Smart Capabilities

Positional Accuracy Improvement

(PAI)

What is PAI and why is it needed?

Mapping, as a means of communicating information about things in the real world, has been around for thousands of years. Its value has always been determined by how effectively and unambiguously it achieved this communication. An early map showing in general terms where fertile land was to be found relative to a major river and nearby mountains was completely "fit for purpose" at that time.

However, in a world where differential GPS provides centimetre-level accuracy and every smartphone includes a GPS receiver, there is an expectation that reference maps and associated data are maintained at a much higher level of positional accuracy to be fit for purposes.

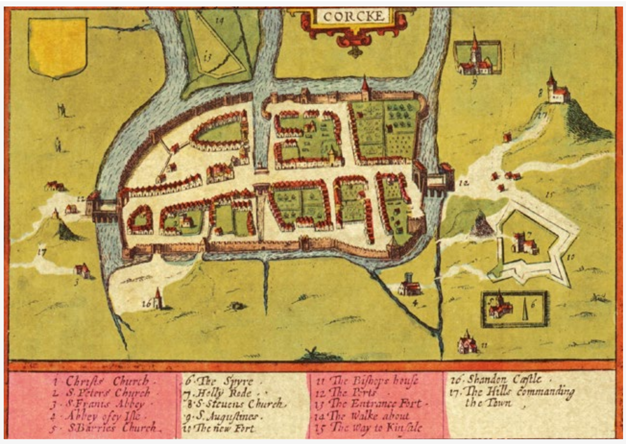


Figure 1:
Cork City 1617

Figure 2:
Cork City today

A map's accuracy has always been a function, both of the methods used to survey and capture the data and of the way in which this information is presented. Historically, as today, whenever an area was fully surveyed for the first time, the best available survey technology and practice were used to achieve the highest level of accuracy that was reasonably achievable.

However, when an area was re-surveyed, it was common practice to record only real-world changes within the context of the existing map information. This often involved adjusting the mapped representation of a feature that had been collected with higher absolute positional accuracy so that it is presented better relative to existing features originally collected with lower absolute positional accuracy. When everyone worked with paper maps, this process was perfectly acceptable: the context at any point in time was, typically, the extent of the map sheet being viewed. As a result, the relative accuracy of features on a map has, historically, been seen as more important than absolute accuracy.
That has now changed. The ubiquity of highly accurate GPS data sets the benchmark for absolute positional accuracy.

Coupled with the move from paper maps to a digital environment (where multiple geospatial datasets can be viewed and analysed together), the legacy of the traditional map maintenance approach has manifested itself in the variability of absolute positional accuracy across datasets.

As spatial datasets from various sources are integrated more frequently, it becomes essential that the positional accuracy of the data be fully understood, and improved, to avoid misinterpretation. Many reference maps were originally captured on paper at different scales. When these were digitised to create the first digital datasets, they effectively locked-in a positional accuracy deficit.

The result is that many reference base maps have needed positional improvement. Some have already been done, others are in the process. The far-reaching consequence is that any other datasets that were captured with reference to the original reference map now need to be updated to maintain their spatial relationship with the improved reference map. Examples include databases of asset locations (such as sewers, utility cables etc.) and records of land boundaries.

When a geospatial dataset is derived from a reference (base mapping) dataset – either by copying the geometries or digitising against the reference – a topological relationship, such as shared boundaries or connected endpoints, is created between the user's dataset and the reference dataset. If the reference dataset is then upgraded with improved geometry positions, the user's data needs to be re-synchronised with this improved reference to maintain the topological relationship. This is the process of **Positional Accuracy Improvement** or **PAI**.

Absolute or relative?

The term "Position" can refer to either absolute or relative position. Absolute position can relate to either an individual vertex on a feature geometry, or to a complete geometry relative to a given reference system or dataset. Relative position can be a single vertex within a geometry, or a complete geometry relative to other geospatial feature geometries in the vicinity.

Likewise, **"Positional Accuracy"** is considered from the same perspectives with **Relative Positional Accuracy** defined as "the difference of the distance between two defined points in a geospatial dataset and the true distance between these points within the overall reference system". **Absolute Positional Accuracy** is defined as "the distance between a defined point in a geospatial dataset and its true position in the overall reference system".

Ultimately, in a PAI process, the aim is to improve the absolute positional accuracy of the feature geometries in a dataset without compromising either the relative accuracy of vertices within a feature geometry (i.e. its shape) or the relationship between feature geometries and the reference data (i.e. the topology).

PAI is not a re-survey

We must always remember that a PAI is not the same as a re-survey. PAI takes geometry data that is known to have quantified errors in its absolute position. It transforms this data so that the absolute positional accuracy is improved, while maintaining the same feature identifiers and attribution. The most important consideration is that the complete process, and the data used to drive it, must provide a well-behaved, trusted and fully understood transformation of the dataset that will deliver enhanced value for the user.

It is also important to recognise that PAI is not always the best solution. If the source data errors are too random to effectively identify and eliminate outliers, the only viable solution may be to re-survey key datasets and then use automated conflation techniques to upgrade associated datasets.

This paper outlines a comprehensive, fully automated solution for Positional Accuracy Improvement (PAI) that can be applied to any customer data scenario.

How is PAI delivered?

Process and concepts

Not every element of the PAI solution will be used in every scenario, but by developing a comprehensive approach it is then possible to rapidly select those components that are necessary and relevant to a given situation. A typical PAI Solution is divided into three distinct stages, some of which have sub-stages that may apply in particular circumstances:

1

Managing Shift Vectors

- Generate or acquire Shift Vectors.
- Post-process and sanitise the Shift Vectors.
- Classify Shift Vectors including Zero Shift and Outliers.

2

Transforming the Data

- Snap Source Data to Original Reference.
- Shift Source Data using an intelligent triangulation approach.
- Constrain the shape or location of Immutable Geometries during Shift.
- Snap/Conflate Target Data to New Reference.

3

Validating and Reporting on Results

- Analyse geometry change (Source to Target) and report exceptions.
- Analyse connectivity and relationship to reference data and report significant differences.
- Analyse Target Data with New Reference Data and suggest automated improvement opportunities through conflation.
- Generate Outlier Influence Area geometries.

In this paper, we explore these stages in some detail and show how they link together to provide a complete solution. There are many technical solutions to PAI and they tend to share some key concepts:

- Shift Vectors (sometimes called Link Vectors) are captured or generated. These describe the difference in location between identical points in the original and improved reference data.
- Automated (or semi-automated) processes are used to apply the effect of the Shift Vectors to the user's datasets.
- The transformed datasets are validated to confirm the success of the process and to identify any anomalies requiring further attention.

When to start a PAI project

Any specific PAI adjustment is done once only and should be implemented in a manner that minimises organisational downtime and disruption.

There is never a "right time" to undertake a PAI project. The challenge for a data custodian is to implement it in a way that minimises risk, disruption and cost – and ensures that the resulting enhancement more than compensates for these.

A PAI process

The Flowchart in Figure 3 shows the key processes and high-level artefacts involved in any PAI process. It provides an overview of how the elements of the solution relate to each other.

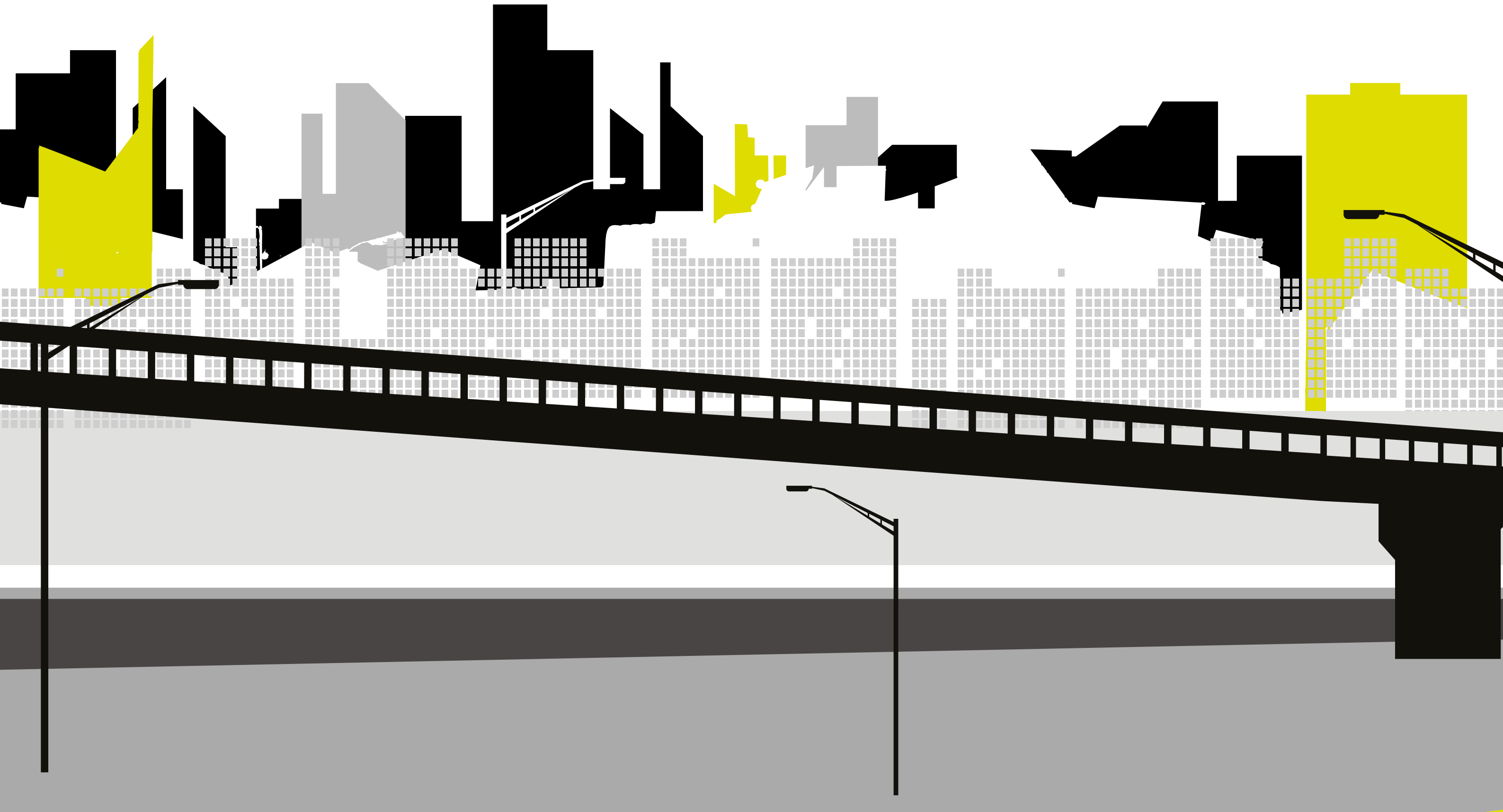
In the following sections, we explore the key concepts in more detail with a particular focus on how data flows through the process and how to manage it to get the best results.

Managing shift vectors

Shift (or Link) Vectors are the key input for a PAI. A collection of Shift Vectors distributed over the area of interest for a dataset is sometimes called a Shift Vector Field. By definition, Shift Vectors should provide a mapping from an identified point in the original reference to the same point in the improved reference.

Selection of Shift Vectors

There are a number of characteristics of good shift vectors that are helpful to consider when assessing their suitability for a PAI process. Shift Vectors may be provided by a third party or they may be digitised or automatically derived from reference data. Whatever the source, it is important to ensure that the Shift Vectors are fit-for-purpose before they are used to transform feature data. The following notes on Shift Vector selection provide guidance on a number of key considerations in determining if a Shift Vector dataset is appropriate to use and the filtering required to eliminate Outliers that would distort the process. Where the quality of the provided Shift Vectors is such that it cannot be refined to address these issues it may be necessary to consider re-survey or conflation to transform the feature data.



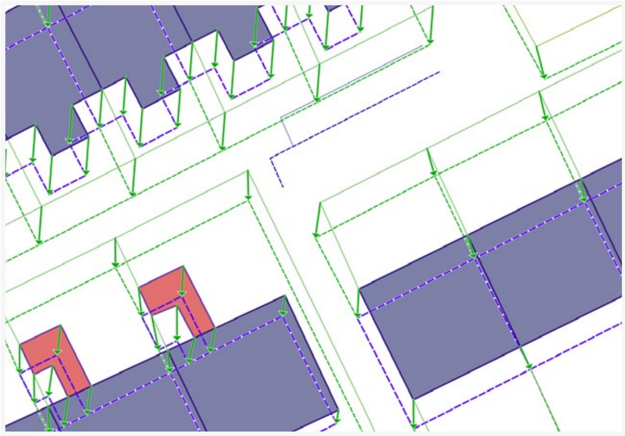


Figure 4:
Well behaved Shift Vectors

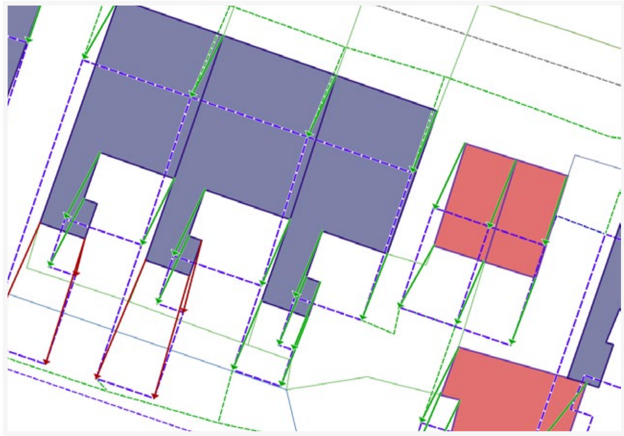


Figure 5:
Shift Vectors representing Real-World Change to be excluded

Useful Shift Vectors will typically vary in density across an area of interest, but would be expected to vary slowly, if at all, in terms of magnitude and direction over their extent. They represent a correction for a systematic error and should not exhibit any significant random behaviour. Where random variations are noted in the Shift Vector Field, the source of variation must be identified so that the rogue shift vectors can be removed from the process. Shift Vectors located at points where features meet are the most reliable as their locations are least ambiguous. The next most valuable are those located at readily identifiable points such as fence or building corners.

In Figure 4, the Shift Vectors shown in green with the source dataset are well behaved with only moderate variation across the area of interest. Also, all the points originate at either topology nodes or geometry corner-points. Shift Vectors should not be associated with real-world change. Only points that represent the same real-world point in both the original and improved data should be joined by a Shift Vector. A common error is to correctly link the front corners of a building in both versions of the reference map, but to then link the original back corner of the building to the corner of an extension that has been added in the newer, improved map.

In Figure 5, the green Shift Vectors are correct whereas the red ones relate to real-world change and should not be used. Any Shift Vectors associated with real-world change should be removed as they are not related to the underlying systematic error that is being corrected and will corrupt the transformation process.

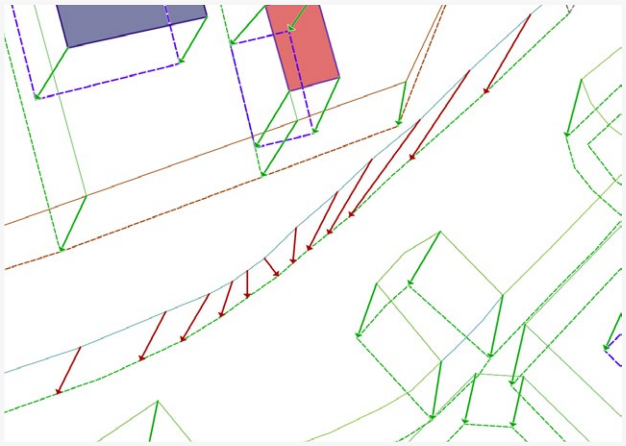


Figure 6:
Ill-defined Shift Vectors (Red)

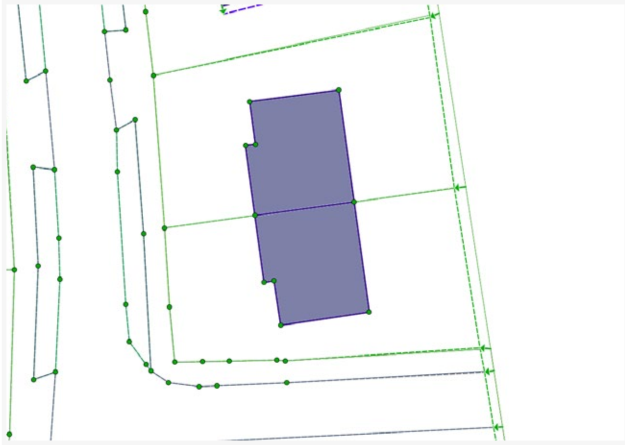


Figure 7:
Zero Shift Vectors (Green Dot)

Shift Vectors that are defined with reference to points that are difficult to identify on features in the real world should be avoided. A corner of a building or fence that is unambiguous and unchanged in both the Original and New reference data is ideal for defining a Shift Vector. A point that is part way along a curved hedge or verge is not a good candidate because the source and destination are not well defined. In the best-case scenario where Shift Vectors are captured along such features, they are simply replicating what has been captured at better control points in the vicinity and are not adding any useful information to the Shift Vector Field. However, in cases where they are captured at random (as shown in Figure 6) they can introduce significant, localised corruption to the Shift Vector Field.

Figure 6 shows appropriate Shift Vectors in green, based on identifiable feature points, and Shift Vectors in red that are not appropriate because they are based on random points along a verge.

Zero Shifts are Shift Vectors of zero length (as shown in Figure 7), as the start and end are at the same location. It is possible for a systematic error to be zero if it occurs between areas of opposing shifts, but a Zero Shift is different. They are usually the result of updates in the Original reference map being applied with the correct absolute positional information where the relative positional accuracy was not an issue. An example where this might be found is when a new housing development has been added in an area that previously had no other infrastructure. Correct use of Zero Shifts in a PAI is essential to ‘lock in’ positions that are already correct and thereby avoid the risk of downgrading the absolute positional accuracy of geometries in the affected area.

Whether Shift Vectors are captured manually or generated using an automated process, they should reflect the guidance here if they are to successfully transform user’s datasets from the Original to the New map reference. In many cases where the Shift Vectors are supplied by others they will have been captured or generated in a manner that is not consistent with this guidance and will require post-processing to classify each vector based on its origin and to identify Zero Shift vectors and Outliers in the data.

Filtering Shift Vectors

Whether Shift Vectors are digitised individually using a manual data entry process, or generated automatically by comparing the Original and New reference map data, they must be processed to ensure that Non-Topographic Shift Vectors and Outliers are excluded. Figure 8 shows an example of Original reference data together with the manually captured Shift Vectors in an area of interest.

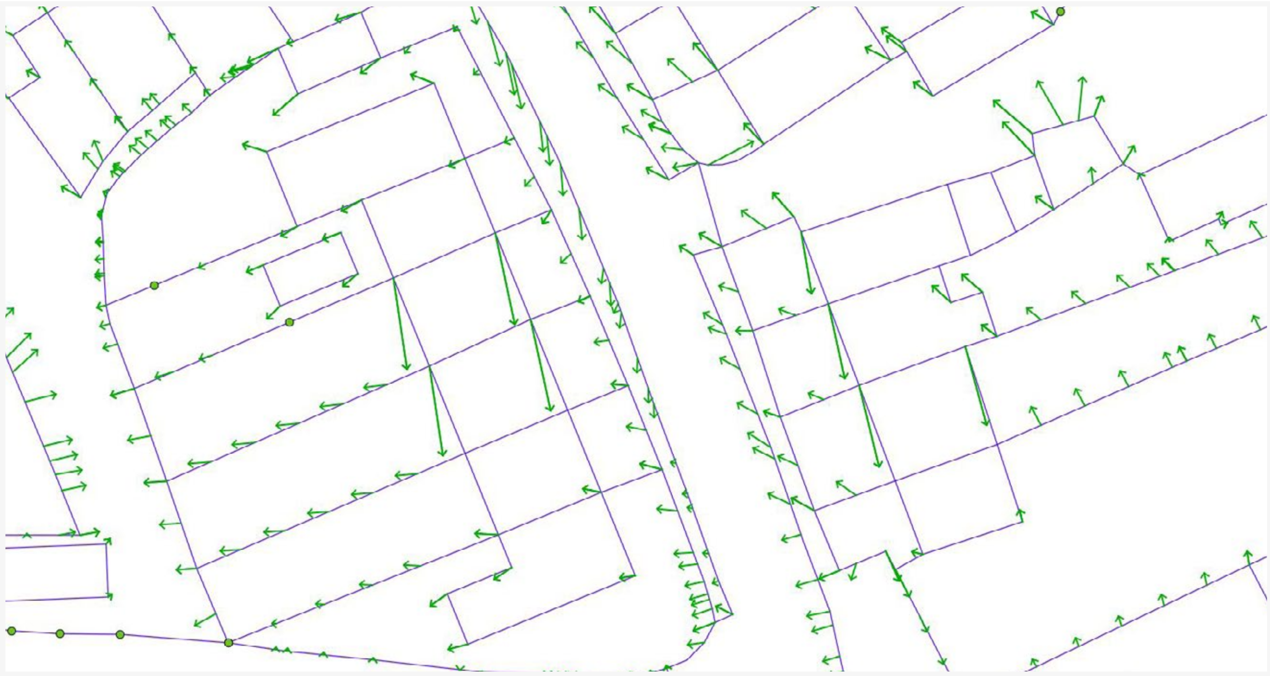


Figure 8:
Original Shift Vectors

Non-Topographic Shift Vectors are those that should not have been captured in the first place as they have no relationship with real-world features.

- The primary source of invalid Shift Vectors is road and water centrelines. These are not real-world artefacts and are not points that can be unambiguously identified in both the Original and New reference map data. They should not be used in any process.

Outliers are those Shift Vectors that are not consistent with a well behaved, smoothly varying, Shift Vector Field. They are typically the result of one or more of the following:

- Capture of Shift Vectors that represent a combination of systematic error and real-world change (e.g. a new extension on a building that didn't exist in the original reference map).
- Capture of Shift Vectors at locations that are not identifiable points in the real-world (e.g. random points along a curving hedge or verge).
- Capture of random error as well as systematic error in Shift Vectors.

Where the Original reference data is available, the Shift Vectors should be analysed to determine where each one originates in that reference data and to classify them accordingly. Any that don't originate on a real-world feature should be flagged as Non-Topographic. Those that do originate on real-world features should be automatically classified based on whether the origin is at a topological node, a readily identifiable corner point, or a potentially ambiguous point in the reference data.

Figure 9 shows the classified Shift Vectors for the example area. If the number of Shift Vectors identified as originating at topological or corner points is adequate then the Shift Vectors originating from the potentially ambiguous source points in the Original reference data can be excluded from the process.

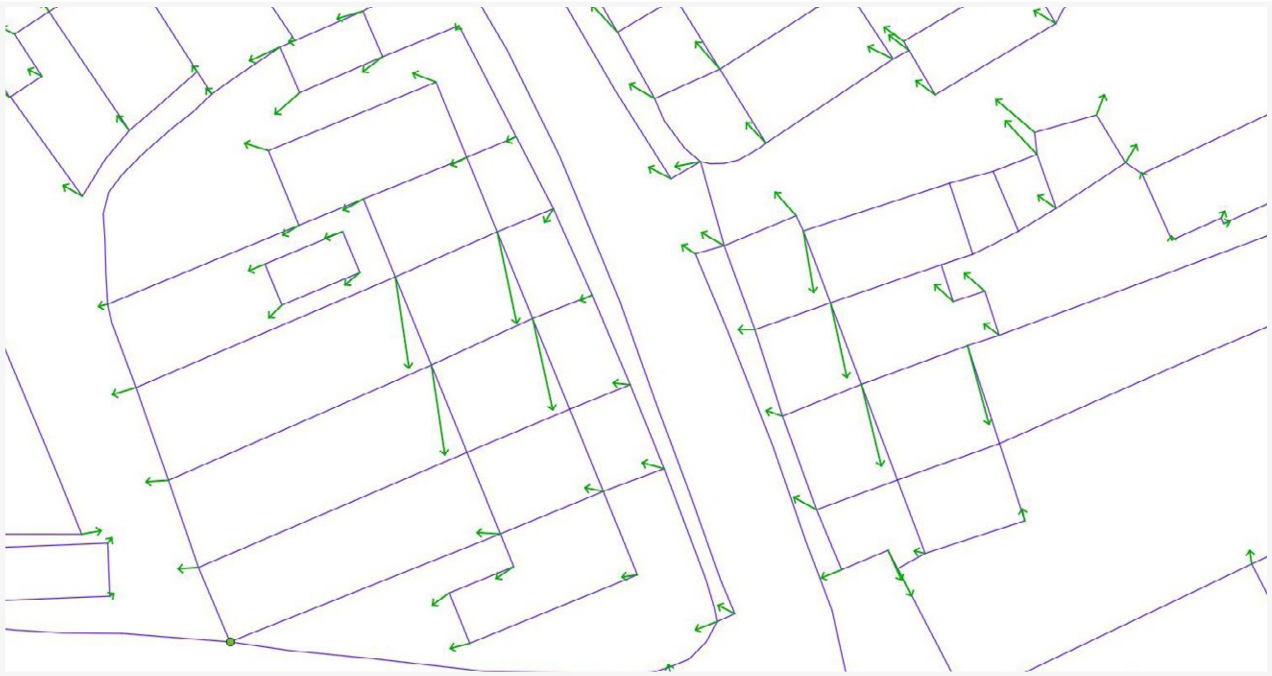


Figure 9:
Shift Vectors filtered by origin

The inclusion of Outliers is a significant potential source of distortion in a PAI process. These must be identified and excluded from the correction of the systematic error. However, they must be retained outside the process so that they can be used to identify areas where real-world change may have occurred. This will support an orderly review and any necessary, manual update of corrected geometry in these areas.

Outliers are identified by computing the vector difference between individual Shift Vectors and the average Shift Vector profile in an area of interest. The vectors are filtered to identify all those where the magnitude of difference is greater than a specified value. The choice of filter threshold will always be a function of density of Shift Vectors and the rate of change across the domain. The Shift Vectors shown in Red in Figure 10 have been automatically identified as Outliers in our example.

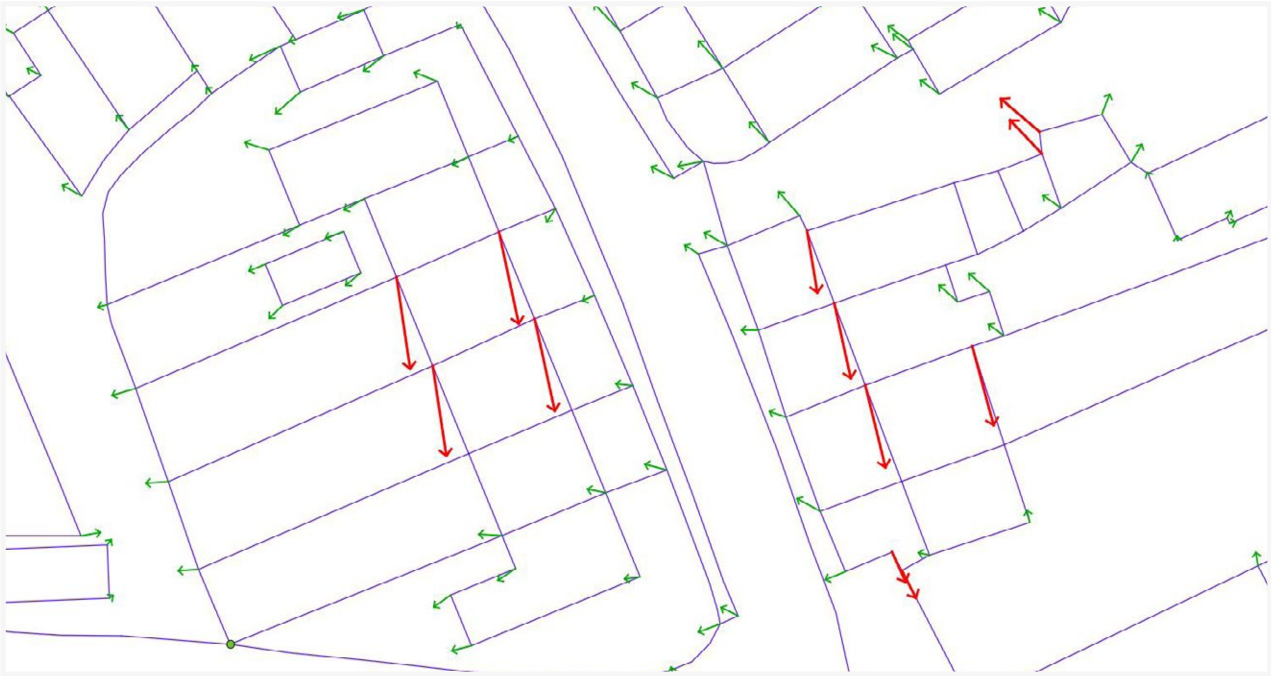


Figure 10:
Shift Vectors with Outliers isolated

Deriving Shift Vectors

In situations where Shift Vectors have not been provided, but where the Original and New reference map vector datasets are both available, it is possible to derive the Shift Vectors from the reference data, automatically.

An effective process is based on the analysis of reference datasets to identify “before” and “after” versions of features that belong to specific feature classes where the real-world locations are readily identifiable. The key steps in the process are:

- **Create a sparse, preliminary Shift Vector field** over the area of interest using identifiable points. This can be done using updated field survey data that records the true position of key points. Identification of areas where Zero Shift applies is important in this step.

- **Use this preliminary Shift Vector field to move complete polygons** (or topology faces) from appropriate feature classes in the Original reference data so that they can be paired with matching features in the New reference data by using fuzzy geometric matching processes.

- **Analyse the paired versions of geometry** to record real Shift Vectors for identifiable vertices on those polygons, ensuring that the process deals with slight differences in shape but ignores significant differences due to real-world change.

- **Classify the derived Shift Vectors** based on the nature of the source point in the Original reference data.

Implemented properly, this process will only generate valid Shift Vectors that represent systematic error correction and should not include any non-topographic or Outlier Shift Vectors.

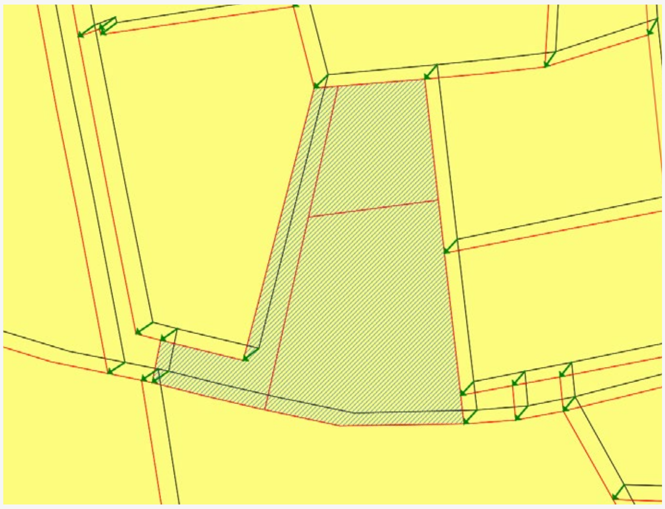
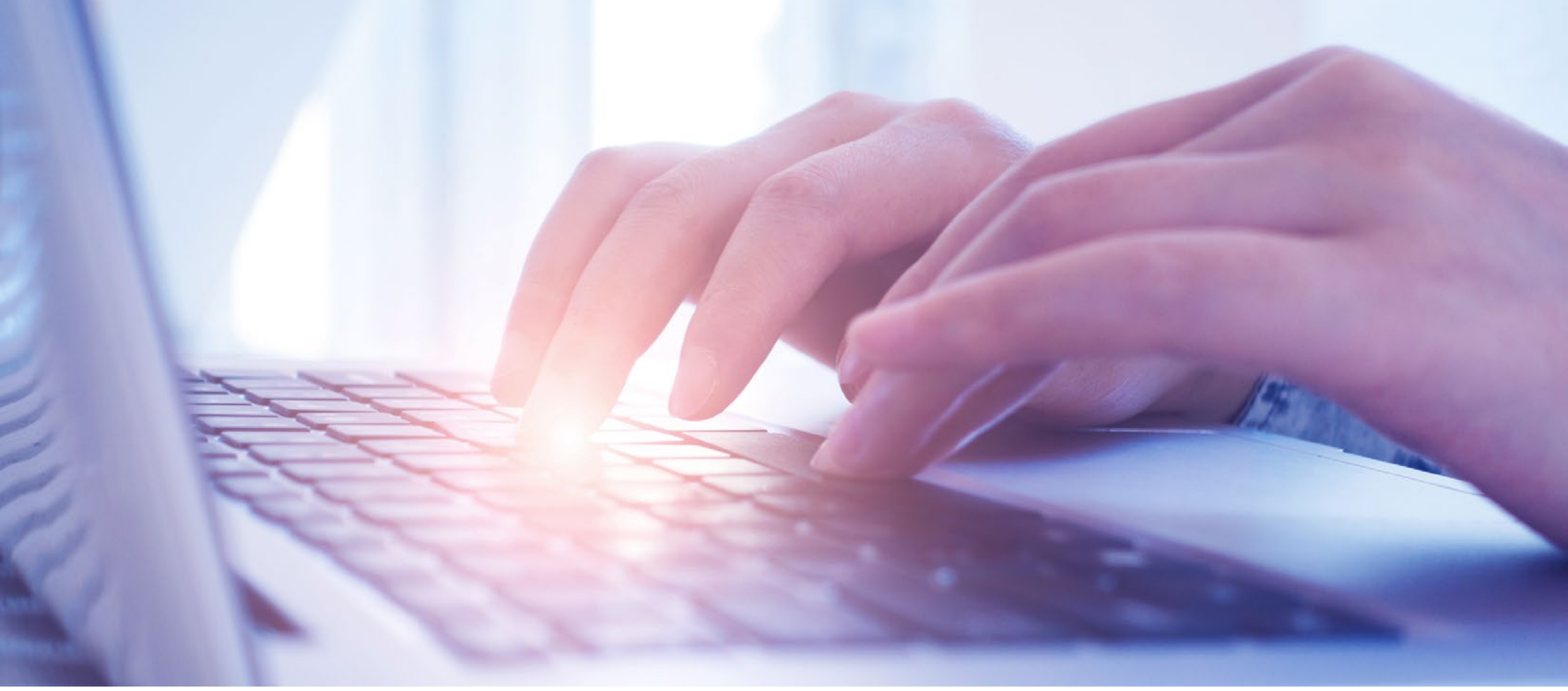


Figure 11:
Derived Shift Vectors

Figure 11 shows an example of original reference data, outlined in black, overlaid by new reference data (outlined in red) with new subdivisions highlighted. The automated process of deriving Shift Vectors has correctly identified the relationship between key points on the original and new data while ignoring real-world changes recorded in the new data as well as potentially ambiguous locations along features.



Transforming the datasets

Once a complete set of Shift Vectors has been classified, and any Outliers have been isolated, they can be used to transform the datasets to be migrated from the Original reference map to the New reference map.

This part of the process can also include steps that will improve the alignment of the dataset to the reference data so that the transformed data is not just shifted but is more accurately aligned with the reference data.

The key steps in the process involve:

- **Optional snapping of the dataset to be transformed to the Original reference data** if this is available. If the snapping tolerance chosen is small enough, the modification to the dataset geometry will be: (a) an improvement in the position of geometry vertices so that edges are aligned with reference data features against which they would have been originally digitised, and (b) the introduction of new vertices on the dataset geometry where these were missed during the original digitisation. The effect of this step is to bring the source version of the data to the best initial state for accurate transformation.

- **Shifting the dataset to remove the systematic errors using filtered Shift Vectors.** In this step, a Triangulated Irregular Network (TIN) is created for both the X and Y elements of the Shift Vectors. The use of TINs means that the shift to be applied to any geometry vertex is linearly interpolated from the Shift Vectors in its vicinity. This ensures that the adjustment to shared vertices on adjacent features is identical and that topological relationships are fully preserved throughout the dataset. A key requirement when transforming a dataset using TINs is the handling of immutable shapes such as buildings. These are special cases where, for example, the source data would have included many fixed right angles which need to be preserved following transformation.

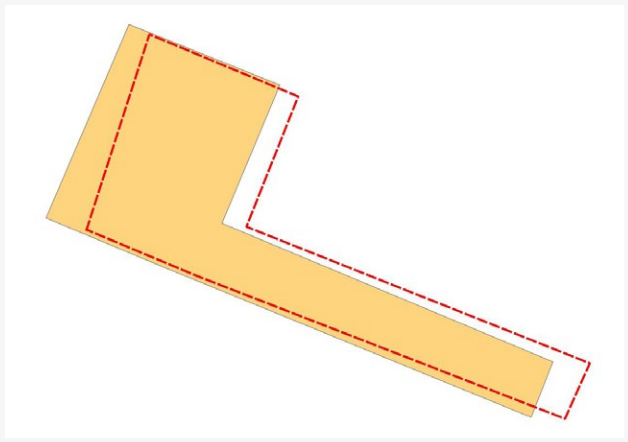


Figure 12:
Provisional transformation of immutable object

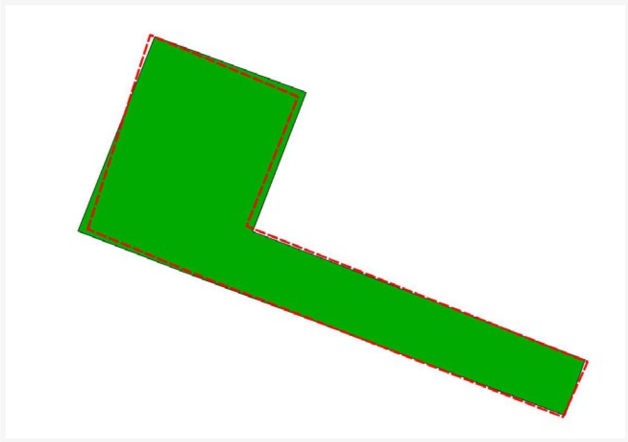


Figure 13:
Final transformation of immutable object

In Figure 12 we have a building in the user's data that, if transformed on a vertex by vertex basis using the Shift Vectors, would produce a distorted result. Using advanced rules logic, the building can be transformed so that the shape is not distorted but is translated and rotated. The final result minimises local error, as shown in Figure 13, while keeping all the surrounding data connected so as not to introduce gaps or overlaps.

- **Optional snapping of the shifted dataset to the New reference data,** if this is available. If every vertex in the Original reference data has a one-to-one mapping with a corresponding vertex in the new reference data, and if a valid Shift Vector was available for each of these vertices, then the transformed dataset will have the same level of alignment with the new reference data as the source dataset had with the Original. In reality, this is unlikely to be the case and very small differences between the locations of vertices on the transformed dataset and adjacent new reference geometry will exist where Shift Vectors were not specified. For this reason, an optional post-process can be applied to the transformed dataset to snap it to the new reference data using a small tolerance.

As described earlier, the identification and isolation of Outliers in provided or digitised Shift Vectors is crucial if an acceptable transformation result is to be achieved. This can be demonstrated by looking at how mapped features are transformed using our sample Shift Vectors.

Figure 14:
Source Data with Original Reference Map

Figure 15:
New Reference Map with all Shift Vectors

Figure 14 shows feature data to be transformed aligned with the Original reference map. Figure 15 shows this data overlaid with the New reference map and all the Shift Vectors that were digitised. These were then filtered to identify the Outliers (in red) that are the result of errors in the manual digitising of the Shift Vectors.

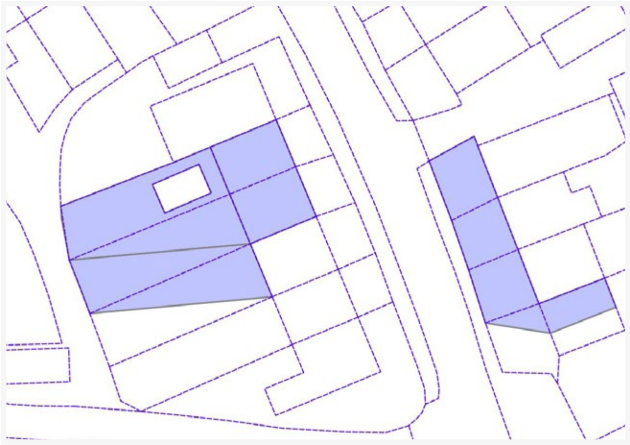


Figure 16:
Results using Filtered Shift Vectors with Outliers

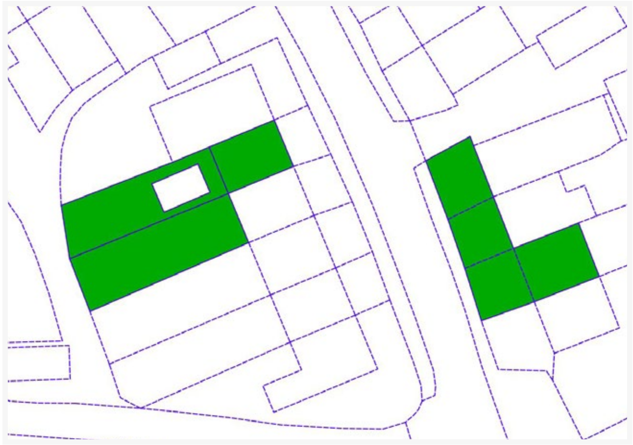


Figure 17:
Result with Snapping and excluding Outliers

Figure 16 shows how the data to be transformed would be corrupted by using all of the supplied Shift Vectors. Figure 17 shows how the result is significantly improved when the Outliers are excluded and when the Snapping of the results to the New reference data is included in the process.

Validating and reporting on results

Data owners need to have confidence in the transformation process and trust that it has correctly transformed the data to remove systematic errors.

They also need to be alerted to the existence of any anomalies in the data so that they can deal with these issues in an efficient and controlled manner. Knowing that a process has automatically fully transformed 99% of your data is of limited value if it cannot also identify the 1% to be reviewed and revised.

The final stage in the automated process is validation and reporting on the results. This stage generates materials which are of particular use to the data owner in reviewing and possibly revising transformed data in areas where anomalies have been detected. The purpose of this stage is to provide full confidence in the automated process and to provide accurate guidance to facilitate efficient review and revision, where required.

When the transformation has been completed, the dataset will have been transformed using Shift Vectors that represent the underlying systematic error to be removed. The dataset geometry may also have been snapped to the New reference data.

However, there are likely to be a number of anomalies in the transformed data that will require review and possible revision by the data owner. These are not situations where the transformation has produced an incorrect result, but are situations where either the provided Shift Vectors or the New reference data included real-world changes that would have distorted the dataset during transformation if they had not been isolated in the Processing Shift Vectors operations.

The transformation process is completed by analysing the data and by generating validation reports and anomaly identification datasets, ordered by significance, which can then be used to deal with what should be a small number of items requiring human intervention.

The reports include the following:

Geometry size

Polygon geometries in the source dataset are compared to their transformed versions and any cases where the area has changed by more than a specified percentage are listed in a tabular report and included in a spatial dataset to aid review. A similar process is applied to Line/Polyline geometries.

Angle changes

Within a geometry, changes in internal angles larger than a certain size can be important, particularly for buildings. Typically, the tolerances will vary depending on the type of data: a building would have strict tolerance for maintaining the angles, but a river would have looser tolerances.

Connectivity changes

The connectivity between a feature and its surroundings, and between a feature and its underlying reference data, is also important. Establishing that all pipes are still connected as before, or that a parcel overlaps the same set of topography polygons - and in the same proportions – adds greater confidence in the result and proves that the nature and quality of the resulting data has not been negatively affected by the shift. This report allows users to inspect any features which have failed this test.

Outlier report

Each Outlier Shift Vector that was identified and excluded from the transformation process is potentially a marker for a location where real-world changes have been recorded in the new reference data. This information is of particular value to the data owner as it can be used to identify transformed geometries that require review and possible revision. It can also indicate the specific area where that revision is likely to be required.

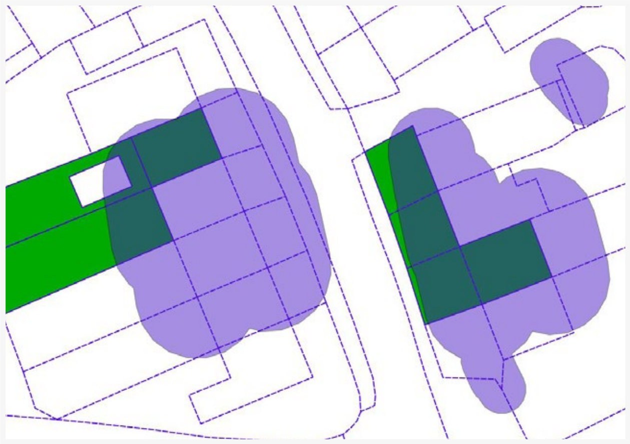


Figure 18:
Outlier Report showing impact of error in Shift Vector digitising

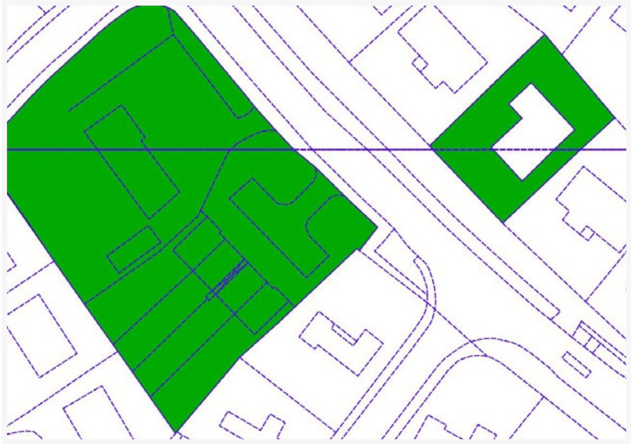


Figure 19:
Transformed data with no Outliers

Typically, the Outlier Shift Vectors are buffered and the resulting polygons are unioned with any other Outlier polygons that they intersect. These derived geometries provide a spatial dataset of Outlier Markers to indicate the potential area of interest for review of real-world changes. This can then be used to rapidly navigate to areas of interest where the feature data can be reviewed and modified as required.

Figure 18 shows the new reference data together with the transformed data where the effect of the Outliers has been excluded and overlaid by Outlier. In this case the Outliers have been identified as being a result of an error in the Shift Vector digitising and no further modification is required on the transformed data. Figure 19 is an example of an area that is not affected by Outliers and the transformation is complete.

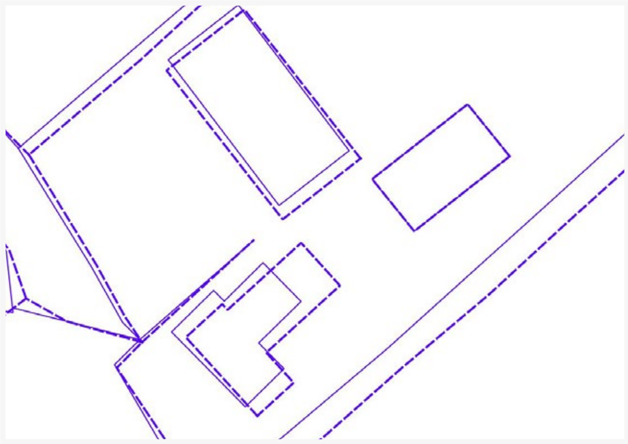


Figure 20:
Shift Vectors, Zero Shift and Real-World Change

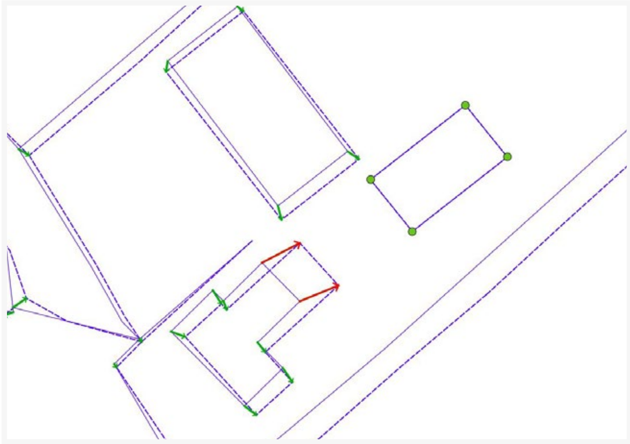


Figure 21:
Processed Shift Vectors

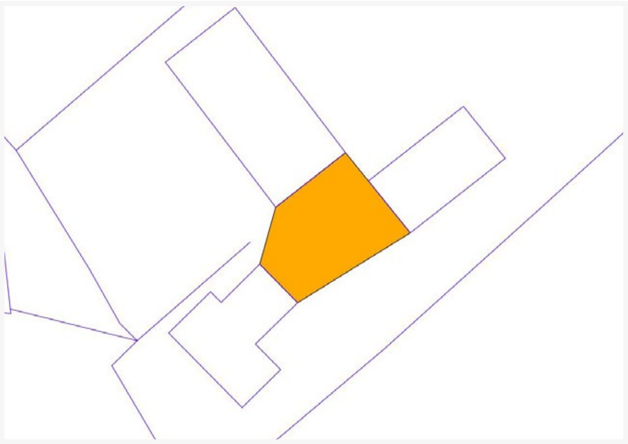


Figure 22:
Source data to be transformed

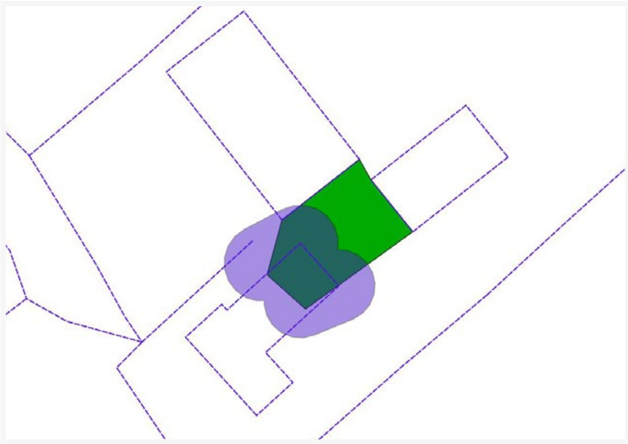


Figure 23:
Transformed data with Outlier Marker for Real-World Change

This example demonstrates how a Real-World Change is identified and reported. Figure 20 shows the Original (solid line) and New (dashed line) reference data including shifts due to the following:

- Removal of systematic errors.
- Zero shift where a building has been added to the Original reference data in the correct position.
- Some real-world change where a building has been extended since the Original data was recorded and before the New data was captured.

In Figure 21, we can see how the Shift Vectors have been classified and filtered showing normal Shift Vectors as green arrows, Zero Shift Vectors as green points, and detected Outliers as red arrows. Figure 22 shows the Original reference data and some user data possibly representing a paved area between the buildings. In Figure 23, the transformed user data is shown overlaid on the New reference data with an automatically generated Outlier Marker. The marker indicates the area where real-world change was detected and where manual intervention is required to investigate how the user data is impacted by the real-world change and to make any necessary changes to the transformed user data.

Conclusion

PAI can be an extremely powerful and effective tool for transforming data that was originally captured against a superseded reference map and now needs to be upgraded to properly align with a new reference map.

However, many attempts at using this approach end in failure because of a combination of using the wrong inputs, not having an intelligent, automated process, and not having an adequate understanding of the results.

The approach described here deals with each of these to provide a trusted and reliable solution that can be successfully implemented in an extremely wide range of circumstances.

The key elements of this approach are:

- Pre-processing of captured or generated shift vectors to ensure that they are appropriate and to identify outliers that may be a result of digitising errors or of captured real-world change.

- Use of TIN models to apply interpolated transformation of objects on a vertex by vertex basis across the domain and use of specialist rules to preserve the shape of immutable objects.

- Analysis of results and generation of tabular and geometric reports that allow users to review the transformed data and resolve any detected anomalies in an orderly and controlled manner.

Done correctly, and maximising the use of automation, implementing a positional improvement on operational data doesn’t require significant production down-time or disruption. Normal operations can continue whilst a snapshot of the data is pre-processed and results are analysed. This cycle can then be repeated as often as required until the data owner is happy that the automated processes have been fully tuned to deliver the desired result. It is then simply a case of re-running the process with the latest version of the data and returning the improved data to production with minimum impact on operations.

When applied correctly, an automated PAI can deliver real data improvement cost effectively and efficiently with minimum disruption to operations.

Contact us to learn more:

info@1spatial.com
www.1spatial.com