

A01.3

Delivering geo-statistics on-line: lessons from the 2001 Census

John Maslen, Managing Director, GeoWise Ltd
Garnett Compton, Project Manager, General Register Office for Scotland

Introduction

Geo-statistics, also termed spatial statistics, are defined for this paper as being values of a variable that are distributed in space. They can also be distributed in time. There is a whole sub-discipline of statistics that has been built up around spatial analysis including exploratory data analysis, structural analysis and making predictions. These techniques have been adopted widely in areas such as epidemiology and are more recently being used in areas such as crime pattern analysis.

This paper concentrates on one type of geo-statistics, those relating to people and their activities. This data tends to be geo-referenced by an address of residence or perhaps a work place. Significant issues arise for this type of information in relation to confidentiality and the wish to minimise the risk of disclosure of personal details. Traditionally the approach to get round this problem has been to aggregate statistics to 'higher' geographic areas. This means that geo-statistics about people are very often spatially referenced to polygonal areas of varying sizes. On-line delivery approaches therefore need to be able to offer users simple ways to search for areas of interest and also more flexible tools that allow them to browse areas and build up selections. Interactive web mapping software can offer significant benefits in terms of area searching and also thematic map presentation.

Public Sector Context

Within the context of UK Modernising Government initiatives to encourage the public sector to e-enable their service delivery, this paper concentrates on one delivery mechanism, the web browser. However this does not limit the discussion to a specific type of audience or application. It is possible to deliver a range of applications through a web browser from simple read-only reports to complex analytical and transactional operations suited only to users with specialist skills. While many high-end, specialist software products are not suited to web delivery, even these often have web-enabled versions to address, for example, 'dashboards' for management level reporting. An increasing number of large organisations are seeing the benefits of light-weight intranet applications for relatively sophisticated users which can be rolled out to any number of staff at minimal marginal extra cost and wherever they are located.

Within the UK public sector there is also increasing recognition that:

- Joined-up solutions require data to be shared amongst departments and external partners - web technologies allow this to be carried out effectively and in a controlled manner;
- Local citizens and voluntary organisations are important customers and should be kept informed of how their public services are performing at a local area level;
- Communities can be empowered where local area statistics are made available and decision-makers can be held more accountable.
- Finite resources can be spatially targeted in much smarter ways as geo-statistics become increasingly available for small areas;

- Patterns and trends can be spatially analysed at relatively fine resolutions in order to aid understanding;

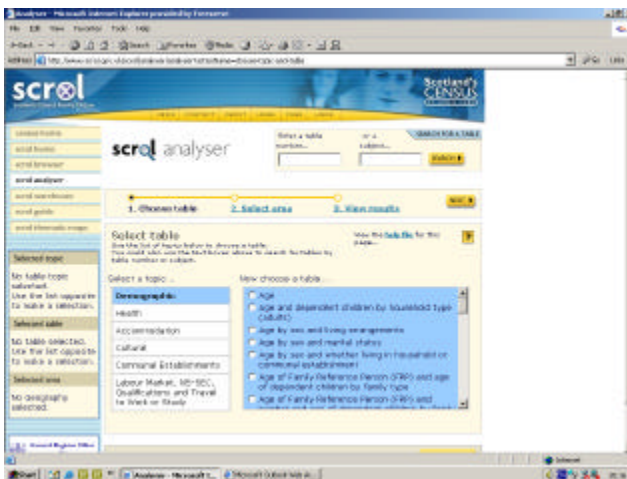
All these factors make a strong case for public sector bodies using and delivering local level geo-statistics to their staff and external customers in a way that is appropriate to their needs. This is now starting to happen with the birth of new tools like CensusInside (www.censusinside.com) allowing organisations to set up their own web-based delivery tools internally and benefit from more widespread use of this dataset.

This paper examines the issues of on-line delivery of geo-statistics through a detailed case study, the development of an Internet application to deliver Scotland's 2001 Census. It will identify specific issues that arose during this development and highlight areas where lessons have been learnt. In the light of the continued growth in volume of geo-statistics made available by public bodies, particularly through the UK Neighbourhood Statistics and Census Access initiatives, the case study raises questions of how this information is best delivered to maximise its value to users and minimise the risks of misuse.

Case Study: Scotland's Census Results On-Line (SCROL) www.scrol.gov.uk

The on-line delivery of geo-statistics of this nature can be traced back to the early beginnings of the Internet when the US Bureau of the Census set up the Tiger Map Service (<http://tiger.census.gov>). In the UK census data has tended to be made available as raw statistics through registered Census Agents or embedded into desktop applications like SASPAC. The CASWEB application developed by the team at MIMAS (www.mimas.ac.uk) was a ground-breaking UK project starting in 1997 to use Internet-based tools including interactive mapping to deliver census tables to the academic community. The Census Access project to make the data free at the point of use has meant the UK can finally start catching up with its North American counter-parts.

In June 2000 GROS initiated the SCROL project. With a fixed delivery schedule, its objective was to develop a system to disseminate the results of 2001 Census in Scotland with the aim of improving the use of, and access to, the Census. The project utilised the services of external contractors in both the pilot and operational phases to support an internal IT Team with minimal previous experience of developing such applications.



Two key dissemination channels were chosen, CDs and the Internet. They were considered to be complimentary approaches with users looking to carry out more extensive data manipulation likely to opt for CDs while the web site would cater for more ad hoc use. It was felt that experienced users would make use of both channels while new user would probably 'discover' the census through the web. This has proved to be the case.

The web site makes all the Census results for all areas freely available. It is aimed at a very wide range of potential users, such as those working as public sector elected representatives, researchers and academics, business users, community workers, voluntary groups and school students. Many of these users had made contact with GROS customer support over previous years. As with many information-rich web sites, the challenge was to assess what these users are likely to want.

Once a pilot application had been developed over the first 5 months, workshops were carried out with groups of specific types of users to gauge reaction. These were managed by web usability experts. As a result of these, a set of user scenarios were drawn up to identify specific requirements for content and functionality. For example:

- An MP wants to know the latest figures on population and households in their Ward and compare them with other 'comparison areas'.
- A local citizen acting on behalf of a voluntary group needs to define a custom area to produce a population figure for a grant application.
- A school student needs to research the health of their local area in relation to the national average.

In terms of system design, estimates were made of potential traffic volumes and worse case scenarios based on discussions with other census agencies around the world and also other UK government agencies. The experience of the 1901 census site having to be re-launched several months later due to system overload within days of launch was still fresh in people's minds. A best estimate of potential traffic levels and user interaction processes was developed with an over-riding principle that the whole system could be scaled up rapidly should the need arise.

Considerable discussion took place as to the best approach to host the site. The high costs of improved bandwidth internally, the need for 24x7 support and the flexibility to upgrade bandwidth in line with demand meant that an external hosting solution was favoured.

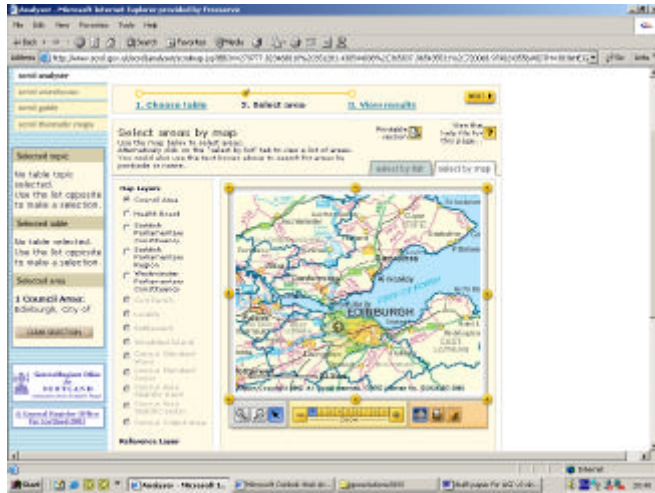
Given the range of users and varying degrees of Census experience, GROS identified three design principles for the operational application to ensure that the Census results were easy to access, use and interpret:

- The data needed to be much more user friendly and remove, as much as possible, the Census jargon which, traditionally, only experienced Census users have been able to easily follow and understand. This would enable the user to more easily access the results without the support of a customer service line to guide them through the process or help them translate their need as happened with previous Censuses.
- The complexities of Census geography needed to be simplified and the geographies made more visual to help users identify the geographical area for which they wanted results.
- The results, presented in the format of tables, were designed based on the requirements of key users. However, these were traditional users of Census data and these tables were likely to be too detailed for the casual user who would require quick access to key indicators, rather than wading through potentially complex tables searching for the results of interest.

The design and functionality of the site were developed with these three principles in mind. In doing so, a site was developed that enabled a user to:

- quickly access profiles for an area with a minimum of one click from the home page. Once a profile was retrieved, users could change the type of profile, change the area, or compare the results for one area against another area of their choice. Specific profiles were designed to cover the key indicators within five main topics (Population, Households, Health, Education and Employment).
- access more detailed tables and results by retrieving a pre-defined table through either a search or list facility. These facilities were developed to make this process more intuitive reducing the need for prior knowledge of the make-up of the tables, yet still making it easy to use for users who do have the prior knowledge.
- select an area of interest using a list facility or an interactive map and gazetteer. The interactive map allows users to visualise Census areas and the different boundaries, group areas of interest or simply roam using standard pan and zoom functions.
- view commentary and text (supporting information) which helps users understand and make effective use of the results. It contains Census classifications and definitions for each variable, information on the One Number Census, Edit and Imputation, Geography and Disclosure Control

To manage the demand of users and the internal resource of loading and quality assuring data, the site was rolled out in three core phases. The first phase provided access only to scrolbrowser (the section of the site where profiles are held); the second phase provided access to some of the more detailed statistics (80 out of 300 tables); and the last phase provided access to the vast majority of tables (270 out of 300 tables). Further roll-outs are planned to accommodate the remaining 30 tables and new functionality. This rolling programme has meant that traffic has been managed effectively and, provided users are kept informed, has meant the site retains its interest. Most of the components within www.scrol.gov.uk are now live.



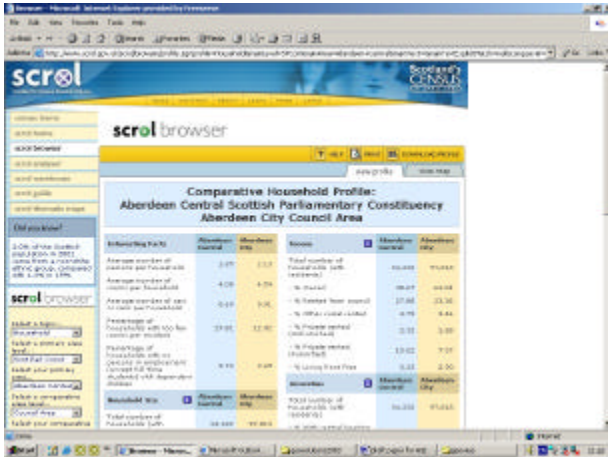
Lessons learnt

1. Traditional data outputs may not be appropriate for a much broader audience –in the case of the Census, tables were not designed for the non-specialist. Opening up the Census to new users requires much more thought than simply providing smart access to all the traditional tables.
2. Try to find out what users want even if the site is catering for a general purpose 'citizen' audience - in most cases they want content delivered fast that is reliable and interesting. Do not be led by technology.
3. Pilot applications are very useful for developing concepts, refining options, testing technology platforms and generating valuable user feedback –it can be necessary to throw away elements of a pilot in order to move forward operationally with acceptable risk.
4. Focus Groups of different sections of the community can be useful to test pre-production versions although these exercises need to be carefully managed by experts.
5. Do not under-estimate size of projects to get large, statistical or geographic databases on line, especially if they are likely to be high profile sites with significant user traffic – there are many 'cost elements' to consider, not least hosting options.
6. Developers need to understand the information being made available and its limitations –in the case of these types of applications they also need to ensure they understand basic rules for the presentation of geo-statistics.
7. Focus on delivering the core functionality well and, once bedded in, consider the need for more advanced functionality. Phased roll-outs of functionality and content and 'soft' launches allow you to manage expectations (internally and externally) and retain better control.
8. Consider involving a range of experts in multiple areas for the development of the application eg. usability, security, platform architecture design, web design etc.
9. Getting in-house IT staff up-to-speed can present a v steep learning curve unless they have worked on large web projects - if you have fixed deadlines then this is high risk and there are alternatives eg. get external contractors in to work alongside internal IT staff to ensure skills transfer.
10. Sites like these do not conform fully to eGIF standards – the question is, how much conformance do you need?
11. A new site needs marketing – make sure these costs are built in.

- Ensure you consider how the site will change working procedures – it may generate more work for internal staff rather than deflect ‘everyday phone queries’ which many might believe.

Conclusions

The SCROL project raised a number of important issues that are not only applicable to on-line geo-statistical reporting but also to any major Internet application looking to deliver rich data content to a broad audience.



The key message is to try and understand user needs, as diverse as they may seem.

Techniques like user scenario modelling to build up realistic sets of questions offer a useful approach. Consider who your priority audiences are and ensure their needs are fully met initially. In many cases the needs of the majority are surprisingly simple – while tools are being developed for exploratory data analysis (see www.commongis.com) many users simply want easy and fast access to area specific profiles summarising and ranking key indicators (such as composite indexes) to help their locational decision-making. They may

want to know the proportion of houses in an area that are second homes or the state of health of people within a neighbourhood. Increasingly they may want to know how everyone in their area travels into the nearby town every morning.

As our technological limitations diminish, we should bear in mind that ‘content remains king’ and smart functionality is not going to replace the need for time investigating user needs. In this vein the paper concludes with two examples of current best practice for delivering useful content using relatively simple on-line tools. The first screenshot below comes from www.census.gov.uk, the ONS site for delivery of the census in England and Wales. This profile report offers a rich array of high level census information including national comparisons and ranks. Below is the CensusAtlas site (www.censusatlas.org.uk) for simple dissemination of council level census data using SVG technology to deliver geo-statistics in a dynamic yet intuitive and visually attractive manner. ONS are also using this technology for thematic mapping on their site for the same reasons.

