

Data Processing Tools for OmniTRANS

The quality of a Transport Model depends on the standard of its data. To help ensure this there is merit in processing the 'raw data', where possible, within the framework of the model for which it is to be used.

Transport modelling data is very expensive to collect, equally expensive to process, yet is often validated, manipulated and transformed in software packages which do not understand the inherent relationships within transport data and are so processed out of context.

A good example is that of Roadside Interview (RSI) Survey data. This is collected in vast quantities in often difficult conditions and undergoes much processing before it can be used to build trip matrices for the transport model. There is scope for error entering the data set at each of these stages, yet ultimately this data is used to provide the base upon which models are calibrated. It goes without saying that if this base data is of poor quality then the task of model building will indeed be difficult.

OmniTRANS provides an excellent environment for not only building a Transport Model, but also for processing many of the required data inputs.

To facilitate this, Minnerva has developed a set of tools to undertake some of the more common data processing tasks in an efficient and productive manner.

Postcode – Grid Reference – Zone

Coding: Converting interview survey data to usable records for matrix building, whether it is from Roadside, Public Transport or Household Interviews is a task that has long been with us.

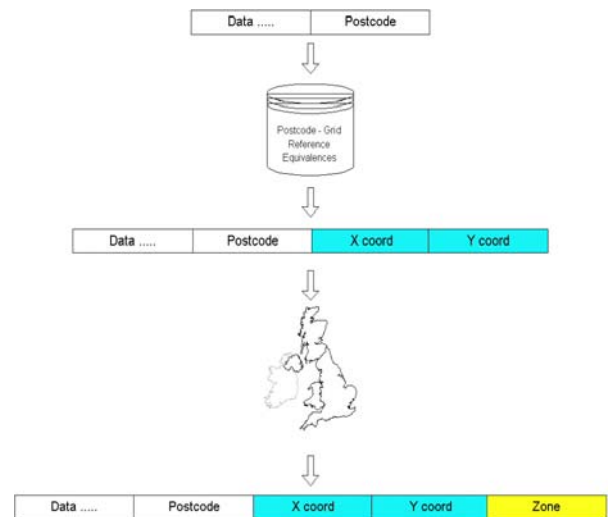
Interview records typically contain one or more addresses which must be allocated to the traffic zones used in the transport model. Given the address postcodes, the corresponding grid references are determined. They are then used to identify

which zone they fall in using a 'Point-in-Polygon' process. Using the toolkit, this task is approached as follows: Firstly, an OmniTRANS project is established which, at a minimum, must contain the boundary definitions for the zones used in the model. Typically these are generated in a GIS and imported easily into the OmniTRANS project.

Then, a proprietary data set is required containing the postcode – grid reference equivalences for the study area. Typically in the UK this would be either the Ordnance Survey's AddressPoint® or the Royal Mail's Postcode Address File® (PAF). *Note: Non UK postcode structures can be supported on demand.*

These files can be very large, up to half a million records for an 'average' sized UK county, so an initial task is to 'pre-process' them for efficient downstream interrogation when the survey data is processed. This is done using an OmniTRANS 'class', provided as part of this toolkit, and the output is a set of files which comprise the master 'gazetteer' used for record matching.

These are supplemented by two additional gazetteers. The first is for valid postal addresses that fall outside the study area where it is sufficient to append grid references at a coarser level of detail; the second deals with non-postal or 'generic' addresses that commonly occur within the interview records; for example: the Town Centre, the bus stop outside the Post Office, and so on.



Once these gazetteers have been prepared, the survey data records can be processed. They are matched against the postcodes in these files and if a match is found the grid reference is appended.

For a variety of reasons not all postcodes will be matched perfectly, the most problematical being the 'partial' postcode.

A method of dealing with these is provided using the observed distribution of successfully matched full postcodes. A feature of this process is that different distributions can be constructed and used with different data sets; for example if Roadside Interview data records are split by mode (car and goods vehicles) then different observed 'successful' distributions can be constructed and used to distribute the partial postcode records associated with each of these modes.

The output of this process is a dataset of records with appended grid references and 'quality' flags which indicate the nature of the match that has been made. They are now ready for zone coding.

Zone Coding: This is done using a 'Point-in-Polygon' tool provided as part of the toolkit. It takes the zone boundary

definitions in the OmniTRANS project (the polygons) and processes each of the grid references (the points) found in the survey records. The output is an extended survey record which contains the appended zone(s) for each point.

The method is intelligent in that once a grid reference has been 'zoned' it is added to a table of matched 'grid references-to-zones'. This table is searched before a grid reference is subjected to the Point-in-Polygon process making it faster to extract data.

The resulting set of data records is then ready for matrix building which the user does by means of the standard classes provided as part of the OmniTRANS Data Manager.

RSI Data Validation: It is prudent when building RSI matrices, by survey site, to subject them to 'reality' checks so as to be confident that the trips recorded going through each site could indeed sensibly have been observed there.

Minnerva has developed a methodology and associated 'Plug-in' to undertake this. The process is premised on the notion that a 'select link analysis' for a survey site will identify the set of Origin-Destination pairs for which trips would 'reasonably' be expected to be observed. The issue then relates to other Origin-Destination pairs which are detected going through the survey site; are they reasonable or in error?

The Plug-in is used to mark those O-D pairs which are not part of the select link O-D set, but local knowledge deems to be

'reasonable' – maybe because of network routing issues. By combining these two definitions of 'reasonable', trips between O-D pairs which *do not* fall into this category can be identified as being in error and separated out for further investigation.

Other classes are available related to survey data processing; these include:

- on-board bus survey expansion factor processing
- RSI expansion factor processing
- serial number matching for postal surveys to control sheets

Once survey data records have been validated and passed as 'fit for purpose', a class is available to build multidimensional trip matrices quickly and easily.

By encapsulating the various data processing steps within the OmniTRANS 'job' structure, processing is mechanised, repeatable and auditable. This makes it less prone to error than when using other tools.

Network Manipulation Tools: A useful starting point for developing an OmniTRANS network is to import a digital network from a proprietary source. An example is the Ordnance Survey's Integrated Transport Network.

For the network to be effective in modelling terms it is necessary to add Zone Centroids (nominal loading points for trips into and out of a zone) and then Zone Centroid Connectors (dummy links which connect the centroids to the network).

The toolkit provides two classes to do this. The first generates the centroid by locating

it in the 'centre of gravity' of the zone. This requires the zone boundary data referred to previously.

The second class is used to generate the centroid connectors in an intelligent way; to avoid linking in to one-way systems, junctions or links belonging to a user defined category (i.e. motorway).

Used in collaboration, these two classes provide a very efficient way of adding these attributes to a network.

Data Import/Export Tools: There are other types of object which can be used in an OmniTRANS model for which classes are available to handle both import and export.

Although editing tools are provided as part of the OmniTRANS' graphical user interface, if an OmniTRANS project is being developed by more than one person, possibly at different sites, the need often arises to transfer objects from one project to another, or to 'batch' import data from external sources. Examples include:

- importing count data from count databases. This can be per link direction, time period, and vehicle class, as well as other attributes which define the count site
- defining screenlines using these count sites
- transferring this data to another copy of the project by export and re-import
- importing public transport stop data from external sources

Model Analysis Tools: To help in the analysis and understanding of a model's outputs, classes have been developed to simplify the:

- reporting of assignments: total travel time and loads by link type for highway assignment or by mode for public transport assignment.
- plotting variables such as link counts, link flows, zonal data, trip end data and matrix row/column totals, together with a regression analysis. If doing a link based plot, the standard errors per link can be saved as a 'load' so they can be plotted as bandwidths.

