the near future prior to a pilot implementation and trials in one Local LR Office. Only when this is in place will LR be able to fully report on the expected success rates of what will have been a two year project costing £0.5m.

It would be easy to say that none of this work would have been necessary and no costs would have been incurred had the OS not undertaken the PAI programme. This would be quite wrong as it conveniently ignores equivalent historic milestones that mark the development, evolution and quality improvement of the OS map from the paper County Series and first National Grid maps through to 'overhaul' and the early digital maps to LandLine and now MasterMap. The housekeeping task of realigning Geographic Information Assets was easier to avoid or defer when the media was paper. The digital age and the demands of business and GIS users for up-to-date information forces managers to review the currency of its GI Assets and it is that factor, perhaps more than PAI, that drives the change.

Colin R Smith
Telephone: 01752 635818 (Direct Dial)
Email: colin.smith@landreg.gsi.gov.uk
Web: www.landregistry.gov.uk

Information and Acknowledgements

The material contained in Section 7 (The Tenet Solution) was kindly provided by Paul Starkey, Technical Director of TENET ITMapping Ltd, North Heath Lane, Horsham, West Sussex RH12 5UX
Telephone: 01403 273173
Email: paul.starkey@tenet.co.uk
Web: www.tenet.co.uk

Rebecca Wadd, Account Manager, Ordnance Survey, Romsey Road, Southampton, SO16 4GU
Telephone: 023 8079 2079
Email: rwadd@ordsvy.gov.uk
Web: www.ordnancesurvey.co.uk

Tom Timms, Director, APIC, The Stables, Cannons Mill Lane, Bishop's Stortford, Hertfordshire, CM23 2BN
Telephone: 01279 466966
Email: tom.tims@apic-eads.co.uk
Web: www.apic-eads.com

Jamie Justham, Director, Dotted Eyes, Hanbury Court, Harris Business Park, Stoke Prior, Bromsgrove, B60 4DJ
Telephone: 01527 556920
Email: jamie.justham@dottedeyes.com
Web: www.dottedeyes.com