Did the dream become reality? Twenty five years of using GIS for Census analysis

Derek Bond*
University of Ulster at Coleraine, U.K.
d.bond@ulster.ac.uk

Abstract

In the last twenty five years Geographic Information Systems (GIS) have moved from being expensive, mainly research orientated, tools to computer applications that reach into many peoples everyday lives. This paper considers whether the predictions made a quarter of century ago about the joint dependency of GIS and population censuses have been realised. When GIS software, hardware and data capture were expensive there was a need to justify the investment. Similarly the cost of population censuses was high and there was a need to justify the expense involved in such an activity. Thus the functionality (often at that stage vapourware) offered by GIS was used to justify continuing with population censuses and the benefits that could be derived from using population censuses as 'ground truth' justified the continuing development of GIS functionality. This paper looks at what actually happened and if either argument has stood the test of time.

Keywords: GIS development; Population Censuses; Ground Truth

1 Introduction

With the decreasing cost of storage and the increasing processing power per unit cost, various forms of Geographic Information Systems have become common place. GIS applications such as satellite navigation and street views are now freely available However, in the early days of GIS the hopes were for much more complex uses of the technology than these simple consumer applications. In particular GIS was seen as a way of justifying the expensive operation of conducting population censuses as it would provide the 'ground truth' required by other attribute data. This paper briefly looks at whether the hoped for benefits and opportunities that GIS technology was seen to offer have been met. This is done by considering the issues that where highlighted in a review paper by Bond, Flowerdew, and Brown (1999).

2 Review of perceived issues

In Worrall and Bond (1996) an extended definition of GIS was given:

'.. a computer-based system for capturing, storing, manipulating, analysing and visualising spatially referenced data and integrating it with other computer based information; a toolkit for the modelling and analysis of complex research, management and planning problems; and a system to support decision makers by enabling them to structure problems and identify potential solutions for evaluation'

(Worrall and Bond (1996, p.366))

.

In the early years it was forecast that the real potential of GIS was that it was likely to lead to the development of new spatial analytical techniques that would utilised both the attribute and topographic data. This would allow for rigorous analysis of the spatial dimension of many real world problems. Until then most spatial issues were dealt with by visualisation or simple spatial correlation coefficients. To this end there was much discussion on the need for vector rather than raster topographic data. The idea was that the topographic data would become an integral part of the new techniques that would be developed. However, very accurate vector data was expensive to obtain needing high quality digitising. So the issue was whether there was attribute data of a quality that could make use of such digital data. Outside of the countries that had moved to a registered based system the only dataset that seemed to fit this requirement was the census of population. Thus it was argued that Censuses and GIS, though both expensive, could be justified on the synergy and value added (Bond 2000).

Even with such detailed data there were many issues. Firstly the issue of confidentiality meant that data could not be released except in aggregate form, so the modifiable area unit problem attracted much attention. The use of dasymetric mapping to help address this problem had attracted little attention (Langford and Unwin 1994). More emphasis was being placed on the role of artificial intelligence and the possibility of geographic analysis machines (GAMs) (Openshaw 1992). Many felt that small area estimation techniques could provided spatially disaggregated data, for decision making purposes, using surveys and the population census for 'weight' (Pfeffermann 2002). Then, given the prevalence of the New Public Management(NPM) paradigm in the public sector at the time (Bond and Ramsey 2013), the issue of cost recovery for data was seen as a major handicap in the development of effective GIS. Finally the issue of visualisation was attracting considerable attention (Hearnshaw and Unwin 1994). The main issue was whether non-specialists could correctly interpret spatially presented information.

3 NPM and Open Data and Software

A big fear in the 1990s was the move towards agency based structures in the professional public sector would lead to 'silos' of information for which even other public sector bodies would have to pay the market price (Worrall and Bond 1996). Such a fear proved unfounded, as while most National Statistics Offices (NSOs) and other bodies responsible for data production restructured in-line with the New Public Management paradigm, few adopted a cost-recovery model. The general argument was that public funds were used to collect the data and so the public should not be charged twice (Bond and Ramsey 2013). In recent years the move has been the other way, with a growth in the open data movement, its impact on GIS is discussed in Bond (2013).

An interesting development has been the increase in what is known as 'volunteer geographic data' (Haklay 2010). This can be described as data provided by individuals rather than formal organisations. As well as open data there has been a growth in open source GIS software (Steiniger and Hunter 2013). Whether these developments have addressed the issues of cost is open to debate, in particular the issue of quality has still to be decided (Ellul 2012).

4 Cognitive issues

The issue of spatial literacy continues to be researched, though mostly in relation to pedagogic issues (Nazari and Webber 2011). There would appear to be more interest in cognitive issues in the area of remote sensing (Burwell, Jarvis, and Tansey 2012). It is often assumed that users are spatially aware and the issue is the usefulness of spatial interfaces as part of information portals (Huang 2011). The impact of software such as Google maps on spatial literacy is, apparently as yet, an unresearched area. The role of spatial visualisation still seems to be a neglected area for NSOs (Bond and Ramsey 2009). This raises the issue of whether public sector statistics bodies have the expertise to develop spatial interfaces to or provide spatial representation of their data (Robinson, Yu, Zeller, and Felten 2008).

5 Spatial analysis

Whilst there have been considerable developments in both the spatial analysis of remotely sensed and medical imagery data there has been little development in the area of statistical inference using spatially referenced socio-economic data. A few papers use techniques that take account of the spatial aspects of the data. For example, Fotheringham, Kelly, and Charlton (2013) used geographically weighted regression to explore demographic aspects of the Irish potato famine and Lofters, Gozdyra, and Lobb (2013) described cancer mapping using traditional Moran I indicators. There is a considerable volume of work looking at linking spatially referenced data. An interesting example is Li and Weng (2007), they linked both socio-economic and remotely sensed data to develop a 'quality of life index'. Others have tried to develop historical, spatially disaggregated series by interpolating data (Gregory, Marti-Henneberg, and Tapiador 2010).

6 Defining spatial units

The modifiable area unit problem continues to attract attention (Flowerdew, Manley, and Sabel 2008). The issue of how to define area units has also received considerable research (Cockings, Harfoot, Martin, and Hornby 2011). This has been an ongoing problem in countries that do not use register based systems. For some the issue is of high political significance, for example, in Shuttleworth, Lloyd, and Martin (2011) the implications for the highly segregated population of Northern Ireland is considered.

7 Dasymetric Mapping and Small Area Statistics Estimation

In recent years the use of dasymetric mapping to help address the modifiable area unit problem has come to the fore. However there are still many issues to be resolved (Langford 2013). Most of the research in the area is concerned with linking socio-economic data to more finely defined remotely sensed data to infer the likely spatial distribution of the attribute data.

Recent examples of dasymetric mapping of census data have generally been concerned with trying to model the spatial distribution of the data at a finer resolution than enumeration districts. For example, in Batista e Silva, Gallego, and Lavalle (2013) CORINE land use data is used to produce a European wide population grid at 100 metres by 100 metres resolution. In Bhatta (2010) the use of remotely sensed data for exploring urban growth and sprawl is discussed. Briggs, Gulliver, Fecht, and Vienneau (2007) uses both CORINE land cover and DMSP light emissions data to model population density. Issues surrounidng the use of fine level satellite imagery data is discussed in Lung, Luebker, Ngochoch, and Schaab (2013). One interesting use is given in (Stimson and Shyy 2013) where voter patterns are inferred using a mixture of small area estimation and dasymetric mapping. Small area estimation continues to be developed, especially inetresting developments are in the area of multi-level and state-space modelling (Pfeffermann 2013). Though, in many cases the survey data to which the population is giving weight is of limited size to be safe to use (Asthana and Gibson 2008, Vallejo-Torres, Morris, Carr-Hill, Dixon, Law, Rice, and Sutton 2009).

8 Conclusions

In this paper the issue of whether the predicted joint benefits of GIS and population censuses have materialised in recent years is considered. Whilst GIS has moved from being an expensive research tool to the general public domain, the killer applications such as satellite navigation and street views have not required population censuses. The hoped for added value that GIS would give to population censuses has not developed. Population censuses still provide 'ground truth' in modelling the spatial distribution of socio-economic attributed data but whether this can continue to justify their costly existence is questionable. Whether the rise in open data will provide less costly alternatives is an area for future research.

References

Asthana, S., and A. Gibson (2008): "Health care equity, health equity and resource allocation: towards a normative approach to achieving the core principles of the NHS," *RADICAL STATISTICS*, 98, 6–26.

Batista e Silva, F., J. Gallego, and C. Lavalle (2013): "A high-resolution population grid map for Europe," *JOURNAL OF MAPS*, 9(1), 16–28.

Bhatta, B. (2010): "Analysis of Urban Growth and Sprawl from Remote Sensing Data," in *ANALYSIS OF URBAN GROWTH AND SPRAWL FROM REMOTE SENSING DATA*, Advances in Geographic Information Science, pp. 49–63.

Bond, D. (2000): "The Role of Population Censuses in a Data Rich Environment," in *INSEE-EUROSTAT SEMINAR ON CENSUSES AFTER 2001 (Paris, November 2000)*, pp. 9–13. INSEE.

——— (2013): "The challenge of open data and GIS," in *ISI WORLD STATISTICS CONGRESS, Hong Kong*. STS78.

Bond, D., R. Flowerdew, and P. Brown (1999): "GIS and Spatial Statistics: Recent UK experiences and developments," in *BULLETIN OF THE INTERNA-*

TIONAL STATISTICAL INSTITUTE, vol. Tome LVIII, pp. 245–248. Available on-line at http://www.stat.fi/isi99/proceedings/arkisto/varasto/bond0484.pdf.

Bond, D., and E. Ramsey (2009): "Evaluating the public's perception of NSO websites using a non- standard mixed methodological approach," in *STATISTICS: INVESTMENT IN THRE FUTURE 2, Prague*.

——— (2013): "Impact of New Public Management Theory on NSOs: An International Case Study," in *59th ISI WORLD STATISTICS CONGRESS, Hong Kong*. STS70.

Briggs, D. J., J. Gulliver, D. Fecht, and D. M. Vienneau (2007): "Dasymetric modelling of small-area population distribution using land cover and light emissions data," *REMOTE SENSING OF ENVIRONMENT*, 108(4), 451–466.

Burwell, C., C. Jarvis, and K. Tansey (2012): "The potential for using 3D visualization for data exploration, error correction and analysis of LiDAR point clouds," *REMOTE SENSING LETTERS*, 3(6), 481–490.

Cockings, S., A. Harfoot, D. Martin, and D. Hornby (2011): "Maintaining existing zoning systems using automated zone-design techniques: methods for creating the 2011 Census output geographies for England and Wales," *ENVIRONMENT AND PLANNING A*, 43(10), 2399–2418.

Ellul, C. (2012): "Can Free (and Open Source) Software and Data be Used to Underpin a Self-Paced Tutorial on Spatial Databases?," *TRANSACTIONS IN GIS*, 16(4), 435–454.

Flowerdew, R., D. J. Manley, and C. E. Sabel (2008): "Neighbourhood effects on health: Does it matter where you draw the boundaries?," *SOCIAL SCIENCE & MEDICINE*, 66(6), 1241–1255.

Fotheringham, A. S., M. H. Kelly, and M. Charlton (2013): "The demographic impacts of the Irish famine: towards a greater geographical understanding," *TRANSACTIONS OF THE INSTITUTE OF BRITISH GEOGRAPHERS*, 38(2), 221–237.

Gregory, I. N., J. Marti-Henneberg, and F. J. Tapiador (2010): "Modelling long-term pan-European population change from 1870 to 2000 by using geographical information systems," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A-STATISTICS IN SOCIETY*, 173, 31–50.

Haklay, M. (2010): "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets," *ENVIRONMENT AND PLANNING B-PLANNING & DESIGN*, 37(4), 682–703.

Hearnshaw, H., and D. Unwin (eds.) (1994): *VISUALIZATION IN GEOGRAPHIC INFORMATION SYSTEMS*. Belhaven Press, London.

Huang, K. H. (2011): "A GIS-Interface Web Site: Exploratory Learning for Geography Curriculum," *JOURNAL OF GEOGRAPHY*, 110(4), 158–165.

Langford, M. (2013): "An Evaluation of Small Area Population Estimation Techniques Using Open Access Ancillary Data," *GEOGRAPHICAL ANALYSIS*, 45(3, SI), 324–344.

Langford, M., and D. Unwin (1994): "Generating and mapping population-density surfaces within a Geographic Information-System," *CARTOGRAPHIC JOURNAL*, 31(1), 21–26.

Li, G., and Q. Weng (2007): "Measuring the quality of life in city of Indianapolis by integration of remote sensing and census data," *INTERNATIONAL JOURNAL OF REMOTE SENSING*, 28(1-2), 249–267.

Lofters, A. K., P. Gozdyra, and R. Lobb (2013): "Using geographic methods to inform cancer screening interventions for South Asians in Ontario, Canada," *BMC PUBLIC HEALTH*, 13.

Lung, T., T. Luebker, J. K. Ngochoch, and G. Schaab (2013): "Human population distribution modelling at regional level using very high resolution satellite imagery," *APPLIED GEOGRAPHY*, 41, 36–45.

Nazari, M., and S. Webber (2011): "What do the conceptions of geo/spatial information tell us about information literacy?," *JOURNAL OF DOCUMENTATION*, 67(2), 334–354.

Openshaw, S. (1992): "Some suggestions concerning the development of artificial-intelligence tolls for spatial modeling and analysis in GIS," *ANNALS OF REGIONAL SCIENCE*, 26(1), 35–51.

Pfeffermann, D. (2002): "Small area estimation - New developments and directions," *INTERNATIONAL STATISTICAL REVIEW*, 70(1), 125–143.

——— (2013): "New Important Developments in Small Area Estimation," *STATISTICAL SCIENCE*, 28(1), 40–68.

Robinson, D., H. Yu, W. P. Zeller, and E. W. Felten (2008): "Government Data and the Invisible Hand," *YALE JOURNAL OF LAW AND TECHNOLOGY*, 11, 160–175.

Shuttleworth, I. G., C. D. Lloyd, and D. J. Martin (2011): "Exploring the implications of changing census output geographies for the measurement of residential segregation: the example of Northern Ireland 1991-2001," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A-STATISTICS IN SOCIETY*, 174(1), 1–16.

Steiniger, S., and A. J. S. Hunter (2013): "The 2012 free and open source GIS software map - A guide to facilitate research, development, and adoption," *COMPUTERS ENVIRONMENT AND URBAN SYSTEMS*, 39, 136–150.

Stimson, R. J., and T.-K. Shyy (2013): "And now for something different: modelling socio-political landscapes," *ANNALS OF REGIONAL SCIENCE*, 50(2, SI), 623–643.

Vallejo-Torres, L., S. Morris, R. Carr-Hill, P. Dixon, M. Law, N. Rice, and M. Sutton (2009): "Can regional resource shares be based only on prevalence data? An empirical investigation of the proportionality assumption," *SOCIAL SCIENCE & MEDICINE*, 69(11), 1634 – 1642.

Worrall, L., and D. Bond (1996): "Geographic information systems, spatial analysis and public policy: the British experience," *INTERNATIONAL STATISTICAL REVIEW*, 65, 365–380.