The Elements of a Computational Infrastructure for Social Simulation

Mark Birkin[1], Rob Procter[2], Rob Allan[3], Sean Bechhofer[4], Iain Buchan[5], Carole Goble[4], Andy Hudson-Smith[6], Paul Lambert[7], David de Roure[8], Richard Sinnott[9]
[1] School of Geography, University of Leeds; m.h.birkin@leeds.ac.uk (corresponding author) [2] School of Social Sciences, University of Manchester [3] Daresbury Laboratory, Science and Technology Facilities Council [4] School of Computer Science, University of Manchester [5] School of Medicine, University of Manchester [6] Centre for Applied Spatial Analysis, UCL [7] Applied Social Science, University of Stirling [8] Electronics and Computer Science, University of Southampton [9] National e-Science Centre, Glasgow

Abstract

The applications of simulation modelling in social science domains are varied and increasingly widespread. The effective deployment of simulation models depends on access to diverse data sets, the use of analysis capabilities, the ability to visualise model outcomes, and to capture, share and re-use simulations as evidence in research and policy-making. We describe three applications of e-social science which promote social simulation modelling, data management and visualisation. An example is outlined in which the three components are brought together in a transport planning context. We discuss opportunities and benefits for the combination of these and other components into an e-infrastructure for social simulation, and review recent progress towards the establishment of such an infrastructure.

Keywords: social science; simulation models; e-infrastructure; scenario planning.

1. Introduction

Recent years have seen a big upswing of interest in simulation for the natural and physical sciences. Some commentators have emphasised how simulation embodies the transition from science 'in vitro' to science 'in silico' and argue that this signals the emergence of a new scientific paradigm where the focus is on data exploration, rather than hypothesis or experimentally driven research: "*The world of science has changed, and there is no question about this. The new model is for the data to be captured by instruments or generated by simulations before being processed by software and for the resulting information or knowledge to be stored in computers. Scientists only get to look at their data fairly late in this pipeline. The techniques and technologies for such data-intensive science are so different that it is worth distinguishing data-intensive science from computational science as a new, fourth paradigm for scientific exploration*." (Bell, Hey and Szalay, 2009)

This trend towards simulation-based methods can also be seen across a range of social science disciplines. Social simulation applications can be characterised as wide-ranging in both method and subject domain; policy and theory relevant; short- and long-term. They are typically data-rich; evidence-based; often computationally intensive, and are of communal interest. Both micro-simulation and agent-based modelling – for which Epstein (2007) has coined the term 'generative social science' – have been widely adopted within economics, sociology and geography. Simulation models have also provoked high levels of interest in healthcare research, anthropology and political science. For example, in economics simulation approaches have been applied to derive the nature of market behaviour from the interactions between individual consumer and producer 'agents'. These approaches are characterised by an acceptance of disequilibrium behaviour and heterogeneity, both in contradiction of traditional normative economic assumptions (Arthur, 1999). In the field of

epidemiology, similar agent-based models have been used to model the transmission of illness across a population, with alternative policy interventions to control resistance (i.e. vaccination) or social mixing (Ferguson et al., 2005). Long-term projections of the composition and behaviour of populations is increasingly of great importance to planners and policy-makers with an interest in the increasing mass, ethnic diversity and age profile of the population in developed countries – for example, the implications of increased dependency of a growing elderly and retired population on a shrinking cohort of economically active workers (http://www.lse.ac.uk/collections/MAP2030/).

Social simulation has been an important focus for the UK National e-Social Science programme (Halfpenny et al., 2009). In particular, three projects – DAMES (http://www.dames.org.uk), MoSeS (http://www.ncess.ac.uk/research/geographic/moses) and GeoVUE (http://www.ncess.ac.uk/research/geographic/geovue) – have targeted complementary stages of the simulation research lifecycle: data management and integration; mathematical modelling and analysis; and visualisation respectively. Building on these foundations, the National e-Infrastructure for Social Simulation (NeISS – http://www.neiss.org) project is undertaking a coordinated programme of productionising services and community engagement to create a sustainable, world-class social simulation capability, supporting the entire research lifecycle and with potential for major impact, both in the UK and internationally.

The NeISS seeks to provide production-quality services for social simulation that will enable researchers to take full advantage of data intensive methods. Exploiting the capacity that the new generation of simulation tools and infrastructure have for generating large volumes of data is not, by itself, sufficient to maximise their benefits: the aim of NeISS is to support the entire simulation research lifecycle, from data management and population generation, through to model building and execution, visualisation, analysis, publication and archiving of the entire process for future discovery and re-use. Further, NeISS aims to build community capacity in the use of social simulation by promoting a sharing of simulation methods, models and results among social scientists, and broadening the user base to include commercial and public sector decision-makers and the general public.

In this paper, we discuss the motivations for providing computational support for social simulation, and then describe some of the important developments to date in the NeISS project. An example scenario will be articulated and we consider the scope and benefits of an integrated e-infrastructure platform.

2. NeISS Motivation and Overview

The benefits of the NeISS can be viewed from a number of perspectives, but what is most important is the way in which it aims to bring together a range of services in order to provide integrated support for the whole simulation research lifecycle. The NeISS aims to introduce social scientists to new ways of thinking about social problems, and provide new services, tools and research communities to support them. These tools will be capable of being deployed for a diverse range of social research domains: they will enable users to create workflows to run their own simulations; visualise and analyse results, and publish them for future discovery, sharing and re-use. NeISS will facilitate development and sharing of social simulation resources within the social science community, encourage cooperation between model developers and researchers, and help foster adoption of simulation as a research method in the social sciences, and as a decision support tool in the public and private sectors.

The computational requirements of simulation can be rather extensive. In work on the transmission of avian bird flu, Ferguson et al. (2005) observed that "to do these runs quickly, the model needed ... huge amounts of memory — 20 times that found on a typical PC. In fact, the team hooked up ten high-powered computers in parallel, but even then the final runs took more than a month of computer time". The suitability of social simulation models for massive parallelisation has also been demonstrated previously (George et al., 1997). Computational support for decision-making with social simulation has been explored through a series of projects funded through the UK e-Social Science programme (Birkin et al., 2005; Townend et al., 2009).

Visualisation is intimately linked to simulation as the means by which complex model outputs can be easily digested and interpreted. In addition, the visual representation of geographic information – as maps – provides a standard for decision-support in applied contexts, from emergency planning to retail management. McEachren et al. (2005, p300) explain how visual simulation supports emergency planning in "transportation support, search and rescue, environmental protection and fire fighting" and go on to argue that since decision-making is typically a multi-step, multi-participant process then collaborative environments have major importance. The potential for computational steering, with a (real-time) interaction between computational simulation and policy intervention (e.g. Brodlie et al., 2004), also deserves further exploration in the context of social policy. Work in the GeoVUE node has combined the development of a shared repository for spatial data (MapTube) and more innovative and experimental approaches through mechanisms like 'Second Life'.

The importance of simulation modelling as a means for combining evidence at alternative levels of aggregation, or from complementary sources, has long been recognised in the domain of micro-simulation and a range of techniques for data merging (e.g. Merz, 1986) have been developed to support micro-simulation. Increasingly, the relationship between theory and evidence, and the issues of model validation and verification, are seen as crucial questions for agent-based social simulation (Manson and O'Sullivan, 2006). Social simulations typically require the use of multiple data resources that are heterogeneous in both scale and formats, and subject to varying access constraints (while a simulation itself can also generate further data resources relevant to subsequent analysis). Moreover, it is increasingly recognised that the most important data has a rich, complex longitudinal character (Wolfson, 2009) that is especially open to further permutations and enhancements. Tools supporting access to and further enhancement of heterogeneous social science data resources are a key component of the DAMES project and have been incorporated into the NeISS.

There are other e-Research components that are potentially of great value for social simulation, but whose potential has not yet been widely recognised. These components may be established in other fields of e-Research but not yet deployed within social simulation; they may be still evolving as e-Research tools; or they may be manifestations of practices that are undertaken within social simulation, but not yet translated into an e-infrastructure environment. Examples include secure, role-based access to simulations, the integration of service components into repeatable (plug-and-play) workflows, and the provision of repositories for the archiving and sharing of social simulation models and components.

Finally, ensuring the widest possible take up of NeISS requires raising awareness of social simulation among academic research groups, public and private sector organisations; understanding the requirements of different types of user; building capacity and skills through training programmes; and providing services to facilitate public engagement in research and policy-making.

3. Social Simulation Services

The UK e-Social Science programme has supported development in three areas which are highly relevant to social simulation and can provide the elements of a much more comprehensive e-infrastructure. These are simulation modelling, data analysis and visualisation. Here we present a brief review of progress in each area.

3.1 Social Science Modelling and Analysis

Methods developed in the MoSeS project can support the generation of synthetic populations, demographic forecasting and activity analysis. A Population Reconstruction Model (PRM) is used to produce synthetic individual and household populations for small geographical areas. In the context of a specific city – Leeds – this means that 705k individuals are aggregated into 302k households, and each household is allocated to one of 2,400 census output areas. The individuals are synthesised from the UK Census Sample of Anonymised Records (http://www.ccsr.ac.uk/sars/) and therefore represent real people. The PRM uses UK census data for small areas to ensure that real variations are also represented at local level, for example, heavy concentrations of young people, students and ethnic minorities in the central areas, while more affluent families are much more likely to be found in the northern suburbs and commuter villages.

The Population Forecasting process produces annualised 'future' versions of the household and individual database for a 30-year planning horizon, starting in 2001. These projections were generated using dynamic models of individual and household demographics to accommodate ageing, morbidity and mortality; as well as household formation and dissolution, new births, and the migration of individuals and households between small areas in the city, as well as immigration and emigration elsewhere in the UK and the Rest of the World (Wu et al., 2008; Birkin et al., 2009).

A dynamic micro-simulation model is used which applies demographic transition rates to individual people and households. These transitions include ageing, migration, household formation and dissolution (leading to the creation and elimination of households), fertility, illness and mortality (leading to the creation and elimination of individuals). The dynamic transition rates are estimated in relation to longitudinal data drawn largely from the British Household Panel Survey (BHPS)[1], but also using other sources, including ONS ward-level Vital Statistics, International Passenger Survey and pupil census data (see Wu et al., 2008, for more details).

The advantages of this dynamic micro-simulation approach are well-known, for example, allowing efficient representation of highly disaggregate large-scale populations, and facilitating flexible aggregation to various levels of spatial and sectoral detail (van Imhoff and Post, 1998). The synthetic data generation approach also facilitates detailed modelling without incurring threats to privacy or the risks associated with the use of personal data.

An example application of population recreation with dynamic forecasting has been prepared in association with Leeds City Council Social Services. A 'Needs Analysis' for social care was conducted to identify high risk groups. In many cases, for example to identify multiple deprivation and co-dependency, synthetic reconstruction is required simply to compute current distributions. Estimates for the base year of 2006 and projections to 2016 and 2031 are shown in Table 1. Some

key policy questions here are whether existing utilisation is a fair reflection of need, or whether provision drives uptake; and how future provision can be optimised or improved in relation to expected change, particularly in relation to an ageing and more ethnically diverse population.

Activity Modelling: the implications of population change for service provision in cities is an important motivation for the MoSeS project. Such understanding requires that demographic projections be combined with models of service utilisation and uptake. A powerful approach to this problem, with applications to services as diverse as education, health care, retailing, utilities and emergency services (Birkin et al., 1996) is the application of spatial interaction models which combine demand for services with the quality, quantity and accessibility of provision to generate modelled flows of products, patients, students and customers within a city region. Such flows can then be aggregated to provide indicators of equity and efficiency in relation to current or future service networks. An example of this technology was provided in an early e-Social Science demonstrator project – Hydra – which considered the optimal location of cancer screening services under alternative scenarios of patient demand and the economics of provision (Birkin et al., 2005). Further examples are introduced in Section 4 below.

	Day Care		Demographics				
PCT	Attendance	Places	Elderly	LLTI	Codeps	Ethnicity	Deprived
2006							
Leeds NW	514	639	21464	12837	1931	1200	1506
Leeds NE	298	138	14465	8720	1323	1136	1022
Leeds E	514	587	20271	12152	1843	942	1241
Leeds S	400	431	17328	10438	1475	596	1151
Leeds W	343	277	13685	8096	1168	488	946
2016							
Leeds NW	?	?	23297	12234	1662	2146	1104
Leeds NE	?	?	15457	8182	1199	1962	798
Leeds E	?	?	22108	11747	1611	1856	989
Leeds S	?	?	18972	9980	1316	1167	792
Leeds W	?	?	14020	7304	997	960	643
2031							
Leeds NW	?	?	29061	16015	2308	4719	1191
Leeds NE	?	?	19662	10766	1638	3789	913
Leeds E	?	?	26628	14709	2240	4518	1058
Leeds S	?	?	25742	13974	2048	3112	921
Leeds W	?	?	16971	9202	1427	2414	700

Table 1: Demographic forecasts to support the delivery of social care.

3.2 Social Simulation Data

An exciting opportunity afforded by more powerful computational infrastructural provision for simulation modelling is the enhanced capacity to conduct multiple analyses in response to plausible

variations in the representation of the underlying data. Examples might include re-running simulations using different categorisations of measures of ethnicity or educational qualifications, or using different weighting factors for population subgroups. Across the social sciences, while it is recognised that changes to the data underlying an analysis may have significant consequences upon subsequent results, a lack of suitable tools to manage and analyse the large volumes of data has meant that nevertheless, it has hitherto been relatively uncommon for researchers to try out numerous different data permutations, or to systematically document the options available to them and the choices they made.

The skills and issues associated with (potential) variations in the underlying data are often referred to as those of 'data management'. They can usually be described as concerning the possibility of recoding or restructuring variable measures, or of enhancing data by linking it with other relevant resources. The DAMES project seeks to raise capacity in, and to raise standards in the documentation of, a number of data management tasks associated with quantitative datasets in social science research.

An example of the application of data management to social simulation would be in understanding the long run impact of demographic changes upon the structure of social inequality. These are likely to be non-trivial because many other consequential social structural outcomes are strongly differentiated by age and birth cohort. For instance, due to recent expansions in further and higher education, younger age cohorts in contemporary Britain have much higher average levels of educational qualifications, but it might be expected that the long term benefits of those qualifications, as these cohorts age, may not be as strongly differentiating as has been the case in the past. Other simulation studies have addressed similar themes (e.g. O'Donoghue, Leach and Hynes, 2009). An innovation of the NeISS project are that its infrastructural resources will (i) support numerous alternative measurement representations of key data on educational qualifications and related socio-economic outcomes, and (ii) allow the development of the application in a dynamic spatial model which are sensitive to local level variations which themselves are of consequence to the age and socio-economic distribution (e.g. Birkin, Wu and Rees, 2009).

3.3 Social Science Visualisation

We argued in Section 2 above that visualisation is an essential component of a social simulation infrastructure. MapTube (http://www.maptube.org) was initially developed as part of the GeoVUE project to allow rapid sharing of thematic maps within an environment that allows simple visual analysis and data overlay. Data is converted via the GMapCreator software suite that allows users to quickly and easily convert geographical boundary data (shapefiles) to a MapTube overlay. In essence GMapCreator creates a tile-based raster version of any vector dataset thus allowing data, which may be subject to file size restrictions or indeed copyright issues, in vector format to be shared online. Created tiles are hosted on external servers with MapTube collating data via use of XML. MapTube acts as a content creation tool as well as a portal for thematic geographic information.

MapTube's custom tile renderer was developed to extend beyond the ability to convert Shapefile file format. Indeed, it is possible to convert almost any format containing geographic information into a map on the system. This is ensured via the creation of geographic boundaries directly on the server so any data linked to boundaries such as postcode districts, census wards, output area and such like can be mapped via a simple link to the raw data file. This allows near real-time data collection and visualization with the addition of any relevant geographic tag in the collection process. This aspect of

MapTube has been used extensively by traditional media with surveys carried out by BBC Radio 4, BBC South, BBC North West and BBC East all powered by the MapTube system. We are extending this feature set via the NeISS project to create a web based service allowing custom surveys to be setup and mapped in real-time with output compatible with higher end GIS systems.

Initially aimed at social scientists and policy makers, this crowd sourced data has notable potential. The system makes use of social networking systems, such as Twitter, to collect data. Currently in beta testing a demonstration map integrated with Twitter to display live weather information shaded by postcode has attracted over 20,000 responses a day. The Twitter feed is geo-located and linked to the outlines located on the MapTube server allowing districts to be shaded according to the response. The map updates itself every three minutes allowing for time tagged data collection and display. Figure 1 illustrates the demonstration map with Tweets overlaid with the shaded data.

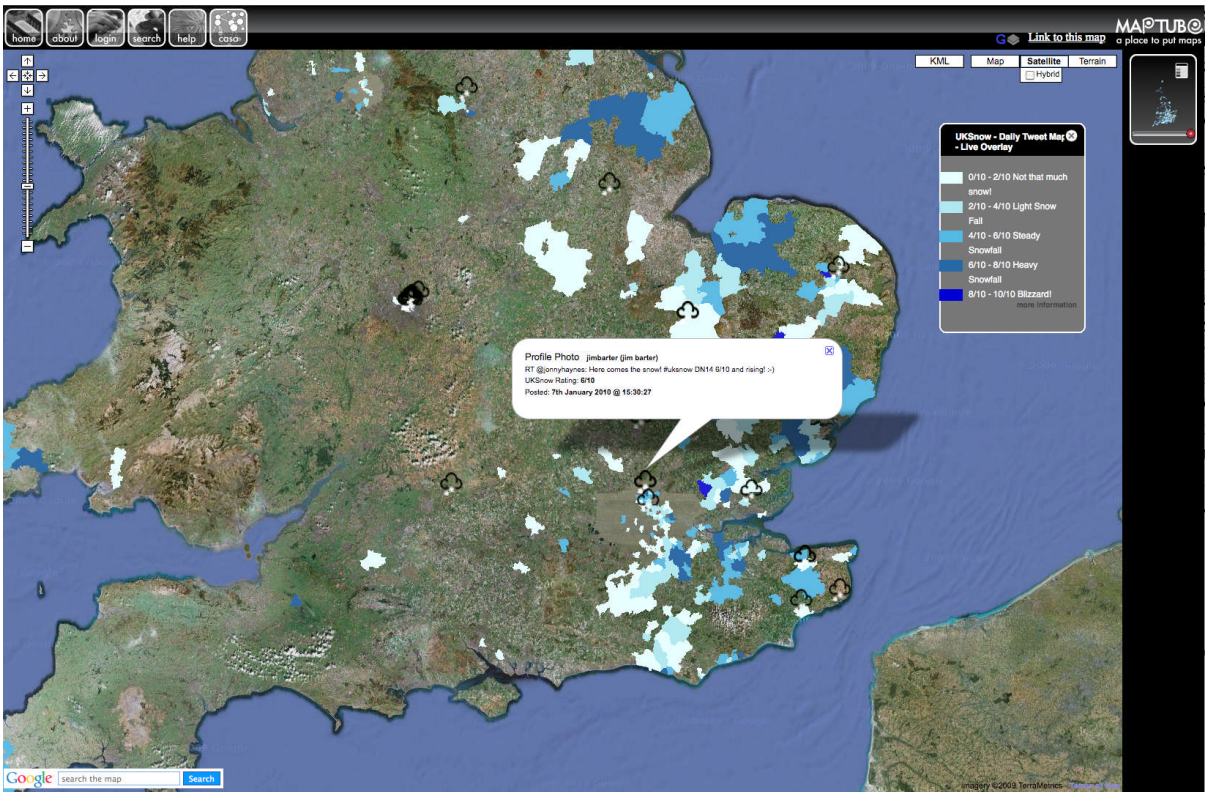


Figure 1: Snow reports from January 2010: live twitter feed to MapTube.

4. An Example: Congestion in Leeds in the Year 2031

In this section we describe an example of the application of the NeISS service components to a problem involving the development of the transport system in Leeds from the period 2001 to 2031.

4.1 Population Reconstruction and Transport Generation

We used the PRM described in Section 3.1 to generate a base population of the city of Leeds for the year 2001.

Transport demand was estimated from the combination of individual level data from the SARs with travel demand data from the British Household Panel Survey (now enhanced to the UK Household Longitudinal Study (http://www.iser.essex.ac.uk/survey/ukhls)). It includes many behavioural and activity variables, including for example political voting patterns, leisure activities and membership of sporting clubs, income and expenditure, as well as duplicating most census variables.

We have used a synthetic matching procedure in order to link individual records from the PRM to a similar record in the BHPS on the basis of five shared characteristics (age, ethnicity, gender, social class and economic activity). Although this procedure is well-defined and robust, it is also arbitrary and a number of questions merit substantial further exploration, concerning aggregation, variable and data selection. For example, should higher (or lower) levels of aggregation be applied to variables like ethnicity? Which variables should be selected for matching, and could alternative statistical matching algorithms be deployed? Should an alternative source such as the National Travel Survey (Department of Transport, 2009) be used? Furthermore questions also arise in whether the procedure is free from biases between rural and urban areas due to differences in accessibility and travel mode variations. By providing an infrastructure to support multiple analyses after the re-specification of data resources, the DAMES contribution to NeISS (as described in section 3.2 above) seeks to address such questions.

4.2 Trip behaviour

A simple social simulation model of urban travel behaviour is used to generate estimates of the volume of flows on the traffic network which arises from our estimates of travel demand as described in section 4.1 above. The model assumes a simplified transport network which has both a public transport and a private transport component. A coarse spatial zoning into 33 census wards was adopted and the distances between each pair of zones generated from a simplified set of routes between neighbouring wards. Trip destinations were estimated as a function of workplace locations as captured within the 2001 Census Special Workplace Statistics (SWS). For this purpose we used a doubly-constrained spatial interaction model which was calibrated in order to produce estimates of the average trip distance by both public and private means and a model split between trips equivalent to those derived from the 2001 census for the city of Leeds (Birkin et al., 2009).

4.3 Model indicators and visualisation

The outputs from the trip modelling process are based on three key performance indicators (KPIs): the level of congestion as measured by the average time spent in travel (to the workplace) by residents of each area; the level of pollution given as a linear function of the number of private and public trips in each ward normalised by the physical area of each zone; and the number of road accidents given as a linear function of the total trip distances in person kilometres by both public and private transport in each ward.

In order to visualise the outputs of the modelling process, we export data in CSV format files of the KPIs into the MapTube map generator, as described in Section 3.3 above. Three sets of outputs are shown at Figure 2 below with a *Google Maps* representation which indicates a broadly concentric pattern of pollution with its focus in the centre of Leeds; and thumbnails of both average journey time and road accidents. Not surprisingly, the distribution of pollution, road accidents and accessibility all follow a strongly urban-rural pattern.

Note that although we have presented these outcomes as a simple and limited set of indicators (3 KPIs and 33 wards: 99 values in total), one of the beauties of the simulation modelling approach is that extensive disaggregation is a straightforward matter. For example, if we were to examine the trip patterns of eight different age groups with four ethnicities, four social groups in both public and private modes, and to disaggregate spatially down from 33 census wards to 2,400 output areas then we would now have more than 600,000 values (8 x 4 x 4 x 2 x 2,400) values for each indicator which is roughly equivalent to the population of Leeds. The opportunity and challenge of flexible aggregation to appropriate levels of detail is a well-known and important property of social micro-simulation modelling in this style (e.g. van Imhoff and Post, 1998).

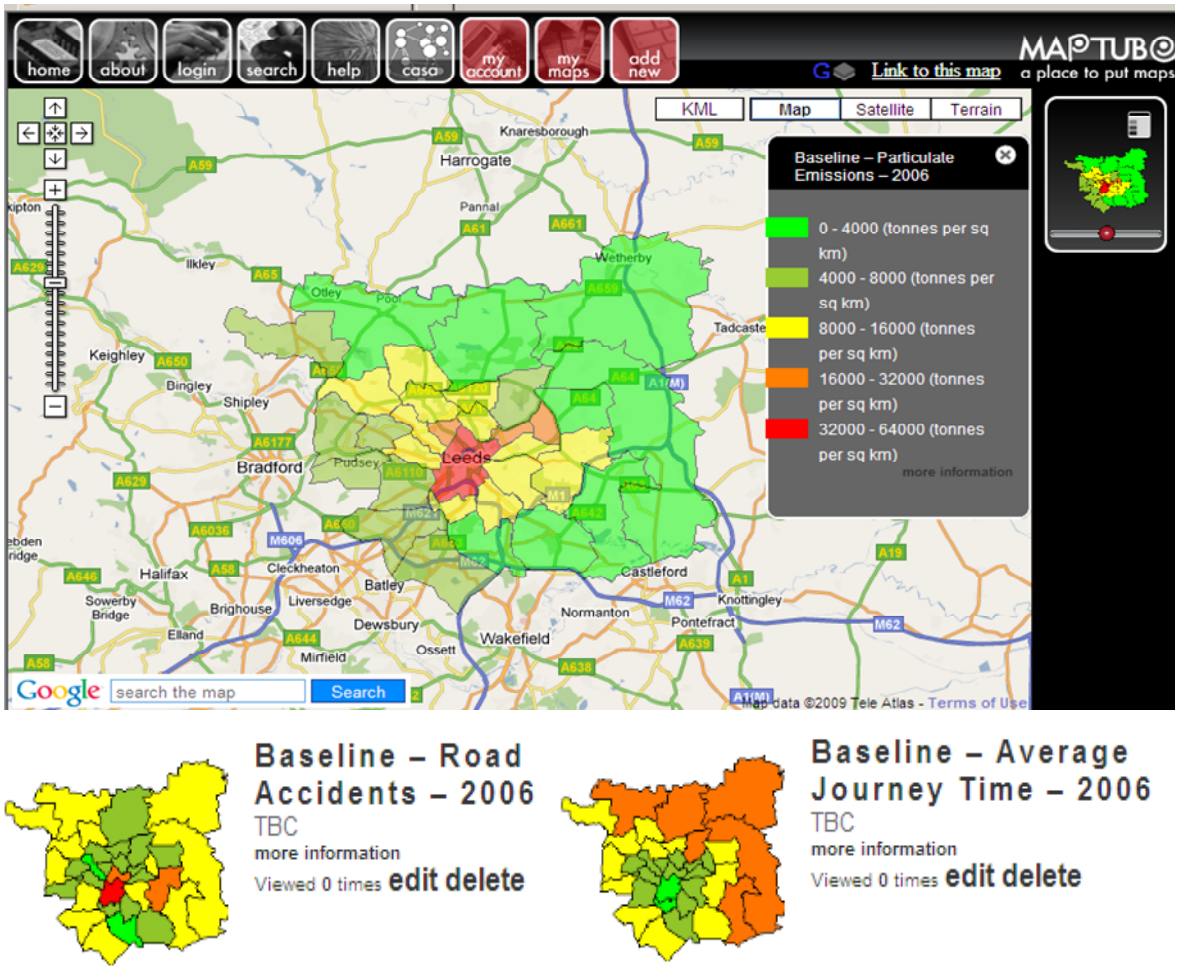


Figure 2: Baseline performance indicators for transport analysis.

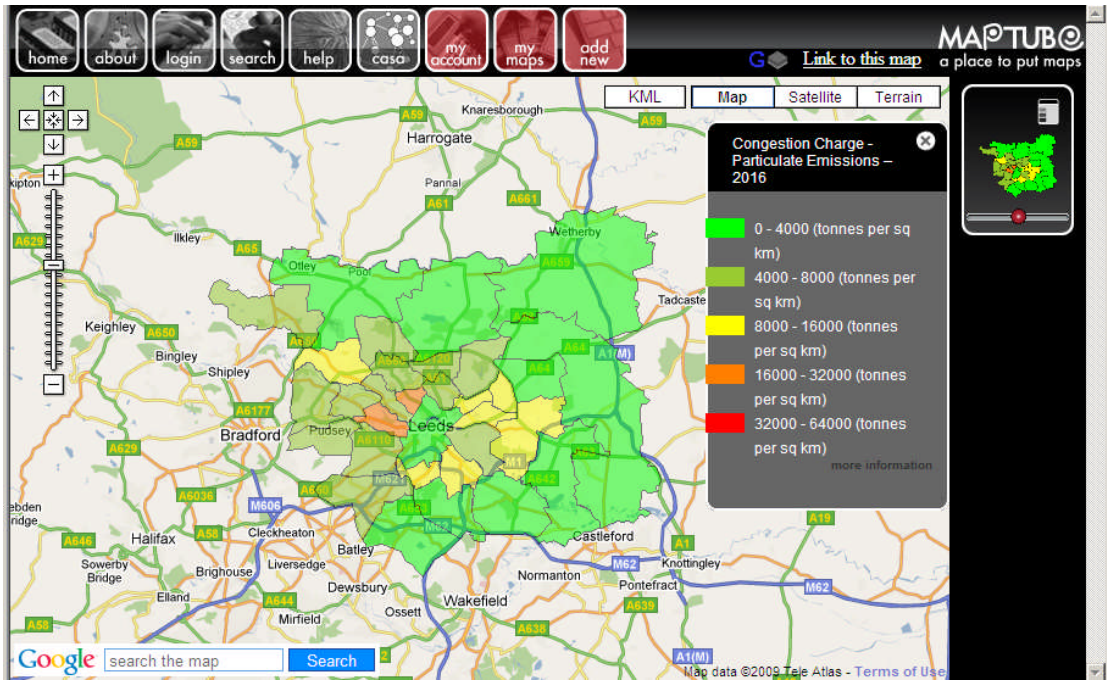
4.5 4 Model scenarios

The real power of social simulation becomes evident when alternative scenarios and future projections are examined. In this work, we performed the following simulation experiments. First, we projected the population of Leeds over a thirty-year period from 2001 to 2031.

Second, we ran ‘what if?’ simulations of trip behaviour assuming:

i. No change in the underlying transport network
ii. The introduction of a Supertram into Leeds between the years of 2011 and 2014
iii. Adoption of a congestion charge in the central areas from 2009 onwards
iv. The introduction of uniform road pricing in all areas of the city from 2009 onwards

Figure 3 shows the effect of the road pricing and congestion charging scenarios on the pattern of air pollution which was seen earlier. Road pricing effects a general softening in the impacts of air pollution by encouraging a city-wide redistribution from private to public transport, while congestion charging has big benefits for the central areas but some displacement of externalities to neighbouring zones.



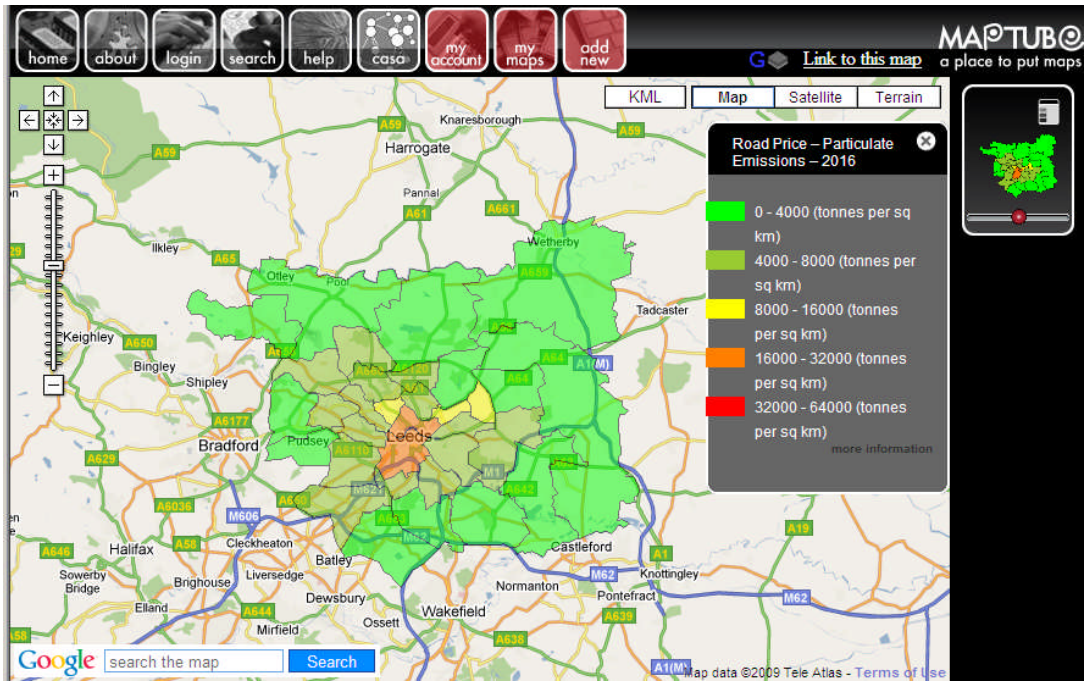


Figure 3: Impact on air pollution of alternative transport scenarios.

A fuller set of scenario illustrations can be found at the MapTube website, at http://maptube.org/map.aspx?mapid=773 up to http://maptube.org/map.aspx?mapid=791.

5. Discussion

In section 4, we looked at an example of a social simulation model in the context of transport scenario analysis for the city of Leeds. The example combines building blocks from existing e-Social Science activities in data management, spatial modelling and visualisation. At present these components are loosely coupled to the extent that interactions are essentially flows of data between them. In this section, we consider the framework for the articulation of a more tightly coupled e-Infrastructure which combines these elements, and we use this as the basis for a discussion of development priorities for the NeISS.

A generalised description of a social simulation scenario is provided in Figure 4. The combination of constituent services into a tightly-coupled infrastructure could provide the capability for repeated simulations across different local areas (i.e. repeat models for congestion charging in Leeds, Manchester, Nottingham etc.); for revised scenario analysis across different time periods (e.g. impact of road charging in 2012, 2015, 2018, ...); and even the execution of different models at various levels of refinement (e.g. coarse-grained analysis of alternative road-pricing structures as the basis of selection of a candidate policy; fine-grained analysis of impacts across different occupations, age bands, ethnic groups for local neighbourhoods). Flexible data inputs could support various different substantive application domains: thus in moving from transport to health care simulations one might substitute health-based constraints in the population reconstruction; merge lifestyle and morbidity variables in place of trip behaviour; and apply interaction models to patient flows or GP visits rather than journey-to-work trips, but the overall simulation architecture would be very similar.

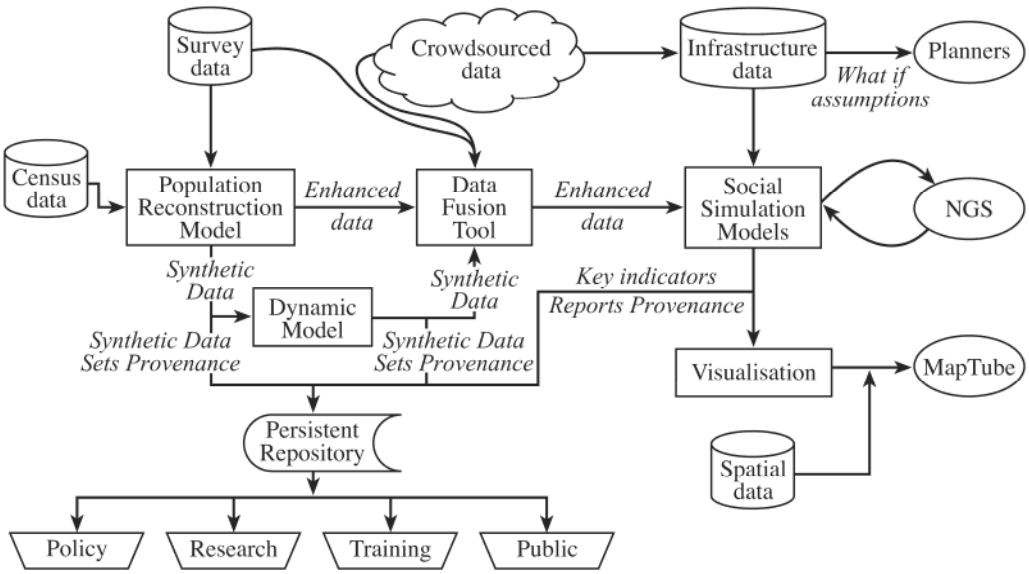


Figure 4. The components of an e-infrastructure for social simulation.

For the transport application, the social simulation tools of Figure 4 comprise a set of spatial interaction models. The development of a more extensive set of tools including analytics such as pattern detection is envisaged within NeISS and in the longer term the provision of standards to allow external simulation tools to be imported and integrated could greatly extend its capability. The integration of NeISS with the UK National Grid Service (NGS) creates the possibility for the definition and execution of simulations across a wide-ranging parameter space or group of 'what if?' scenario options (for example, show me projections of my local area under assumptions of high/medium/low economic growth; free movement/restricted migration etc) where such model variants might be executed under something like a parameter sweep of the type popular in e-Research applications in the physical sciences and realised in projects such as nanoCMOS (Walker et al., 2009).

With standardised licensing and security it would be possible to access European data repositories and integrate distributed datasets. This would allow the investigation of, for example, social research problems that require modelling of population movement across European boundaries. Given the computational intensity of these activities, it is likely that tackling such problems will call for access to international scale compute resources. We observe that the European Grid Initiative (http://web.eu-egi.eu/) architecture of federated National Grid Initiatives maps well to a wide range of problems involving spatially-based analysis where the requirement for scalability can be satisfied by federating national-scale simulations.

Another promising application area for the NeISS lies in 'near real-time' forecasting in support of emergency planning and management. Emergency planning demands prediction of the impact of foreseeable events on populations, physical and social infrastructure, and the formulation of ways to deal with them. Emergency management requires access to up-to-date situational data and the means to revise predictions on a moment-by-moment basis. For example, planning for and managing the outbreak of an epidemic may require decisions about optimal locations for mass immunisation facilities and the capacity to track and predict the probable patterns and rates of infection transmission. Effective emergency planning and management calls for multi-faceted models that can capture relevant features of population behaviour and physical infrastructure, and can be adapted in the light of real-time data. Planners need to be able to explore different scenarios and use simulation predictions to devise determine optimal plans, test them under realistic conditions and train personnel to react. Managers need to be able to collect situational status information from diverse, distributed sources, to visualise data in meaningful ways and to update simulations so as to predict how an emergency will unfold.

The integration of Taverna functionality into the NeISS (http://www.taverna.com) is underway to allow the development of portlets with JSR 168/286 compliant components that provide access to enactment, management, monitoring and creation of workflows. At present, work is underway on the disaggregation of MapTube thematic visualisations in this framework, so that a single workflow can be used to generate a variety of representations of a single data field according to the selection of colours, shading styles, the number of intervals and boundary selection method. The higher level task of combining workflows in the generation of scenarios will be achieved in part by re-purposing of services from myExperiment (De Roure, Goble and Stevens, 2009), a VRE for the social curation and sharing of scientific research entities, especially workflows and *in silico* experiments, facilitating their integration into a portal. An important focus will be the incorporation of social networking functionality to provide extended archiving and documentation capability for social simulations, and early progress on this theme is illustrated in the twitter application described in Section 3.3. A persistent repository of social simulations is being designed around the notion of Research Objects,

aggregations of resources (data sets, analysis methods, workflows, results, people) that tell a particular story about an investigation, experiment or process and capture key information about the lifecycle of the investigation (for example, provenance information about analyses), facilitating re-use of results and repeatability of experiments.

For many users of the NeISS, access to social simulation resources could be mediated solely through repositories of simulation resources. This is the case for existing demonstrators that have been developed for policy users (Townend et al., 2009), and for Level 2 undergraduate teaching of urban simulation at the University of Leeds. Our collaborators include city councils with an interest in the provision of social services, housing, transportation and economic development, including simulation perspectives on proposed 'eco-towns', as well as the Royal Town Planning Institute (RTPI) with interests in planner education. We see considerable scope for public engagement through this process. For example, in relation to the transport demonstrator, social simulation could be used as a means to enable the public to understand the long-term implications of continuing existing travel patterns and to explore for themselves the effects of possible changes of behaviour. The fourth group of 'research' users in Figure 4 could typically be expected to have a much more wide-ranging interest in the development and deployment of social simulation services.

6. Summary and Conclusions

In recent years, substantial intellectual progress in the domain of social simulation has been complemented by the deployment of social science models, data management and visualisation through web portals. The work of the NeISS project to incorporate and extend these components within an e-Infrastructure for Social Simulation will make them easy to access, use and share for diverse strategic applications including health and social services; transport; land-use and economic development; and potentially for real-time issues such as emergency planning and disaster management. Development priorities include the combination of services into workflows, the creation of persistent repositories, and the deployment of social simulation tools through secure portals. Applications are expected to benefit a heterogeneous user community of researchers, policy-makers and citizens.

Acknowledgements

The research reported in this paper has been funded under the ESRC National Centre for e-Social Science, and through the JISC Information Environments Programme, e-Infrastructure for Social Simulation.

Word count excluding references: 5,300.

References

Arthur, W. 1999 Complexity and the Economy, *Science*, **284**, 107-109.

Bell, G., Hey, T. & Szalay, A. 2009 Beyond the Data Deluge, *Science*, **323 (5919)**, pp. 1297–1298, doi: 10.1126/science.1170411.

Birkin, M., Clarke, G., Clarke, M. & Wilson, A.G. 1996 *Intelligent GIS: Location decisions and strategic planning*, Geoinformation International, Cambridge.

Birkin, M., Dew, P., Macfarland, O. & Hodrien, J. 2005 Hydra: A Prototype Grid-enabled Spatial Decision Support System, *Proceedings of the First International Conference on e-Social Science*, National Centre for e-Social Science, Manchester.

Birkin, M., Wu, B. & Rees, P. 2009 Moses: Dynamic Spatial Microsimulation with Demographic Interactions In A. Zaidi, A. Harding & P. Williamson (Eds.), *New Frontiers in Microsimulation Modelling* (pp. 53-77). Farnham: Ashgate.

Brodlie, K., Wood, J., Duce, D. & Sagar, M. 2004 gViz: Visualization and Computational Steering on the Grid. In *Proceedings of UK e-Science All Hands Meeting 2004*, September 2004.

De Roure, D., Goble, C. & Stevens, R. 2009 The Design and Realisation of the myExperiment Virtual Research Environment for Social Sharing of Workflows. *Future Generation Computer Systems* **25**, pp. 561-567. doi:10.1016/j.future.2008.06.010

Department for Transport, 2009 National Travel Survey, 2002-2006 [computer file]. 3rd Edition. Colchester, Essex: UK Data Archive [distributor], January 2009. SN: 5340.

Epstein, J.M. 2007 *Generative Social Science: Studies in Agent-Based Computational Modelling*. Princeton University Press.

Ferguson, N. 2005 Making the paper: What would happen if a flu pandemic arose in Asia, http://www.nature.com/nature/journal/v437/n7056/full/7056xia.html.

Ferguson, N. M., Cummings, A. T., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., et al. 2005 Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, **437 (7056)**, 209–214. doi:10.1038/7056xia

George, F., Radcliffe, N., Smith, M., Birkin, M. & Clarke, M. 1997 Spatial Interaction Model Optimisation on Parallel Computers, *Concurrency: Practice and Experience*, **9 (8)**, 753-780.

Halfpenny, P., Procter, R., Lin, Y. & Voss, A. 2009 Developing the UK e-Social Science Research Programme. In Jankowski, N. (Ed.) *e-Research, Transformation in Scholarly Practice*, Routledge.

Liu, S., Chen, H., Birkin, M., Mao, B., and Guo, J. (2009) Assessment of the transport policy effect on CO2 emissions using a system dynamic model, *Air Quality 2009*, Istanbul.

McEachren, A., Cai G., Sharma, R., Rauschert, I., Brewer, I., Bolelli, L., Shaparenko, B., Fuhrmann, S. & Wang H. 2005 Enabling collaborative geoinformation access and decision-making, *International Journal of Geographical Information Science*, **19 (3)**, 293-317.

Manson, S. & O'Sullivan, D. 2006 Complexity theory in the study of space and place. *Environment and Planning A*, **38**:677-692

Merz, J. (986 Structural Adjustment in Static and Dynamic Microsimulation Models, in Orcutt G., Merz, J. & Quinke, H. (Eds.), *Microanalytic simulation models to support social and financial policy*, Amsterdam: North-Holland, Elsevier, 423-446

O'Donoghue, C., Leach, R. H. & Hynes, S. 2009 Simulating Earnings in Dynamic Microsimulation Models. In A. Zaidi, A. Harding & P. Williamson (Eds.), *New Frontiers in Micro-Simulation*, Farnham, Ashgate.

Townend, P., Xu, J., Birkin, M., Turner, A. & Wu, B. 2009 Modelling and Simulation for e-Social Science, *Philosophical Transactions of the Royal Society A*, **367**, 2781-2792.

Van Imhoff, E. & Post, W. 1998 Microsimulation methods for population projection. *Population: An English Selection*, **10**, 97–138.

Walker, J., Sinnott, R., Hilder, J., Stewart, G. & Tyrrell, A. 2009 *Optimised Generation of Electronic Standard Cell Libraries with Variability Tolerance through the nanoCMOS Grid*, submitted to Philosophical Transactions of the Royal Society A, July.

Wolfson, M. 2009 Orcutt's Vision 50 years on, in Zaidi, A., Harding, A. and Williamson, P. (eds) *New Frontiers in Microsimulation Modelling*, pp21-30. Farnham: Ashgate.

Wu B., Birkin M. & Rees P. 2008 A spatial microsimulation model with student agents, *Computers Environment and Urban Systems*, **32**, 440-453.