

# The wider use of data

**W**hen thinking about ICT, most local authorities tend to concentrate on how it can be used to improve their services and communication with the public and internally; but there is another aspect that can potentially lead to substantial benefits. This is the use of data within operational systems as part of the evidence base for monitoring and making strategy, policy, and decisions on practical action and intervention.

Local authorities, amongst other large organisations, have been described as 'data rich, information poor' – meaning in part that their operational databases are a rich source of potentially useful data about many aspects of their area of operation, but this is seldom turned into useful information.

## Database intelligence

Operational databases such as Council Tax, Housing Benefits, Electoral Roll, Planning, Building Control, Social Services delivery, Supporting People, Transport Pass administration, School admissions – indeed, potentially all service areas – routinely hold many records on individual transactions. These are typically incidents or episodes of relevance for some function of the local authority, like collecting Council Tax, paying benefits, getting a bus pass and so on. The data collected often covers very high proportions of a population in an area. Council Tax covers probably 95%+ of households, for example, and is very much cheaper and quicker to extract than a sample survey that only covers perhaps 5-10%. Most pensioners get a bus pass; most children go through a school admission process.

These records, in anonymised and aggregated form, can then potentially provide intelligence on processes, patterns and trends in that area, both to help monitor the effects of policies or environmental factors, and to add to the evidence base for strategy and policy development. So, for example, a Council Tax account is ended when a household moves out; or when a school admission is made, the characteristics and origins of that student, and data on the school, is all recorded.

The operational service delivery function must come first; it is extremely important that this is not slowed

or distorted by the information collection function, or staff resistance and cynicism will result. A new IT system is not usually needed to enable this to happen, although modern systems do have better reporting, and systems could be designed with this wider information function in mind. Fields could be added or structured so that cleaner and more useful data is collected – again, as long as this does not significantly adversely affect the ergonomics of the process.

---

**'Software applications such as statistical packages, spreadsheets, database systems, Geographical Information Systems (GIS) and visualisation systems can all be used to help make sense of the data.'**

---

## Extracting, cleaning and sorting data

Data can usually be extracted by running a query and exporting the results into a usable tabular format, such as a spreadsheet, comma separated variable or text file. It must be in electronic format; hard copy is useless for this purpose unless for very few records.

The data extracted is usually dirty and distorted, as various operational actions are carried out as required, and again staff and systems should not be constrained or controlled to try to meet information needs if this obstructs operational processes in any way. So the data inevitably needs cleaning, de-duplicating and sorting before it is of any use. This is usually possible by using some of the handy functions in spreadsheets like Excel or databases like Access, such as: edit/replace; extract left, right and mid; pivot tables; cross tabs; and filtering.

Operational staff can also help and advise on this cleaning process, and tracking a few specific records through to see what has happened can be very useful to show what is going on and how the data may be affected. Even so, sometimes the data may be so dirty that it can only be indicative of possible patterns, and needs extensive health warnings and qualification. If the global

**Bob Line,  
Freelance  
Consultant,  
B.Line Housing  
Information,  
highlights data  
mining and  
predictive  
analysis in local  
government...**

totals ring true with operational staff, it is a great comfort that the data is significant.

This is more than just management reporting on the efficiency and outputs of the operational processes, but takes into account the wider possibilities for use of the data.

In this approach, the collection of data is not an addition to service delivery work processes, resulting in extra work – usually 'form filling' – for front line staff, but is carried out using the existing system by extraction or 'mining' of the data.

### Interpretation and analysis

The transformation of raw data into information is a process of a mixture of visualisation, interpretation and aggregation, in which local knowledge and experience is essential. It is best carried out by a combination of data mining specialists – front line staff who know what the data entered into fields really means. Quirks and idiosyncrasies are common – like the ongoing accounts that all have an end date of 2247, or the 'exors' tag for properties in probate.

These processes produce large datasets, and IT methods of handling and analysing them are essential. Many methods are available, and indeed some universities now run courses in data mining techniques. Software applications such as statistical packages, spreadsheets, database systems, Geographical Information Systems (GIS) and visualisation systems can all be used to help make sense of the data. It is best not to be tied to one particular application, but to use the best tool for any particular job with the data, moving it between applications as required.

The queries developed to extract the data can be saved, or shared, between authorities with similar systems, and re-run whenever required so that a series builds up. Trends over time can then be identified, which can, in turn, be much more revealing and powerful than single snapshot surveys.

### Geographical analysis

Some 80% of data held in local authority databases has a spatial location tag of some sort – often a post-code. This makes geocoding and visualisation in GIS possible to give insights into the data. This generally brings no great surprises, but shows patterns that staff, residents and members recognise – but now underpinned by quantified data. But sometimes surprises do arise, giving new insights.

### Visualisation of linkages

Another software application that can be useful is Java Mondrian (see <http://stats.math.uni-augsburg.de/>

Mondrian/). This allows visualisation of the links between different variables, so that concentrations of factors can be seen. Sometimes it is clear that there are no linkages and correlations – which itself is significant and useful, showing that deprivation and problems are dispersed and random, and perhaps therefore should not be tackled by area-based initiatives, but by broadly-based programmes.

These techniques are much better in interactive form when the software functionality can be used, rather than in fixed, static reports – which often sit unread on a shelf anyway. In GIS, this allows zooming, panning, selecting and querying the data, to allow interpretation, build understanding and help consensus. Seeing data on the map of the area, you know can really help facilitate cross-departmental, policy development, resident and community meetings.

The techniques have a solid theoretical basis, best conceptualised by MacEachren (1994) with respect to the use of GIS, but also applicable in other ways.

### Data protection

One of the concerns commonly raised is about data protection. While this should always be taken seriously and considered carefully, the advice of the Information Commissioner's Office is that such data mining activity is permissible as long as no individual can be identified by disclosures of data or a combination of data, and that it does not affect any decisions on individuals.

Section 33 of the Data Protection Act 1998 provides for various exemptions in respect of the processing (or further processing) of personal data for research purposes (including statistical or historical purposes), provided that the processing (or further processing) is exclusively for those purposes and, also, that the following conditions are met:

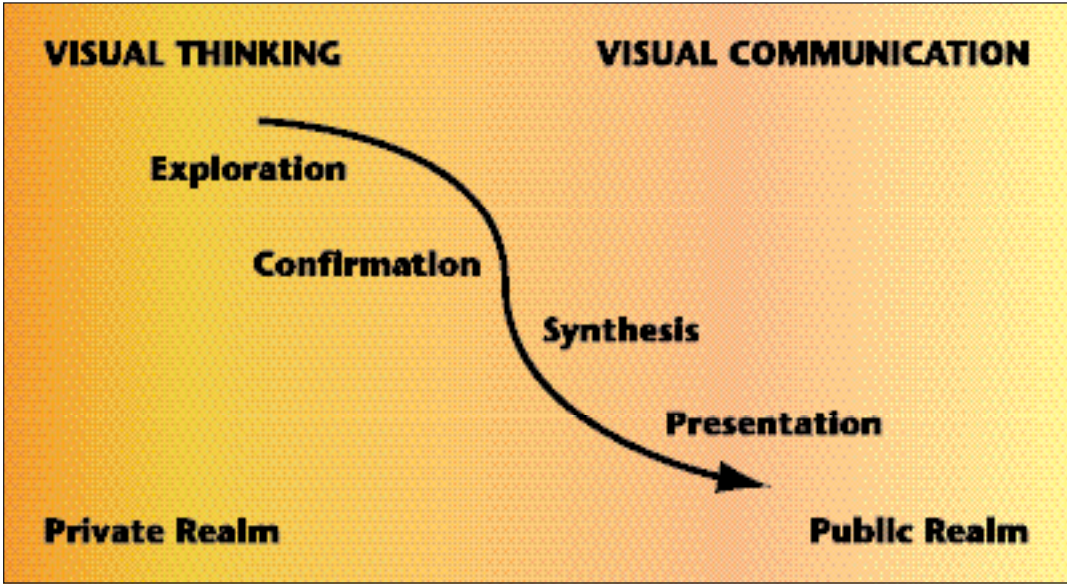
- That the data is not processed to support measures or decisions relating to particular individuals; and
- That the data is not processed in such a way that substantial damage or substantial distress is, or is likely to be, caused to any data subject.

### Predictive analysis

While all this can be useful and revealing in its own right, the ultimate objective is to be able to use the data and the understanding derived from it to help make better policy.

In most cases, the data held in operational systems is historical – it is what has already happened, but because the systems are live, it can potentially be what has happened yesterday. Even real-time tracking is

**'The transformation of raw data into information is a process of a mixture of visualisation, interpretation and aggregation, in which local knowledge and experience is essential.'**



**‘The ultimate objective is to be able to use the data and the understanding derived from it to help make better policy.’**

theoretically possible, with direct links into operational systems, if perhaps still some way in the future.

But the prediction of future trends needs more than just up-to-date data. Data alone is just numbers, even if masses of them can now be extracted relatively quickly and cheaply. Better understanding leading to more reliable predictive capability needs some or all of the following:

- Extrapolation from past patterns;
- Understanding of the social and/or cognitive processes giving rise to patterns;
- Building hypothetical models of social processes;
- Scanning and anticipation of other social, environmental, economic or political factors that could affect the future pattern.

The first of these is a well established method, but the past is not always a reliable indicator of the future; indeed, for some factors, the more a trend continues, the more chance that there will be a rapid collapse or correction. Techniques of multiple regression can show which variables may be linked, but it is important to try to distinguish between cause, effect and collinearity – factors varying together, but not directly causing each other.

With data more readily available, surveys and more qualitative forms of research can instead be focused to find out why people choose options or behave in certain ways, but this is more prone to differences of interpretation, subjective position and opinion.

A more sophisticated approach is to try to understand the processes going on leading to any patterns or trends revealed in the data. This is something that often needs to be done more thoroughly and rigorously; several social studies disciplines like education and housing

have, in the past, been criticised for being under theorised, so that, often, research reports are merely empirical description, like a form of glorified journalism.

As W Edwards Deming has said: ‘Without theory, experience has no meaning. Without theory, one has no questions to ask. Hence, without theory, there is no learning’ (<http://curiouscat.com/management/demingsystempk.cfm>).

**‘With data more readily available, surveys and more qualitative forms of research can instead be focused to find out why people choose options or behave in certain ways.’**

While acknowledging the more complex debates about modernist and postmodernist, and so on, the outcome is that most social processes are very complex and interrelated, with so many variables that can affect them in different ways and to different degrees. Building predictive models that capture all this well enough can be very difficult, and partial knowledge and understanding can be as bad as none at all. But IT modelling can maybe help here as well. One form of programming that captures complex interactions between objects is gaming technology. The ‘Serious Games Initiative’ in the US is focused on uses for games in exploring management and leadership challenges facing the public sector (<http://www.seriousgames.org/>).

It may yet be some way off before we can show and model complex patterns like a city’s economic interactions or housing markets ‘on-screen’, but perhaps it is closer than we think.



**Bob Line**  
**Freelance Consultant**  
**B.Line Housing Information**  
**52 South Knighton Rd**  
**Leicester LE2 3LP**  
**Tel: 0116 270 6441**  
**bline@stayfree.co.uk**  
**homepages.stayfree**  
**.co.uk/bline**

